Intelligence 49 (2015) 32–43

Contents lists available at ScienceDirect

Intelligence

Defrauding Cyril Burt: A reanalysis of the social mobility data

Gavan Tredoux *

Xerox Palo Alto Research Center, United States

ARTICLE INFO

Article history:
Received 31 July 2014
Received in revised form 4 December 2014
Accepted 4 December 2014
Available online xxxx

Keywords:
IQ
Socio-economic class
Sir Cyril Burt
Academic fraud
History of science
Statistics

ABSTRACT

In the last comprehensive review by Mackintosh et al. *Cyril Burt, Fraud or Framed?* (London: Oxford University Press, 1995) of the fraud charges posthumously leveled against the once eminent psychologist Sir Cyril Burt, Mackintosh and Mascie-Taylor asserted that statistical anomalies they detected in his social mobility data of 1961 provided crucial evidence of guilt. The anomalies included apparent departures from normality in some parts of the data, incommensurate cell totals, and suspicious uniformity within IQ bands across fathers and sons. It is shown here that the departures from normality were a natural consequence of unavoidable rounding when inverting the cumulative normal distribution to construct the class IQ bands used in the tables. Elementary procedures are given, known since at least the 1930s, which could have been used by Burt to simultaneously preserve both the normality of his IQ data and the desired population proportions of occupational classes. Other anomalies first noticed by the statistician Donald Rubin are explainable as artifacts produced by fixing marginal totals in the presence of rounding to IQ scores, then using the same weighting procedures to conform to margins. The grounds given by Mackintosh and Mascie-Taylor for finding fraud in Burt's social mobility data are therefore dismissed.

In 1961 Cyril Burt published a paper arguing that social mobility was a predictable consequence of mismatches between cognitive ability and the intellectual demands of occupations, and of parental intelligence regressing to the mean among descendants in the presence of a stable class structure (Burt, 1961). He provided an illustrative example of this effect drawn from his earlier work when he was employed by the London County Council as its official psychologist. The paper and its data caused no stir at the time it was published.

Charges that Burt's social mobility data were "suspiciously perfect" were first made by the psychologist Michael McAskie in concert with his colleagues at the University of Hull, Clarke and Clarke (1974).[1] Soon after, they amplified these claims, moving rapidly from vague trouble that they reported in determining how the data was collected and treated from Burt's references (Clarke & McAskie, 1976), to outright charges in *The Times* of definite fraud (Clarke, Clarke, & McAskie, 1976). Burt's biographer Leslie Hearnshaw (1979) endorsed most of these charges, relying chiefly on an unpublished analysis by McAskie.[2] In the meantime, the psychologist Dorfman (1978) had proclaimed in a lead article in *Science* that his own extensive statistical analysis of Burt's social mobility had "shown, beyond reasonable doubt" that Burt "fabricated data on IQ and social class".

Dorfman's charges were immediately rebutted, notably by Burt's pupil Banks (1979), followed by two prominent

* Xerox PARC, Building 128, Phillips Road, Webster, NY 14580, United States. Tel.: +1 1445852655753.
E-mail address: Gavan.Tredoux@xerox.com.

[1] The charges first appeared under the authorship of the Clarkes, but as they acknowledged, they really owed them to McAskie. The original charges are dealt with in detail at the end of this analysis.

[2] Although this analysis was referenced by Hearnshaw as "awaiting publication" it was apparently never published, and cannot be traced. Exhaustive searches of McAskie's entire published output turned up empty. His output appears to consist of just four items, three of which concerned Burt directly and one indirectly. Joynson (1989) also reported that repeated requests to McAskie for a copy of the analysis met with no response (p. 199).

http://dx.doi.org/10.1016/j.intell.2014.12.002
0160-2896/

statisticians, Stigler (1979) and Rubin (1979), at first separately and then jointly (Rubin & Stigler, 1979). They argued that Dorfman's statistical methods were flawed and misinformed. As the “Burt Affair” erupted in the late seventies, the social mobility charges receded into the background, overshadowed by allegations about Burt's twin data, his IQ trend data, his apparently missing research assistants, and a host of supposed psychological defects, autobiographical quirks and economies with the truth that his biographer Hearnshaw claimed to detect. Exhaustively detailed accounts of the broader “Burt Affair” are given by Joynson (1989) and Fletcher (1991), followed up by the group presided over by Mackintosh (1995). Only the claims about the social mobility data will be considered here, for the rest the reader is referred to those sources.

In their detailed re-examinations of the Burt Affair, Joynson and Fletcher both deferred to Stigler and Rubin's rebuttals of Dorfman's analysis. Subsequently, when the broader Burt case was re-opened by Mackintosh and panel, it was indeed found that most of the *original* charges against Burt could not be sustained beyond reasonable doubt: “… the most suspicious feature of Burt's later claims and papers are usually *not* those actually identified by his initial critics” (Mackintosh, 1995, p. 147). As a result, the social mobility data rose again in importance. In the split verdict that emerged from the book, both of the contributors who found unambiguously against Burt, Mascie-Taylor and Mackintosh himself, cited the social mobility data as problematic enough to convince them that Burt was guilty. These were not the only grounds cited by Mackintosh — he also cited the failure of Burt to identify the secular increase in IQ scores now known as the “Flynn Effect”, a charge we will not consider here — but they formed a key component of his anti-Burt findings. Mascie-Taylor was *only* concerned with the social mobility data in his contribution. Therefore it is worthwhile to re-examine the data and weigh the evidence and arguments produced, which were in the end *not* precisely those made by Dorfman.

The main argument used by Mascie Taylor and Mackintosh, which we will return to in full later, took the following schematic form:

1. The data presented could not have been normalized, and so must be treated as raw empirical data.
2. The data contained cryptic departures from normality and other peculiar anomalies and mysterious regularities revealed only by computation and comparison.
3. Therefore, the data was fabricated since at the very least the departures from normality would not be expected.

The methods used attempt, in essence, to assess how plausible it is that the data just happens to exhibit regularity of various kinds and irregularity of other kinds. Although Dorfman had originally argued that the IQ data was “too normal” overall, Mascie-Taylor did not accept that argument, since he believed that the conclusion depended on the sample size, which was not stated by Burt. The cryptic departures from normality and other anomalies mentioned above were actually first noticed in passing by Donald Rubin, who described them only as “suspicious” (Rubin, 1979), but they are made to work much harder here. Mackintosh himself placed great emphasis on the above anomalies: “The critical problem with these IQ data is not their perfect normality … it is the departures from normality … those departures are not random, but show every sign of fabrication” (Mackintosh, 1995, p. 147). He speculated that Burt gave the game away by pushing assessments of intelligence in directions that suited him: “assessments were ‘adjusted’, i.e. moved from one side of a borderline to another, to give the answers he wanted” (Mackintosh, 1995, p. 146). The appealing feature implicit in this argument is that the resulting departures from normality are not immediately obvious and therefore we would not have to add stupidity to the fraud charges. Mackintosh sternly warned that “no trust can be placed in data such as these obtained by someone who knows in advance what results he wants” and that “I do not believe it possible to draw … a hard and fast distinction between adjustment and fabrication of data” (Mackintosh, 1995, p. 147). As we shall see, arguments that depend critically on the idea that the data was not transformed, that it just happens to exhibit both regularity and irregularity, are entirely mistaken.

Before examining these arguments for fraud in more detail it is useful to first understand the role that the data in question played in Burt's paper, as this throws a great deal of light on his methods and possible motives. Burt's claims for the data turn out to be surprisingly modest. In response to an ongoing debate that he had been conducting with some sociologists, Floud and Halsey, about the influences of hereditary factors on occupational social mobility, he undertook to show how regression of intelligence to the mean across generations, in the presence of a stable cognitive class structure, necessarily implies substantial social mobility between cognitive classes. These classes defined by Burt are not the usual socio-economic status (SES) classes, but are based instead on the intellectual demands of the occupation, which is how he preferred to think of the problem. The generational social mobility that Burt derived would, he asserted, be over and above the redistribution caused by an imperfect match between intelligence and class *within* a generation. His main concern was to show how substantial this mobility would be, since his opponents had doubted the magnitude and importance of the effect. To reinforce his point, Burt offered an illustrative example using data he had collected from 1913 onwards when conducting investigations as the official psychologist of the London County Council (L.C.C.), the product of “cross-sectional surveys of pupils in London schools, initiated primarily for the purposes of educational or vocational guidance and selection” (Burt, 1961, pp. 3–4). This is the first of two data sets in the paper, the second being a longitudinal study with “subsequent inquiries … carried out at intervals over a period of nearly fifty years … from 1913 onwards” (Burt, 1961, pp. 3–4) which has not played any role in the fraud charges against Burt and will not be considered here.

Burt's demonstration of mobility induced by IQ-to-class mismatching and regression to the mean depends on some additional facts which he explicitly lists and takes as agreed upon by all. Those relevant to our purposes are extracted below (Burt, 1961, pp. 4–5):

1. “During the period covered by our inquiries the population, from which our samples are drawn, and to which we intend out conclusions to apply, greatly increased in numbers”.

2. "During the last half century the proportional number of children in the population steadily declined and that of the elderly increased".
3. "During the period for which the information is available there has been no great change in the average level of intelligence".
4. "The amount of individual variation about the average level of intelligence has apparently remained fairly constant; certainly it has not declined".
5. "There are appreciable differences in the average level of intelligence in the different socio-economic classes, and in spite of the remarkable improvement in material and cultural conditions, the differences have altered hardly at all during the period in question."

A modern reader may or may not agree with some of these assumptions in the light of more recent evidence, but that is beside the point. The first two assumptions led Burt to present his data, as we will see below, in terms of numbers per 1000 rather than absolute totals. Burt is also explicit about his view of the "points of disagreement" between himself and his sociologist critics (Burt, 1961, p. 5):

1. "Dr. Floud and Dr. Halsey … deny that the apparent differences between the class-means for general intelligence are in any degree due to innate differences; and both contend instead for 'a hypothesis of near-randomness in the social distribution of innate intelligence.' This implies that the means for all the classes would be approximately the same."
2. "Dr. Halsey … criticizes both the amount of social mobility which I had assumed and the length of time over which I assumed it had operated".

Burt's use of his first (cross-sectional) data set is aimed primarily at the second point of disagreement, to show how easily a large amount of mobility may be generated by intelligence regressing to the mean and by IQ-to-class mismatches, although he does address the first point in passing. He warns his readers upfront that his "data are too crude and limited for a detailed examination by a full analysis of variance" and that "it is my purpose to keep, so far as possible, to the simplest and most intelligible methods of comparison, relying largely on the percentage methods favoured by the sociologists themselves" (Burt, 1961, p. 9). He proceeds "to examine, not only (as is usually done) the class-means, but the entire frequency distributions", which he gives for each cognitive class that he defines. Since this data will be examined here in great detail, it is reproduced in full below, along with other subsequent tables (Burt, 1961, p. 11).

The cross-sectional data set consists of intelligence assessments of parents (fathers, it would seem) and children (sons, presumably), cross-classified by occupational cognitive class, and Burt warns that "For obvious reasons the assessments of adult intelligence were less through and less reliable" (Burt, 1961, p. 9). In one of his earlier publications, referred to elsewhere in the discussion and the first item in his list of references, assessments that may have been used here were described in more detail (Spielman & Burt, 1926). They were based on 20 to 40 minute interviews conducted with the parents, supplemented by overt and covert tests during the conversation where there was some doubt about the estimates, which is why he warned that they were less reliable.[3] The assessments for the children (sons) are based instead on more straightforward intelligence tests, and their "occupational classes" are obviously just those of their parents (fathers) since they were at school when assessed — this data set is, as noted before, not longitudinal.

Here the occupational classification is explicitly "based, not on prestige or income, but rather on the degree of ability required for the work" (Burt, 1961, p. 10). Burt had used this classification scheme and the expected population proportions of each class extensively in his earlier work, and he referred his readers to "previous reports" where it is described in more detail. The most important of these is Spielman and Burt (1926), which by an apparent oversight he does not refer to explicitly at this point, but only elsewhere in the text. He also makes a crucial qualification, to which we have added some emphasis because of the confusion it caused long after (Burt, 1961, p. 10):

> "In constructing the tables the frequencies inserted in the various rows and columns were proportional frequencies and in no way represent the number actually examined: from class I the number actually examined was nearer a hundred and twenty than three. To obtain the figures to be inserted (numbers per mille) *we weighted the actual numbers so that the proportions in each class should be equal to the estimated proportions for the total population*. Finally, for the purposes of the present analysis we have rescaled our assessment of intelligence so that the mean of the whole group is 100 and the standard deviation 15."

It is obvious from Tables I and II that the margins are almost identical. The column totals, appearing in the final row of Tables I and II represent the IQ distribution as a whole for fathers and sons respectively and are close, usually differing only by a unit. The row totals, appearing in the second-last column, give the proportional representation of the cognitively-defined occupational classes in the general population per 1000 and are identical. By scaling to the row totals, Burt was actively correcting the unbalanced nature of his original sample with respect to the expected proportions of the classes in the population. The agreement is not surprising, since as called out above, Burt stated that he weighted his data to get this agreement with the marginal totals, and we shall see how he could have obtained the agreement using some simple scaling techniques which were well-known since the 1930s.

The fraud charges later leveled against Burt, namely that he invented the cell entries in Tables I and II, depend entirely on these marginal totals. Dorfman (1978) noticed that the IQ

[3] Burt does not explicitly state whether the estimates for fathers were point estimates or less exact grades. In Gaw, Ramsey, Smith, Spielman, and Burt (1926, p. 75) four grades (A+, A, B, C corresponding to mean IQs 124, 107, 92, 79) were used for *mothers* who were interviewed, but Burt clearly did not use this approximation, since he would have needed at least 6 grades corresponding to his cognitive classes. Note that he does *not* state that he used the 1926 data here — it is referred to only for the occupational classification used. Of course, for each class a normal distribution may be fitted to a grade scheme using the estimated mean and variance for that class, and then proportions for ranges, as in Tables I and II, can be deduced from the fitted normal. It is only the class *distributions* that matter for Burt's purposes here.

Table I
Distribution of intelligence according to occupational class: adults.

	50–60	60–70	70–80	80–90	90–100	100–110	110–120	120–130	130–140	140+	Total	Mean IQ
I Higher professional									2	1	3	139.7
II Lower professional							2	13	15	1	31	130.6
III Clerical				1	8	16	56	38	3		122	115.9
IV Skilled			2	11	51	101	78	14	1		258	108.2
V Semiskilled		5	15	31	135	120	17	2			325	97.8
VI Unskilled	1	18	52	117	53	11	9				261	84.9
Total	1	23	69	160	247	248	162	67	21	2	1000	100

Table II
Distribution of intelligence according to occupational class: children.

	50–60	60–70	70–80	80–90	90–100	100–110	110–120	120–130	130–140	140+	Total	Mean IQ
I Higher professional						1		1	1		3	120.8
II Lower professional				1	2	6	12	8	2		31	114.7
III Clerical			3	8	21	31	35	18	6		122	107.8
IV Skilled		1	12	33	53	70	59	22	7	1	258	104.6
V Semiskilled	1	6	23	55	99	85	38	13	5		325	98.9
VI Unskilled	1	15	32	62	75	54	16	6			261	92.6
Total	2	22	70	159	250	247	160	68	21	1	1000	100

distribution in the column margins closely matched the normal distribution, assuming that the sample size was 40,000, and that the row margins closely matched the expected class proportions given in Spielman and Burt (1926). He invested a great deal of effort in comparing the IQ distribution with many other populations in an attempt to show that it is "too normal" compared with those populations, an empirical proposition that depends on the characteristics of the relevant populations, and cannot be tested directly using formal statistical techniques (by contrast, one can easily show that data is not normal enough using techniques like the chi-squared test). The correspondence with normality (mean 100, standard deviation 15), given to 3 places, is apparent from Table 1 below.[4]

Similarly, Dorfman also went to great lengths to show that the class proportions in the row margins in Tables I and II match those given in Spielman and Burt (1926), effort on his part that might have been avoided if Burt had explicitly called out that source at that point of his paper, though it certainly is noted elsewhere in the text. Dorfman claimed that the marginal agreement he thought he had uncovered was not obvious and concluded "beyond reasonable doubt" that Burt had made up the data to obtain that agreement, though he was not clear about the exact motive for this, since as we shall see none of Burt's stated conclusions actually depend on these marginal agreements. He also noticed that the mean IQs of the occupational classes were highly correlated for fathers and sons, concluding that the agreement there was also produced by inventing the data to suit. Again, Dorfman was not clear about the supposed motive for doing this, since Burt did not use the exact values of these means, a point we will return to.

Two statisticians, Stigler (1979) and Rubin (1979) responded independently to Dorfman's claims. They noted in letters to *Science* that, as we have emphasized above, Burt had explicitly stated that he was weighting the data to get the marginal agreements, and that the IQ normality likely stemmed from the common practice of assuming that IQ data was approximately normal and scaling to that expectation. They also pointed out that Burt did not state his sample size and that Dorfman's assumption that it was 40,000 was based on misreading Burt's text. Stigler noted that high correlations between class means are in fact common in the bivariate normal distribution. Dorfman's analysis contained, they both argued, no evidence that Burt had committed fraud. Rubin did detect some anomalies in the data, which we shall examine and resolve at length below, but did not himself draw the conclusion that those anomalies proved anything beyond reasonable doubt.

Dorfman (1979a,b) to Stigler and Rubin contained an argument that became the foundation of the guilty verdicts given by Mascie-Taylor and Mackintosh. Dorfman first appealed to other authorities who had supposed that Burt had used a sample size of 40,000, but as Mascie Taylor pointed out this is simply not supported by Burt's text, so it must be discarded. Dorfman next argued that Burt *could not possibly* have weighted his data to match *both* the IQ distribution in the column margins, and the class proportions in the row margins, and offered an algebraic "proof" of this claim. Mascie-Taylor found this "proof" convincing, and therefore treated Burt's data as if it was not adjusted to the margins (with important consequences when considering Rubin's anomalies) but this is misconceived from the start. Dorfman's argument depends entirely on an assumption that Burt used *row weights* to match the class proportions, but would thereby have disturbed the normality of the IQ column margins, which is certainly true, but irrelevant. Likewise, he argued that if Burt had first matched the row margins and then adjusted the column margins back to normality, he would have disturbed the row margins. This analysis and its "proof" depend entirely on supposing that Burt used either *row* or *column* weights, but he could simply have

[4] Tables original to this discussion and not reproduced from Burt (1961) are numbered in Arabic numerals. Burt's tables are numbered using the Roman numerals that Burt used for them and are those used in previous discussions.

Table III
Distribution of intelligence according to occupational class: adults.

Rescaled	VI	V	IV	III	II	I	Total
	50–91	91–103	103–115	115–127	127–141	141+	
I					2	**1**	3
II			*1*	**15**	*14*	1	31
III	1	15	38	**56**	12		122
IV	16	86	**114**	38	4		258
V	53	**178**	84	10			325
VI	**191**	46	*21*	*3*			261
Total	261	325	258	122	32	2	1000

Table IV
Distribution of intelligence according to occupational class: children.

Rescaled	VI	V	IV	III	II	I	Total
	50–91	91–103	103–115	115–127	127–141	141+	
I			1	1	1		3
II	1	4	11	9	**6**		31
III	11	28	51	**20**	12		122
IV	46	66	**75**	62	8	1	258
V	91	**122**	84	23	5		325
VI	**112**	105	36	7	1		261
Total	261	325	258	122	33	1	1000

used individual *cell* weights to simultaneously match both margins. It is possible that both Dorfman and Mascie-Taylor were led astray here by reading (emphasis added) "*Finally*, for the purposes of the present analysis we have rescaled our assessment of intelligence", in Burt's description above, to mean the last of several sequential steps. Instead, one might read "finally" to refer not to steps but rather to the statements made, as in *the last thing I am going to say about the data*.

Regardless of the source of Dorfman and Mascie-Taylor's confusion, methods for finding cell weights to simultaneously match both row and column margins were well-known and widely used since the 1930s, particularly in adjusting census data. W. Edwards Deming described several methods for finding cell weights in his classic text Statistical Adjustment of Data (1938), which though it went through several editions may not have been read widely enough in psychology departments.[5] Burt, who avidly consumed and reviewed statistical literature, was presumably familiar with that text and prior papers, and may have used any one of the methods Deming describes, which range from least-square solutions using Lagrange multipliers to iterative approximations. One simple method given by Deming, which is still very popular today, is called "Iterative Proportional Fitting", though it has many other names depending on the domain it is applied in (see Bishop, Fienberg, & Holland, 1975; Pukelsheim & Simeone, 2009, where formal conditions for convergence are given). Others have called this procedure "raking", or "scaling" etc. One first weights by row, dividing each cell in a row by the current row total and multiplying by the desired row total. Then one weights by column, using current and desired column totals, and repeats the process again, alternately by row and by column, until the result is close enough. A cell in position (i, j) then has an individual weight produced by multiplying the successive row weights used for row i by the successive column weights for column j. As Deming pointed out, usually only a few iterations of this procedure are required until the result is accurate enough. When whole numbers must be used, as they must be in Burt's Tables I–VI, rounding up or down and, if necessary, just adjusting a margin here and there by a unit will give a good enough result for most purposes. Inspecting Burt's margins it seems likely that he did something like that. We cannot, of course, be sure exactly which procedure Burt used here, but we only have to find one or more possible procedures that he *could* have used to adjust his data to the margins, to refute the contention that his data *must* have been unadjusted.

The following example shows the use of Iterative Proportional Fitting for a simple 2 × 3 matrix, with column margins 100, 50, 4 and row margins 70, 84. The procedure starts with an arbitrary initial matrix of values, which is then fitted to both margins using four iterations, stopping when more iterations would produce little difference in the result. Note that the final result has been rounded to integers, and unit adjustments of the margins were not needed here.[6]

```
Start          40     7    23    70

               21    60     5    84

              100    50     4   154

Iteration 1   66.103     5.334988   3.2994162        70

              33.897    44.665012   0.7005838        84

              100.000   50.000000   4.0000000       154

Iteration 2   63.28241   4.774207 3.2251411          70

              36.71759 45.225793 0.7748589           84

              100.0000 50.000000 4.0000000          154

Iteration 3   62.5001    4.631397 3.2039827          70

              37.4999   45.368603 0.7960173          84

              100.000   50.000000 4.0000000         154

Iteration 4   62.29402   4.59462   3.1983679         70

              37.70598 45.40538   0.8016321          84

              100.0000 50.00000   4.0000000         154

Rounded       62     5    3       70

              38    45    1       84

              100   50    4       154
```

[5] In his book *Fourier Analysis* (1989), the Cambridge mathematician T.W. Körner adorns his discussion with some chapters on Burt's social mobility data that have not been widely noticed. He states that "There do exist methods for fitting data to preassigned marginals but they are computationally tedious and bear little resemblance to Burt's description of his procedure" (Körner, 1989, p. 439). However, Burt states only his outcomes, the weights, and not how he found them, which is trivial and not at all computationally tedious using Iterative Proportional Fitting (no Fourier Analysis needed).

[6] See the appendix for an implementation of this procedure in the statistical package R.

Table V
Adults. Percentage in each group whose intelligence is below, above, or equivalent to that of their occupational class.

	Below	Equivalent	Above	Number
Class I–III	46.2	45.5	8.3	156
Class IV–V	26.6	50.1	23.3	583
Class VI	–	73.2	26.8	261
Total population	22.7	55.4	21.9	1000

Table VI
Children. Percentage in each group whose intelligence is below, above, or equivalent to that of their occupational class.

	Below	Equivalent	Above	Number
Class I–III	75.5	16.8	7.7	156
Class IV–V	34.8	34.3	30.9	583
Class VI	–	42.9	57.1	261
Total population	32.1	33.5	34.4	1000

To obtain his means for each class, Burt would have used the weighted sum of the actual point values in his data set, where each group of values in a cell receives the cell weight found by the above procedure, divided of course by the total count. It is easy to check that he did *not* use a less exact method like simply summing the expected value of the normal distribution over each range multiplied by the final cell count for the class in that range, since though the values obtained are close they are not exactly the same (see Table 2 below).

Nor did he use the alternative method of just taking the mid-points of the intervals (rather than the expected values) with some reasonable choice for the interval greater than 140 (like the expectation 144.627 or say 150), since that procedure also does not produce the results given, though they are again not far off (see Table 3 below). Mascie-Taylor repeats Dorfman's argument that this failure to match the mid-point approximation proves that Burt could not have scaled the data to match the column margins (Mascie-Taylor, 1995, pp. 88–9), but this is a non sequitur. After scaling, each class mean would be obtained from the weighted sum over the point values and need not exactly match either approximation.

Burt was now in a position to point out that his data exhibited "an overall regression averaging 0.52" (Burt, 1961, p. 14). In the light of the general qualifications he offered about the quality of his data, this cannot be read as a confident claim about the population as a whole, for which a more definite statement of his sample size would be required.

Burt went on to make two more general points about regression to the mean. The first of these was that the regression meant both a decrease from higher IQ classes, and an increase from lower IQ classes, which he argued was hard to explain purely in terms of cumulative socio-economic advantage or disadvantage. The second of these was that if the social class structure stayed relatively stable, which he took as agreed on by all, then an interchange between classes would have to take place, otherwise the means of the classes would converge to the population mean over time, as would their variances to the overall population variance, and the class structure would melt away.[7]

The next step of Burt's argument was to transform his data into a format where it is immediately obvious how much social mobility could be produced by IQ alone. Inspecting the tables it is clear that there has been a generational regression to the mean between the IQ scores of the fathers and the sons, within occupational classes. The class means for sons are more close to the overall mean (100) than they are for fathers, and moreover there is an increase in variance among the sons for each class, when compared with the variance of their fathers. The regression to the mean across the generation for each class can be seen more clearly in Table 4, where the regression coefficients for each class are calculated from ($IQ_{son} - 100$) / $IQ_{father} - 100$) and the approximate standard deviations are estimated from the midpoints of the classes and the stated counts and means.

There is, for both fathers and sons, a considerable overlap in the range of IQs within each occupational class, with the potential for social mobility within the current generation to achieve a better match between the intelligence required for an occupation and the intelligence of those who actually pursue that occupation. Burt wanted a simple way to convey the amount of mobility induced by IQ mismatches, one that required no formal knowledge of statistics to understand. He therefore constructed a *new* set of hypothetical cognitively-defined occupational classes that would need to exist if there had to be a *perfect* match between the IQ required for an occupation and one who pursued it, with no overlap of cognitive ability between classes. This would then result in a simple spillover effect, with numbers of people who would have to be transferred between classes to match actual and desired numbers. He obtained his new set of cognitive classes using a construction that will be shown here in fine detail, since his description of it was somewhat elliptical and it produces some simple side effects that were noticed by Rubin and then entirely mistaken by Mascie-Taylor and Mackintosh as evidence of fraud.

The result that Burt required was a set of cognitive classes which matched the population proportions set by his row margins, and could therefore fit in the right amount of people without overlap. To get these new classes, he derived new IQ boundary points between the classes by inverting the right tail of the cumulative normal distribution, since his data was cut off below by IQ 50 and in theory extended upward without limit. Consider his cognitive class I, which as its row margin specified, he expected to contain 3 out of 1000 people, a frequency of 0.003. The problem is to find X so that 3 in 1000 have IQ > X. One finds this easily by inverting the right tail of the cumulative normal distribution for the probability 0.003, so that X is, to three places, 141.217. Then for class II, which has 32 out of 1000 people in it, class I and II combined have $32 + 3 = 34$ in 1000. Finding X so that IQ > X has 34 in 1000 people yields 127.375, and so on. Fig. 1 shows the case for the cumulative proportion 0.156, which corresponds to the IQ 115.166.

Continuing for the other classes, we find the following IQ ranges with the desired numbers of individuals in them (Table 5).

[7] It may be relevant that the Clarkes enter into the discussion at this point. Burt refers to a paper by them (Clarke, Clarke, & Brown, 1960) which, he says, invents a concept called "egression from the mean", something like Darwin's "spontaneous variation", in order to explain why the class means does not approach the population means asymptotically. He points out that no such concept is called for.

Table 1
Normal distribution of column margins.

IQ range	50–60	60–70	70–80	80–90	90–100	100–110	110–120	120–130	130–140	140+
Normal (100,15)	3.401	18.920	68.461	161.281	247.507	247.507	161.281	68.461	18.920	3.830
Fathers	1	23	69	160	247	248	162	67	21	2
Sons	2	22	70	159	250	247	160	68	21	1

Table 2
Expected values of the normal distribution by IQ range.

Range	50–60	60–70	70–80	80–90	90–100	100–110	110–120	120–130	130–140	140+
Normal expectation	56.545	66.230	75.895	85.543	95.182	104.818	114.457	124.105	133.770	144.627

However, IQs are traditionally given in whole numbers, so Burt had to *round* the IQ boundaries found above. He chose to round 90.396 *up* to 91, while he rounded the rest *down*. This produces the IQ bands given in the last row above (141, 127, 115, etc.). This procedure was noticed by Dorfman (1978), but he apparently failed to comprehend the second part of this procedure. Burt had fixed the column margins for his data, and could then reclassify his data into the new IQ ranges, but he then had to *rescale* his data to match the desired margins, probably using one of the techniques listed above such as Iterative Proportional Fitting or an equivalent. He had to do that because the rounding unavoidably produced intervals containing numbers which were either *too large*, or *too small*, for their desired frequencies. This is apparent from Table 6 below which gives the numbers per 1000 that the normal distribution predicts for the rounded intervals versus the exact intervals, and the resulting pathologies that must be corrected by refitting the cell values to the new margins.

After rescaling to the rounded margins, Burt exhibited Tables III and IV below for the fathers and sons, classified according to the new ranges guaranteeing an exact match between IQ and class size. He was careful to mark both Tables III and IV as “rescaled”, as our added emphasis shows:

Notice above that the margins differ slightly for the sons and fathers, which would have been produced by rounding after weighting the cell entries and pragmatically choosing to adjust the margin by a unit to accommodate the result. We will return below to the other features which are shown highlighted in Table III above.

Burt then noted that the diagonal entries, in bold, give the number who are correctly placed in their cognitive class in both Tables III and IV. To achieve a perfect allocation, the remainder would have to be reassigned to other classes, producing the following Tables, V and VI, of reallocations, simplified by some combination of classes.

Burt's analysis of his cross-sectional data set was directed ultimately at the construction of Tables V and VI above, with their elegant numerical demonstration of the *amount* of social mobility that perfect allocation of individuals to appropriately defined classes would produce. It is noticeable that the mobility of the sons is increased by the effect of regression to the mean, as they require more redistribution than the fathers due to their increased variance within each class.

To summarize then, the procedure that Burt followed produced *transformed empirical data*, and involved the following steps, which we will take to be a plausible reconstruction of his methods. We suppose that Burt:

1. Normalized his empirical IQ data for fathers and sons, obtaining column margins for both that were approximately normal.
2. Fixed the proportions of the overall population that he expected his cognitive occupational classes to take up, forming his row margins.
3. Simultaneously scaled his empirical data to match both the above margins, using any one of many available techniques for doing so, to form the weighted cell entries in Tables I and II.
4. Calculated his class means from the individual data points, weighted by the cell weights found above.
5. Constructed IQ bands large enough to contain entire occupational classes and rounded the boundaries to whole IQs.
6. Reclassified his original data into the new IQ bands.
7. Necessarily scaled the reclassified data to match the new column margins and almost identical row margins, forming Tables III and IV as a result.
8. Calculated using percentages the social mobility implied by the above, both within a generation and across the single generation in his data, forming Tables V and VI.

Table 3
Means calculated from expect values over, and mid-points of, intervals.

	Approximate class mean using range expectation	Approximate class mean using range mid-point and 150 for 140+	Weighted mean
I	137.389	140	139.7
II	128.821	130	130.6
III	115.172	115.738	115.9
IV	105.940	106.163	108.2
V	97.671	97.585	97.8
VI	85.945	85.421	84.9

Table 4
Regression of class means.

Class	I	II	III	IV	V	VI
Fathers' mean IQ	139.7	130.6	115.9	108.2	97.8	84.9
Std. dev.	7.077	7.096	9.337	10.101	9.930	10.920
Sons' mean IQ	120.8	114.7	107.8	104.6	98.9	92.6
Std. dev.	12.502	11.212	13.625	14.430	13.865	13.757
Regression	0.524	0.480	0.491	0.561	0.500	0.490

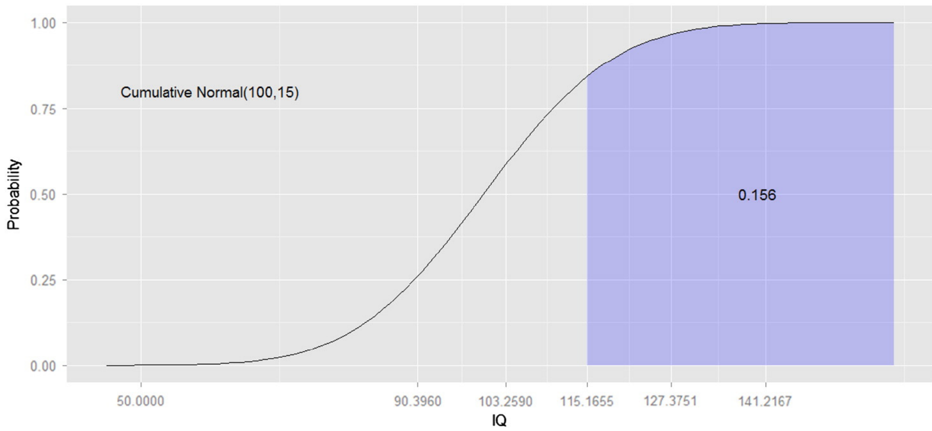

Fig. 1. Construction of the new IQ intervals.

We may turn then to the supposed anomalies in this data mentioned previously. When inspecting Burt's data during his interchange with Dorfman in the pages of *Science*, Rubin (1979) noticed three features of the data that he found hard to account for. These anomalies would, as noted above, be critical to the "guilty" verdicts offered by Mascie-Taylor and Mackintosh (1995), who interpreted them as evidence that Burt had deliberately and fraudulently manipulated his data, but was "found out" by these anomalies. We will deal with each anomaly in turn and show that each is either a statistical artifact produced by the interval construction methods outlined above, or an unsurprising outcome deliberately and explicitly created. Since Rubin noticed them we will call them Rubin's Anomalies, though one should point out that Rubin did not conclusively argue that they indicated malfeasance by Burt, only that "Although Dorfman's statistics do not provide any evidence that Burt fabricated data, there may be such evidence in Burt's tables", and that "the patterns … are suspicious" as are "the excellent fits of the IQ margins in Burt's Tables I, II, III and V" (Rubin, 1979, p. 245).

Rubin's First Anomaly. By combining the column margins of Tables I and III, and likewise II and IV, above, Rubin deduced a narrower band of IQ data than Burt had explicitly presented. Consider the fathers. Since there are data in Table I for the IQ range 50 to 90, adding up the column margins gives a total of 1 + 23 + 69 + 160 = 253. From Table III we see that the range 50–91 is totaled at 261. This implies that the band 90–91 has a count of 261–253 = 8 in it. Rubin pointed out that the normal distribution **N**(100, 15) predicts a much larger number in this range, namely 21. Similarly for the sons. This is the deviation from normality that Mackintosh referred to above, and it has the appearance of being cryptic since one has to do some calculation to reveal it. However, this rests on a simple mistake. As we have seen, when Burt was finding his IQ ranges by inverting the cumulative normal distribution, he had to round his intervals because IQs are given in whole numbers. The corresponding interval is *not* 90–91, but instead 90–90.396. It follows directly from the construction explained above that the normal distribution **N**(100, 15) predicts 8 for the interval 90–90.396, but readers can calculate this directly if they like. The "departures from normality" are merely statistical artifacts produced by the fixed marginal totals interacting pathologically with the whole number IQ ranges. Since this had nothing to do with Burt's argument, he did not comment on this cryptic side-effect. It is a bad mistake to suppose that these "deviations from normality" constitute proof of fraud or deceit.

Rubin's Second Anomaly. There is also a peculiarity in Table III for the fathers, which Rubin spotted. If one combines the cell counts in the ranges 103–115 and 115–127 for class VI, one obtains a total of 21 + 3 = 24. However, Table I lists a total of 19 + 1 = 20 for class VI in the *larger* range 100–130. This cell count inconsistency is highlighted and emphasized in Table III above, along with another inconsistency that Rubin did *not* notice for class II, which sums over 103–127 to 15 + 1 = 16 in Table III, but the larger range 110–130 previously summed to 13 + 2 = 15 in Table I. Rubin suggested this might be evidence that Burt had fiddled with the data, but also supposed that it might have been a typo or miscalculation. Mascie-Taylor follows Rubin in calling out this anomaly, though it is not clear if he accepts the

Table 5
IQ ranges calculated by inversion of the cumulative normal distribution.

IQ range	>141.217	141.217–127.375	127.375–115.166	115.166–103.259	103.259–90.396	90.396–50
Band frequency	0.003	0.031	0.122	0.258	0.325	0.261
Per 1000	3	31	122	258	325	261
Cumulative frequency	0.003	0.034	0.156	0.414	0.739	1.000
Rounded IQ range	>141	141–127	127–115	115–103	103–91	91–50

Table 6
Pathologies due to rounding.

Interval	Normal capacity	Desired capacity	Pathology	Difference
50–91:	274	261	Bigger	+13
91–103:	305	325	Smaller	−20
103–115:	262	258	Bigger	+4
115–127:	123	122	Bigger	+1
127–141:	33	31	Bigger	+2
>141:	3	3	Equal	0

typo explanation or not. However, looking more closely at Table III one notices that the row and the column totals still add up correctly in both cases! This is unlikely to have been the direct result of any typos. The most likely cause is instead another by-product of rounding the IQ intervals, reclassifying and then rescaling. As noted above, most of the IQ intervals were made too large by rounding, with one too small. When reclassifying his data using the new intervals, starting again from the raw data and *not* from the weighted cell counts in Table I, the initial cell counts would probably have been too large in places and too small in others, due to the rounded intervals being either too small or too large for the desired column totals, or because of unevenness uncovered in the redistribution of the reclassified data. These cell counts would then have to be weighted differently to match the desired marginal totals. For this reason *the counts in the cells of* Tables III/IV *are not comparable with* Tables I/II. The anomalous larger counts are most likely just artifacts of the rescaling procedure interacting with the reclassification and the need to round to whole numbers (in the case of the unit difference for class II), or they may have been produced by arithmetic errors during the initial and intermediate steps of the rescaling process, or by some combination of the two — one cannot tell. It is of course all too easy to make elementary mistakes when doing this kind of work by hand. It is unlikely that Burt noticed this at all, and it has no important bearing on his argument.

Rubin's Third Anomaly. The last anomaly noted by Rubin is pervasive regularity in the marginal counts for IQ bands. He showed this effect using Table 7 of marginal counts obtained by combining Tables I, II, III and IV, with the normal distribution totals given for comparison:

This regularity is even more striking in the following form, also given by Rubin, when some of the bands are combined (Table 8).

It is in fact easy to understand this. It did not arise by chance. Those who have followed the detailed procedure given above for constructing Tables III and IV will not be at all surprised by the regularity shown, since those tables were constructed to have those column totals, modulo rounding. In particular, for the IQ bands involving those found by inversion like 103–115, the column marginal totals were inherently specified in advance to produce this agreement, with the IQ band widths chosen to yield (approximately) the right totals, with Tables III and IV then rescaled to meet those marginal totals almost exactly. In so far as there is minor disagreement between some of the marginal totals in Tables III and IV for fathers and sons, that is likely because of the effects of rounding cell entries to whole numbers after scaling or fitting, and adjusting the marginal total by a unit or so where needed. Combining bands removes those unit effects and reverts back to the cumulative proportion that the bands were constructed to meet, e.g. 34 for 127+. For the other band totals (90 and below) shared coherence to normality suffices, and note that the bands 50–70 and 70–90 have simply been chosen for display in Table 7 *because* they display coherence. Most of the regularity in the marginal totals shown above is present purely by construction. The entries given for the normal distribution, which create the impression that the bands resemble each other more than they do the normal, do not account for the *rounding* of the IQ intervals constructed, and are once again the wrong intervals to compare with — they should simply be ignored.

We may return then to Mascie-Taylor's (1995) summary of his findings, which was clearly endorsed by Mackintosh:

1. "If Burt had normalized his data using the individual IQ scores of fathers and sons then his data would no longer have remained normal when the new weights from Spielman and Burt were used. Alternatively, if Burt normalized the data by cell means then the overall mean IQ of an occupational class would be equal to the weighted sum of the midpoints of each cell range. This is also not correct since only one out of the nine classes gives the correct mean. Thus neither of the straightforward interpretations of normalizing data can reproduce Burt's column totals.
2. Rubin was able to show, when combining the information from Burt's Tables I–IV, that the data were not as normal as hitherto supposed; there were too few counts in the range of 90–1 — only 8 when 21 would be expected. Furthermore Rubin's analyses showed that fathers' and sons' IQ distributions were even more similar to one another than either was to the normal distribution. Even more suspiciously, he noted that digit discrepancies in one IQ band were regularly 'corrected' in the next. He also showed after reconstructing each occupational class using the narrower IQ categories 'a blatant inconsistency' in class VI. In Burt's Table I there are 20 with IQs greater than 100, while in Table III in the same class, 24 occur with IQs greater than 103. As Rubin acknowledges, this inconsistency might be evidence of Burt's fabrication but it could be due to some type of recoding, computational, or even, I suppose, typographical error" (p. 93).

On the basis of these points, Mascie-Taylor concluded that "There is no doubt, in my view, that Burt deliberately concealed information — on the sample size, on where the row totals came from, on when the information was collected. Even if he did not fabricate the data, then he was deliberately deceptive. But there really are good reasons to believe that the data were fabricated" (Mascie-Taylor, 1995, p. 93).

From the detailed reconstruction we have given of Burt's methods, the charge of fabrication can definitely be dismissed for all the points raised by Mascie-Taylor in his summary. Simultaneously fitting row and column margins, including the column margins fixing the approximately normal distribution of the IQs, is easily achievable using methods in use since the 1930s, while Rubin's anomalies rest for the most part on simple

Table 7
Similarity between distributions of fathers and sons.

IQ	50–60	60–70	70–80	80–90	90–91	91–100	100–103	103–110	110–115	115–120	120–127	127–130	130–140	140+
Fathers	1	23	69	160	8	239	86	162	96	66	56	11	21	2
Sons	2	22	70	159	8	242	83	164	94	66	56	12	21	1
Normal	3	19	68	162	21	226	79	168	94	68	55	13	19	4

misunderstandings of Burt's inversion of the cumulative normal distribution to fit stated proportions. The additional claims made about an impression of *deception* by Burt are less precise and inherently subjective. Nevertheless we will consider them as best we can.

It is certainly true that Burt did not explicitly state his sample size, in contrast to the modern practice of always stating sample sizes, but it is important to understand that the sample size did not form any part of his argument. Burt did *not* perform an analysis of variance or make any statistical inferences that depend formally on the sample size through the parameters of probability distributions. Reiterating from above, he wished to "so far as possible, avoid unfamiliar methods and formulae" (Burt, 1961, p. 6) and cautioned that "The data are too crude and limited for a detailed examination by a full analysis of variance" (Burt, 1961, p. 9). He goes on to say that he was tailoring the presentation to the intended audience: "it is my purpose to keep, so far as possible, to the simplest and most intelligible methods of comparison, relying largely on the percentage methods favoured by sociologists themselves" (Burt, 1961, p. 9). Hence he used the less formal methods we have described in detail above, leading to the percentages in Tables V and VI. Mascie-Taylor (1995, p. 82) accepted that Burt had a sample size of the order of at least 1000 based on other references in his work, which is also implied by his presentation in proportions of 1000. This is certainly large enough to exhibit, at least within the sample itself, the magnitude of mobility he was arguing that Floud and Halsey had underestimated. Assessing whether this can plausibly be generalized to the broader population would be illuminated more by more definite knowledge of the *representativeness* of his sampling than by any more precise knowledge of the exact sample size. We can only judge this after all this time in the light of subsequent replications, a topic that is returned to in closing.

The idea that Burt sought to conceal the source of the expected population percentages of his occupational classes is definitely refuted by the inclusion of the source (Spielman & Burt, 1926), in which those percentages appear, as the very first reference in his list appearing at the end of the paper, a list that was *not* organized alphabetically. He explicitly states that the "occupational classification is much the same as that used in previous reports" and that it "has been described by Carr-Saunders and Caradog Jones" (Burt, 1961: pp. 9–10) citing the exact page and table in which the classification is given there, together with the population percentage estimates. Referring to Spielman and Burt (1926, Table III, col. 5, p. 13) one sees at once that the percentages given there are indeed the same as those used by Burt (1961, Tables I and II, p. 11) and reproduced in Carr-Saunders and Jones (1937, Table XXXI, p. 56). As with the issue of the sample size, a simple query to Burt at the time would have clarified this, so what, exactly, could he hope to achieve by "concealing" either of these?

Mackintosh (1995, p. 145) offers a more sharply defined theory, asserting that Burt "would have invited ridicule had he acknowledged that his estimates of the distribution of occupational classes were from Spielman and Burt (1926)". However, there is nothing to ridicule in Burt's use of these proportions. He explains in detail that the sizes were chosen to match estimated cognitive demands, by approximating observed proportions of grammar school scholarship winners, those transferred to Central schools, those in public or preparatory schools and so on (Burt, 1961, p. 10). Following the chain of references, one is given clear explanations of the construction of proportions from occupational labels, e.g. "proportions given have been computed primarily from the figures given in the Census returns for London" (Spielman & Burt, 1926, p. 15). Given that when Burt's *cross-sectional* data was collected he had classified his data accordingly, he was bound to use the best population estimates he had for that period. When considering the second *longitudinal* data set, which included follow up data collected later and has played little role in the discussion so far, Burt explicitly raised and dealt with this question (Burt, 1961, pp. 16–17).

> "The type of work has changed appreciably: the number of those engaged in manufacturing and in professional and administrative work of various kinds has increased: the number engaged in agriculture, in the extractive industries … in domestic work, and in the distributive trades has diminished; moreover, the amount of prestige attaching to different types of occupation has altered. Nevertheless, these further changes are hardly relevant to our present problem, as we have formulated it, although in a more intensive study the bearing of all the varying conditions I have mentioned should undoubtedly be systematically examined. …".

Burt also argued that (almost) any occupational classification would do for his purposes (Burt, 1961, p. 17, fn. 1):

> "It has been objected that any figure for social mobility … is bound to vary with the lines of division adopted in classifying occupations. However, as long as the basis of the classification remains unaltered, changes in the line of division will not seriously affect the estimated figure unless the lines of division become so few and the resulting classes so large that the amount of movement is obscured".

Table 8
Similarities when bands are combined.

IQ	50–70	70–90	90–103	103–115	115–127	127+
Fathers	24	229	333	258	122	34
Sons	24	229	333	258	122	34
Normal	22	230	326	262	121	36

Regardless of whether one agrees with these arguments or not, and Mackintosh clearly does not, it is impossible to maintain that Burt was trying to conceal these issues when he aired them himself.

Given the high rate of claim resurrection in the Burt Affair, it is worthwhile to examine, before closing, two arguments that were not offered by Mackintosh and Mascie-Taylor. The first of these is the original charge made by Michael McAskie (as repeated by the Clarkes) in 1974, that Burt's finding of an overall regression of 0.52 from the IQs of fathers to sons was "suspiciously perfect". It is sufficient to consider Burt's finding in the light of other research. In a comprehensive review of 111 studies of familial intelligence (excluding Burt's data), Bouchard and McGue (1981) estimated a weighted average correlation of 0.42 between the IQs of individual parent–offspring pairs reared together. In a meta-analysis of a subset of 45 of those studies, David Caruso (1983) derived a corrected correlation of 0.57, controlled for restriction of range, measurement error and other artifacts. Thus Burt's admittedly rough estimate of a correlation of 0.52 between parents and sons is squarely within the ranges reported in the published literature. McAskie's "suspicious" argument may therefore be discarded.

The second is a different argument for "fabrication" made by Dorfman that was not endorsed by Mascie-Taylor or Mackintosh. This claim involved the high correlation between the mean IQ scores of fathers and sons (0.9987, to be exact), and the appearance of linearity in the relationship between the two sets of scores. Dorfman based his argument on the graphical representation reproduced as Fig. 2 below, where the line shown is an equation he believes Burt used to "fabricate" the respective means, and is essentially the same as the fitted linear model between the means. This may *look* impressive, but it is not. Consider first the correlation between the means. If one fixes the means of the fathers, making only the assumption that the means of the sons would be ordered by class, with class VI having the lowest mean and class I the highest, and then makes 100,000 random draws of "means for the sons" from the normal **N**(100, 15) distribution, then the average correlation between the means of the fathers and the sons is approximately 0.95. In reality there are far tighter constraints on these sets of means than the simple ordering assumption made, since they have been weighted by simultaneously fitting to almost identical marginal totals, using the process described previously. Moreover, the averaging process removes variation that would otherwise attenuate the correlation.

From a different angle, suppose that each class mean of the sons is allowed to randomly regress from the class mean of the father by a coefficient of anywhere between 0.4 and 0.6 (Table 4 shows that Burt's class mean regressions fall within this range). Generously, let the random variation be uniform so that no particular value is favored. Then the average correlation of the class means over 100,000 draws is 0.9943. More generously still, if we let the regressions vary randomly but uniformly between 0.35 and 0.65, the average correlation over 100,000 iterations is still as high as 0.987.[8] Clearly the high correlation reported by Burt, on a data set with considerable regularity imposed on it, is not in itself persuasive.

[8] R code used in these simulations is available in the Supplementary materials for this paper.

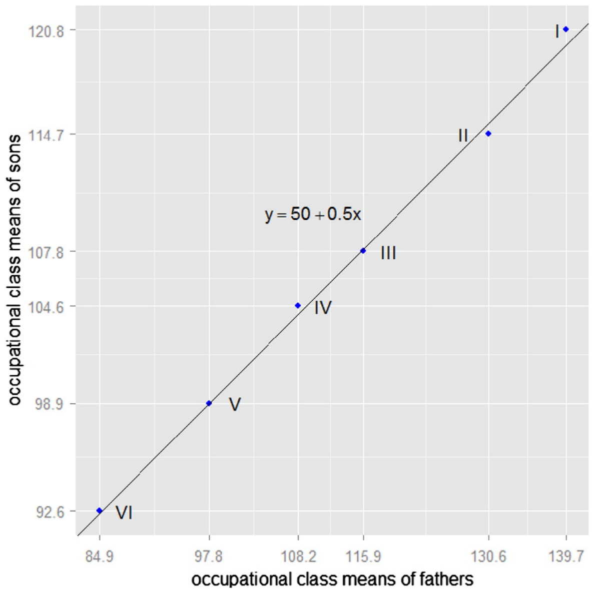

Fig. 2. Regression of class means.

To see why the graphical evidence is unpersuasive, consider the plot in Fig. 3 below, which is based on the procedures used above for simulating random associations between class means, with regressions ranging uniformly but randomly from 0.4 to 0.6.

The points have been superimposed to show the restricted freedom that any line through them could enjoy. Burt's values are included in blue. They show that linearity is more or less what we would expect a visual presentation to suggest. Moreover, Dorfman gives no reason why he believes Burt was interested in "fabricating" the means to achieve the linear agreement he identifies. This (approximate) linearity plays no role in Burt's argument, and is nowhere mentioned by him. This leaves us with no plausible motive on his part. Nor do the small deviations from linearity which we see in most of the class means make much sense from the point of view of fabrication. Burt would have to introduce these deviations deliberately, but again, to what end? Certainly not to hide apparent linearity, for that remains. There is in this, as there is in all of the arguments adduced from this data, only informal suggestion. Supposed "patterns" or "deviations" in the data are pointed out and the audience is invited to conclude that something is deeply significant in all of it. Since the audience typically does not have the expertise, opportunity or zeal to perform the sort of detailed examination required, their suggestibility is amplified into prejudice. This argument occurs in the absence of a sound theoretical framework for making formal deductions, with no plausible account of motives. Replication is more informative, which Burt (1961) explicitly called for from "fresh investigators" (p. 23).

Since Burt's paper a number of studies have confirmed his tentative findings, including Waller (1971) and Nettle (2003), using the more conventional notion of class in terms of status, rather than Burt's cognitive ability classification. Waller analyzed 1960s data from Minnesota with a sample size of 131 fathers and 173 sons, finding a significant correlation of 0.368 between father–son difference on IQ and father–son difference

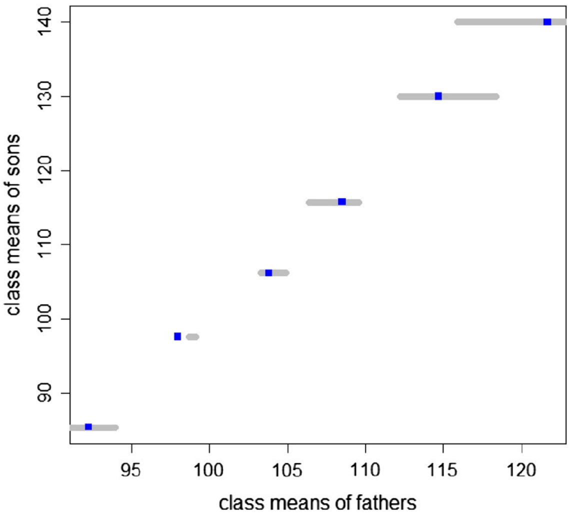

Fig. 3. Unsurprising linearity of the class means. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

on attained social class. Nettle, using longitudinal British National Child Development Study (NCDS) cohort data up to the year 2000, from 4529 fathers and 5038 sons, found comparable correlations between "General Ability" (GA, a close proxy for IQ) and "class trajectory" of the sons (up, down or stationary), within each parental class of origin The correlations varied between 0.31 and 0.39 by paternal class of origin, and the GA score of sons at age 11 was a stronger predictor of attained class at age 42 than parental class, which correlated at 0.26.

No persuasive arguments have thus been offered, statistical or empirical, that Burt's social mobility data was fabricated. Once his procedures are properly understood, the balance of the evidence lies heavily in Burt's favor.

Acknowledgments

The author would like to thank Prof. Thomas Bouchard Jr. for comments and encouragement based on a draft of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.intell.2014.12.002.

References

Banks, Charlotte (1979, Feb. 2). Burt's intelligence fraud. Letter to the Editor. *New Statesman and Society*, *150*.

Bishop, Yvonne M. M., Fienberg, Stephen, & Holland, Paul W. (1975). *Discrete multivariate analysis*. Cambridge, MA: MIT Press.

Bouchard, Thomas, & McGue, Matthew (1981). Familial studies of intelligence: A review. *Science*, *212*(4498), 1055–1059.

Burt, Cyril (1961). Intelligence and social mobility. *British Journal of Statistical Psychology*, *14*, 3–24.

Carr-Saunders, A. M., & Jones, David Caradog (1937). *A survey of social conditions in England and Wales* (2nd ed.). Oxford: Clarendon Press.

Caruso, David (1983). Sample differences in genetics and intelligence data: Sibling and parent–offspring studies. *Behavior Genetics*, *13*(5), 453–458.

Clarke, Alan D. B., & Clarke, Ann M. (1974). *Mental deficiency*. London: Methuen.

Clarke, Alan D. B., Clarke, Ann M., & Brown, R. I. (1960). Regression to the mean. *British Journal of Psychology*, *51*, 105–118.

Clarke, Ann M., Clarke, A. D. B., & McAskie, Michael (1976, 13 November). *Heredity and intelligence: Sir Cyril Burt's reputation. Letter to The Times.*

Clarke, Ann M., & McAskie, Michael (1976). Parent–offspring resemblances in intelligence: Theories and evidence. *British Journal of Psychology*, *71*, 172–173.

Deming, W. Edwards (1938). *Statistical adjustment of data*. New York: John Wiley (Dover reprint, 1964, of the 1943 edition).

Dorfman, Donald (1978, September 29). The Cyril Burt question: New findings. *Science*, *201*(4362), 1177–1186.

Dorfman, Donald (1979, April 20a). Burt's tables. *Science*, *204*(4390), 246–254.

Dorfman, Donald (1979, October 12b). Burt's data: Dorfman's analysis. *Science*, *206*(4415), 142–144.

Fletcher, Ronald (1991). *Science, ideology and the media: The Cyril Burt scandal*. New Brunswick, NJ: Transaction.

Gaw, Frances, Ramsey, Lettice, Smith, May, Spielman, Winifred, & Burt, Cyril (1926). *A study in vocational guidance*. London: H. M. Stationery Office.

Hearnshaw, Leslie (1979). *Cyril Burt psychologist*. London: Hodder and Stoughton.

Joynson, Robert (1989). *The Burt Affair*. London: Routledge.

Körner, T. W. (1989). *Fourier analysis*. London: Cambridge University Press.

Mackintosh, N. J. (Ed.). (1995). *Cyril Burt: Fraud or framed?*. London: Oxford University Press.

Mascie-Taylor, C. G. N. (1995). Intelligence and social mobility. In N. J. Mackintosh (Ed.), *Cyril Burt: Fraud or framed?*. London: Oxford University Press.

Nettle, Daniel (2003). Intelligence and class mobility in the British population. *British Journal of Psychology*, *94*(4), 551–561.

Pukelsheim, Friedrich, & Simeone, Bruno (2009). On the iterative proportional fitting procedure: Structure of accumulation points and L_1-error analysis. *Technical report*. Augsburg: Institut für Mathematik.

Rubin, Donald (1979, April 20). Burt's tables. *Science*, *204*(4390), 245–246.

Rubin, Donald, & Stigler, Stephen (1979, September 21). Dorfman's data analysis. *Science*, *205*(4412), 1204–1206.

Spielman, Winnifred, & Burt, Cyril (1926). The estimation of intelligence in vocational guidance. In F. Gaw (Eds.), *A study in vocational guidance* (pp. 12–17). London: H. M. Stationery Office.

Stigler, Stephen (1979, April 20). Burt's tables. *Science*, *204*(4390), 242–245.

Waller, Jerome H. (1971, September). Achievement and social mobility: Relationships among IQ score, education, and occupation in two generations. *Social Biology*, *18*(3), 252–259.