The Galton Laboratory,
Rothamsted Experimental Station,
Harpenden, Herts.

December 7, 1939

My dear Taylor,

I have been looking at the German data we got out the other day, and I am afraid the general conclusion must be that some of the German work suffers from error-like disturbances to a greater extent than our most alarming Sectors!

Before discussing them, I want to make a point which influences me, though I may not have emphasised it sufficiently in discussing these things with you; that is, that a mere failure to record a small percentage of A or B reactions will make very little difference to the observed discrepancy in the number of AB's, because this is based on a comparison of the AB's observed with those expected from the numbers of A's and B's found, so that, if we find too few A's we automatically expect too few AB's. There may, in fact, be an important error in the gene ratios estimated from such data, though scarcely any effect ~~from~~ the apparent deficiency of AB's.

I have just worked out an example to illustrate this. Maizels and Janet Vaughan talk of mis-classification in perhaps 2% of cases. Suppose we start with a theoretical population and misclassify a proportion by transferring, not 2%, but 6 1/4%, or 1 in 16 from A to O and from AB to B.

	O	A	B	AB
True frequencies	100	96	21	8
As classified	106	90	21.5	7.5

If from the apparent numbers of O, A and B we reconstruct the expected number AB, we find 7.379 expected against 7.5 observed, or a discrepancy of only 1.6%. Of course, for none of the data we have been discussing has a deficiency of 1% or 2% come even near to significance.

If we make the same test, misclassifying 1 in 16 of the B's, the result is: expected 7.475 against observed 7.5, a deficiency of only 1/3rd of 1%. We cannot, therefore, easily explain the difference between AB expected and observed, when these are at all large, by postulating even a gross failure of the reagents, supposing this to affect A's and AB's or B's and AB's equally.

You are, of course, on much stronger ground if you rely on the specific weakness of the A reaction in AB subjects, and on the worker, whether using weak reagents or not, being sufficiently inexperienced not to allow for the difference in strength to be expected.

What all this is the prelude to is that the German data show deviations in the number of AB's recorded which are extravagently great compared with anything we know how to account for, even on the lowest opinion of the technique they may use. For example, you may remember taking out a largish East Prussian investigation divided into four groups, I think according to the language, or perhaps the surnames of the people. This lot of 6,000 people in all gives the following comparison:

	Women	Men
No. expected	207.673	148.702
No. observed	276	239
difference	+ 68.327	+ 90.298
~~std. error~~ S.E.	17.934	15.084

The excess is more than 30% in one case, nearly four times the standard error, and over 60% in the other, about 6 times the standard error.

Two other lots, one from Serbia, in six sections, and one of Greeks go in the opposite direction, also with deviations so large compared with their standard errors that they are not due to random sampling.

I am really most puzzled as to what to make of these large discrepancies, and wonder if you see any possibility of explaining them by bad technique.

On the whole in the data we extracted, covering about 20,000, the errors, as I suppose them to be, of the different published collections largely compensate, and the sex difference is the reverse of what I expected on the basis of the English data, there being an apparent excess of men and deficiency of women. Still, nothing can be made of that unless we have, at least some idea of why the different series should show such large discrepancies.

Yours sincerely,

It was [illegible]