7th March 1934.

Whately Carrington Esq.,
Calandstraat 64,
Rotterdam,
HOLLAND.

Dear Whately Carrington,

On your first point my own logical position is that we can test any definite hypothesis, e.g. indentity of two personalities, in the sense that we can give the facts a chance of contradicting it if it is in fact untrue. This contradiction will never be absolute, but in the nature of a probability statement to the effect that some appropriate features of the facts would very rarely occur if the hypothesis were true.

Further, by continued experimentation we can increase ad lib the opportunity of disproving the hypothesis, and we like the hypothesis better and better the greater the number and the more relevant the quality of the facts which have failed to disprove it. But all this of course is not to say that the facts can ever prove the truth of the hypothesis, or even give in general a measure of the probability of its being true. All this so far is entirely free from the considerations which arise in statistical estimation.

One way of extending these notions in the way you want is to let the hypothesis have an arbitrary element, or mathematically, to let the probabilities of any assigned event be known only as a function of some unknown parameter. The test of significance will then usually give some such result as the hypothesis is disproved (P less than .01 say).for all values of the parameter lying outside certain limits, but is not disproved for values lying between them. A good deal now turns on the appropriateness of the test employed, but there is a limited class of cases admitting of what I call sufficient estimation where tests can be designed which are absolutely appropriate, or as Neyman and Peason put it, which are most powerful tests for discriminating every pair of the series of hypothesis considered. In these cases and where there is only one variable parameter it is possible to set up a probability statement about the value of the unknown parameter provided that, or assuming the mathematical form of the hypothesis is correct and when the observations are continuous variates, i.e. measurements rather than frequencies.

I am saying all this to make it clear that in passing from straight tests of fully specified a priori hypothesis to variable hypothesis and estimation, one is taking a biggish step into much more difficult country.

For example, the statement XY have p per cent in common opens up the question how are we to determine experimentally how much all men have in common. Would this be increased if the Fuegians or the Pygmies were to die out? Is this common part the same percentage of different individuals, or is it nearly 100% of sum and about 5% of others? In different fields, e.g. mathematics or pygnacity, one could imagine very different answers to these questions. All that is essential, however, is that our parameter should be something in terms of which the probability of any particular observation is expressible. This is analogous to the restriction favoured by some Physists that physical laws should be stated in terms of observable quantities only.

I think I understand what you have done to the Leonard complex, but I don't think I should say that P is the chance that the personalities are different when P is based on $\frac{W}{WP}$.

If on the series of words tested there was no correlation, or in other words, similarity between L and F, then P would not often be less than

.05 and the fact that it is less shows that L and F do respond to some extent similarly to the series of words chosen, but how similarly would they respond if they were identical? For this we need OPW to compare with WP. If WP were no bigger than OPW then the similarity would be as great as would be expected from identical personalities as shown by actually trying the same personality on different occasions, and there would be no evidence that the personalities would be different.

From the lower table on page 2, I should judge that F is rather aberrant from the other three, and J rather central, but I should like to see the test of significance first of the whole P x W with 3 multiplied by 74 degrees of freedom against O x P x W which for our purpose is error. This I think is the direct test for your question 3. One would then be in a position I think to judge whether such personality differences as are established could be expressed in terms less than 4 distinct personalities.

I do agree that the results are most exciting.

Yes I should like to meet you again on the 25th of April. Let me know about times etc.

(A) I have only worked the z distribution for the 5% and 1% points, as these are really sufficient for tests of significance. A rough idea of P beyond these may be obtained by extrapolating log P as though it were linear in Z. Pearson's incomplete Beta function tables are fuller, but would need some translations before use, so I don't advise trying them.

(B) When $\frac{W}{WP}$ is less than 1, Z is the negative. One can then reverse the numbers of degrees of freedom and take 1-P instead of P, but as I explained before such values ought not in reason to be significant, and if not the personalities show no similarity in their response to different words.

(C) Certainly both $\frac{W}{WP}$ and $\frac{WP}{OWP}$ might be significant. The latter would then show that there were really sceptical differences between the personalities, and this I think is the important test. The tests are quite independent and it is not possible to infer one from the other. I take it that the human race might be so constituted that we all had a good deal in common and $\frac{W}{WP}$ was always significant, but $\frac{WP}{OWP}$ might then be different for people of different temperaments, but not significant for identical twins.

Yours Sincerely,