13 June 1933.

Dr. Whately Carington,
Calandstraat 64,
ROTTERDAM.

Dear Dr. Whately Carington:

I find myself in entire agreement with the preamble, and have no doubt whatever that entirely valid and sensitive test of significance can be made. On a separate sheet I have outlined the arithmetical analysis, which has now become fairly familiar under the name of analysis of variance. Its principle is simple; your 600 readings are capable of yielding 599 "independent comparisons", most of which per se mean nothing, but that is all there is. The data are grouped so that like effects reinforce one another in every way that seems to have an experimental meaning, and the comparisons are seen to fall in 7 classes, with a definite number in each, these being now classes of comparisons of the same kind, and with its own and a different meaning. A large and compact class is "pure error", i.e. discrepancies wholly independent of the phenomena you are after; there are 360 degrees of freedom, or independent comparisons in this class, and these supply the basis for judging whether other comparisons are

numerically significant or not. In other words for calculating the probability that they should have occurred by chance, without differentiation of personalities, etc.

This probability, which you probably will not really want, though you will want to know whether it is very small or not, has no probable error, because it is not an estimate of anything. It is a function of your observations only, in relation to a particular hypothesis, that there is no differentiation between the personalities in their reaction to test. As in other cases the hypothesis is taken to be disproved if the probability is very small, but data exist only to disprove hypotheses, never to prove them.

I ~~am not~~ must also myself writing an essay on inductive reasoning, and only wish to offer to handle the data at any later stage, as the whole thing interests me greatly, if this seems of any use to you.

Yours sincerely,

There are ~~50~~ 25 sets of 6 readings for each of 4 personalities. For each word there are 6 x 4 readings, divisible into 4 sets of 6 each. Are they also divisible in 6 sets of 4 each? I.e. is there a correspondence between the first test for W and the first test for X? Are they taken on the same day or sitting? If not, should the fact of order be considered, as it should be if the reading may be ~~obtained~~ affected by the same word having been already used on previous occasions? I assume that the 6 sets of readings should be differentiated for some such reason.

If so there is a straight analysis of variance, which will include all the information derivable from correlational methods.

	Degrees of freedom	Sum of Squares	Mean Square
Personalities	3	$\frac{1}{6}\sum^{4}(A-\bar{A})^2$	
Occasions	5	$\frac{1}{4}\sum^{6}(B-\bar{B})^2$	
Remainder (error)	15		
Total	~~23~~	$\sum^{24}(x-\bar{x})^2$	

To the 23 degrees of freedom there corresponds the sum of squares of the deviations of all the 24 readings from their mean. If A is the sum of the 6 readings for any one personality, then the sum of the squares of the deviations of the 4 personality totals from their mean, divided by 6 will give the 3 degrees of

freedom for personalities. If B is the sum of the 4 readings for any one occasion, or ordinal number of test, we have similarly $\frac{1}{4}\sum_{1}^{6}(B-\bar{B})^2$ for occasions. The remainder is best obtained by subtraction from the total, and gives 15 degrees of freedom for estimate of error.

These 15 degrees of freedom may be regarded as the pooled inconsistencies of the 4 states, with 5 degrees of freedom each, deducting 5 for allowance for the differences between occasions.

Analysed in this way each word gives an independent test of personality, by comparing the mean square for 3 against 15 degrees of freedom. Or for all words 75 against 375, making a very precise test.

Taking now the totals of all 25 words, you have 24 totals, which may be analysed in exactly the same way as for each word, dividing each sum of squares by a new factor 25. The 3 degrees of freedom for personalities will measure the extent to which the average reaction to all the words differs in the different states; these will be part of the 75 degrees of freedom for individual words, leaving 72 degrees of freedom for the differential response of the different personalities to different words. Treating the other items in the same way you have the full analysis:-

	All words together		Differences among words	
Personalities	3	W x P	72	
Occasions	5	W x O	120	
O x P	15	W x O x P	360	
		W	24	
	23		576	599.

which together with the 24 degrees of freedom for words only (summing all personalities and occasions) accounts for all the 599 independent comparisons possible among your 600 readings.

You will want to see how all this is related to the calculation you had in mind. The chief bother with these is that they bring in the differences between different words from the first. Thus your consistencies, if applied to all 4 personalities, would really give W against w x O; while the resemblances are W against partly W xO and partly W x P. This last W x P is clearly of immense and direct importance, but you want to compare it with W x O x P to get its consistency on different occasions. P alone will also be important to you on measuring the differences between different personalities in general ~~routine~~ reaction, i.e. to all words. It may be compared with W x P to see if the differences are due to your particular list of words, and with O xP for consistency on different occasions.

You will see the analysis treats all personalities on exactly the same level, and this I think you will appreciate as an advantage.

With respect to probabilities these are obtainable from the ratios of the mean squares in each test, taking account of the degrees of freedom. It is not usually of any interest to evaluate low probabilities below .01, as this is enough to show that a real difference is there, and if so whether one gets 10^{-6} or 10^{-12} depends merely on how long one goes on trying. The tables for the z test have therefore been drawn up for P = .05 and P = .01, so as to make it easy to discriminate doubtful and definite significance.

I should not touch the method of ranks, but indeed using simple sums and sums of squares you can obviate all correlational notions altogether. If you have any doubt as to procedure, send me a few sets of results.