

AN ANALYSIS OF THE FACTORS IN A TYPICAL TEST OF INTROVERSION-EXTROVERSION¹

BY J. P. GUILFORD AND RUTH B. GUILFORD

UNIVERSITY OF NEBRASKA

IN a recent survey² of the literature upon the subject of introversion-extroversion it was found that there is considerable agreement among psychologists upon the existence of a distinction known as introversion-extroversion. Different authors in defining this alleged dimension of personality disagreed somewhat as to just where the emphasis should be placed. Some based the distinction almost entirely upon the direction of interest of the individual, whether predominantly outward upon the environment or inward upon the self. Others made the distinction mainly upon the way in which emotional reactions ran their usual course, whether freely expressed in overt action, or whether inhibited and shunted through implicit channels. Still others emphasized the social aspects of introversion-extroversion. Tests and rating scales have probably favored this aspect more than any other, attempting to detect whether or not the individual responds quickly and easily to social situations. The reader of the literature on introversion-extroversion is left without being enlightened as to whether these three aspects, intellectual, emotional, and social, really belong to one main dimension of personality, or whether introversion-extroversion is coincident with any one or two of them. To quote a part of the conclusions of Guilford and Braly,³ "We need very much to know whether there are such constellations of habits, tendencies or dispositions which can be called extroversion and introversion. A measurement of the frequency of coincidence of the various partials which have been suggested for these two constellations, those items of the Freyd list, for example, will help to determine whether we are talking about such entities as extroversion and introversion, and to decide just what they do include. The technique of Spearman for testing for general, group, and specific factors may be applicable in this case, and may solve the riddle of personality traits in general. Having established the reality of such traits as extroversion and introversion, we are ready to look for simple objective tests and for some physiological basis for them."

From a practical standpoint, also, it is absolutely imperative that we understand the make-up of introversion-extroversion before we can secure a valid measure of it. In some brief studies in which it was essential to obtain some kind of rating of a number of subjects for introversion-extroversion, the present tests and rating scales proved to be almost worthless.⁴ The three tests that were used, the Laird, the Marston, and the Neymann-Kohlstedt, did not correlate very highly with one another.⁵ And when the three were correlated with more objective measurements such as rate of fluctuation of an outline cube and height of the patellar reflex, some coefficients were negative and some were positive. The two studies just referred to were inspired by a desire to test experimentally McDougall's theory of introversion-extroversion. The authors of the second study were forced to the conclusion that,⁶ "One thing is certain, a better agreement upon the description and the rating of introversion-extroversion will have to be reached before any adequate proof or disproof of McDougall's theory, or any physiological theory, can be made." Such a state of affairs demands that the whole concept of the aspect of personality here under question should be carefully re-examined and if possible empirically justified. That is what we hope to accomplish in this and in a following article.

In the first article, we shall take what we regard as a typical test or rating scale of introversion-extroversion of 36 items, and subject it to two different procedures of statistical analysis. Spearman's technique,⁷ or rather Dodd's version of it,⁸ will be applied first to see whether there is a general factor running throughout the entire list of 36 items and whether there are any 'group' factors in addition which need to be assumed in order to account for all the correlations between items. Since group factors were actually indicated by this technique, we then applied Thurstone's method of multiple factor analysis⁹ in order to determine how extensive these factors are and how many there must be to account for the correlations. In the second article, we shall apply Thurstone's method of similar reactions¹⁰ to the same data. Since Spearman's technique did indicate a fairly widespread 'g' factor in the test, involving all but a very few of the items, we assumed that the items could be allocated along a continuum, which by the common assent of a number of authors who have contributed the test items, may be called the continuum of introversion-extroversion. The scale values of the items as determined by the method of similar reactions, would serve as weightings for any future reactions to the test items. It was

thought that in this manner an empirically accurate measurement of introversion-extroversion might be set up. It was possible, of course, to find similar loadings for the test items by carrying out Spearman's technique. This was done, and a comparison of the two sets of weights, those from Spearman's method and those from Thurstone's method, should be of interest from a statistical point of view. The one set of weights also serves as a check upon the validity of the other.

In our choice of a 'typical test' of introversion-extroversion, we did not use any one of the several now in general use. We chose rather to devise one of the usual Yes-No type that should be based upon the universally accepted descriptions of introversion-extroversion now available. First, we combed the writings of Jung¹¹ in order to find his own descriptive phrases. We used the Freyd list,¹² itself a kind of symposium of descriptive material. Of the tests now in use, we considered the Laird,¹³ the Marston,¹⁴ the Neymann-Kohlstedt,¹⁵ and the Northwestern.¹⁶ There are other tests, but they are almost always variants of the Freyd list or of the Laird test. Altogether we found 75 un-repeated items of description, each one supposed to be diagnostic by at least one of the above writers. Fifty-six of them were considered diagnostic by two or more authors. Only two of the items were given by all six authors, five of them were given by five, thirteen were given by four, sixteen were given by three, and twenty were given by two different authors. This gives some idea of the agreement and lack of agreement, although it may be supposed that some of these writers would accept descriptive items not included in their tests, and it is true that some of the items so included are mentioned in varied form a number of times. From this list of 75 items, we selected 35 as representative. We attempted to select those items upon which there was most agreement, and those which appeared to duplicate each other a minimum amount. As it happened, all of the items so selected were mentioned by at least three of the six authors.

The question might be raised as to how we know that any 'g' factor found in these 35 items would be the introversion-extroversion that we are talking about. The only criterion we have is the common judgment of those who have tried to define and to describe this very thing. So far as we know there is no better court of appeal at the present time. There are, however, two other possible criteria, both more objectively defined than the one we have used. One is the McDougall test, the rate of fluctuation of an ambiguous figure, and the other is the use of pathological

subjects. We rejected the former because, although we may be inclined to give some credence to McDougall's physiological theory, it has been shown in several previous experiments that a pool of test items similar to the ones given in the above test may correlate very poorly with the McDougall test. Until McDougall's notion of the nature of introversion-extroversion agrees better with that of a half dozen other writers, we are hardly safe in using his test as a criterion.

There has been some tendency recently¹⁷ to assume that dementia præcox patients are extreme introverts and that manic depressives are extreme extroverts, and to use the reactions of these two psychotic groups as criteria of this dimension of personality. We believe that without further examination of the matter this assumption is not warranted. The depressed type of insanity surely shows more of the so-called introverted traits than extroverted. At least, if one does not wish to prejudice the question as to which of the diagnostic traits are introverted and which extroverted at the start, one can find numerous examples of opposition between the manic and the depressed upon the items of the usual test. One can find, also, cases of paranoid dementia and even catatonic that exhibit glaringly opposite characteristics to the simple or hebephrenic dementia præcox. We also know, from personal experience in giving such tests orally to a large number of psychopathic patients, how little dependence can be put upon their answers, even in the most seemingly normal cases. For one thing, as anyone who has administered association tests to such patients will testify, a verbal stimulus does not have the same meaning to the average psychopathic patient that it has to the average individual who is mentally healthy. If these were the only sources of error, perhaps the testing of psychopaths with introvert-extrovert items would not be entirely worthless. But what is an even more serious matter, the cultural level of the average patient in a public hospital is so remote from that of the college student, for whom the typical test is devised, that he cannot interpret all questions in the same way. His answers "Yes" and "No" may be promptly given, but additional spontaneous remarks or responses to further questions of the examiner reveal that the subject may have entirely missed the point.

We also question the main assumption that the two large groups of the insane exhibit extremes of introversion-extroversion. Hunt and Guilford administered the fluctuating cube test to a number of both groups.¹⁸ According to McDougall's theory and the above assumption, the dementia præcox patients should exhibit a rate of

fluctuation more rapid on the average than that for normal individuals and the manic-depressive group should show a markedly slower rate than normal. The latter was found to be true, with little doubt. But the average for the dementia præcox group was almost identical with that for non-pathological subjects. The paranoid dement group exhibited extreme variability, from one individual to another and in the same individual at different times. If there is any connection between the rate of fluctuation of an autoline vube and introversion-extroversion, then it is going too far to assume at the present time that the dementia præcox and manic depressive groups give us two clear cut extremes of the scale. Any test that is validated on this assumption might be regarded as a differential diagnostic test for these two pathological groups, but it may be nothing more than that. One is not justified, without further proof to support the assumption, in calling such a scale of items a test of introversion-extroversion.

The 35 items that we chose for our test are given below. The numbers in parentheses at the right represent the number of authors, among the six from whom the items were selected, who mentioned the items as diagnostic. The wording of the items was given much consideration, guided by a few general principles. We attempted to let an affirmative answer indicate introversion as often as it did extroversion. Whenever possible, the questions were framed so that either an affirmative or a negative answer would be equally desirable ethically and personally. It was attempted to frame as many as possible in the form of "Do you like —?", assuming that it is easier for an individual to judge what he likes or dislikes than to judge what his behavior has been, and also easier for him to be more honest about it. Other statements, where possible, began, "Are you inclined to —?", so that no one would be forced apparently into making a more extreme judgment than he cared to. It is assumed in this study that with few possible exceptions the questions are not completely disjunctive; that probably a large group of individuals would distribute normally over the entire Yes-No range for any one question. The milder judgment, "Are you inclined to —?", permits an either-or response from those who are nearer the ends. The instructions that were given to the subjects are also reproduced here.

INSTRUCTIONS

Below you will find 36 questions which are to be answered either "yes" or "no". Read each question in turn. Think what your behavior has usually been and underline either "yes" or "no", whichever answer describes your behavior better. If you

cannot decide, then guess. *Be sure to answer every question.* There is no implication of right or wrong in any of these items.

1. Yes No Do you express yourself better in speech than in writing? (3)
2. Yes No Are you inclined to limit your acquaintances to a select few? (3)
3. Yes No Do you generally prefer to take the lead in group activities? (4)
4. Yes No Do you prefer to read about a thing rather than experience it? (4)
5. Yes No Do you like work which requires considerable attention to details? (5)
6. Yes No Are you generally very particular about your personal property, *i.e.*, do you take very good care of your things? (3)
7. Yes No Are you inclined to be considerate of other people's feelings? (4)
8. Yes No Are you inclined to act on the spur of the moment without thinking things over? (4)
9. Yes No Have you ever kept a personal diary of your own accord? (3)
10. Yes No Do you work much better when you are praised? (4)
11. Yes No Do you like to change from one type of work to another frequently? (6)
12. Yes No Are you inclined to study the motives of others? (4)
13. Yes No Do you day-dream frequently? (4)
14. Yes No Do you prefer to work with others rather than alone? (5)
15. Yes No Are you inclined to worry over possible misfortunes? (4)
16. Yes No Are you frequently somewhat absent-minded? (4)
17. Yes No Do you like to persuade others to your point of view? (3)
18. Yes No Are you inclined to keep in the background on social occasions? (4)
19. Yes No Are you more interested in athletics than in intellectual things? (3)
20. Yes No Do you usually dislike to change opinions you have already formed? (5)
21. Yes No Do you like to speak in public? (5)
22. Yes No Do you prefer to work things out for yourself rather than accept suggestions from others? (4)
23. Yes No Do you have frequent ups and downs in mood, either with or without apparent cause? (3)
24. Yes No Are you inclined to be slow and deliberate in movement? (4)
25. Yes No Are your feelings rather easily hurt? (4)
26. Yes No Do you enjoy getting acquainted with most people? (3)
27. Yes No Are you inclined to keep quiet when out in company? (6)
28. Yes No Do you adapt yourself easily to new conditions, *i.e.*, to new environments, situations, places, etc.? (3)
29. Yes No Do you like to confide in others? (3)
30. Yes No Do you express such emotions as delight, sorrow, anger, etc., readily? (4)
31. Yes No Are you inclined to think about yourself much of the time? (3)
32. Yes No Do you like to have people watch you when you are working? (3)
33. Yes No Do you frequently rewrite social letters before mailing them? (3)
34. Yes No Do you like to sell things? (3)
35. Yes No Do you get rattled easily in exciting situations? (3)
36. Yes No Are you a male? (3)

The subjects were practically all undergraduates, distributed among the four classes. There were 930 altogether, including 430 men and 500 women. The test was given a second time, one month later, to 277 of the same subjects, 163 men and 114 women. This was for the sake of determining something about the reliability of each item and of the test as a whole when the items were later properly weighted. Realizing that there might be sex differences in the proportion of subjects responding "Yes" or "No" to each question, and that the correlation between each item and every other one might be influenced by such sex differences, we had to take account of sex in some manner. To carry through all the

computations for the two sexes separately made an enormous task doubly enormous. We let sex be merely an additional item in the test and correlated sex with everything else. This is item number 36. Other writers have found that men are inclined a little toward extroversion and women toward introversion. We can therefore treat sex as one more diagnostic item along with the rest. For any given correlation between two items, then, if there is a significant sex difference involved one can simply partial out the factor of sex.

It is too sanguine for anyone to hope that a "Yes" or "No" response given by an individual to any one of these questions really measures that individual's position on the scale of that question. This would assume that each subject was a good judge of his own behavior, that he was not self deceived due to any motive of self defense or of wish fulfillment, and further that if these two things were true he would tell the truth about himself. We can only hope that in the long run his reactions do come in the neighborhood of the truth about himself. With no pretense of any knowledge as to how truly each reaction measures an individual, we may merely assume that we are applying 36 verbal stimuli to 930 subjects and that to each stimulus one of two reactions will occur. In correlating any item, then, with any other one, we are merely finding the likelihood of a similar response to those two items. But we can take at least one more logical step. Having found that two items bring a similar reaction in a large proportion of the cases, we may assume that the responses to those two items arise from similar properties or an identical property of the individual, and hence they may be said to measure the same variable in the same direction. Conversely, if two items bring forth opposite reactions in a large proportion of the cases, if they are negatively correlated, in other words, then it may be said that they measure the same property but in opposite directions. In either case, nothing may be concluded as to the property being measured; merely that two stimuli have indicated a single property of the individual. In the final analysis, this is probably the only way in which a test item of any sort may be said to measure. If, therefore, we speak of any item in the test as measuring introversion-extroversion, it is for the sake of labeling a property of personality in order to facilitate discussion.

The first step in the direction of finding the intercorrelations between the 36 items was to tabulate every pair of items in a four-fold table. For example, if we want to find r_{12} , the correlation between items 1 and 2, we make the four following tabulations:

Let a_{12} = the number of cases responding "Yes" to both items
 b_{12} = the number of cases responding "Yes" to item 1
 and "No" to item 2
 c_{12} = the number of cases responding "No" to item 1 and
 "Yes" to item 2
 d_{12} = the number of cases responding "No" to both items

$$p_{12} = a_{12} + b_{12}$$

$$q_{12} = c_{12} + d_{12}$$

$$p'_{12} = a_{12} + c_{12}$$

$$q'_{12} = b_{12} + d_{12}$$

From such a four-fold table one might either compute a tetrachoric coefficient or a coefficient of contingency. The latter, with two corrections to be mentioned later, is said to yield a value that is equivalent to the Pearson r .¹⁹ This seemed the more expedient thing to do. According to Kelley, the coefficient of contingency can be obtained directly from ϕ , the product moment correlation between two point distributions, when there is a four-fold table. ϕ is found from the four-fold table by means of the formula,²⁰

$$\phi = \frac{ad - bc}{\sqrt{p \ q \ p' \ q'}}$$

in which the notations are the same as those given above. But this formula assumes that the measurements are in the form of non-continuous or point varieties, as, for example, classes of eye color. We have assumed some kind of a continuous distribution for the responses to the items in the test. The responses are merely forced into two categories by the conditions of the test. Kelley gives two corrections by which the proper value of C , the coefficient of contingency may be obtained from ϕ . The first is the correction in ϕ itself for the number of cells used. With a few transformations, which we will not take the space to present here, the corrected ϕ for the four-fold table reduces to,

$$c\phi^2 = \phi^2 - \frac{1}{N}$$

C can then be found by means of the formula,²¹

$$C = \sqrt{\frac{\phi^2}{1 + \phi^2}}$$

This coefficient needs the further correction for class index. The general formula is,²²

$$mC = \frac{C}{\frac{r_{x\chi}}{r_{y\gamma}}}$$

Assuming that our distributions are normal, as most human traits are, the value of r_{χ} and of r_{γ} for four-fold table both equal .798,²³ and the product of these is .637. It was a simple matter, therefore, to divide each obtained C by this common factor. There are two items in the list, however, in which a distribution of any kind is very doubtful. Item 36, for males or females, is obviously more like a point designation, although there are some who claim that the two sexes are not at all pure types and that they overlap in their make-up considerably. Item 9 (keeping a diary) is doubtful. It would seem at first thought that a sharp line of demarcation could be drawn between those who have and those who have not kept diaries. The qualification was added to the question, "of your own accord", however, and this might have opened up the possibility of expressing different degrees of willingness to keep a diary. Being in doubt, we retained the assumption of a point distribution for item 9. As it turned out, almost all correlations with 9 were zero so that no amount of correction would have made any difference for this item in most of its correlations. All the other items, aside from 9 and 36, were given the correction mentioned above.

The 630 coefficients, which are equivalent to Pearson r 's are not presented, since the cost of their publication would not be warranted. The reader who is interested in the individual coefficients may secure them from the writers by request. None of the coefficients are very large, the range being limited, with two exceptions, between $-.50$ and $+.50$. The majority of them look to be insignificantly small. Yet, due to the large number of subjects (930), the coefficients are probably very reliably determined. If they were ordinary Pearson coefficients, the probable errors would range between .016 and .022. The limits of the correlations obtained under the conditions of our test are therefore rather narrowly defined, and we have taken any coefficient greater than $\pm .10$ as being rather indicative of the true amount of relationship between that pair of items.

In order to see whether the correlations can be accounted for by assuming only a 'g' factor and 's' factors, we did not attempt to use Spearman's tetrad differences. With 630 coefficients, the number of tetrad differences would be enormous and the task prohibitive. We have preferred to depend upon Dodd's coefficient of equiproportion to make the same test of the table of correlations. The latter process involves first the determination of the correlation of each item with an assumed 'g' factor, (r_{ag}), Spear-

man has given us the formula for this.²⁴ The r_{ag} coefficients for our test items will be found in the second row of Table I. These coefficients are strictly correct only when there are no group factors in the table of items. But they are approximately correct when the influence of group factors is relatively small. We are safe in assuming that this is true in these data, since all the raw coefficients are low.

TABLE I

Item Number	1	2	3	4	5	6	7	8
Coefficient of Reliability	.82	.73	.92	.72	.84	.85	.87	.80
Correlation with the g-factor	+ .26	— .50	+ .44	— .33	.00	.00	.00	+ .11
Weight of the g-factor	+ .27	— .67	+ .54	— .37	.00	.00	.00	+ .11
Sum of Corrected Correlations	4 26	6 25	5 35	4 51	2 51	3 41	3 64	4 99

TABLE I—Continued

9	10	11	12	13	14	15	16	17	18
1.00	.80	.77	.81	.76	.88	.76	.78	.90	.75
.00	— .07	.00	+ .09	— .20	+ .27	— .35	— .35	+ .22	— .51
.00	— .07	.00	+ .09	— .22	+ .29	— .39	— .39	+ .23	— .69
2.40	3 27	4 07	2.88	4 11	4 16	5 56	4 94	4 44	6 34

TABLE I—Continued

19	20	21	22	23	24	25	26	27	28
.80	.51	.93	.65	.80	.94	.84	.89	.87	.81
+ .07	— .13	+ .36	— .03	— .20	— .27	— .32	+ .43	— .47	+ .39
+ .07	— .13	+ .41	— .03	— .21	— .31	— .35	+ .53	— .60	+ .46
4.10	4.11	5 06	4.13	5 24	5 12	5 48	5 44	5 91	5 40

TABLE I—Concluded

29	30	31	32	33	34	35	36
.75	.71	.61	.71	.90	.91	.80	1.00
+ .12	.00	— .19	+ .26	— .08	+ .34	— .41	+ .21
+ .12	.00	— .20	+ .28	— .08	+ .38	— .49	+ .22
4.42	6.05	4.98	4.07	2.41	4.21	6.07	4.14

Knowing all the correlations with the 'g' factor, it is possible to partial out the effect of this factor in the correlation between every pair of items. This yields the residual or 'specific correlation' over and above that caused by 'g'. Like the values for r_{ag} , they, also, are approximations, but they are probably nearer the truth than the r_{ag} values. The mean of these specific correlations, disregarding signs, is .137. The probable error of these coefficients has a maximum value of .022, obtained from the formula, $.6745/\sqrt{N}$.²⁵ This is the probable error when r_{12g} has a value of zero. The mean is 6.2 times the probable error of these coefficients, which indicates that the specific correlations are real, and that the table of items contains group factors.

Before we attempt to see what these group factors are and through which items they extend, let us try to guess something more definite and perhaps more useful about the assumed 'g' factor. Two of the correlations with 'g' are as high as .50, and many are above .30. Six of them are zero and therefore the items involved cannot possess the 'g' factor. The average specific correlation is very small, indeed not so far from zero, (.137), so that the error in the approximation of r_{ag} should be practically negligible. We can learn at least which ones are negative and which ones are positive, and therefore to which end of the introversion-extroversion scale each affirmative answer belongs. We should be able, then, to see whether the *a priori* judgments of the writers on the subject have been correct. We should be able, furthermore, to compute an approximate weight for each item and to see which items are most diagnostic. Most tests heretofore have merely checked a subject's reaction to an item as indicating introversion or extroversion. All items have been weighted equally. An important refinement in scoring such tests would consist of a differential weighting of the "Yes" and "No" answers according as they reveal a certain amount of the 'g' factor in the test.

We have therefore computed the weights of the items according to another of Spearman's formulæ.²⁶ These weights are given with their proper signs, in the third row of Table 1, Fig. 1 illustrates in graphic form the allocation of each item along the introversion-extroversion continuum. These weights are assigned to the "Yes" answers to the 36 items only. To the "No" answers we may simply reverse the signs of the weights. And herein lie two new assumptions. One is that the distributions of individuals upon the different items are equal in range. The other is that the distributions are not skewed. No one knows whether or not these assumptions are correct. But we insist that it is better to make guesses based upon recognized assumptions than to proceed as has often been done in the past by making guesses based upon ignorance alone.

One can see in Fig. 1 that items 18 (inclined to keep in the background on social occasions), 2 (inclined to limit acquaintances to a select few), 27 (inclined to keep quiet when out in company) and 35 (gets rattled easily in exciting situations) indicate most strongly introverted tendencies. The social aspect, which we have said before tends to dominate all the tests, comes out with strongest diagnostic value. It may be that the frequency with which this aspect obviously appears among the different items actually favors this result. As a group these social items would tend to fortify

one another, constituting the most general group factor in the test. This would seem to be borne out at the other end of the scale where we have indicators of social aggressiveness. Here we find items 26 (enjoys getting acquainted with most people), 3 (prefers to lead in group activities), 28 (adapts readily to new conditions), 21 (likes to speak in public), and 34 (likes to sell things) leading the list. But on the other hand, items 28 and 35, which appear among the leaders at both ends, have no necessary social implications. And following closely behind at either end of the scale are items also without any apparent social factor. If these

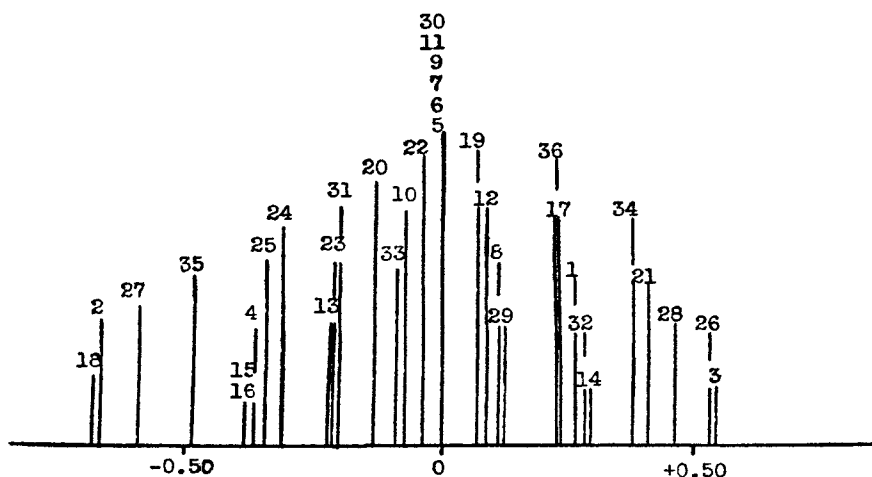


FIG. 1. Showing to what extent each item is weighted with the g-factor

latter items did not partake of the same 'g' factor that is found in the more extreme social items, their weights would approach zero.

The few items which do receive weights of zero are worthy of note. They are 5 (liking work involving attention to details), 6 (being particular about one's personal property), 7 (being considerate of other people's feelings), 9 (keeping a diary), 11 (liking to change type of work frequently), and 30 (expressing emotions readily). And yet these have been rather favorite items in the usual test. It may be, of course, that the Yes-No type of test does not yield valid results on such questions as these. But from a practical standpoint of measurement of personality, such items yield nothing, apparently, in the way of diagnosing introversion-extroversion.

Are there any items that yield an opposite diagnostic reaction to that generally accepted? The results are gratifying to the arm-

chair psychologist in this respect. With the exception of those items mentioned above that are seemingly not diagnostic at all, and possibly one more item which may be reversed to the expectation of some writers, there is remarkable agreement between theory and fact. The doubtful exception is item 23 (having frequent ups and downs in mood). Those who hold that manic depressives are extreme extroverts might expect an affirmative answer to item 23 to indicate extroversion. It is on the introvert side in Fig. 1. It may be that the qualification 'frequent' is the key to the situation. The introverted individual, suppressing overt emotional reactions, is more easily thrown into a continued emotional state or mood than is an extroverted individual. But this would necessitate a distinction in our description of all alternations of mood; between those of short duration, the transitory, and those of longer duration, the more permanent.

The loadings of the different items as represented in Fig 1 might very well be used in scoring the test as a whole for the 'g' factor in it. The validity of test scores so obtained can be estimated by finding the correlation between the best weighted pool of items and the 'g' factor. When this is computed according to Spearman's formula, we get an r_{pg} equal to .871. This is far below the standard set by Holzinger²⁷ for the correlation between a 'g' factor and the best weighted pool of tests, but as compared with the usual multiple coefficients of correlation, or with coefficients of validity found in practice, the value is not so low. It is undoubtedly higher than the similar measure of validity that would be found for tests of introversion-extroversion in which the test items are unweighted in making up the pool.

In order to see how the weighting of test items would work out in practice, we selected some test papers at random from those subjects who had taken the test twice. Among these subjects there were 60 men and 60 women. The loadings as given in Fig. 1 apply only to the "Yes" answers. For the "No" answers we simply reversed the signs. Then, in order to eliminate negative scores, we added a constant number to all the loadings. The loadings then ranged from 0 to 1.4, or by elimination of decimals, from 0 to 14. The total scores for the 60 men ranged from 157 to 285, with a mean of 213. The total scores for the 60 women ranged from 133 to 265, with a mean of 211. The coefficients of reliability, test-retest, were .767 for the men and .869 for the women. When both men and women were combined, the coefficient of reliability was .810. Tryon has shown²⁸ that the coefficient of reliability measures directly the percentage of the test measurement that is produced

by the true thing we are attempting to measure. The actual correlation between the thing we are trying to measure, in this case the posited 'g' factor, and the test scores is equal to the square root of the coefficient of reliability.²⁹ For the two sexes combined, this correlation would be exactly .90. This should be equivalent to the coefficient r_{pg} which was obtained in another manner above. As a matter of fact, the two are very close together. We would have predicted a value lower than the obtained .87 because the correlations with 'g' were probably spuriously higher than the truth. From the self correlation of the test we find that r_{pg} is slightly higher than .871. This might be taken as reassuring the validity of the computation of r_{pg} with Spearman's method. On the other hand, the coefficient of reliability was found by correlating test scores with retest scores. Although a month elapsed between the two trials, memory and mental set for self consistency on the part of the subjects might have played some part in raising the coefficient of reliability. This factor which works toward an apparently increased reliability is offset by another which tends to lower it. After a month's time the same Ss are often tested in a quite different mood and certain more permanent changes in personality may also have occurred.

If the weighted test scores do actually measure from 75 to 85 per cent introversion-extroversion, as obtained from the coefficient of reliability, and the remaining percentage something else, the test would undoubtedly be more valid than the majority of personality tests, and even better than many tests of intelligence. But there is a feasible method of improving upon this. The lack of perfect reliability of the test as a whole is caused by the unreliability of the separate items themselves. It is possible for the same subject to change his answer a second time, even to those questions that are most heavily weighted. Such an event makes a sizable difference in a final test score. It will pay us, therefore, to study the reliabilities of the separate test items.

Taking the responses of the 277 subjects who had been given the test twice, and tabulating the four-fold tables just as we did for the intercorrelation of test items pair by pair, we then found the coefficients of contingency with the same necessary corrections. These coefficients will be found in row one of Table I. They range from .514 to 1.00. Only three, however, are below .70, and seven more are below .80. The reliability of the total test, apparently, is no higher than that for the average test item. There is only one practical way out of this difficulty, and that is to increase the number of items in the test, preferably by repeating the better,

most highly weighted items, a number of times in slightly varied form. In a psychophysical experiment, in which two alternative responses are permitted to a given stimulus, one would never think of stopping after only one application of the stimulus. There are good reasons, of course, why a repetition of the same stimulus here would not have the beneficial effect that it has in the psychophysical experiment. But a lengthening of the test, such as has been done in the Northwestern University test, by giving the same item over again as many as five times, should raise the reliability considerably. And if the items, in their various forms, are all properly weighted, this should have the effect of raising the values of r_{pg} to a really acceptable degree.

Let us now turn to the question of group factors. The large number of specific correlations indicate that there are some. But one can hardly guess merely from an examination of these specific coefficients just how many there are nor just how extensive each one is. Since the so-called 'g' factor was found to have no weight in six of the items, it might be more appropriate to call it a broad group factor rather than a 'g' factor in the sense of a universal. It is, however, the strongest of the factors present, whether regarded as a universal or merely as a group factor. The other factors that exist in the table must be less important. Spearman's methods do not enable us readily to go on and to analyze the table of correlations further. Certain methods might be devised for doing this. We might repeat with the specific correlations the process which was applied to the raw coefficients, that is, assume a new 'g' factor and find correlations of items with it. The next most extensive factor would thus be revealed. New specific correlations could be determined with both the first two group factors held constant, and so on, repeating the process until all significant group factors had been found. The process would be a kind of 'fractional distillation'. We have not attempted to carry this out because the procedure lacks mathematical proof and because when there are specific correlations involved, the correlations with the general factors are only approximated, and so, also, are the specific correlations. If this error could be avoided, the method would seem worth trying.

Thurstone's method of factor analysis as first introduced* enables us to survey systematically a table of raw coefficients in order to detect all the group factors present, no matter how

* Since this paper was written Thurstone has made considerable progress in developing his method and in simplifying his procedures. The fundamental theory, however, remains essentially the same.

general or how limited in extent. It is this procedure that we employ next. When the method is followed out in full it is possible to weight every variable for each general factor that it measures. We have not taken the trouble to go this far with the data at hand at the present time. This will be done at a later time after the method has been tested further and the work of computation has been simplified. It is sufficient for our purposes here to be able to estimate the number of group factors, to determine which test items contain them, and to guess something about the nature of the individual factors. At the beginning of this report, it was suggested that the usual test items tend to cluster about three important aspects of introversion-extroversion, the intellectual, the emotional and the social. Will a correlational analysis of the test items reveal any such broad group factors?

Thurstone's method is most accurately applied when all coefficients are corrected for attenuation. The method assumes, as originally proposed, that the self-correlations are 1.00. All inter-correlations, to be made comparable with this assumption must therefore be corrected for attenuation. Our 630 coefficients were so corrected. We do not present these corrected coefficients, but we do give in Table I, row 4, the absolute sum of the coefficients for every item. These are important in determining the items most heavily 'loaded' with the various factors.

Item 18 has the largest sum, disregarding signs, and therefore item 18 furnishes the reference or criterion for all those items containing the most important group factor which we shall call factor 'a'. The next step involves the selection of all those items that correlate positively with item 18. There are 12 such items, including 2, 4, 13, 15, 16, 20, 22, 23, 24, 27, 31, and 35. In scrutinizing this list in the test, it is not easy to select a name which designates this group. Judging from item 18 and from those items that correlate most strongly with it, we might say that it is a tendency to fear the environment, to shrink away from it and from the necessity of acting to it, especially the social environment. But this is a rather acceptable definition of introversion itself. Indeed, it will be seen in Fig. 1 that item 18 is the most introvert of them all, and prominent in the partial list of items given above are items that follow closely behind item 18 in Fig. 1. None of the items on the extrovert side in Fig. 1 are represented in the list. Only three items on the introvert side in Fig. 1 are not included. These three items are: 10 (works better when praised), 25 (feelings easily hurt), and 33 (rewrites social letters). These three fall

on the introvert side in Fig. 1 probably because they correlate positively with other items than 18 that are on the same side.

But what about those items that correlate negatively with item 18? The correlation of such items with number 18 would have been positive if we had merely changed the sign of the verbal responses of our subjects. Of the items on the extrovert side in Fig. 1, the following correlate negatively with number 18: 1, 3, 8, 12, 14, 17, 21, 26, 28, 29, 30, and 34. Again three items are conspicuous by their absence. They are: 19 (more interested in athletics than in intellectual things), 32 (likes being watched while at work), and 36 (male). It is worth noting that there is no sex difference upon item 18. If this item is the keystone of introversion-extroversion, then there is no significant difference between the two sexes in this important dimension of personality. However, sex does correlate with other items which belong on the negative or positive side in Fig. 1 so that it must carry some weight for factor *a*. It is quite significant that those items which come at the zero point on the scale in Fig. 1, *i.e.*, 5, 6, 7, 9, and 11, do not correlate with item 18 at all. Since they correlate neither with item 18 nor with the pool of tests of which item 18 is the keystone, they are probably of no diagnostic value for factor *a*.

What, then, of the six test items that correlate with the pool and that do not correlate with item 18? We refer to items 10, 25, 33, and 19, 32, and 36. Should these be eliminated from a test of introversion-extroversion? They have substantial weights as illustrated in Fig. 1. According to Hull, a test which correlates zero with the criterion may still be diagnostic if it correlates highly with other tests that do correlate with the criterion. One could not say that any of the other items correlate *highly* with any criterion in this case, but relatively speaking, the six items just referred to have significant correlations with items that do correlate with the criterion. They may therefore be retained as diagnostic.

The next step, according to Thurstone's procedure, is to eliminate from further consideration all tests that correlate higher than .50, positively or negatively, with item 18. Now there is only one such item in the whole list, namely, number 28 (adapts readily to new situations). If this rule were followed throughout the remainder of the process, we would find as many group factors, lacking only two, as there are items. This may represent the actual state of affairs. But the limits of $\pm .50$ seem to have been arbitrarily chosen by Thurstone. There are many tables of intercorrelations with few coefficients higher than $\pm .50$. In

this study we have lowered those limits to $\pm .30$. This reduces the number of group factors that need to be postulated, but it raises the question as to just how many group factors there actually are among these 36 items. By lowering the limits still more, we could reduce the number of postulated factors again. Using the limits of $\pm .30$, as we did, there proved to be no less than 18 group factors, including the first one which has already been discussed, namely factor *a*.

Factor *b*, the next most extensive one, had as its most representative item, number 35 (gets rattled easily in exciting situations). The items positively correlated with it are found on the introvert side in Fig. 1, and in addition, factors 6, 9, 11, and 30 which were weighted zero for factor *a*. The items negatively correlated with it appear on the extrovert side. Only three items, 7 (considerate of other people's feelings), 8 (acts on the spur of the moment), and 14 (prefers to work with others), show no correlation whatever with item 35. From the point of view of the *number* of items correlated with 18 and 35, the latter has more claim to contain the most representative factor. This is true even if we eliminate from consideration all coefficients lower than $\pm .10$. Any coefficient higher than those limits has been considered significant for this particular study. But the sum total of the correlations with 18 is higher than that for 35, and hence the *a* factor is to be considered the strongest one in the table. And some of the correlations with item 35 may be due to factor *a*.

What can be said of the nature of factor *b* from an examination of the items most strongly associated with it? It would seem to be a kind of sensitiveness to the environment. It is hard to distinguish this constellation of reactions from those containing factor *a*. But if we note the items in this group not found associated with item 18 so strongly, we find an emotional element coming in. For example, item 25 (feelings easily hurt) is rather heavily loaded with factor *b*, but it contains little or nothing of factor *a*. Item 15 (inclined to worry) contains both, but much more of factor *b*. Item 29 (likes to confide in others) and 30 (expresses emotions freely) both contain factor *b* but not *a*. All these reveal an emotional element. The reference item, 35, may be a consequence of this very thing. A person who is over-sensitive to the many incongruous demands of his environment feels unprepared to meet them and hence often responds in a confused, incoordinated, emotional manner.

The third factor centers around item 30 (expresses emotions freely). Here is another emotional factor, but different from factor *b*. Among those items most strongly associated with 30

are: (acts on the spur of the moment), 23 (frequent ups and downs in mood), 25 (feelings easily hurt), 29 (likes to confide in others). Among those negatively associated with it are: 18 (inclined to keep in the background), 24 (slow and deliberate), and 32 (inclined to keep quiet when out in company). Surely the outstanding thing about this grouping of reactions is impulsiveness; a quick and ready response to environmental changes. This has been considered by many as the *sine qua non* of extroversion. It bulks large in McDougall's physiological theory. But curiously enough, this item stands at the zero point in Fig. 1, and is weighted zero for factor *a*. And in the table of raw coefficients, it correlates positively, or negatively, with both introvert and extrovert items.

This is probably the secret of the failure of those who have attempted to correlate the rate of fluctuation of an ambiguous figure with the typical test of introversion-extroversion. The typical test far transcends this group factor *c*, or *impulsiveness*. This factor, which we would expect to find related to the rate of fluctuation of an ambiguous figure, is too far obliterated in the typical test to give any significant correlations between the two types of measurement. If only those items that fall within this group were used, and if they were properly weighted for the *c* factor, we would reasonably expect a correlation with fluctuation tests. But if group factor *a*, of which item 18 is the key, is the thing that the great majority of writers are meaning by introversion-extroversion, then McDougall is talking about something quite different but nevertheless something that may be as significant. We believe, therefore, that the analysis which we have undertaken has already thrown some light upon the difference between McDougall's notion of introversion-extroversion and that of the majority of writers.

Of the three large aspects postulated at the beginning of this analysis, two have already been discovered, in addition to a third which was not postulated. The social factor that was looked for may be identified with factor *a*. The emotional factor may be identified with *c*. The latter was found, however, to be almost uncorrelated with introversion-extroversion. What of the third factor, the direction of interest, the one that is probably paramount in Jung's conception? This one is indicated in the very next factor that appeared in our analysis. The key item is number 31 (thinks about self much of the time). Among those items most strikingly associated with it are: 2 (inclined to limit acquaintances), 13 (daydreams), 16 (absentminded), and 25 (feelings easily hurt). Two items more strongly negatively associated with it are: 7 (considerate of other people's feelings), and 6 (particular about

personal property). Aside from the negative correlation with the last mentioned item, the picture might be one of Narcissism. With the exception of item 8 (acts on the spur of the moment), all the items associated with item 31 are on the expected side of the scale in Fig. 1. Several items correlated with 31, however, are at the zero point in Fig. 1, *i.e.*, items 5, 6, 7, and 30. On the other hand, other items in the test which we would expect to indicate direction of interest, such as 4 (prefers to read about a thing), and 19 (more interested in athletics than in intellectual things) do not correlate at all with the key item. It may be that there are no items in the list appropriate for bringing out the difference in direction of interest as posited by Jung's conception. But if this is so, such appropriate items have never yet been included in any test of introversion-extroversion.

It is probably not worth while here to discuss the remaining 14 group factors in as much detail as we have the four more important ones. It may suffice to give the key item for each one. Only those other items that correlate with the key item to the extent of $\pm .10$ may be regarded with certainty as belonging to that group. The remaining key items, in the order of their importance, are as follows: 21 (likes to speak in public), 17 (likes to persuade others), 1 (expresses self better in speech than in writing), 34 (likes to sell things), 14 (likes to work with others), 36 (male), 22 (prefers to work things out alone), 19 (more interested in athletics than in intellectual things), 11 (likes to change type of work frequently), 32 (likes to be watched while working), 6 (particular about personal property), 10 (works better when praised), 5 (likes detailed work), and 33 (rewrites social letters). The last two factors extend through only six to eight items.

Granting that each item in the test tells us something definite about personalities, and the apparent reasonableness of many of the correlations indicates that they do, we are able to grasp some idea of the intricacy of the whole affair. But every reaction to an item depends to some extent upon several strata of dispositions in the individual. It is first of all an expression of some deep temperamental trait, perhaps of an hereditary nature. It is partly determined by very old personal habits, emotional, social, or attitudinal. It may be determined to some extent by habits of a more recent origin and of a less permanent sort. And lastly, it may be influenced by the present mental sets of the individual.

It was hoped in this study, that the use of over 900 subjects might reduce the errors of measurement to variable ones, although this, we know, can never be fully realized. For example, when we examine the items correlated with the sex of the subjects, we

might wonder how much the sex differences are due to the traits obviously intended in the item and how much they are due to a different mental set in the two sexes. To cite a few traits that seem on the basis of this test to be correlated with being a male, we find: 'indifferent to praise', 'feelings not easily hurt', 'inhibits expression of emotions', and 'not easily rattled'. Are girls merely more ready to confess to these traits which might seem to suggest shortcomings? Have men any more real claim to these virtues than women? Are not men more likely to prefer to think that they are Spartan-like, as these four traits imply, or more likely to claim these traits? It is usually found that women coöperate more fully in carrying out instructions given them, and hence even if they recognize the weakness of some of these traits they would probably be more honest about themselves in answering the questions.

The use of such questionnaire types of tests, therefore, is not a perfect technique for analyzing personality into its underlying components. But assuming valid measurements of a variety of aspects of personality such as these test items imply, the use of Spearman's or Thurstone's or any other accepted technique of analysis of such measurements for their components seems to offer a fruitful approach to a very intricate problem. The primary purpose of this study was to examine the typical Yes-No type of test of introversion-extroversion critically and empirically. In so far as the test items have yielded valid measurements of different aspects of personality, and in so far as current methods of factor analysis permit, the analysis of those aspects of personality represented in the typical test of introversion-extroversion has been accomplished.

SUMMARY AND CONCLUSIONS

A typical test of introversion-extroversion of the Yes-No variety was prepared after a thorough survey of the literature on the subject and of the tests now in use. The test was given to 930 students, 277 of whom repeated it after an interval of a month. The frequency with which every one of the 36 items was responded to in a certain way in conjunction with a certain response to every other item was obtained. That is, for every combination of items, taken two at a time, the frequencies for 'Yes-Yes', 'No-No', 'Yes-No', and 'No-Yes' combinations of reactions were tabulated. From these were obtained coefficients of contingency between pairs of items, which when corrected gave equivalent Pearson coefficients of correlation.

By means of the Spearman-Dodd technique, the correlation of

each item with an assumed 'g' factor was obtained and also the corresponding weight for each item. The positive and negative weights agreed with the traditional armchair opinions about the same test items. That is, the so-called extrovert items and the introvert items as groups tended to cling together at two ends of a scale. Several items long supposed to be diagnostic, however, have zero weights. All others form a continuous graded series of traits, from extreme introversion to extreme extroversion. A test of introversion-extroversion based upon these weights had a self correlation, test-retest, of .81. The validity of the test, as denoted by the correlation of the pool of items with the 'g' factor was .87.

By computing specific correlations, with the 'g' factor held constant, however, it was revealed that the test is full of group factors. The weights obtained by the Spearman method are therefore only approximations. Thurstone's method of multiple factor analysis was applied to the table of intercorrelations and at least 18 group factors were found to be present. Most important of these were four factors named as follows: (a) a tendency to fear the environment, to shrink away from it (this is probably what most writers mean by introversion); (b) an emotional sensitiveness to the environment; (c) impulsiveness; and (d) interest in self. All of these except (c), impulsiveness, are clearly correlated with the 'g' factor that was found by Spearman's method, and may therefore be regarded as aspects of introversion-extroversion. Factor *c* cannot be so regarded. It is suggested that factor *c* is what McDougall has in mind as introversion-extroversion, and that the rate of change of the fluctuating cube will measure this variable.

Is there a single dimension of personality to be called introversion-extroversion, and to be measured by the several standard tests of this trait? It has been shown that one can force most of the items of this test, and perhaps of more extended tests, onto a single continuum. But in reality, our analysis would seem to show that such a procedure is largely fictitious, and that personality is an extremely multidimensional affair. It is possible to find several closely allied dimensions including several like three of the four listed above, and to project them upon a single more inclusive continuum, and to give a name to that larger 'variable' thus created. That is apparently what armchair psychology has done in the case of introversion-extroversion. But one who wishes to approach a real analysis of personality and its basic causes will not be misled by names. One could apply the term introversion-extroversion to any particular test of the usual type,

but we would have to keep in mind the composition of that test and the weighting of responses to its items, just as we have learned to distinguish between the Binet I.Q., Otis I.Q. and NIT I.Q. As in the case of intelligence tests, scales of introversion-extroversion may have their own practical value when correlated with practical criteria. But let us remember that the usual scale bearing the name does not refer to any real dimension of personality any more than the usual intelligence test measures a single real variable of mental ability. In both cases a true understanding of the ingredients of ability or personality is to be had only by a rigid analysis of the factors which enter into a large number of single tests or measurements.

¹ This study was made possible by a Grant-in-Aid from the Social Science Research Council. We wish to express here our gratitude for the generosity of the Council.

² Guilford, J. P., and Braly, K. W., Extroversion and Introversion *Psychol. Bull.* 1930, 27, 96-107.

³ *Ibid.*, 105.

⁴ Guilford, J. P., and Braly, K. W., Some Experimental Tests of McDougall's Theory of Introversion-Extroversion. *J. Abnor. & Soc. Psychol.*, 1930, 25, 382-389. Guilford, J. P., and Hunt, J. M., Some Further Experimental Tests of McDougall's Theory of Introversion-Extroversion. *J. Abnor. & Soc. Psychol.*, 1931, 26, 324-332.

⁵ Guilford and Hunt, *op. cit.*, 329.

⁶ *Ibid.*, 331.

⁷ Spearman, C., *Abilities of Man*. 1927, Appendix.

⁸ Dodd, S. C., The Coefficient of Equiproportion as a Criterion of Hierarchy. *J. Educ. Psychol.*, 1928, 19, 217-229.

⁹ Thurstone, L. L., Multiple Factor Analysis. *Psychol. Rev.*, 1931, 38, 406-427.

¹⁰ Thurstone, L. L., Theory of Attitude Measurement. *Psychol. Rev.*, 1929, 36, 222-241.

¹¹ Jung, C. G., *Collected Papers on Analytical Psychology*, 1916, *Psychological Types*, 1920; *Problems of Personality*. Studies in Honor of Morton Prince, 1925.

¹² Freyd, M., Introverts and Extroverts. *Psychol. Rev.*, 1924, 31, 74ff.

¹³ Laird, D., *Personal Inventory*. C2, Colgate University.

¹⁴ Marston, L. B., *The Emotions of Young Children*. Iowa Studies in Child Welfare, 1923, 3.

¹⁵ Neymann, C. A., and Kohlstedt, K. D., A New Diagnostic Test of Introversion-Extroversion. *J. Abnor. & Soc. Psychol.*, 1929, 23, 482-487.

¹⁶ Gilliland, A. B., and Morgan, J. J. B., An Objective Measure of Introversion-Extroversion. *J. Abnor. & Soc. Psychol.*, 1931, 26, 293-303.

¹⁷ This was the basic assumption of Neymann and Kohlstedt, *op. cit.*, and also of Gilliland and Morgan, *op. cit.*

¹⁸ Hunt, J. M., and Guilford, J. P., Fluctuations of an Ambiguous Figure in Dementia Praecox and in Manic Depressive Patients. *J. Abnor. & Soc. Psychol.*, 1933, 27, 443-452.

¹⁹ Kelley, T. L., *Statistical Method*, 1924, 266.

²⁰ Kelley, *op. cit.*, 259.

²¹ *Ibid.*, 266.

²² *Ibid.*, 267.

²³ *Ibid.*, 268.

²⁴ Spearman, *op. cit.*, Appendix xvi.

²⁵ Dodd, *op. cit.*, 221.

²⁶ Spearman, *op. cit.*, Appendix xix.

²⁷ Holzinger, K. J., *Statistical Resumé of the Spearman Two Factor Theory*, 1930, 21.

²⁸ Tryon, R. C., The Reliability Coefficient as a Per Cent with Application to Correlation Between Abilities. *Psychol. Rev.*, 1930, 37, 140-157.

²⁹ Garrett, H. E., *Statistics in Psychology and Education*, 1926, 272f.

³⁰ Thurstone, *op. cit.* (see footnote 9 above).