THE
JOURNAL OF MENTAL SCIENCE
[*Published by Authority of the Royal Medico-Psychological Association.*]

No. 334 [NEW SERIES No. 298.] JULY, 1935. VOL. LXXXI

NOV 16 1935

Part I.—Original Articles.

THE RESEMBLANCE OF TWINS WITH REGARD TO PERSEVERATION.

By ELLA PRATT YULE, M.A., Ph.D.

From the Department of Social Biology, University of London.

SECTION I: INTRODUCTION.

THE social behaviour which is the material of psychological observation at any given time is the product of a certain genetic constitution on the one hand, and of a complex process of social, educational, familial and, to some extent, uterine environment on the other. The social behaviour of human beings cannot be subjected to direct methods of genetic analysis. The genetic psychologist may be compared to the astronomer, in so far as he has to rely on Nature to perform the experiments which he himself is unable to undertake. To the category of such experiments belongs the phenomenon of twin production. In the production of monozygotic twins Nature provides us with individuals of the same genetic constitution. Such differences as they exhibit arise uniquely from the operation of differences in the environment in which they are placed. In ordinary circumstances they share many more features of environment than any other two individuals selected at random from the population. When reared together, the extent of their dissimilarity can only throw light on such differences of nurture as may operate on members of the same family unit and of the same age. If they are reared apart, the measure of resemblance exhibited by monozygotic twins may throw light on the influence of social environment in a less restricted sense. With a sufficiently large sample of twins reared apart, a satisfactory standard of comparison for assessing the extent of resemblance due to genetic constitution, plus the effects of a common uterine environment, would be provided by the degree of resemblance existing between their foster sibs in families into which they have been adopted. Such data cannot be collected easily. As yet they do not exist in sufficient quantity to yield satisfactory conclusions.

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press

If we confine our attention to the effect of differences of nurture such as operate on individuals brought up together in the same family unit, three standards of comparison are available : these are the resemblance of dizygotic twins of the same sex, the resemblance of dizygotic twins of unlike sex, and the resemblance of sibs of widely different ages and of both like and unlike sex. Two individuals belonging to any of these specified groups have, in virtue of common parentage, more genes in common than any two individuals selected at random from the population. In virtue of genetic segregation they have fewer genes in common than two monozygotic twins.

With respect to the influence of environment, they form a series. The environment of two individuals of a mixed sex pair will be less alike than the environment of two individuals of a like sex pair, in so far as sex tends to prescribe a specific routine of family upbringing and education. The environment of two ordinary sibs will be less alike than that of two twins in so far as differences of age determine the content of experience. Dizygotic co-twins or two sibs of like sex have the same sex environment, but the latter have a different historic environment. Dizygotic co-twins of like sex or dizygotic co-twins of unlike sex have the same historic environment, but the latter do not share the same sex environment. Monozygotic co-twins and dizygotic co-twins of like sex have the same sex environment and also the same historic environment in this sense. However, it would be rash to attribute the greater resemblance of monozygotic twins to the *direct* effect of their greater genetic similarity alone. The very fact that two individuals differ in genetic constitution may force them into different situations, and consequently expose them to widely differing environmental influences. Thus the existing genetic similarity in a pair of monozygotic twins may make their common environment more homogeneous than the environment of a pair of dizygotic twins. Similar genetic constitutions may tend to lead the two genetically similar individuals to seek, or find themselves in, similar environments. Thus the closer similarity of monozygotic co-twins may in part be attributable to the effect of environmental factors reinforcing genetic similarity.

The use of the twin method for assessing the relative contributions of genetic and environmental factors to differences in behaviour requires careful consideration, for the reason that it is impossible to draw a hard and fast line of demarcation between genetic and environmental factors as productive of separate effective differences in behaviour. It is, in short, difficult, and at present impossible, to distinguish with any finality between the effects of genetic and the effects of environmental factors on observed differences in behaviour. The issue is further complicated because what psycho-analysts call " protest " may also call for consideration. In objective terms *protest* may be described as the reaction of one individual against the personality of another individual who resembles him closely in physical and mental constitution. In the terminology of psycho-analysis, every individual feels the need not only to

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press

differentiate himself from his environment, but to be unique within the range of his own environment. If this has any objective basis in the study of social behaviour, it must be taken into account in discussing individual differences in the members of a twin pair. Either or both of the members of the pair may react to a " felt need " for differentiating himself in some way from an individual who resembles him so closely. If protest is a factor, it is a factor which will most probably enter into the consideration of individual differences in a group of monozygotic twins. Should it be a significant element in the situation, environmental factors, instead of reinforcing the degree of similarity resulting from a common genotype, will clearly lessen it. Such considerations suffice to emphasize that the issues involved in the application of the twin method of the problems of Nature and nurture as they affect social behaviour are less simple than when it is used to throw light on differences of physique.

The utility of the twin method is not confined to investigation of the problems of Nature and nurture. The study of twin data provides an independent avenue for exploring the value of tests used in the measurement of differences in behaviour, and directs attention to similarities which could provide the basis for testing characteristics, methods for the description of which have not hitherto been standardized. For example, the twin method provides one way of deciding whether a given test gives a constant index of some aspect of an individual's behaviour. In a pair of monozygotic twins we are presented with two individuals who have the greatest possible degree of human resemblance. If this is so, such resemblance must be expressed in behaviour. Another issue upon which twin resemblance may throw light arises when trying to detect constant features of behaviour characteristic of an individual. The comparison of the performances of individuals who are as much alike as possible with those of individuals who are as different as individuals generally are, is then the most likely avenue for getting a clue about what facets of the behaviour of individuals are characteristic.

Hitherto work on the psychological resemblance of twins, that is, the resemblance of their behaviour patterns, has been limited by the precision of available methods of description and measurement. The only extensive work on twin resemblance is that which is based on the use of the intelligence quotient. The reader may refer to the works of Stocks, Merriman, Lauterbach, Wingfield, and of Herrman and Hogben. The intelligence quotient unquestionably provides a quantitative measure of a definite aspect of human behaviour amenable to equally objective or quantitative measurement. Among the few aspects of human behaviour which can be studied with any degree of precision and objectivity, the characteristic called by Spearman " perseveration " seemed to the present writer the most fruitful.

The reasons for adopting perseveration tests as a measuring instrument of individual differences in behaviour in the present investigation were : (1) That it was claimed for perseveration tests that they provide a method which does

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press

not rely on the subject's testimony to his (or her) behaviour, (2) that the aspect of behaviour measured does not depend on I.Q., and may be considered as falling within the category of temperament, personality or character. Thus an investigation into the resemblance of twins with regard to their performance in these tests may be expected to throw some light on the relative contributions of genetic and environmental factors to differences in temperament. Further, since previous work with perseveration tests has produced varying conclusions, not only about the nature of what is measured, but about the reliability of the tests as a measuring instrument, it was thought that the application of the twin method in the analysis of individual perseveration scores would provide an independent method of assessing the reliability of the tests themselves. The main objective of the investigation was thus to discover whether it is possible in this way to measure an aspect or aspects of behaviour other than those measured by intelligence tests, or, to state the problem from another angle—whether it is possible to detect uniformities in behaviour patterns other than those belonging to the category of intelligent behaviour.

Section II: Nature of the Materials.

Two groups of subjects were tested:

(*a*) A group of 60 unrelated London elementary school-children.

(*b*) A group of 115 pairs of twins.

All individuals tested in the (*b*) group in this investigation belonged to the group previously tested in the investigation into "The Intellectual Resemblance of Twins", by Herrman and Hogben (1932). For purposes of that investigation, the group had been tested by the Otis Advanced (Form A) Group Test. Thus, whenever reference is made in the present investigation to I.B.* or I.Q. of subjects, it will be understood that the reference is to the I.B or I.Q. of the Otis Intelligence Scale. Since the previous investigation had necessitated a physical diagnosis of the genetic type of the pairs of twins in Group (*b*), the requisite data were at hand for the present investigation. Full particulars of the methods which were used by Herrman and Hogben in the diagnosis and classification of the pairs of twins into genetic type groups, including details of the finger-print method, can be obtained in the report on the previous investigation (see *op. cit.*, pp. 109–114).

The (*b*) twin group of the present investigation had thus been classified into genetic type groups, viz., (1) monozygotic, (2) dizygotic twin pairs. For purposes of this investigation a clear demarcation between dizygotic like-sex pairs and dizygotic pairs of unlike-sex has been made, giving the three groups used in discussion of twin resemblances in "p" score as—

(1) Monozygotic.

(2) Dizygotic, like sex.

(3) Dizygotic, unlike sex.

* *Vide infra*, Section IV (*c*).

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press

It is particularly to be noted that the individual pairs referred to as composing groups (1) and (2) above are selections from the " refined " groups of the investigation by Herrman and Hogben, that is, group (1) is selected from group " M ", and group (2) from group " D " (see *op. cit.*, p. 113). In the present investigation, groups (1) and (2) are composed of such pairs of twins as exhibit only striking similarity or dissimilarity. Group (2) of the present investigation, for example, is composed of dizygotic pairs of like sex chosen in virtue of the fact that they are as unlike as could be found within the limits of the available material.

The social status of subjects in both (*a*) and (*b*) groups in this investigation was comparatively uniform. All subjects either were, or had been, elementary school pupils in the London County Council area. The age range in the (*a*) school group was small, being 9 years 5 months to 12 years 6 months. The age distribution of the (*b*) twin group is given in the following table :

TABLE I.—*Age Distribution.*

Age (years, months).	All.	Twin-pairs.	Age (years, months).	All.	Twin-pairs.
10–10.5	2	1	14–14.5	30	15
10.6–10.11	12	6	14.6–14.11	10	5
11–11.5	14	7	15–15.5	16	8
11.6–11.11	18	9	15.6–15.11	6	3
12–12.5	26	13	16–16.5	4	2
12.6–12.11	20	10	16.6–16.11	..	..
13–13.5	16	8	17–17.5	..	..
13.6–13.11	48	24	17.6–17.11	2	1

SECTION III : METHOD OF TESTING.

(a) *Description of the Test.*

No initial hypotheses about the nature of " p " or what was measured by an individual score in a " p " test have been taken for granted in this inquiry. Using Spearman's terminology, " p " stands for *perseveration*, which is defined as " the hindrance effect exercised by any activity upon a subsequent activity ". The writer has preferred to talk of " p " factor rather than of perseveration, and to regard the existence of the " p " factor as a fact which previous investigators have derived from correlations for a set of tests of a certain kind. In his paper on " Standard p-Tests ", Dr. Stephenson has classified " p " tests into three main types. He gives an admirably clear analysis of the differing factors which a " p " score resulting from the application of any of the three types of " p " test would tend to emphasize. It seemed clear to the writer that the selection of one type in preference to another on other than utilitarian

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press

grounds could only involve the existence of an *à priori* hypothesis concerning the nature of " p ". Hence, on grounds of utility and ease of administration, the form of " p " test used by Dr. Stephenson has been employed in the present investigation.

In the Stephenson form, the " p " test falls into four sections, X_1, X_2, Y_1, Y_2:

In X_1 the subject is asked to write, e.g., forward S's (S S S . . .).

In X_2 he is told to repeat (S S S . . .).

In Y_1 he is asked to write (2 2 2 . . .).

In Y_2 he is asked to write (S 2 S 2 S 2 . . .).

In the present investigation thirty seconds were allowed for each of the four sections equally, and each subject was permitted to make four, and only four, sample or trial letters before commencing X_1, Y_1 and Y_2. Instructions were that each subject was to write with as much speed as was compatible with reasonably well-formed letters or figures. Every effort was made to eliminate individual differences in the apprehension of what was required, by as much previous explanation as seemed called for.

Standard " p " test forms, analogous to those used by Stephenson, were adopted to eliminate as far as possible individual differences in the size of letters or figures, and also make for ease in scoring. The Stephenson form of " p " test has many advantages in compactness, standardization, ease of scoring and administration and time required over the form of " p " test used, for example, by Pinard (1932). Eight were selected, viz., S, Δ, e, Z, 9, 2, 6, G. Each test occupied roughly four minutes, the time taken to give the battery, allowing for explanation and a rest pause, being forty minutes.

(b) *Scoring.*

Of two possible methods of scoring, viz.,

(*a*) $(X_1 + X_2) - (Y_1 + Y_2)$,

(*b*) $\dfrac{(X_1 + X_2)}{(Y_1 + Y_2)}$,

(*a*) was discarded on the ground that if it were adopted the factor of speed would complicate the evaluation of the " p " score. Method (*b*), i.e., $\dfrac{"x"}{y}$, was therefore adopted.

The tests were scored both individually and as a battery, that is, " p " score $\frac{x}{y}$ S, $\frac{x}{y}$ Δ, $\frac{x}{y}$ e . . . $\frac{x}{y}$ G were found, giving the individuals' total " p " score as: $\frac{x}{y}$ S + $\frac{x}{y}$ Δ + . . . $\frac{x}{y}$ G = " p ",

and also $\dfrac{x\,S + x\,\Delta + x\,e \;\ldots\; x\,G}{y\,S + y\,\Delta + y\,e \;\ldots\; y\,G}$ or $\dfrac{\text{Total } x}{\text{Total } y}$ = " p ".

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press

In what follows, the second method is referred to as Battery " p ". Further points in connection with the scoring will follow with the discussion of reliability and abnormally high scores.

Section IV: Reliability and Autonomy of the Tests Employed.

For work of this kind a test should fulfil three criteria:

(*a*) It must be reliable in the statistical sense, i.e., must be capable of yielding consistent results on successive occasions.

(*b*) The method of scoring should yield an index which is independent of age.

(*c*) The results should not correlate significantly with those of any other known test; otherwise its existence does not fulfil a specifically independent function.

(a) *Statistical Reliability.*

Table II summarizes briefly the results of retesting the (*a*) school group. The interval between the test and subsequent re-test was seven days.

Table II.

Test.	Coefficient of correlation of test and re-test scores.
1 . S .	·426 ± ·110
2 . Δ .	·411 ± ·112
3 . e .	·330 ± ·120
4 . Z .	·016 ± ·134
5 . 9 .	·522 ± ·098
6 . 2 .	·517 ± ·098
7 . 6 .	·523 ± ·098
8 . G .	·391 ± ·114
9 Battery ($\frac{x}{y}$S + $\frac{x}{y}$Δ . . . $\frac{x}{y}$G)	·555 ± ·093
10 Battery ($\frac{\text{Total } x}{\text{Total } y}$S . . . G)	·791 ± ·065

In using the whole series of tests collectively as a battery, the coefficient of correlation for test and re-test is based on two methods of scoring, viz.:

(1) $\frac{x}{y}$S + $\frac{x}{y}$Δ + $\frac{x}{y}$e . . . $\frac{x}{y}$G and

(2) $\frac{\text{Total } x}{\text{Total } y}$(S . . . G) (see 9 and 10).

From inspection of this table it seems quite clear that " p " tests 9, 2 and 6 show significantly higher reliability than Z or e.

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press

On account of the occurrence of certain individuals who obtained abnormally high scores, the general picture which is disclosed by first-hand acquaintance with the data is one of higher reliability than the statistical indices reveal. The method of scoring adopted does, in fact, lend itself to the occasional occurrence of abnormally high scores.

The scoring of the tests as a battery has the effect of levelling the range of the scores to so great an extent as almost to eliminate the effect of very high scores. In the (*b*) twin group data, the range of battery " p " is 1·00–1·90, though there is one outstanding score of 2·69. Battery scores in the (*a*) school group never fall below 1·15, though in two cases in the test, and seven in the re-test, they rise above 2·00. It is probable that group testing with " p " tests produces a greater number of abnormally high scores than individual testing, since in the group test there can be much less adjustment of the explanations to suit the requirements of the individual, and many more extraneous accidental factors must be operative.

Of several conceivable reasons which might be offered to explain the occurrence of these abnormal scores, one is defective technique in the administration of the tests—that is to say, the test may prove to yield more consistent results when the technique of administration has been more carefully standardized. For the present investigation, however, we are only concerned with the reliability which can be claimed for the method as it was actually applied. There is no guarantee that abnormal scores are accidental, and would disappear on more extensive testing or further retesting.

As a check on the internal consistency of the test series, Table III gives the split-half reliabilities.

TABLE III.—*Split-half Reliability.*

$$\text{I} = \text{Test} \qquad a = (\tfrac{x}{y}\,\text{S} + \tfrac{x}{y}\,\text{e} + \tfrac{x}{y}\,9 + \tfrac{x}{y}\,6)$$

$$\text{II} = \text{Re-test} \qquad b = (\tfrac{x}{y}\,\Delta + \tfrac{x}{y}\,\text{Z} + \tfrac{x}{y}\,2 + \tfrac{x}{y}\,\text{G})$$

Test.	r.
AI, AII	·666 ± ·075
BI, BII	·465 ± ·105
AI, BI	·581 ± ·089
AII, BII	·403 ± ·126
AI, BII	·540 ± ·095
AII, BI	·446 ± ·108

(b) *Effect of Age on Scores obtained.*

Table IV gives age and " p " score coefficients of correlation for each half of the test and for battery " p ". It is based on the " p " score of the (*b*) twin group (226 subjects).

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press

TABLE IV.

Age range 10–18 years.

Age — "p" $(\frac{x}{y}S + \frac{x}{y}\Delta + \frac{x}{y}e + \frac{x}{y}Z)$. $+ \cdot 01 \pm \cdot 066$

Age — "p" $(\frac{x}{y}9 + \frac{x}{y}2 + \frac{x}{y}6 + \frac{x}{y}G)$. $- \cdot 06 \pm \cdot 066$

Age — "p" $\frac{\text{Total } x}{\text{Total } y}(S \ . \ . \ . \ G)$. $- \cdot 079 \pm \cdot 066$

The figures indicate clearly that the correlation with age is so small that it may be assumed not to affect the interpretation of the results. This justifies what has already been stated in connection with the method of scoring.

Pinard (1932) reports "that there is a steady decrease in P from 12–8 years (but this decrease may have been influenced by the speed at which the younger subjects worked)". In calculating this result Pinard was working with the (X—Y) "p" score, not with the "*X*" method, as for the data of the present investigation, on the grounds previously stated, viz., that allowance would have to be made for speed.

(c) *The Autonomy of the "p" Test.*

Table V gives the coefficients of correlation for I.B.* and I.Q. with "p" score for both Group (*a*) school, and (*b*) twin.

TABLE V.

Group.			. . . "p" score.		
a. (55)	I.B. — "p"	$(\frac{x}{y}S + \frac{x}{y}\Delta \ . \ . \ \frac{x}{y}G)$	Test	$+ \cdot 074 \pm \cdot 159$
	I.B. — "p"	(,, ,, ,,)	Re-test	$- \cdot 191 \pm \cdot 152$
	I.Q. — "p"	(,, ,, ,,)	Test	$+ \cdot 062 \pm \cdot 157$
	I.Q. — "p"	(,, ,, ,,)	Re-test	$- \cdot 163 \pm \cdot 156$
	I.B. — "p"	$(\frac{\text{Total } x}{\text{Total } y}S \ . \ . \ G)$	Test	$+ \cdot 022 \pm \cdot 159$
	I.B. — "p"	(,, ,, ,,)	Re-test	$- \cdot 102 \pm \cdot 158$
	I.Q. — "p"	(,, ,, ,,)	Test	$+ \cdot 019 \pm \cdot 159$
	I.Q. — "p"	(,, ,, ,,)	Re-test	$- \cdot 124 \pm \cdot 158$
b. (226)	I.Q. — "p"	$(\frac{x}{y}S + \frac{x}{y}\Delta + \frac{x}{y}e + \frac{x}{y}Z)$	Test	$+ \cdot 01 \pm \cdot 066$
	I.Q. — "p"	$(\frac{x}{y}9 + \frac{x}{y}2 + \frac{x}{y}6 + \frac{x}{y}G)$	Test	$- \cdot 03 \pm \cdot 066$
	I.Q. — "p"	$(\frac{\text{Total } x}{\text{Total } y}S \ . \ . \ G)$	Test	$- \cdot 009 \pm \cdot 066$

* On the Otis Scale I.B. is a measure of the individual increment or decrement of score from the norm for persons of his own chronological age, the latter being in all cases expressed as 100. I.Q. is obtained by dividing the mental age—i.e., the age for which the score is obtained by the individual is the norm—by the chronological age.

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press

It is clear from inspection of this table that whatever is measured by " p " is not the same thing as that which is measured by I.Q. The consistent appearance of the negative correlations with re-test scores may or may not be significant.

(d) *Distribution of " p " Score with Reference to Sex.*

Apart from age, another factor which enters into an assessment of the homogeneity of the material is sex. The following table summarizes the data on sex differences.

TABLE VI.—*Sex Differences in " p ".*

Group.	Number in group.	Sex.	Score.
Monozygotic	32	F.	144·45 ± 16·54
	22	M.	143·08 ± 15·24
Dizygotic, like sex	40	F.	147·64 ± 14·18
	36	M.	151·42 ± 12·95
Dizygotic, unlike sex	46	F.	150·27 ± 17·43
	46	M.	159·63 ± 21·05
All subjects	118	F.	147·81
	104	M.	153·28

[*Note.*—In this table " p " = $(\frac{x}{y}\,S + \frac{x}{y}\,\Delta \ldots \frac{x}{y}\,G)$.]

The data do not disclose any significant connection between sex and " p ". It may be noted that Pinard also found no sex difference in " p " in boys and girls between the ages of 8 and 14 years of age.

SECTION V: RESEMBLANCE OF TWINS.

Table VII summarizes the data bearing directly on twin resemblance. In the preparation of this table the scores obtained in individual tests have not been utilized separately, but have been compounded according to their comparative reliability as estimated previously.

The scores in tests 9, 6 and 2—being the tests with the highest reliability (·555, ·517 and ·523)—were summed to give the first sub-group, those of tests

TABLE VII.—*Twin Resemblance*

Group.	Number of pairs.	Tests 9, 2, 6.	Tests S, Δ.	Tests e, Z, G.	Battery.	Battery with Spearman's correlation reliability.
Monozygotic.	27	·638 ± ·114	·278 ± ·177	·709 ± ·096	·615 ± ·119	0·78
Dizygotic, like sex	38	·345 ± ·143	·147 ± ·158	·292 ± ·148	·395 ± ·137	0·50
Dizygotic, unlike sex	48	·082 ± ·145	·016 ± ·144	·433 ± ·177	·214 ± ·137	0·27
Unrelated pairs	112	—·034 ± ·099			—·099 ± ·093.	0·12

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press

S and Δ (reliability ·426 and ·411) give the second sub-group, and those of tests e, Z, G with reliabilities of ·330, ·391 and ·016 give the last group (see Table II, reliability). Battery scores given in this table are taken as $(\frac{\text{Total } x}{\text{Total } y}\text{S} \ldots \text{G})$. The biological classification of the twin groups has been explained already in Section II.

The most striking feature of these results is that for tests which have highest reliability, the heterogeneity of the group is most clearly shown by the differing indices for inter-twin resemblance, as in the battery, or for the summed scores of tests of 2 and 6.

Since the study of twin resemblance provides an independent method of assessing the reliability of the tests used (see Section I), it is instructive to examine its bearing on the two different systems of scoring adopted in using the battery as a whole. When the battery score was calculated as $(\frac{x}{y}\text{S} + \frac{x}{y}\Delta \ldots \frac{x}{y}\text{G})$, the correlation coefficients for the twin groups were :

Monozygotic	$\cdot 386 \pm \cdot 163$
Dizygotic, like sex	$\cdot 450 \pm \cdot 129$
Dizygotic, unlike sex	$\cdot 120 \pm \cdot 111$

This might be taken as an indication that the more reliable battery score is given by $(\frac{\text{Total } x}{\text{Total } y}\text{S} \ldots \text{G})$, which is given in Table VII.

With regard to the reliability of the " e " test, it is necessary to mention that the " e " test scores of the (*b*) twin group, as compared with the " e " test scores of the (*a*) school group, contain no abnormal scores :

	Below 1·00.	1·00–2·00.	2·00–3·00.	3·00–4·00.	Outlying scores.
Twin group	10	185	31	..	..
School group I	2	39	10	3	1
,, II	0	42	10	1	2

Now, when the (*b*) twin group scores for tests " e " and 2 were taken together, the correlation coefficients of the twin groups were :

Monozygotic	$\cdot 651 \pm \cdot 111$
Dizygotic, like sex	$\cdot 355 \pm \cdot 141$
Dizygotic, unlike sex	$\cdot 144 \pm \cdot 141$

Thus the " e 2 " test scores seem to give significantly varying group indices.

The consideration of these facts strengthens the impression that the reliability of the tests as calculated from the (*a*) school group, tested by the group method, may not be of the same order of reliability as the reliability of the " p " test scores of the (*b*) twin group, tested by the individual method. It must be remembered that the 55 test and re-test " p " scores obtained from the

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press

(*a*) group constitute a small sample on which to base conclusions as to reliability of a series of tests. Moreover, it is generally accepted that a group test allows for the intervention of many extraneous or incalculable factors.

Section VI: Summary and Conclusions.

Three main conclusions emerge from consideration of the data in this investigation. These are that—

(1) With respect to "p", monozygotic twins are more alike than dizygotic twins.

(2) Dizygotic twin pairs of unlike sex are less alike than dizygotic twin pairs of like sex.

(3) There is no sex difference with respect to "p".

In the light of the third conclusion, the second would suggest that environmental differences within the family unit influence the "p" score of an individual, in so far as such differences as are associated with the different methods of upbringing of a boy and a girl are significant, and must be taken into account.

With regard to the first conclusion, it would be premature to discuss the genetic significance of the result for two reasons. In the first instance, the present status of the test must be considered. Previous discussion of the reliability of the test will have shown that many points with regard to the internal construction, uniformity and technique of administration require closer investigation. In the second place, a discussion of the genetic significance of the result must take into account the difficulties of interpretation alluded to in the introduction. On the one hand it may be true that the existing genetic similarity in a group of monozygotic twins accounts for the greater resemblance of monozygotic co-twins with regard to whatever may be measured by "p". On the other hand, it may be that the existing similarity is only partly the direct result of genetic similarity in so far as genetic similarity will tend to produce environmental similarity.

The limitations imposed on the interpretation of the results by such considerations as these only affect inferences about the relative contributions of genetic and environmental factors. Given that genetic similarity exists between monozygotic twins, we expect it to be expressed, directly or indirectly, in a degree of similarity in their behaviour patterns. Results in the present investigation establish the fact that this is true of the kind of behaviour pattern measured by "p". The "p" tests, in short, do detect and measure a specific uniformity (or uniformities) of behaviour. Thus, the outstanding conclusion which emerges from the investigation is that this type of test does provide an index characteristic of an individual, and moreover an index which measures an aspect, or aspects of behaviour other than those measured by intelligence tests.

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press

ACKNOWLEDGMENTS.

Acknowledgment is made to Dr. W. Stephenson, who so kindly extended help and advice on the administration of the tests ; to Prof. Lancelot Hogben, under whose supervision the investigation was carried out ; and also to the Metcalfe Selection Committee and the London School of Economics for funds enabling the research to be undertaken.

Bibliography.—Stocks, P., "A Biometric Investigation of Twins and their Brothers and Sisters", *Annals of Eugenics*, 1930, iv, pts. 1 and 2.—Merriman, "The Intellectual Resemblance of Twins", *Psych. Monographs*, 33, Princeton, U.S.A., 1924.—Lauterbach, "Studies on Twin Resemblance", *Genetics*, 1925, x, No. 6.—Wingfield, *Twin and Orphans* (Rent), 1928.—Herrman and Hogben, "The Intellectual Resemblance of Twins", *Proc. Roy. Soc. Ed.*, 1932.—Stephenson, W., "Standard 'p' Tests : An Alternate Theory for "p" Factor",—Pinard, "Tests of Perseveration, their Relation to Character", *Brit. Journ. Psych.*, xxiii, pt. 1.—Spearman, C., *The Abilities of Man*.

https://doi.org/10.1192/bjp.81.334.489 Published online by Cambridge University Press