

AN EXPERIMENTAL INVESTIGATION OF "DESIRABILITY" RESPONSE SET IN A PERSONALITY QUESTIONNAIRE

Sybil B. G. Eysenck & Hans J. Eysenck
Institute of Psychiatry (University of London)
(Received 4 April 1963)

It is well known that questionnaires are subject to errors arising from the "response styles" of individual subjects; the two main response sets identified are those of acquiescence and desirability (1). Several studies have shown that acquiescence response set plays only a relatively minor part in the questionnaire measurement of neuroticism and extraversion, provided items of the kind used in the Maudsley Personality Inventory (2) are employed (3,4,5). The present study is concerned with the problem of "desirability" in relation to the same two dimensions of personality. The questionnaire employed in this investigation contained 75 questions, to be answered by ringing either the "YES" or the "NO" answer; no "?" answers were permitted. Twelve questions were scored for the N variable, 54 questions for the E variable, and 9 questions formed a "lie" scale (L); the items for this scale were taken from a longer scale used by Eysenck (6) and by Gibson (7). It was adapted from the M.M.P.I. lie scale. The N items were all taken from the M.P.I.; the E items were partly taken from the M.P.I., but other items were added from several studies carried out in an attempt to improve and extend the M.P.I. (8,9,10). Each of the items used for the N and E scales had been shown to have high loadings on the appropriate factor in at least two, and sometimes as many as ten, independent factor analyses; not all of these analyses have been published.

The experiment consisted in having each questionnaire filled in twice, with an interval varying from one or two days to several weeks. On the first occasion, subjects were simply instructed to complete the inventory; on the second occasion, they were instructed to fill it in in such a way as to give the best possible impression of themselves, i.e. to put themselves in the best

light. They did not know on the first occasion that they would be required to fill in the questionnaire again. In other words, on the second occasion they were asked to fake responses, and the main purpose of this article is to compare the "truthful" response and the "fake good" responses. Our anticipation was that faking would shift the N scores towards a lower level, would leave the E scores relatively untouched, and would increase the level of "lie" responses. It was also anticipated that while the mean E score would remain unchanged, there would be a tendency for both high and low E scores to shift responses towards a more average level, thus reducing the variance of E scores.

Ten groups of subjects in all were studied, giving a total of 873 Ss in all. The composition of these groups is shown in Table 1; it will be seen that the majority of Ss were University students (675, in all). All the Ss except for one group of U.S. students and a group of U.S. parents, teachers and housewives, were British; all the British Ss were English except for the Belfast group, who were of course Irish.

TABLE 1

	n
(1) Belfast students	163
(2) Non-student group (Housewives, evening classes, etc.)	85
(3) Working-class group	57
(4) Sheffield students	38
(5) Exeter students	89
(6) Non-student U.S. group (Housewives, parents, teachers etc.)	57
(7) Belfast law students	66
(8) U.S. students	185
(9) Welsh students	32
(10) Manchester students	101
	<hr/>
TOTAL:	873

TABLE 2

MEANS

	N	AGE	SEX		A			B			A-B		
			M	F	E	N	L	E	N	L	E	N	L
1.	163	18.822	70	93	28.748	6.736	1.601	30.454	1.632	7.325	-1.706	5.104	-5.724
2.	85	33.341	36	49	26.541	5.953	1.788	29.059	1.941	5.800	-2.518	4.012	-4.012
3.	57	41.491	50	7	28.772	7.035	2.807	31.018	4.632	4.860	-2.246	2.403	-2.053
4.	38	18.895	18	20	30.579	7.026	1.316	34.316	2.395	7.500	-3.737	4.631	-6.184
5.	89	19.820	33	56	28.112	6.213	1.416	29.966	1.944	5.449	-1.854	4.269	-4.033
6.	57	37.579	17	40	26.491	4.982	1.737	32.211	1.386	6.456	-5.720	3.596	-4.719
7.	66	20.727	48	18	29.288	6.773	1.515	32.652	2.318	6.697	-3.364	4.455	-5.182
8.	185	19.286	73	112	30.730	6.708	1.411	32.897	1.178	7.795	-2.167	5.530	-6.384
9.	32	23.875	15	17	27.719	5.031	1.281	30.969	0.875	5.969	-3.250	4.156	-4.688
10.	101	22.792	52	49	27.624	6.495	1.653	32.337	2.079	5.248	-4.713	4.416	-3.595

STANDARD DEVIATIONS

1.	163	2.261	70	93	8.220	3.226	1.451	5.891	1.641	1.993	10.455	3.237	2.340
2.	85	10.672	36	49	8.814	3.387	1.559	5.803	2.607	2.530	9.643	2.966	2.628
3.	57	13.791	50	7	8.604	2.988	1.931	7.749	3.074	2.532	11.189	4.697	3.476
4.	38	1.843	18	20	8.494	3.192	1.276	9.367	2.626	2.501	14.757	3.582	3.021
5.	89	2.443	33	56	9.449	3.379	1.629	7.541	2.465	2.884	11.253	3.704	2.909
6.	57	11.260	17	40	8.289	3.254	1.653	5.621	2.266	2.529	8.525	3.417	2.944
7.	66	5.058	48	18	8.372	3.042	1.756	6.607	2.513	2.219	10.687	2.962	2.833
8.	185	3.637	73	112	7.353	3.202	1.385	4.949	1.958	1.748	8.175	3.133	2.074
9.	32	3.087	15	17	7.722	2.706	1.276	4.617	1.100	2.559	9.281	2.919	2.507
10.	101	3.232	52	49	7.516	3.142	1.439	6.834	2.444	2.823	9.001	3.686	3.128

Means (top half) and S.D.s (bottom half) of E, N and L scores for 10 groups of Subjects under conditions A and B, and A - B.

Table 2 gives the means and the S.D.s, for E, N and L, of all the groups for both administrations of the questionnaire; these two administrations are referred to as A and B, respectively. Also given are the differences (A - B) between the two administrations. Analyses of variance were carried out to determine whether the differences between groups were significant; the results are indicated in Table 3. It is clear that the groups are differentiated with

TABLE 3

ANALYSIS OF VARIANCE

	(A)	(B)	(A - B)
E	2.944*	4.907**	1.433 NS
N	3.017*	13.312**	5.570**
L	4.931**	18.611**	21.003**

NS = Not significant

* P < .05

** P < .01

respect to E and N, but only slightly; with such small numbers the F values given suggest that differences may be safely disregarded. When we look at the B scores we find much higher F values for N and L, and a somewhat higher F value for E as well. The groups are obviously rather different in their "faked" responses, particularly for N and L. For A - B, only L responses show a really high F value.

Inspection of the A scores reveals that the U.S.A. students are the most extraverted group; this might not come as a surprise. On the other hand, the U.S.A. non-student group (mostly Palo Alto housewives and other wage-earners) has the lowest E score; this might not have been anticipated. However, the F ratio is so low as to indicate that search for significance (in the psychological sense) would seem to be useless. The same is probably true of N,

where the working class group has the highest score; this is in accordance with other studies (11). This group also has the highest L score, which is unexpected; usually high L scores go with low N scores. This difference is startlingly large, and almost entirely responsible for the significance of the F ratio. No previous data on class comparisons are available for "Lie" scale scores, so that we cannot tell whether this finding is unusual.

The working class group is again outstanding with respect to the B neuroticism scores, where they have far and away the highest scores, and also with respect to the Lie scores, where their score is now the lowest, by a long chalk. It would appear that this group is responsible for the main differences discovered between our ten groups, both for the A and the B administrations; a discussion of the possible reasons for this will be postponed until later.

Difference scores (A - B) for E range between 1.706 and 5.720; in other words, there is a slight tendency for Ss to make themselves out as more extraverted when "faking good"; the extent of this tendency is only about $\frac{1}{3}$ S.D., which is almost negligible from the psychological point of view, although the unanimity of the shift (it is shown by all ten groups) leaves little doubt about its validity. For N the difference scores range from -2.403 to -5.530; all groups agree in faking a lowering of their neuroticism scores by over one S.D. This is a very large shift, although one might perhaps have expected an even larger one. On the L scores, too, there is a strong and unanimous shift towards greater lying, from between 2.053 to 6.384 points, corresponding to between 3 and 4 S.D.s; this is a large shift for all groups except the working class group. On the whole, these figures bear out our expectation of a large shift for N and L, and a small or non-existent shift for E. It will also be noted that the predicted shrinkage of the variance for E scores from A to B does actually take place for all groups but one. Variances for N are lower, and for L higher, under condition B as compared with condition A, indicating a shift towards greater uniformity among Ss with respect to N, and towards greater heterogeneity with respect to L.

The details of change from condition A to condition B are shown in Figs. 1, 2 and 3 for E, N and L respectively. In these figs. we have plotted the A scores of Ss in several of the groups against their change scores (A - B). Consider first Fig. 1. It will be clear that Ss with high E scores on A have lower E scores on B, while Ss with low E scores on A have higher E scores on B; the direction of the regression is indicated by the two lines enclosing roughly the middlemost 90% of cases. Out of 260 cases, only 51 (20%) go counter to this trend, the majority because of the slight general trend towards higher E scores under condition B. Fig. 2 points to quite a different situation for N scores; here nearly all the Ss show a trend towards lower B scores. There are only few scattered Ss in the quadrant below the abscissa (19 out of 260, to be precise; most of these had a 0, or anyway a very low score on condition A, so that any change might be considered due to the natural and inevitable unreliability of the scale). Fig. 3 shows exactly the reverse picture for the L scores; with very few exceptions (14 out of 260) there is a shift towards higher scores. While we have presented these trends only for four out of our ten groups, mainly in order to prevent confusion, the tendency in the other groups is identical, as will be obvious from the figure in Table 2.

We may now turn to a consideration of the relationship between the Lie scale and the E and N scales. Fig. 4 shows that an increased tendency to lie (assuming that it is this which is measured by the L scale) is associated with introversion; L scorers of zero have high E scores (30.6) while Ss scoring 7 or above have E scores of 24.6. The regression does not depart significantly from linearity and is highly significant by analysis of variance. E scores under condition B show no significant relation to L, and it follows of course that change scores (A - B) show a tendency for liars to have negative change scores, i.e. to pretend to greater extraversion; this tendency is highly significant by analysis of variance.

Liars pretend to less neuroticism than do non-liars, as is shown in Fig.5. There is a drop from a score of 7.1 (L score zero) to a neuroticism score of

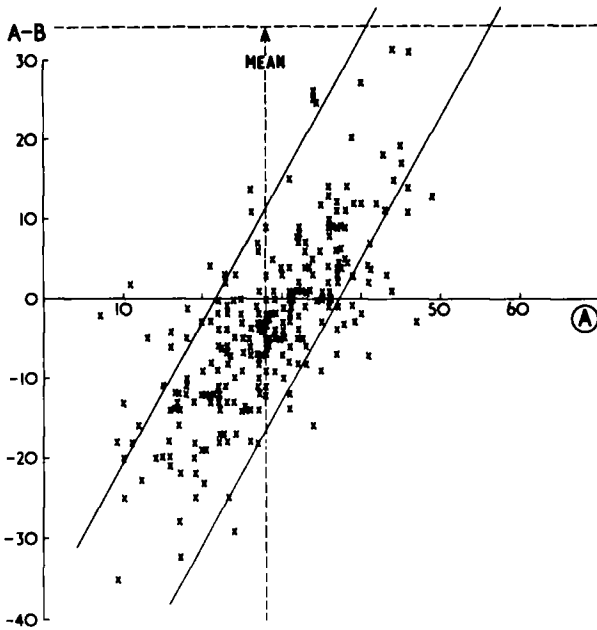
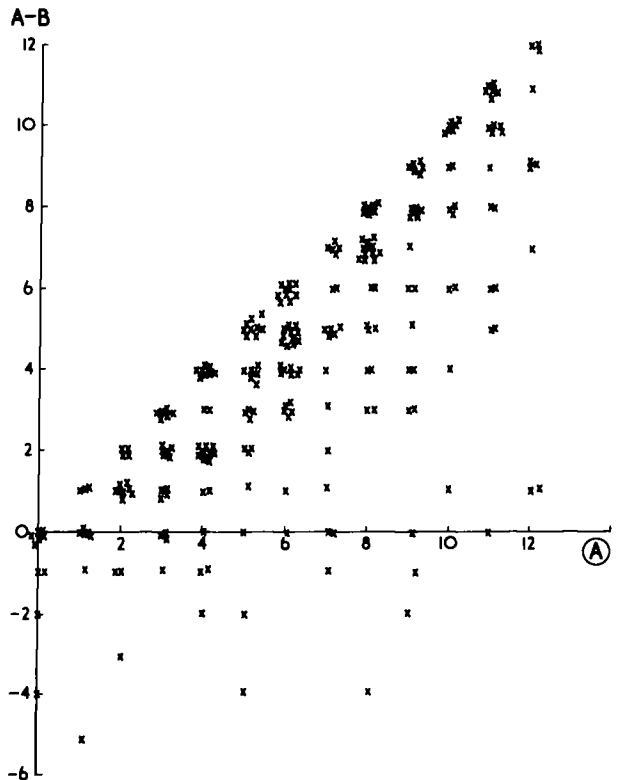


FIG. 1

A scores on extraversion plotted against change scores (A - B) for Sheffield, Exeter, Manchester, and Wales student groups.

FIG. 2

A scores on neuroticism plotted against change scores (A - B) for Sheffield, Exeter, Manchester, and Wales student groups.



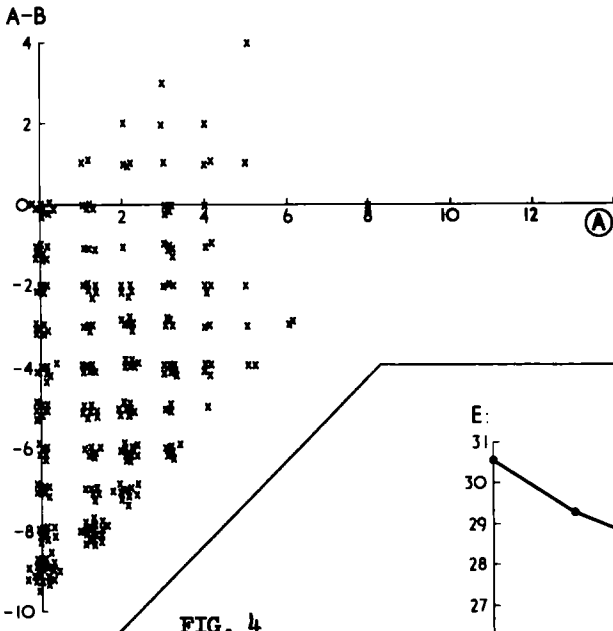


FIG. 4

Extraversion scores for subjects having various Lie Scale scores.

FIG. 3

A scores on Lie Scale plotted against change scores (A - B) for Sheffield, Exeter, Manchester, and Wales student groups.

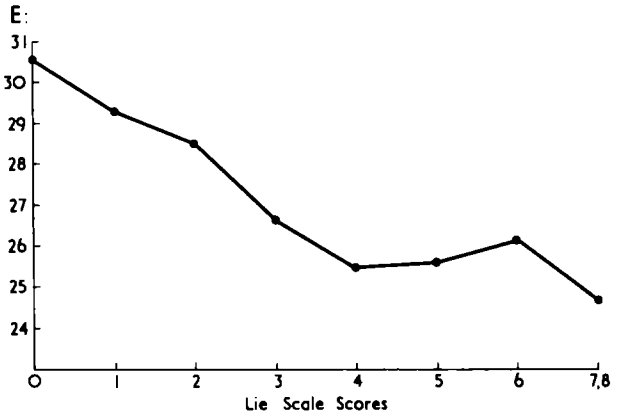


FIG. 5

Neuroticism scores for subjects having various Lie Scale scores.

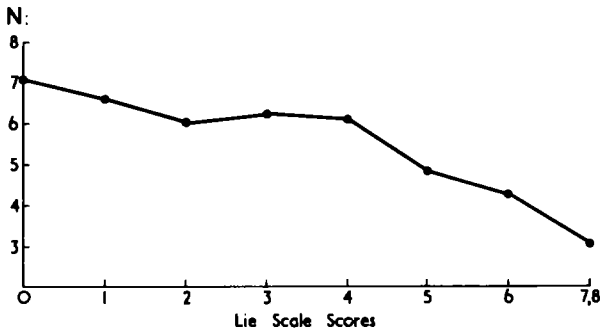


FIG. 6

Neuroticism change scores (A-B) for subjects having various Lie Scale scores.



3.0 for Ss with lie scores of 7 or above; this regression is linear and highly significant by analysis of variance. There is a non-significant relation between N and L under condition B, and a highly significant linear one between L and change scores (liars change least). This last relation is demonstrated in Fig. 6.

Discussion

We must first of all deal with the odd behaviour of the working class group which, while on the whole showing the same tendencies as the other groups, changed much less than the others with respect to N and L. Our own interpretation would link this with Orne's (12) concept of the "demand characteristic" of the testing situation, which may be assumed to be rather different for working class subjects, unfamiliar with the very notion of an "experiment", as compared with relatively sophisticated University students. Indeed, while the "faking" instructions were accepted readily by all the other groups, the working class Ss showed clearly that they thought this to be a crazy idea. This curious behaviour should certainly be followed up, and the adequacy of this hypothesis be tested; our data do not enable us to judge its value.

As regards the main enquiry, we find that no hypothesis is tenable which would ascribe a large portion of the variance of the scores obtained under condition A to desirability response sets, i.e. a tendency for Ss to give answers which were socially accepted, and making Ss out to have desirable characteristics. If this were so, then little change should have taken place from condition A to condition B; yet this change was quite large, particularly on the L scale. With a possible "desirability" score of 9 points, Ss only averaged a score of 1.5 under condition A; under instruction to "fake good" they produced a change to almost 6.5 equal to between 3 and 4 S.D.s. Clearly desirability played relatively little part in their answers to the questions. Changes in N scores point in the same direction. These results are in good agreement with the independent evidence showing that E and N scales are valid when compared with external criteria (13;4); no such validity could be

demonstrated if a large portion of the variance were due to desirability responses.

This does not mean, of course, that desirability responses are completely absent in our questionnaire answers; such an hypothesis would be rendered untenable by the relations demonstrated between L scores and the E and N scales. It will be remembered that "liars" pretend to less neuroticism than do non-liars; non-liars only catch up when instructed to "fake good". Extreme liars have already put themselves into the best possible light under condition A; they have no room left for "faking good", and accordingly do not change under condition B. These results can be explained best if we assume that the L scale is a good measure of "desirability" response set (faking good) on the part of Ss; only relatively few Ss engage in this practice to any significant extent, and these can be spotted by their high L scores. For the great majority, desirability response set may be disregarded. It is clear from Fig. 5 that there is little change in N score while L scores are between 0 and 4; it is with scores of 5 and above that there appears to be a strong tendency to "fake good". (This agrees reasonably well with Gibson's figures (7) who shows, with a scale twice as long, that a score of 10 is indicated as a cutting-off score; it also agrees with Eysenck's original suggestion (6, see also 14) to use 10 as the out-off point with the 18-item scale.) Only 46 Ss have scores on the L scale of 5 or above, thus leaving 95% of the sample as being relatively unaffected by desirability response set. (475 Ss, or well over half, have 2 scores of 0 or 1.)

The relation between L and E is a little puzzling if we follow this line of thought. We have found that under "faking good" conditions there is a slight trend towards extraversion, and one would have thought that liars would follow this trend and have higher E scores. The facts are actually opposite; liars have low E scores, and change correspondingly more towards extraversion under instruction to "fake good". We may here have a genuine relation between introversion and tendency to give desirable responses, which is only slightly attenuated by the somewhat greater desirability of extraverted responses. In

other words, we would ascribe the relation between lying and low N scores to the effects of the former on the latter (liars fake good), and we would ascribe the relation between lying and low E scores to the effects of the latter on the former (introverts lie more). This hypothesis cannot be proved to be correct on the basis of our data; it would require an independent investigation, specially planned for the purpose, to make an evaluation possible. An alternative hypothesis might be framed in terms of the greater socialization postulated by Eysenck (2) as characteristic for introverts. The "Lie" scale items all refer to unusually good behaviour patterns, such as never lying, cheating, losing one's temper, et cetera. It is possible that there may be genuine differences here between introverts and extraverts which might account for the observed relations between L and E. In support we might quote such investigations as those of Fine (15) showing that extraverts are guilty of more traffic violations and accidents than introverts, and of S. B. G. Eysenck (16) showing that unmarried mothers are more extraverted than married ones. The relationship between psychopathic behaviour and extraversion has been confirmed quite frequently (17), as has that between tender-mindedness and introversion (18). This alternative hypothesis might therefore repay investigation.

Summary

A questionnaire measuring extraversion, neuroticism and tendency to lie was administered to 10 separate groups of Ss under ordinary conditions and later under instruction to "fake good". The results suggested that intermediate degrees of extraversion and low degrees of neuroticism are generally preferred; that "desirability response set" played some part in the answers, but not an unduly large part; and that this response set could be measured by means of the lie scale. It was also suggested by the data that introverts are more likely to show "desirability response sets" than extraverts.

References

1. D. N. Jackson & S. Messick. Response styles on the MMPI: comparison of clinical and normal samples. J. Abnorm. Soc. Psychol., 65, 285-299 (1962)
2. H. J. Eysenck. The Maudsley Personality Inventory. London: University of London Press (1959).
3. H. J. Eysenck. Response set, authoritarianism and personality questionnaires. Brit. J. Soc. Clin. Psychol., 1, 20-24 (1962).
4. S. B. G. Eysenck & H. J. Eysenck. The validity of questionnaire and rating assessments of extraversion and neuroticism, and their factorial stability. Brit. J. Psychol., 54, 51-62 (1963).
5. S. B. G. Eysenck & H. J. Eysenck. Acquiescence response set in personality questionnaires. Life Sci. 2, 144-147 (1963).
6. H. J. Eysenck. Fragebogen als Messmittel der Persönlichkeit. Z. F. Exp. u. Angew Psychologie, 4, 291-335 (1953).
7. H. B. Gibson. The Lie Scale of the Maudsley Personality Inventory. Acta Psychol. 20, 18-23 (1962).
8. H. J. Eysenck. The questionnaire measurement of neuroticism and extraversion. Riv. di Psychol. 50, 113-140 (1956).
9. H. J. Eysenck & S. B. G. Eysenck. A factorial study of an interview-questionnaire. J. Clin. Psychol. 18, 286-290 (1962).
10. H. J. Eysenck & S. B. G. Eysenck. On the Dual Nature of Extraversion. Brit. J. Soc. Clin. Psychol. 2, 46-55 (1963).
11. H. J. Eysenck. A short questionnaire for the measurement of two dimensions of personality. J. Appl. Psychol., 42, 14-17 (1958).
12. M. T. Orne. On the social psychology of the psychological experiment. Amer. Psychologist, 17, 776-783 (1962).
13. S. B. G. Eysenck. The validity of a personality questionnaire as determined by the method of nominated groups. Life Sci. 1, 13-18 (1962).
14. A. R. Jensen. The Maudsley Personality Inventory. Acta Psychol. 14, 314-325 (1958).

15. B. J. Fine. Introversiion-extraversiion and motor vehicle driver behavior. Percept. Mot. Skills, 16, 95-100 (1963).
16. S. B. G. Eysenck. Personality, and Pain Assessment in Childbirth of Married and Unmarried Mothers. The J. Ment. Sci., 107, 417-430 (1961).
17. H. J. Eysenck. The Structure of Human Personality. London: Methuen (1960) (2nd. Ed.).
18. H. J. Eysenck. The Psychology of Politics. London: Routledge and Kegan Paul (1954).

Acknowledgements

We wish to gratefully acknowledge the help of the following in contacting and testing the various groups:

Dr. John Beloff (The Queen's University of Belfast);
Dr. J. N. A. Ridyard;
Dr. A. Martin;
Mr. K. Warwick;
Dr. G. W. Granger;
Dr. G. Claridge;
Mr. A. Dennington;
Dr. Peter McKellar (University of Sheffield);
Dr. Richard Lynn (University of Exeter);
Dr. A. Broadhurst;
Dr. A. Milner (Queen's University of Belfast - Law Faculty);
Dr. Judson Finley (Arizona State University, U.S.A.);
Prof. and Mrs. Staats (" " ");
Dr. John Gibbs (University of Cardiff);
Dr. F. Warburton (University of Manchester).

We also wish to acknowledge the support of the Bethlem Royal and Maudsley Research Fund & Dr. A. E. Maxwell for statistical advice and help.