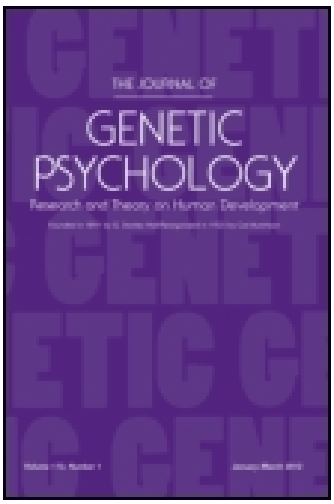

The Journal of Genetic Psychology: Research and Theory on Human Development
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/vgnt20

The Determination of Personality Inventory Factor Patterns and Intercorrelations by Changes in Real-Life Motivation

W. Michaelis [a] & H. J. Eysenck [a]
[a] Institute of Psychiatry, University of London , England
Published online: 04 Sep 2012.

To cite this article: W. Michaelis & H. J. Eysenck (1971) The Determination of Personality Inventory Factor Patterns and Intercorrelations by Changes in Real-Life Motivation, The Journal of Genetic Psychology: Research and Theory on Human Development, 118:2, 223-234, DOI: 10.1080/00221325.1971.10532611

To link to this article: http://dx.doi.org/10.1080/00221325.1971.10532611

The Journal of Genetic Psychology, 1971, **118**, 223-234.

THE DETERMINATION OF PERSONALITY INVENTORY FACTOR PATTERNS AND INTERCORRELATIONS BY CHANGES IN REAL-LIFE MOTIVATION*

Institute of Psychiatry, University of London, England

W. MICHAELIS[1] AND H. J. EYSENCK

A. INTRODUCTION

It is well known that scores on personality inventories can be changed intentionally by "faking good" or "faking bad"; there are many studies in which instructions to do either have produced changes in the required direction. The early work done in this field (2, 12, 13, 17, 18, 19, 21), as well as the more recent studies (6, 7, 8), on the Eysenck Personality Inventory (E.P.I.) show that personality traits relating to N (neuroticism) can be changed in the direction of greater stability by instructions to "fake good"; that L (Lie Scale) scores can be increased dramatically by the same instructions; and that E (extraversion) scores remain reasonably unaffected. Real-Life motivation to "fake good" seems to produce similar changes (14, 16, 20); when some of the respondents are filling in the inventory under the impression that the scores will count towards their acceptance or rejection for some position, while other respondents have already been accepted or are in safe tenure, then the applicant groups are found to have lower scores on scales related to N than the tenure group, while no differences were found for scores on scales related to E.

These studies, while interesting, do not permit us to consider in sufficient detail just what the changes are which are produced by real-life motivation. In the first place there is the possibility that motivation will affect the factor structure of the items which go to make up each of the scales involved; factor analyses of the kind which precede the construction of personality inventories are usually carried out in groups not motivated to present themselves in a particularly good light, and it is quite possible that "fake good" motivation might alter the factor structure of the items making up the various scales employed. In the second place, the intercorrelations between scales may be altered by changes in motivation; it seems unlikely that the correlation between N and L, for instance, will remain unaltered by a change in motivation which

* Received in the Editorial Office, Provincetown, Massachusetts, on September 10, 1969.

[1] Of Kiel University, Germany.

would be expected to increase L scores and decrease N scores. In the third place, it would seem that even if no changes take place in the factor patterns as a function of motivation, dramatic changes might occur in the size of the loadings; motivation to "fake good" should increase loading on L items, to take but one obvious example.

The present experiment was undertaken to investigate the differences in mean scores, factor structure, and intercorrelations between four personality scales, applied to applicant and tenure groups very similar in most other respects. The situation made use of has been described in "Experiments in Motivation" (3); essentially the high motivation (HM) group is made up of candidates for an apprenticeship course run by one of the largest motor manufacturing concerns in England, a course which is very much in demand, and for acceptance to which a series of selection tests is given (mostly of a perceptual-motor type, but also including intelligence tests). The inventories are given as part (ostensibly) of this selection procedure, but are not in fact used for determining the candidate's success or failure. The low motivation group (LM) is made up of apprentices already accepted at least six months earlier, and actually working for the company. These apprentices know that the scores on the inventories are used only for experimental purposes, and that results will not be communicated to the company; similar experiments have been going on for some 10 years or so, and the apprentices know all about the practice and are fully aware of the independence of the research organization from management. This situation is uniquely favorable for obtaining truthful answers, as the apprentices are not required to rely on the (unsupported) word of the psychologist that the results will be secret, but know from long-continued experience that this secrecy is in fact observed with respect to any of the tasks they have been required to carry out by the psychologist. Evidence regarding the marked differences in motivation between the LM and the HM groups on many divergent tasks, from pursuit-rotor reminiscence to eye-blink conditioning, have been reported in the literature (3); there seems little doubt that the HM group is in fact very highly motivated to do well by the situation in which they are placed, while the LM group does not have any motivation other than that intrinsic to the task.

The LM group is not, of course, identical in all important variables with the HM group, having been selected from former HM group members by a special process. Thus, they are about one year older; they are somewhat brighter; and they do better on various hand-eye coordination tasks. It is possible that the interview to which they submit selects out the more stable youngsters, but in view of the known fallibility of the interview this must

remain doubtful. In any case, our prediction is that the LM group would have higher N scores; if selection against N takes place, then the LM group would have N scores which would (without faking) be *lower* than those of the HM group. Thus, any such effects would work against our prediction. The other differences do not appear to be serious; the age difference is very small, and past work has shown that factor structure remains remarkably invariant over much longer periods (6). Differences in *IQ* or in mechanical and hand-eye coordination tests are not likely to influence the outcome, as these have not in the past been found to be correlated with personality (3). On the whole, we would suggest that the experimental group (HM) and the control group (LM) are reasonably well matched for carrying out the comparisons of their inventory responses.

B. The Experiment

1. Population. The experimental group consisted of 339 applicants to the Apprentice Training School of a large motor manufacturing concern; the control group consisted of 225 apprentices already employed in this Training School. The principles underlying this differentiation have already been discussed. Applicants were all school leavers 16 years or so of age; employees are on the average 1.1 years older. All are male, and nearly all are working class or lower-middle class by origin.

2. Test. The test used was the P.E.N., a personality inventory not yet available in published form. This incorporates 20 questions concerned with N (neuroticism), 20 questions concerned with E (extraversion), 20 questions concerned with P (psychoticism), and 18 questions constituting a lie scale (L). The N, E, and L items are very similar to those making up the E.P.I. (4); the P variable is a relatively new scale the construction of which has been discussed in some detail elsewhere (5, 9, 10). Its purpose is to measure potential for psychotic breakdown; it is a scale incorporating personality traits and not psychiatric symptoms. Psychotics have in fact been found to score very highly on this scale, compared to normals, and the scale has been successfully extended to children (11).[2]

3. Application. The test was given as a group test; all applicants are, of course, literate and well able to understand the complete inventory without help. Instructions emphasized the need to answer every question. The HM group were not told explicitly that the scores would be used to decide on

[2] A copy of the P.E.N. inventory and of Table A may be obtained by ordering NAPS Document 01216 from CCM Information Corp.—NAPS, 909 Third Avenue, New York, New York 10022; remitting $2.00 for microfiche or $5.00 for photocopies.

their application, but the placing of the inventory suggested without doubt that they considered it in the same light as the proper application tests which they knew would determine their success or failure. The LM group were told explicitly that the results would not be divulged to the management and, in any case, as already explained, they knew that this would be so from previous experience with our psychologist working there.

C. Results

The mean scores of the experimental and control groups are given in Table 1. Differences on E, as expected, are nonsignificant. Differences on P are significant, but relatively small; the large number of cases involved has obviously blown up this small numerical difference into a very significant statistical difference. It is with respect to N and L, however, that as expected the really large differences occur, with the experimental group having higher L scores (over twice as high as the control group) and lower N scores. The control group has scores which agree quite well with those of the appropriate standardization group (male working-class youths) as given in another article (10); it is the experimental group which deviates considerably from the standardization figures, leaving little doubt about the effectiveness of the motivating procedure.

TABLE 1
Mean P, E, N, and L Scores of HM and LM Groups

	HM		LM		
Dimension	*M*	σ	*M*	σ	*t*
P	1.58	1.57	2.15	2.10	3.54
E	14.70	2.63	14.37	3.41	NS
N	5.42	3.51	8.87	3.93	10.71
L	6.95	3.21	3.10	2.04	17.58

The intercorrelations between scales are reported in Table 2. It will be seen that correlations are similar, but there are also some significant differences. The P *vs.* L correlation is significantly higher for the experimental group; indeed it is nonexistent for the control group. The N *vs.* L correlation similarly is significantly higher for the experimental group, although a sizeable correlation does exist for the control group. The position is reversed for the E scale; the control group shows a significant correlation of —.25 while the experimental group shows an insignificant one of .02. Thus, high motivation produces an increase in the negative correlation between the L scale and the two mental illness scales, as expected; it also reduces to insignificance the (negative) cor-

relation between E and L. All the changes produced thus involve the L scale; the intercorrelations between the other scales are not affected.

A factor analysis was performed on the product-moment correlations between the 78 items on the inventory, for the two groups separately. Twenty-nine factors had eigenvalues exceeding unity, but as our interest lay primarily in the four factors corresponding to P, E, N, and L, only the first four factors extracted by the method of principal components were considered. These were rotated into oblique simple structure according to the Hendrickson-White "Promax" method (15), and the results of this rotation are reported in Table A.[3] Given in this table are the numbers of the items, the loadings of the items on the N, E, P, and L factors for the experimental and control groups, respectively, and the key to the items Y for yes and N for no. Justification for this method of procedure has been given elsewhere (5).

TABLE 2
CORRELATIONS BETWEEN FOUR INVENTORY SCALES

Scale	P	E	N	L
P	/	−.03	.41	−.20
E	.06	/	−.07	.02
N	.38	−.10	/	−.50
L	−.04	−.25	−.36	/

Note: Values for HM group above the leading diagonal, values for LM group below the leading diagonal. Underlining signifies a p level of .01.

The first obvious change in factor structure concerns the percentage of the variance accounted for. This is greater for the control group for N (5.6% *vs.* 4.7%), E (5.8% *vs.* 4.0%), and P (4.3% *vs.* 4.0%)—i.e., for the three personality variables; it is less for the control group for L(3.5% *vs.* 5.1%). Thus the personality factors come out more strongly and clearly in the LM group, the lie factor in the HM group. Are the factors in fact identical? This can be shown by calculating indices of factor comparison, with the use of a method suggested by Kaiser [in (6)]. These comparisons of factor structure are as follows: P = .77; E = .98; N = .90; L = .83. These figures are somewhat lower than are usually found (6), and for P and L the values fall below .90, suggesting that changes in factor composition have taken place as a consequence of introducing the motivating factor. However, on the whole, the effect of motivation has been rather less than might have been feared, and the recogniz-

[3] See footnote 2.

able factor structure of the LM group has been preserved in the HM group. A detailed comparison of item loadings show that while many items have preserved almost identical loadings (e.g., items 2, 14, 42 for N; items 5, 21, 25 for E; items 39, 43, 59 for P; items 4, 56, 64 for L), others have altered beyond recognition (e.g., items 6, 46, 54 for N; items 27, 41, 73 for E; items 19, 23, 77 for P; items 8, 12, 68 for L). Replication of the study would seem necessary to show whether these differences between items (i.e., changed or unchanged loadings) are chance effects, or whether there is something in the actual items which is responsible for their change or failure to change under change in motivation. If such replication should be successful, this would be of some importance as it would provide us with a possible method of attack on the otherwise rather intractable problem of "faking good", and suggest methods of measuring and even countering the effects of this practice.

The factors extracted and rotated are, of course, *oblique,* and the correlations between them are shown in Table 3. When these values are compared with the correlations between inventory scores (Table 2), it becomes clear that scores (e.g., P *vs.* N; L *vs.* N) correlate more highly than factors, although the general pattern is recognizably the same. P *vs.* L presents a reversal of this rule; in the HM group the factors correlate more highly than the scores. (For the LM group both correlations are insignificant.) On the whole it is doubtful if these comparisons add much to our knowledge, and their interpretation is by no means clear.

TABLE 3
CORRELATIONS BETWEEN FOUR FACTORS

Scale	P	E	N	L
P	/	—.03	.21	—.37
E	—.05	/	.00	.05
N	.14	—.06	/	—.16
L	.04	—.03	—.08	/

Note: Values for HM groups above the leading diagonal, values for LM groups below the leading diagonal.

In view of the fact that the Lie scale showed a considerable increase in proportion of variance accounted for with the introduction of motivation, it seemed desirable to perform a separate factor analysis of the 18 items of that scale alone, to discover whether it was in fact reasonable to consider it a single scale, or whether it would break up into several separate scales. The results showed that for both groups two factors did in fact appear; furthermore, indices of factor comparison between the groups were reasonably high (.998

for L_1 and .937 for L_2). However, the second factor is little more than a doublet (made up of items 8 and 44—i.e., "Once in a while do you lose your temper and get angry?" and "Do you sometimes get cross?"); and, consequently, we must conclude that a single factor is adequate to represent the L scale. (It is of some interest that L_1 and L_2 correlate .06 and .35 respectively for the LM and HM groups.) Table 4 shows loadings of the 18 items; the percentage of the variance accounted for is 19 percent for the experimental group and 11 percent for the control group.

TABLE 4

Item	Lie Scale	
	LM	HM
4	.38	.47
8	—.45	—.42
12	—.09	—.50
16	.42	.53
20	—.22	—.35
24	.39	.37
29	—.32	—.40
32	—.18	—.51
36	—.40	—.45
40	.16	.45
44	—.66	—.49
48	.00	—.32
52	.30	.47
56	—.36	—.53
60	.42	.24
64	—.21	—.53
68	.00	—.45
72	—.19	—.04

D. Discussion

The results of this study bear out the anticipation that changing motivation would result in changes in inventory responses and their patterning. These changes throw some light on the nature of the Lie scale, and the motivational factors which affect responses to it. Discussion of these may be related to an hypothesis published by Dicken (1), and an unpublished experiment from the Institute by Furneaux and Lindahl[4] designed to investigate this hypothesis. Dicken suggested that there are three possible causes for high L scores:

(*a*) Deliberate "faking" with intent to deceive the test user;

(*b*) Response in terms of an ideal self-concept rather than a candid self-appraisal;

[4] Personal communication; subsequent quotations are taken from this communication.

(*c*) **Response in terms of an "honest" but inaccurate and uninsightful self-assessment.**

Furneaux and Lindahl made use of the motivational situation also used in our study, but added three further groups, at two other factories, whose motivation they considered to be intermediate between the HM and LM groups similar to our own. Group 1 in their sample was similar (not identical; see below) to our LM group, and Group 5 was similar to our HM group. Their Group 1, unlike our subjects, "took part in the tests mainly because they were paid to do so, and because this gave them an excuse for missing a day's work, by visiting the Institute."[5] The apprentices making up group 2, in addition to coming from another factory, "were not paid, and lost comparatively little work-time by taking part in our experiment, which was conducted in their own factory. One would therefore assume that involvement in the tests was greater in group 2 than in group 1."[6] One might add to this argument that apprentices in the factory from which group 2 was taken had had no experience of psychological testing by outsiders, and consequently might be expected to have some residual qualification in mind regarding the communication of results to management. Group 3 was assured of noncommunication to the management of results, but also came from the factory where no previous testing had taken place; in addition they were not volunteers, as had been members of groups 1 and 2, so that their motivation might well have been raised by more definite suspicions about what the psychologists were up to. Group 4 was made up of selectees, as was group 5, but whereas in group 5 only a very small proportion of applicants was taken, nearly all those in group 4 were taken (this fact was, of course, known to the candidates, and depended on the particular factory at which they were applying). Accordingly, Furneaux and Lindahl argued that the groups were ranged in order from lowest to highest motivation, with group 1 having the lowest and group 5 having the highest motivation. The instrument used for testing was the M.P.I., to which was added a 38-item Lie scale; a score of 10 was used to split the population into "Liars" and "Nonliars."

Table 5 shows the increase in the proportion of liars with increase in (hypothetical) motivation, and the correlation between L and N for each group; correlations with E were not affected, as in our own study. It will be seen that these correlations become increasingly negative, as motivation in-

[5] See footnote 4.
[6] See footnote 4.

creases. Furneaux and Lindahl put forward the hypothesis that in group 1 "only those lacking insight gained high lie scores, and the similarity of the two distributions of neuroticism scores in this group (i.e., that of "liars" and that of "nonliars") suggests that such subjects have little systematic tendency to deny neurotic attitudes and behaviors. Within the group of applicants there must also have been a proportion of subjects gaining high lie scores for the same reason. In addition, however, because of the strong press of the selection situation, a substantial number probably resorted to deliberate lying, or were influenced by an increased tendency to respond in terms of an ideal self-concept. In contrast to lack of insight, these two attitudes do seem to result also in a strong tendency to deny neurotic characteristics. We can in fact postulate a general rule, to the effect that a large proportion of high lie scores, together with a strong negative correlation between scores for lying and neuroticism, will usually be found in groups completing tests under certain conditions of high motivation. Their presence, in fact, might even be used as a measure of motivation."[7] The regular progression of values for the intermediate groups in Table 5 lends some support to this notion.

TABLE 5
PERCENT OF "LIARS" AND TETRACHORIC CORRELATIONS BETWEEN L AND N FOR FIVE DIFFERENT GROUPS OF APPRENTICES IN INCREASING ORDER OF MOTIVATION

Group	Percent liars	r_{LN}
1	38%	.07
2	53%	—.36
3	56%	—.45
4	67%	—.52
5	77%	—.58

Our own data support this view and these results; the HM group has a significantly higher (negative) correlation between L and N (and also between L and P). In addition the proportion of "liars" is much higher in the HM group than in the LM group, as shown in Table 1. Our LM group does not approach the very low value of the r_{LN} correlation reported for Furneaux and Lindahl's group 1, but it should be noted that this group was by far the smallest used by them ($N = 40$), and tetrachoric correlations calculated on such small numbers of *Ss* are notoriously unreliable. Our LM group is in fact much more similar to their group 2, and it should be noted that both these groups were subjected to group testing at the factory, while group 1

[7] See footnote 4.

was in fact individually tested at the Institute, as already pointed out. But apart from this difference, the two sets of data agree remarkably well.

It would seem that our method of manipulating motivation enables us to discriminate two aspects of achieving high scores on the Lie scale: deliberate faking and lack of insight (which term may or may not include the rather imprecise notion of "ideal self-concept"). Increase in motivation increases faking, and also increases the negative correlation between L and N; the presence of such a correlation even in relatively LM groups suggests that motivation in these groups is still appreciable. In HM groups correction for high L scores becomes mandatory if results are to be used at all, and the relatively high correlation found suggests that such correction might be very useful. It may also be suggested that in LM *S*s high L scores are mainly indicative, not of deliberate faking, but rather of lack of insight, and that the L scale might with advantage be used as a separate personality scale, rather than as a mere correction device. Further research is, of course, required to investigate this possibility, but our data strongly suggest the feasibility of this suggestion.

One result of our investigation is rather surprising: namely, the fact that even under strong motivation to put themselves into the best light possible, HM *S*s still have L scores well below the maximum possible (7 out of 18), and N scores well above zero (5.4); one might have expected an almost complete elimination of N (and P), and an almost total L score. Part of this effect may be due to failure to realize just what is "desirable" and what is not, but for the most part it would seem that *S*s in our society are more honest than they are always given credit, and will admit weaknesses and faults even under conditions where these might lead to failure in a severely competitive employment situation. In the LM group lying is hardly a problem at all, except in a few cases, and even there the possibility exists, as mentioned above, that we are dealing with lack of insight rather than wilful faking.

E. Summary

A personality inventory containing scales for N (neuroticism), E (extraversion), P (psychoticism), and L (lie scale) was administered to 339 applicants for a much sought-after training course for industrial apprentices (high motivation group) as part of the selection testing, and to 225 similar apprentices already accepted (low motivation group). The HM group had higher scores on the L scale, and lower scores on the N and P scales, as predicted. Factor analyses of the intercorrelations between the items showed that similar

factors (N, E, P, L) emerged from both groups, but that the L scale accounted for more of the variance in the HM group, while the personality scales accounted for more of the variance in the LM group. In spite of the strong motivation to "fake good," the HM group still admitted many faults and weaknesses, and failed to "fake good" on the L scale to more than a limited extent. Motivation increased the correlation (negative) between L and N, and the data suggest that the L scale might with advantage be used as an empirical correction device, particularly in employment and selection situations.

References

1. Dicken, C. F. Simulated patterns on the Edwards Personal Preference Schedule. *J. Appl. Psychol.*, 1959, **43**, 372-377.
2. Ellis, A. Recent research with personality inventories. *J. Consult. Psychol.*, 1953, **17**, 45-49.
3. Eysenck, H. J., *Ed.* Experiments in Motivation. New York: Pergamon Press, 1964.
4. Eysenck, H. J., & Eysenck, S. B. G. The Eysenck Personality Inventory. San Diego, Calif.: Educational & Industrial Testing Service, 1965.
5. ————. A factorial study of psychoticism as a dimension of personality. *Multivar. Behav. Res.*, Spec. Iss., 1968, 15-32.
6. ————. Structure and Measurement of Personality. San Diego, Calif.: R. R. Knapp, 1969.
7. Eysenck, S. B. G., & Eysenck, H. J. An experimental investigation of "desirability" response set in a personality questionnaire. *Life Sci.*, 1963, **2**, 343-355.
8. ————. Desirability response set in children. *Brit. J. Educ. Psychol.*, 1966, **36**, 87-90.
9. ————. The measurement of psychoticism: A study of factor stability and reliability. *Brit. J. Soc. Clin. Psychol.*, 1968, **7**, 286-294.
10. ————. Scores on three personality variables as a function of age, sex, and social class. *Brit. J. Soc. Clin. Psychol.*, 1969, **8**, 69-76.
11. ————. "Psychoticism" in children: A new personality variable. *Res. in Educ.*, 1969, **1**, 21-37.
12. Faw, V. Situational variations of neurotic scores. *J. Consult. Psychol.*, 1948, **12**, 255-258.
13. Gough, H. G. Simulated patterns on the Minnesota Multiphasic Personality Inventory. *J. Abn. & Soc. Psychol.*, 1947, **42**, 215-225.
14. Green, R. F. Does a selection situation induce testees to bias their answers on interest and temperamental tests? *Educ. Psychol. Meas.*, 1951, **11**, 503-515.
15. Hendrickson, A., & White, P. O. Promax: A quick method for rotation to oblique simple structure. *Brit. J. Stat. Psychol.*, 1964, **17**, 65-70.
16. Heron, A. The effects of real life motivation on questionnaire response. *J. Appl. Psychol.*, 1956, **40**, 65-68.
17. Hunt, H. F. The effect of deliberate deception on Minnesota Multiphasic Personality Inventory performance. *J. Consult. Psychol.*, 1948, **12**, 396-402.
18. Kelly, E. L., Miles, C. C., & Terman, L. M. Ability to influence one's scores on a paper and pencil test of personality. *Charac. & Personal.*, 1936, **4**, 206-215.
19. Mais, R. D. Fakeability of the classification inventory scored for self-confidence. *J. Appl. Psychol.*, 1951, **35**, 172-174.

20. SEARLESS, J. R., & LEONARD, J. M. Experimentation in the mental testing of Detroit policemen. Detroit, Mich.: Detroit Bur. Gov't. Res., 1934.
21. WERMAN, A. G. Faking personality test scores in a simulated employment situation. *J. Appl. Psychol.*, 1952, **36**, 112-113.

Department of Psychology
Institute of Psychiatry
De Crespigny Park
Denmark Hill
London, S.E. 5, England