Brit. J. Pharmacol. (1959), **14**, 512.

CONSISTENT DIFFERENCES IN INDIVIDUAL REACTIONS TO DRUGS AND DUMMIES

BY

C. R. B. JOYCE

From the Department of Pharmacology, The London Hospital Medical College, London

(RECEIVED JULY 24, 1959)

The tendency of some individuals to report changes of physical and mental state after taking pharmacologically inert substances has been investigated experimentally. In a class of healthy medical students, those individuals who reported symptoms and those who did not made significantly different scores on a number of behavioural tests. The likely reactions of the members of a second class (containing none of the previous participants) to dummies were then predicted from their scores on the same tests, some of which were found to be much more efficient predictors than would have been expected by chance. Some implications for further research and for clinical medicine are discussed.

Textbooks and pharmacopoeias often imply that sex, age and weight account fully for the variance of human responses to drugs. A contributor to a recent symposium in which the testing of new drugs in man was discussed hoped that "we never come up against differences in human beings. That would make life intolerable." But it has been known for a long time that people differ in their reactions to alcohol (McDougall, 1929) and to salicylates (Hanzlik, 1913) as well as to a large variety of other drugs (see Clark, 1937, for examples) and even to the administration of inert substances (Wolf and Pinsky, 1954, described some particularly striking observations). Beecher (1955) has surveyed a number of earlier reports. However, very little experimental investigation of factors other than sex, age and weight seems to have been undertaken, although it is very probable that the value of clinical trials and the successful treatment of individual patients could both be increased by considering the aspects of the individual's mentality that influence his responses to drugs.

Responses to inert substances were chosen for the present investigation because placebos or dummies (Gaddum, 1954) are frequently used (not always by intention) in clinical and laboratory investigations; because they often have striking effects that, whether useful or undesirable, are not unimportant; and because it is more convenient to give dummies than drugs in the laboratory. The situations in which drugs or inactive substances are given, and the kinds of reaction that may be provoked, have been discussed elsewhere (Joyce, 1959a).

It will be shown that groups of reactors and non-reactors defined by means of their responses to dummies also have characteristically different scores on external criteria unrelated to such responses occurring in the laboratory; that scores on such criteria can predict future responses with more confidence than chance would allow; that the results obtained with healthy students here resemble those obtained in the clinical study of Lasagna, Mosteller, von Felsinger and Beecher (1954); and that the reaction-tendencies of healthy students to dummies resemble those to certain centrally-acting drugs. Future responses were more reliably predicted from responses that had already been made than would have been expected by chance (Wolf, Doering, Clark and Hagans, 1957).

The experiments fell into two phases: In the first, reactor (R) and non-reactor (N) groups were obtained experimentally, and their scores on psychological tests were compared. In the second, the test scores of a fresh group of subjects were used to predict their subsequent experimental responses. A brief account of the first phase has already appeared (Joyce, 1959a): the second phase was reported to the British Psychological Society (Joyce, 1959b).

METHODS

Each phase of the work (Table I) was performed with a completely different class of preclinical medical students, during their normal courses of

TABLE I
EXPERIMENTAL DESIGN

Expt.		Time
Phase One : A	Practical classes : students 1–30 and 31–59	Jan. 1958
B	8 trials of motion-sickness remedies : students 1–59	Jan.–March 1958
C	Psychological and other tests : students 1–59	Jan.–March 1958
Phase Two : D	Psychological and other tests : students 60–118 Prediction of response tendencies	Dec. 1958–Jan. 1959
E	Practical class : students 60–118	Feb. 1959

PRACTICAL CLASS PROTOCOLS

Expt. A		Expt. E	
Time (min.)		Time (min.)	
0–15	Observe symptoms and signs on self and partner	0–20	Symptom questionnaire. Observe symptoms and signs on self only. Pencil and paper tests
15	Drink unknown solution	20	Drink unknown solution
15–60	Continue observing	20–90	Continue observing. Pencil and paper tests
60–75	Symptom questionnaire	90–95	Short symptom questionnaire
75–24 hr.	Continue observing Symptom questionnaire	95	Subjects informed of treatments
		24 hr.	Symptom questionnaire

instruction in pharmacology and psychology. Each class contained 59 members: those in the first were in their third, those in the second in their second preclinical term at the start of the investigation. In the first class 48 were men and 11 women (age range 17.9 to 34.0, mean 21.0); in the second there were 50 men and 9 women (age range 18.1 to 29.1, mean 20.0).

Administration of Inert Substances

Phase One.—Each student was required, as an exercise in making and analysing clinical observations, to record at three-minute intervals respiration rate, pulse rate, and pupillary diameter in himself and a partner for 15 min. before and 60 to 90 min. after ingesting 25 ml. of an unknown liquid. The liquid only contained dilute proprietary orange juice, but it had previously been indicated to the students that the solution might or might not contain an active substance, whose nature (if present) was to be determined by its actions. The result sheets contained space to record additional signs, symptoms, or changes noticed during the experiment. After 60 min. this sheet was exchanged for a questionnaire upon 25 specific symptoms, and this in its turn for a further sheet on which to continue the record for as long as the subject liked. Twenty-four hours later, he completed a further short questionnaire upon long-term symptoms. As the experimental laboratory was small, two such sessions (one for each half-class) were held, one in the morning and one in the afternoon of the same day. The students had been told in the previous week that their reactions were of experimental interest, and that the class thus also involved a piece of serious research in which they were invited to collaborate. The question of volunteering did not arise. The students appeared enthusiastic, and there were only two absentees. The instructions for the practical class were cyclostyled, and the essential minimum of oral explanation was given when individuals asked specific questions. Students remained sitting throughout the class and the instructions discouraged conversation. Each class was supervised by a different demonstrator: that for the first knew that no solution contained an active ingredient, whereas the second did not; but, once the class had begun, neither demonstrator had any contact with the subjects save to collect and exchange their papers.

During the term in which these classes were held (expt. A), Dr. E. M. Glaser examined the side-effects of 4 substances with central actions on the same subjects (expt. B). For 8 weeks, each student measured his pulse rate and answered written questions (Glaser and Whittow, 1954) about 12 specific symptoms during a lecture at the beginning of which (in all save the first week) he had taken a capsule containing either cyclizine hydrochloride (Marzine, 50 mg.), meclozine hydrochloride (Ancolan, 25 mg.), perphenazine (Trilafon, 4 or 8 mg.), prochlorperazine (Stemetil, 13 mg.), or lactose (100 mg.). The drugs are referred to as A to E and F, and the dummy as D. In the first week, no capsules of any kind were given, and in the eighth the treatment of the second week was repeated. There were thus available reports of "symptoms" in the absence of treatment, symptoms following a dummy, and following each of 4 drugs (of which one was given at two doses and one treatment was repeated). In the same term the same subjects took certain simple psychological tests, and personal and academic information was obtained by examinations and further questionnaires. The latter are collectively referred to as expt. C, and are described below.

Phase Two.—The psychological tests previously used were given, with others, at the start of their second preclinical term to a fresh entry of students (expt. D). From their scores on these tests when compared with the results of Phase One, their probable reactions to a dummy in a subsequent practical class (expt. E) were predicted. The form taken by this practical class differed from that in expt. A, because some results from Phase One had been discussed with the participating students about eight months previously. Although at this time the subjects of Phase Two had not yet entered the College, it was considered unsafe to rely upon the division between senior and junior years as an effective barrier to the diffusion of information. Expt. E therefore formed part of a practical class upon the effect of "stimulant" and "depressant" drugs on autonomic responses and simple intellectual tasks. The subjects were told that control substances would also be employed. A single class was held in a laboratory large enough to accommodate all the subjects

comfortably at the same time. They had previously rehearsed the tasks, and now worked independently to a schedule timed for them orally by a single demonstrator. This required respiration and pulse rates and oral temperature to be measured at frequent specified intervals, for 10 min. before and 90 min. after the ingestion of 25 ml. of an unknown solution (see below). The activity of the palmar sweat glands was also estimated with iodine and starch-paper (Randall, 1946). At intervals the subjects carried out letter cancellations, simple additions, and a pencil and paper reaction-time test (Cherry, 1957) arranged for administration to groups. One administration of each kind of test came before and one after the unknown solution was taken. All the material (sheets for recording symptoms, pulse-rates, etc., and those for the tests) was set out at each subject's place before the experiment began; each was turned over and attended to on a signal from the demonstrator. Two spells of 15 min. each were free of experimental measurements, and the students were then allowed to read books that they had brought for this purpose.

The treatments were composed and allocated as follows. The basic medium of *all* solutions was a 1:100 dilution of concentrated compound decoction of aloes (B.P.C.). Of the substances it contains, only aloes itself is recognized to have any pharmacological effect (Goodman and Gilman, 1955), and the lower therapeutic dose to produce catharsis is 120 mg. The solution used here contained 10 mg., and the basic medium was therefore a dummy or more accurately an "impure" placebo. Such a medium was used in order to disguise the presence or absence of active ingredients and to give a characteristically "pharmaceutical" flavour and appearance. It seemed inadvisable to use orange-juice again. Tincture of belladonna was added to some solutions: to some others, caffeine citrate. These substances were used at two concentrations: the upper or average therapeutic doses (respectively of 1 ml. and 250 mg. in 25 ml.) and a 1:1,000 dilution of each (at which dose neither has any known pharmacological effect). The higher concentrations were given to the 9 students who had been absent from the psychological test sessions (expt. D) or whose results had provided insufficient information to predict their reaction to the dummies. The lower concentrations of belladonna or caffeine and the "control" treatment (containing only the medium) were allocated from a table of random numbers to the remaining 50 students. The solutions were drunk simultaneously, within the space of 2 min.: and the vessels were examined afterwards to see if, as was in fact always the case, they had been emptied.

After the observations described had been completed, the record sheets inquired whether each student thought he had taken the "stimulant" or the "depressant," and, if so, on what evidence. After the papers had been handed in the students were told what drugs had been used (but not the concentrations), and to whom each treatment had been given. They did not know the composition of the medium, and no opportunity was given for them to ask questions about the experiment at this point. Twenty-four hours later they completed the questionnaire on long-term symptoms used in expt. A.

Psychological Tests

It was expected from the clinical study already mentioned (Lasagna *et al.*, 1954) that R group members would be more sociable, less "critical," more prone to report symptoms in response to stress in general and more emotionally labile than members of the N group. It was therefore desirable that the information gathered should attempt to measure these characteristics in addition to intelligence and other standard attributes, and the need to carry out these intentions simply, quickly, and simultaneously for the whole group dictated the choice of tests. All the participants were assured that the information to be obtained was for experimental purposes only, and that it would be treated collectively, anonymously, and in the most strict confidence. This assurance was accepted and has not been betrayed: nor does it appear to have been doubted.

The tests that were used follow, roman numerals indicating the phase in which they were given.

(1) "Intelligence": (*a*) Test A.H.5 for administration to highly selected university and professional groups (Heim, 1956) (I, II); (*b*) a modified form of the Critical Thinking Appraisal (Watson and Glaser, 1952) (II).

(2) "Personality": (*a*) Bernreuter Personality Inventory, a test said to measure "dominance" and "sociability" (Flanagan, 1935) (I, II); (*b*) the Maudsley Personality Inventory (the short form of Eysenck, 1958); this estimates "introversion/extraversion" and "neuroticism" (II); (*c*) a modification of the Moreno Sociogram (Moreno, 1953) in which each student rated his relationship with every other in the class on a five-point scale in which the ranks had the following meanings: 0—name unknown; 1—known by sight; 2—occasional social contact; 3—frequent social contact; 4—a friend; 5—a close friend (I, II); (*d*) the Shapes Analysis Test of ability to handle geometric forms mentally (Cane and Horn, 1951) (I); (*e*) Assessment of Autonomic Activity Awareness in which each subject indicated, by marking off on continuous scales (the left ends of which were labelled "Never" and the right ends "Always" but which were otherwise uncalibrated) the extent to which he was aware of his heart action, body temperature, respiration, digestive tract movements, muscular tension and perspiration when in normal health, when anxious, when frightened, and after a pleasurable experience (Mandler, Mandler, and Uviller, 1958).

Information was also collected about performance in objective-test class examinations in pharmacology and statistics (Joyce and Weatherall, 1957) (I, II); "expectation" and judgment of subjects' own performances in class examinations (I, II); rating of preferences for learning by different methods: lectures, reading, and demonstrations, which have

been classified as "passive" methods (Joyce and Weatherall, 1959) and are less "sociable" than practical classes, seminars and tutorless discussion groups, which are also comparatively "active" as far as the student is concerned (I, II). Students also rated their enjoyment of the courses which they had attended, estimated the use each course had been to them, and recorded on individual charts how each working hour during these courses had been spent (I, II). Further information included statements about tobacco and alcohol consumption; willingness to volunteer for experiments of various kinds; self-rating of artistic, linguistic, mathematical, logical, and critical ability. Students in Phase Two were given a questionnaire, one term before experiments D and E, upon their personal use of drugs over the previous year.

Methods of Analysis

In Phase One, the characteristic distributions of the scores obtained on the tests described above by the whole sample and the R and N sub-groups were examined and an appropriate statistical test of the hypothesis that the scores did not differ was selected: this was usually "Student's" t, sometimes preceded by an analysis of variance. For some tests χ^2 or another so-called "distribution-free" statistic (usually the Wald-Wolfowitz runs test [Siegel, 1956]) was appropriate; these will be identified as necessary. Predicted and observed reactions were compared by χ^2 tests, or Fisher's test of exact probability.

Definition of a "Symptom": Method of Scoring Symptoms: Determination of Reaction Types

The rules to be followed were stated after the information in Phase One was collected and before Phase Two began. A symptom was defined as (1) any voluntary written statement about a physical or mental state implying a *change* in function or behaviour in comparison with the control (pre-drug) period; (2) each positive reply in any questionnaire indicating a similar event; (3) the answer "Yes" to either of the specific questions (in expt. C), "Do you think you have taken [either] the depressant [or] the stimulant drug?" One mark was given to any "symptom" so defined. An individual was held to be a reactor on a given occasion if he had a symptom-score of 2 or more or if he answered "Yes" to either question in (3) above. He was considered to be a non-reactor if he had a score of 0 or 1 or answered "No" to *both* questions in (3) above. In Phase One he was called a *consistent* non-reactor if he reported no symptom either in the practical class (expt. A) or on the occasion in the motion-sickness series (expt. B) when the dummy substance was given, a consistent reactor if he reported at least one symptom on each occasion or if his combined symptom-score for the two was 4 or more, and inconsistent if he fell into neither of these groups. In Phase Two the subject's response was considered to be equivocal if the answer "No" to both questions in (3) was accompanied by a symptom-score of 2 or more. Although inconsistent or equivocal responses are of interest, they have been excluded from detailed consideration below.

Results

Phase One

The number of symptoms reported spontaneously in expt. A and in reply to the questionnaires of expt. B formed J-shaped distributions. The latter contained only 12 questions, whereas the questionnaire in expt. A contained 30; and here the responses were fairly symmetrically distributed about a mean of 6 symptoms/subject (Fig. 1).

The total number of attendances in expt. B and of symptoms reported fell slightly, but steadily. Although there was a rather greater tendency for those who had previously made more reports of symptoms to drop out than for those who had

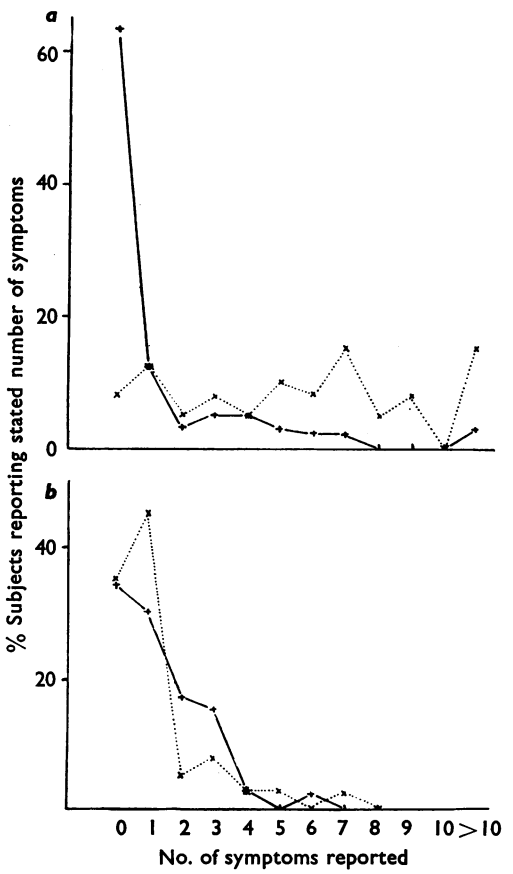

FIG. 1.—Proportion of subjects reporting symptoms. *a*, Expt. A, + —— +, Spontaneous reports (n = 59). X · · · X, Reports to questionnaire on 30 specified symptoms (n = 59). *b*, Expt. B, Reports to questionnaire on 12 specified symptoms, + —— + week 2 (n = 57); X · · · X week 8 (n = 40).

TABLE II

NUMBER OF SYMPTOMS REPORTED FOR EACH TREATMENT IN EXPT. B

Attendances (A) and symptoms reported (S) by whole class (Total), sub-group reacting to dummy (R) and sub-group not reacting to dummy (N); and number of symptoms expected (E) on hypothesis that there are no differences between treatments. D, dummy. For explanation of other treatments see methods.

Treatment:		None	A	B	C	D	E	F	Total	χ^2	P
Total	A	46	50	49	51	54	52	46	348		
	S	53	49	59	60	54	53	68	396	7·64	>0·2
	E	52·3	56·9	55·8	58·0	61·5	59·2	52·3	396		
R	A	24	29	23	28	[31]	28	25	[188] 157		
	S	37	28	28	41	[54]	33	40	[261] 207	6·10	>0·2
	E	31·6	38·2	30·3	36·9	—	36·9	33·0	207		
N	A	22	21	26	23	[23]	24	21	[160] 137		
	S	16	21	31	19	[0]	20	28	[135] 135	6·37	>0·2
	E	21·7	20·7	25·6	22·7	—	23·6	20·7	135		

made fewer reports, this was not significant ($P>0.1$), and there were only small differences in the mean number of symptoms reported for each treatment each week. The distribution of symptoms over treatments for all weeks taken together in the whole group was not significantly different from that to be expected by chance ($P>0.2$), nor were the distributions for the R and N groups taken separately when the responses to the dummy (D) (by means of which the sub-groups were formed) were omitted (Table II).

The mean number of symptoms/occasion reported for the dummy was indeed smaller than that to any drug except A, but the overall differences were insignificant. The number reported after taking drug F was significantly greater than that to the rest combined ($P<0.02$), but as a significance of this order could itself well be attained in one of the set by chance ($P>0.1$) when 6 comparisons are possible (Sakoda, Cohen, and Beall, 1954), the consistency of individual reports about the drug effects was examined on the assumption that the substances did not differ greatly in their pharmacological effects (Table III). The probability that the second reaction could be predicted from the first in any pair of successive occasions more accurately than could be expected by chance was on two occasions highly significant, as was the greater accuracy of the predictions summed over all occasions. The percentage of reactors fell after the start and rose towards the end of the series, but there were no systematic trends in the successful prediction of reactors or non-reactors. Just under two-thirds (63%) of all predictions made were correctly compared with 53% expected to be made by chance. It appeared that, if an interval of a week or longer elapsed between successive attendances, prediction was less successful and, in fact, no better than chance (57% against 56%). However, this sample was too small to justify discussion of the effects upon reaction-tendencies of temporary absence from the experimental situation.

TABLE III

CONSISTENCY OF RESPONSES IN SUCCESSIVE WEEKS

Reactions to drugs or dummy in first week of each pair compared with reactions in succeeding week.

Reactions Observed in First Week of Each Pair	Reactions Observed in Succeeding Week	R	N	% R	% Reactions Correctly Predicted in Succeeding Week: R	N	Total	Chance	P
1	R	22	8						
	N	10	10	64	69	56	64	53	>0·1
2	R	23	11						
	N	11	9	63	68	45	59	53	>0·5
3	R	18	7						
	N	6	12	56	75	63	70	51	<0·02
4	R	9	9						
	N	7	13	42	56	59	58	51	>0·5
5	R	7	6						
	N	6	16	37	54	73	66	53	>0·2
6	R	12	2						
	N	7	14	54	63	88	74	49	=0·0003
7	R	12	6						
	N	10	6	65	55	50	53	51	>0·8
(Total all pairs)	R	81	41						
	N	47	70	54	63	65	63	53	<0·001
Before absence	R	14	8						
	N	5	3	63	71	27	57	56	=0·765

Scores on Psychological and Other Tests

In all, the scores of the R and N sub-groups were compared on a total of 30 measures derived from the tests described above (including some, such as age and intelligence, that were not expected to discriminate between the groups). In a set of this size, about 3 results significant at or better than the 10% level would be expected by chance, of which 1 or 2 would also be significant at or better than 5%, but none at or better than 1%. The number of tests actually observed to reach each level was 12, 6 and 2 respectively

(Table IV: all the results attaining the 10% level or better are tabulated with others of interest). The likelihood of obtaining a constellation of scores such as this by chance is less than one in a thousand (Sakoda *et al.*, 1954).

The Bernreuter Inventory scores showed the R group to be insignificantly more "extraverted" and "sociable" and significantly less "self-confident" and "dominant" in personal relationships than the N group. These measures are known to be highly intercorrelated (Flanagan, 1935) and only the scores on the orthogonal factors of "self-confidence" and "sociability" have been considered here. The smaller degree of self-confidence of the R group was supported by the difference in judgments of their own performance by members of the two groups in the class examination, which was of borderline significance. The R group judged their own performance more severely, and indeed underestimated it: the N group overestimated theirs. There was no difference between the groups, however, on judgments about performances not related to class-work.

The greater "sociability" of the R group was also supported in several other ways: first, the sociograms showed that the R group gave a higher average "vote" to their class colleagues, and were also rated more highly by them; second, the R group indicated before the course began a greater preference for methods of teaching which depended upon group activity; and third, they showed after the course was over that they had valued such methods more highly. Finally, the other scores were taken to indicate that the R group were "less critical" of the teaching that they had received, in that the scores they awarded for their enjoyment of the course as a whole and for their view of the use it had been to them agreed to a significantly greater extent than did those of the N group, whose answers suggested either that they had enjoyed the course more than they had benefited from it, or the contrary, and so perhaps showed that they thought it deficient in one respect or another. The significant tendency of the groups to seek different amounts of further teaching may perhaps also indicate that the N group thought the course of less use. The number of "symptoms" reported in expt. B when no substance, not even the dummy, had been taken, the total number of occasions on which symptoms were reported, and the average number of symptoms/attendance were all greater for the R than for the N group, but the differences were of doubtful significance. There were no differences in the respiration rates or pupillary diameter of

TABLE IV

SCORES OF CONSISTENT REACTORS (R) AND NON-REACTORS (N) ON PERFORMANCE TESTS

All probabilities based on 2-tailed t-test except for those indicated by asterisk (χ^2 or Wald-Wolfowitz runs test): probabilities >0·1 shown to nearest lower 0·1: "X" in Phase Two indicates difference in non-predicted direction.

	Phase One				Phase Two			
	R	N	S.E. $(\bar{R}\text{-}\bar{N})$	*P*	R	N	S.E. $(\bar{R}\text{-}\bar{N})$	*P*
n: total	10–13	11–15	—	—	18	13	—	—
Men	11	13	—	—	11	13	—	—
Women	2	2	—	—	7	0	—	—
Age	20·1	20·6	0·58	>0·3	19·9	20·4	0·52	>0·3
A.H.5: intelligence	33·2	33·0	3·60	>0·9	32·1	34·8	4·15	>0·5
F.1: self-confidence	+6·7	−55·5	20·74	<0·01	−1·1	−28·7	30·10	>0·3
F.2: sociability	−22·8	−11·7	23·71	>0·6	−44·6	−30·1	20·54	>0·4
Sociogram: of others	128·8	120·0	13·25	>0·5	137·0	136·0	14·60	>0·9
Sociogram: by others	125·0	111·4	7·81	0·1>*P*>0·05	139·5	147·5	14·86	X
Shapes Analysis	17·6	18·4	1·94	>0·6	—	—	—	—
Class examination	56·0	57·7	4·78	>0·7	—	—	—	—
Judgment of performance	49·9	62·9	7·25	0·1>*P*>0·05	—	—	—	—
Private study/week (hr.)	8·2	8·4	2·26	>0·9	13·8	13·9	1·75	>0·9
Course use/enjoyment	1·0	0·8	—	<0·05*	1·4	0·8	—	0·1>*P*>0·05*
Requests for more teaching	3·6	3·8	—	<0·05*	—	—	—	—
Group methods preference:								
Before course	5·2	3·6	—	0·05>*P*>0·02*	6·3	4·3	—	>0·3*
After ,,	5·3	3·6	—	0·05>*P*>0·02*	—	—	—	—
Willingness to volunteer	5·7	6·3	0·83	>0·4	5·5	5·5	0·54	>0 9
Resting "symptoms" (no drug)	1·4	0·7	0·45	>0·1	4·8	4·1	0·80	>0·3
Occasions with symptoms	4·4	2·9	0·84	0·1>*P*>0·05	—	—	—	—
Symptoms/attendance	1·4	0·9	0·36	>0·2	—	—	—	—
Resting pulse	76·1	71·5	1·47	<0·01	82·5	81·2	2·92	>0·6
Resting pulse: coefficient of variation	9·9	6·6	1·70	0·1>*P*>0·05	7·7	9·2	1·22	X
% change after "drug"	7·7	5·0	1·38	0·1>*P*>0·05	4·3	7·1	2·00	X

the groups: the resting pulse rate of the R group was significantly higher, however, and showed greater variation: the fall in pulse rate immediately after taking the "drug" was also greater. The last two effects were of borderline significance.

No attempt was made to subject the results to factor analysis, because the tests were "notional" or at most empirical (Eysenck, 1959) rather than founded on theory. But real differences were clearly detected on occasions that were quite distinct from those in which the criterion reaction (to the dummy itself) was demonstrated, and it was this and the prognostic use of such tests that was important rather than the attribution of a label to what was being measured.

In Phase Two the tests were used prognostically. The psychological information about a second set of students was collected first, and from this the probable reaction of each to the administration of an inert substance on a subsequent occasion was predicted. The results of Phase One suggested that information from students whose native language was not English was not strictly comparable with that from their fellows, and such subjects were excluded *a priori* from consideration in Phase Two, although they sat the tests and attended the practical class.

The Bernreuter responses of the R and N groups in Phase One were subjected to an item analysis, and twelve questions to which the responses of the two groups were found to be characteristically different ($P<0.1$) were extracted. One mark was given for each response typical of the N group: and the scores of the fresh groups reacting and failing to react to the dummy in expt. E were significantly different. Scores on this set of questions (the "RN scale") were also used to predict, in advance of expt. E, whether each individual would or would not be a reactor to the dummy. The outcome (Table V) was highly successful, and may be compared with the predictions based upon previous individual responses in expt. B (Table III). No other single predictor (pulse rate, self-confidence, etc.) had better than 60% success, which was not significantly above that to be expected by chance; and no combination of three predictors in a discriminant was found that was as good as the RN scale alone.

TABLE V

PREDICTION OF REACTIONS IN EXPT. E FROM SCORES ON RN SCALE

Probability of this distribution (Fisher's exact test): <0.0002.

		Observed Reaction-type	
		R	N
Prediction: RN scale	R	16	4
	N	2	9

The right-hand part of Table IV gives the scores in Phase Two on those tests that were common to the two phases, and indicates whether or not the trend of the differences seen in Phase One was confirmed. This was the case in all save one of the psychological tests that were significant in Phase One, although the differences in Phase Two did not reach conventionally acceptable levels of significance for individual tests. But the probability of obtaining a set of results showing a consistent trend of this magnitude by chance is very small ($P<0.005$); and even if the results from the three tests upon the pulse rate (two out of three of which were in the unexpected direction) are included, the overall probability is still such that the set of observations is very unlikely to have arisen by chance ($P<0.03$).

Other tests were given for the first time in Phase Two (Table VI). Those who reacted to the dummy scored higher on the "Extraversion" and "Neuroticism" indices of the short Maudsley questionnaire, but not significantly so: and significantly higher on the autonomic questionnaire, thus admitting to a greater awareness (or willingness to report awareness) of such bodily events as bowel movements, sweating and so on, both at rest and under various kinds of stimulation, than did the N group. The groups did not recall significantly different numbers of occasions on which they had made use of drugs during the previous year, nor did they report different

TABLE VI

SCORES OF R AND N GROUPS IN QUESTIONNAIRES USED FOR THE FIRST TIME IN PHASE TWO

All probabilities based on 2-tailed t-tests: probabilities >0.1 shown to nearest lower 0·1. MPI, Maudsley Personality Inventory. BPI, Bernreuter Personality Inventory.

Test	R	N	S.E. $(\bar{R}-\bar{N})$	P
No. of participants	18	13		
MPI (short form)				
Neuroticism	+1·28	−0·08	1·205	>0·2
Extraversion ..	+0·52	−0·15	1·125	>0·5
Autonomic Awareness Questionnaire	90·9	69·0	10·22	$0.05>P>0.02$
BPI "no" answers ..	55·1	61·6	3·67	$0.1>P>0.05$
RN scale ..	6·8	8·8	0·97	$0.05>P>0.02$
Experience with drugs:				
No. taken in previous year	2·1	2·3	0·46	>0·6
Symptoms experienced/drug ..	2·5	2·4	0·42	>0·9
Symptoms expected ..	4·3	3·5	0·97	>0·4

recollections of the drug effects. There was a suggestion that the R group "expected" to experience more side-effects from any future administration of a drug than did the N group.

Save for the Autonomic Awareness test, which consists of interval-free scales that require no verbal response, the questionnaires force a Yes/No or "Don't know" choice to be made for each question. It seemed possible that the R group might be less willing to answer with a categorical "No," and this was found to be true for the Bernreuter Inventory as a whole.

Discussion

The reactions to inert substances of healthy young adult subjects were clearly much more consistent and predictable than would have been expected by chance, whether previous reactions or suitable independent pencil and paper tests were used to make the prediction. This cannot have been due to the fact that the critical experiments were not all "double-blind." It is true and it was inevitable that an experimenter sometimes knew which subjects were receiving the dummy, because these were always in the majority and sometimes the entire group. But the purpose of the critical experiments was to divide the group into those who reacted and those who did not, and the experimenter had no advance information to suggest which subjects would be which in expts. A or B (in the latter of which active substances were also used in a double-blind design) and he was unaware of the *predicted* responses in expt. E, for these had been worked out in coded form. The success of the methods used to predict responses might well have been even greater had the experiments in the two phases resembled each other more closely. The reasons for deliberately varying them have already been discussed and attention has been drawn to the influence of the length and form of questionnaires upon the distribution of the responses, as well as that of whether questionnaires are used at all. The degree of success achieved is therefore satisfactory.

The experiments do not show whether reactions to drugs resemble those to dummies: the very small difference in the number of reactions to the dummy and those to the active agents in expt. B (Table II) suggested that the latter, with the possible exception of drug F, were indistinguishable from dummies, and the fact that these reactions could be predicted as accurately (although by different means) as those in expt. E is therefore not strong evidence on this point. On the other hand the resemblance of such psychological information as has been associated with the specific reaction tendencies of these students to that obtained by Lasagna and his colleagues (1954) from their surgical patients suggests that the clinical and experimental situations can be validly compared. Again, the clinical information was obtained in very different ways (psychiatric interviews, evaluation by nursing staff and Rorschach testing) from those used here (objective measures) and the success of the comparison is therefore encouraging. However, it is as easy and dangerous in the field of "personality" studies to multiply entities as to call different entities by the same name, and a study is now in progress in which the measures derived from healthy subjects are being applied to patients.

No attempt has been made to find "labels" for the differences that have emerged by applying factor analysis without a basis of previous theory: but the results strongly suggest that R and N reactions or reactors constitute the extremes of a continuous distribution with the inconsistent reactors occupying a middle position, for the tests that distinguish significantly between the two groups always, and the remainder frequently, give group mean scores that lie on opposite sides of the mean for the whole group of 59 subjects. Three tests in particular are efficient discriminators: the RN scale, the Autonomic Awareness questionnaire, and the actual response to a dummy: with their aid it seems likely that the factors in the subject and in his environment that combine to produce in him a tendency to react or fail to react can be investigated in the future. The identification of these factors is likely to be worth while for the reasons discussed below, and it is encouraging that they can be studied relevantly in the experimental laboratory.

First, in the clinical evaluation of new remedies. Well-designed clinical trials compare new treatments with the best currently available or, if none yet exists, with a regime that is similar in all respects to the experimental, save that the agent under investigation is omitted. If the treatment investigated is actually effective, a high proportion of reactors (in relation to that present in the population at large) in the experimental groups may tend to decrease the apparent value of the drug. Too many non-reactors in the experimental sample, on the other hand, will tend to cause disappointment when the new drug is brought into general use, because the difference between the results of treatment with the active and control substances has been accentuated in the trial. Self-controlled trials are not always feasible:

and although wherever possible this design should be adopted it will not in fact eliminate the difficulty under discussion. The alternative use of matched groups for treatment and control will not be satisfactory if it does not include a simple test of reaction tendency: indeed, whichever design is used it is probably necessary to show that the proportion of reactors and non-reactors in the group (or groups) resembles that in the general population. It has sometimes been proposed that reactors or even non-reactors should be excluded from clinical trials, but the consequences of this would appear to be even more misleading: though even if this were not so the need to test reaction-tendencies would not be avoided.

Second, in deciding upon the correct treatment for the individual patient. There are some circumstances in which the withholding of a potentially dangerous drug, provided that equal benefit can be obtained from an inert or harmless treatment, is not merely justifiable but probably correct. The truth of this proposition must be examined separately for each patient, but the condition from which the patient to be tried on placebo is suffering should be non-terminal and preferably spontaneously remitting, speedily improved by treatment with the active substance should the placebo fail, and not adversely affected by initial failure to give stronger medication.

The consequences of initially assigning the individual patient to the wrong category must be considered. He may in fact be a reactor, although classified as a non-reactor, or the contrary. In the first case he will be unnecessarily treated with the active drug, in the second incorrectly with the placebo, and the implications of the errors are different. The first course will expose him to the risk of side-effects, addiction, etc.: the latter may cause him to suffer slightly longer than would have been the case had he been correctly placed. No biological predictive instrument can be expected to have an efficiency that approaches 100%, so errors of both kinds will inevitably be made. But such errors are continually and inevitably being made therapeutically, and the use of a rational instrument of the kind outlined here should in some circumstances enable them to be reduced. In fact, errors of misclassification in the experiments reported above amounted to about 37% (Table III) and 19% (Table IV): in both cases, each kind of error contributed about 50% of the total. These errors should certainly be further reduced when the predictive technique is refined.

Third, in investigating adjuncts to other forms of therapy. The reaction to an active drug depends upon the drug itself, the patient's history, his present state, his doctor and other ministrants (Feldman, 1956; Sabshin and Ramot, 1956; Uhlenhuth, Canter, Neustadt and Payson, 1959). The effect of even active drugs can be increased by giving them in the most suitable circumstances, and information about the psychic factors in the patient predisposing him to react favourably is therefore as important to the doctor as knowledge of the patient's body temperature, childhood illnesses and attitude to his job. Such factors may perhaps not be more difficult to estimate reliably than the latter.

Finally, as far as human pharmacology is concerned, this kind of study may illuminate some causes of variation in controlled treatments that are usually ignored or accepted with resignation. The main emphasis in psychopharmacology until now has been on drugs rather than souls: less interest has so far been shown (with notable exceptions such as Eysenck (1957)) in using drugs to display and estimate differences between healthy nervous systems, but the occurrence of reactions to inactive substances is important to studies of both kinds.

I am very grateful to Dr. A. W. Heim, of the Psychology Laboratory, University of Cambridge, to Dr. S. L. Morrison, of the Medical Research Council Social Medicine Unit, and Dr. E. M. Glaser, of the Physiology Department of the London Hospital Medical College, for advice and critical discussions. Special thanks are due to Dr. Glaser for his generosity in making his observations on the first group of students available to me. Mr. L. Hudson very kindly prepared some of the test material.

References

Beecher, H. K. (1955). *J. Amer. med. Ass.*, **159**, 1602.

Cane, V. R., and Horn, V. (1951). *Quart. J. exp. Psychol.*, **3**, 133.

Cherry, E. C. (1957). *On Human Communication: a review, a survey and a criticism.* London: Chapman and Hall.

Clark, A. J. (1937). *Brit. med. J.*, **2**, 307.

Eysenck, H. J. (1957). *The Dynamics of Anxiety and Hysteria.* London: Routledge and Kegan Paul.

—— (1958). *J. appl. Psychol.*, **42**, 14.

—— (1959). *Recent Progress in Psychiatry*, **3**, 118.

Feldman, P. E. (1956). *Amer. J. Psychiat.*, **113**, 52.

Flanagan, J. C. (1935). *Factor Analysis in the Study of Personality.* Stanford: University Press.

Gaddum, J. H. (1954). *Proc. roy. Soc. Med.*, **47**, 195.

Glaser, E. M., and Whittow, G. C. (1954). *Clin. Sci.*, **13**, 199.

Goodman, L. S., and Gilman, A. (1955). *The Pharmacological Basis of Therapeutics*, 2nd ed. New York: Macmillan.

Hanzlik, P. J. (1913). *J. Amer. med. Ass.*, **60**, 957.

Heim, A. W. (1956). *Manual for the Group Test of High-grade Intelligence A.H.5.* London : National Foundation for Educational Research.

Joyce, C. R. B. (1959a). *Proc. Int. Meetg Neuropsychopharmacol.* In the press.

—— (1959b). *Bull. Brit. psychol. Soc.*, **38**, 12A.

—— and Weatherall, M. (1957). *Lancet*, **2**, 402.

—— —— (1959). Ibid., **1**, 568.

Lasagna, L., Mosteller, F., von Felsinger, J. M., and Beecher, H. K. (1954). *Amer. J. Med.*, **16**, 770.

Mandler, G., Mandler, J. M., and Uviller, E. T. (1958). *J. abnorm. soc. Psychol.*, **56**, 367.

McDougall, W. (1929). Ibid., **24**, 293.

Moreno, J. L. (1953). *Who Shall Survive?* Revised ed. New York : Beacon House.

Randall, W. C. (1946). *J. clin. Invest.*, **25**, 761.

Sabshin, M., and Ramot, J. (1956). *Arch. Neurol. Psychiat. (Chicago)*, **75**, 362.

Sakoda, J. M., Cohen, B. H., and Beall, G. (1954). *Psychol. Bull.*, **51**, 172.

Siegel, S. (1956). *Non-parametric Statistics for the Behavioural Sciences.* New York : McGraw-Hill.

Uhlenhuth, E. H., Canter, C., Neustadt, J. O., and Payson, H. E. (1959). *Amer. J. Psychiat.*, **115**, 905.

Watson, G., and Glaser, E. M. (1952). *Critical Thinking Appraisal Manual.* New York : World Book Co.

Wolf, S., and Pinsky, R. H. (1954). *J. Amer. med. Ass.*, **155**, 339.

—— Doering, C. R., Clark, M. L., and Hagans, J. A. (1957). *J. Lab. clin. Med.*, **49**, 837.