Number

The Journal of Parapsychology

CONTENTS

AN ANALYSIS TO DETERMINE A TEST PREDICTIVE OF EXTRA-CHANCE SCORING IN CARD-CALLING TESTS[1]

Charles E. Stuart
Parapsychology Laboratory, Duke University.

Abstract: This analysis originated in an attempt to find an objective test, not involving correctness of identification of card symbols, by which tests of card-calling might be classified into two groups: A, a class in which chance scoring prevailed; and B, a class in which an extra-chance degree of correct card identification was likely to occur. The test decided upon was the degree to which the subject's estimates of his unknown score in a card-calling test was influenced by his knowledge of his just previous score. This influence was measured by a correlation between the first 10 card-calling scores of an experimental session and following estimates; the correlation being designated a T-score. The card-calling score for the subject session was the deviation from chance expectation of the remaining runs in the session. The data consisted of subject sessions in two series. Series I consisted of card-calling tests in which the subjects called the order of cards before they were shuffled randomly; Series II consisted of CC tests in which the subjects called already established random orders of symbols.

When the T-scores were separated into two classes: the mid-range from -.20 to +.20, in which estimate judgments were negligibly affected by previous scores, and other ranges in which some effect of the previous score could be considered probable, it was found that the concomitant card-calling scores of Series I were separated into a significantly extra-chance number of correct calls

1. This report varies in only minor respects from a Ph. D. dissertation of the same title accessible in the Duke University Library. I am deeply indebted to my colleagues, Dr. J. G. Pratt, Dr. J. L. Woodruff, Mr. B. M. Smith, and Miss Lois Hutchinson for generous permission to use some of their observations.

It will be noted that I have avoided the use of the concept of extra-sensory perception in this report. I have done so as a disciplinary exercise, without the intention of any general implication regarding ESP. In a later paper, to be concerned with obvious generalizations of the T-test relation, the question of appropriate terminology will be more amenable to discussion.

for the mid-range class and a chance number of correct calls for the other ranges. The finding was exactly verified in a similar classification of Series II.

When a measure of "success" and "failure" was applied to the individual runs of the T-test, it was found that little evidence of success resulted, but there was some indication of the effect of failure. The efficacy of the A and B classification broke down when the subject overestimated his scores too frequently. It was observed that subjects whose average estimates were equal to or below the chance-expected average score manifested CC scores significantly higher than those whose estimates were above the chance expectation. It is concluded that the absence of "specific score affectability" in the subject session is a psychological condition enhancing the probability of extra-chance scoring. An analysis of variance study of the pre-shuffle card-calling tests of Series I gives individually extra-chance results. It is noted that these results support, therefore, the precognition hypothesis as proposed by Rhine.

Introduction

In 1934 J. B. Rhine reported the results of card-calling tests conducted over a period of several years (13). In these tests subjects were presented with cards bearing one of five symbols. The subject's task was to call or guess the symbol on the face of the presented card. Since the symbols were presented visually screened and in a random order, the number of successful calls expected if the subject could not in any way identify the given symbol could be computed and compared with the number actually observed. If the observed number were sufficiently different from the chance expectation, it could be directly concluded that the subject's calls were related in some orderly way to the random stimuli. A direct inference from this relation would be that the subject had some cognitive experience associated with the card symbols.

Rhine found consistently extra-chance results in the card-guessing tests and proposed the term "extra-sensory perception" to denote the inferred cognitive process. Later investigators utilizing Rhine's procedures observed similar results and generally adopted his terminology. Thus the expression "extra-sensory perception" and its abbreviated form, "ESP," have come to characterize the general field of problems concerned with the

ability of subjects to make extra-chance scores in card-calling tests wherein possibilities of sensory knowledge have been excluded.

The case for the occurrence of an ability of subjects to score at an extra-chance rate in card-calling tests has been fully summarized in *Extra-Sensory Perception After Sixty Years* (16). In that volume the problem has been formulated in the question: "Is it possible repeatedly to obtain results that are statistically significant when subjects are tested for knowledge of (or reaction to) external stimuli (unknown and uninferable to the subject) under conditions that safely exclude the recognized sensory processes?" (16, p. 15) It is noted that this formulation excludes many easily adopted analogies of explanation which assume characteristics of ESP that have not been experimentally established; such as, for example, normality of distribution as an ability, consistency of test performance by the individual subject, etc.

But any attempt to get at the strictly psychological relations of extra-chance card-calling meets with difficulties. This is especially due to the lack of any theoretical system within which to approach the relation fo this field to other fields of problems in psychology which have already been systematically organized. Indeed, although there is considerable agreement among psychologists that these studies belong within the realm of experimental psychology (21), their relation to more fully explored fields of problems remains so obscure that *Psychological Abstracts* lists the literature under the awkward classification of "Psychoanalysis, Dreams, Hypnosis."

The uncertainty of the place of card-calling experiments within the domain of psychology arises in large part from their very recent introduction as psychological experimentation. The card-test procedure instituted by Rhine in 1930 has been the most successful method for the observation of extra-chance performances. But among other findings the ESP investigators have noted that there is great likelihood that the underlying cognitive process is itself erratic and inconstant (16: pp. 312-314). No experimental situation has been discovered in which the conditions may be defined to guarantee with high probability that successful extra-chance scoring may be attained. Only positive results have been amenable to strict interpretation. While it has been possible to set up hypotheses to explain why certain experiments yielding chance results failed to

produce extra-chance scoring, such speculation is usually unconvincing in the face of the simple interpretation of a chance result as signifying absence of orderly relation among the data.

Similarly, even in experimental results significantly extra-chance, much of the statistical material will represent merely chance variation with perhaps only a small unisolable part representative of the functioning of an underlying cognitive process.

The task, then, of qualifying the descriptive definition of an extra-chance card-calling ability by observation of conditions of successful performance and conditions of unsuccessful performance, of noting and establishing concomitant variables of all sorts, is a broad exploratory field of research. The establishment of any significant difference in condition permits more possible insight into fruitful analogies of explanation and relation to other systematized knowledge.

Many conditions favorable and unfavorable to extra-chance scoring performance have been observed with varying degrees of systematic approach (16: Part III). These include such wide ranges of variables as the effects of drugs (16: pp. 287-289), of stimulus distance (12), and of meteorological conditions (2). No one finding is able unequivocally to establish a relation, but each suggests some limitation upon the analogies with which we approach understanding of an inferred cognitive process. For example, the drug effects show the similarity of such processes to other "higher mental processes," the distance effects show that the inferred cognitive process is unlike auditory perception.

The present analysis is another step in this exploratory program. It is concerned to find whether there are consistent individual differences amenable to psychological explanation in the card-calling test situation which vary significantly with the test scores. Whether these differences are temporary, or permanently associated with the individual subject, will be considered of secondary importance.

In order to accomplish this, it is necessary to have observations additional to the card-calling scores during the experiment. The writer conceived the usefulness of asking the subject after each run to "guess" what his score would be upon checking.[2] The estimate

2. Actually, the presentation to the subject of this second task is not an unimportant experimental item, since it provides a

was asked for in such a way as to suggest its relative unimportance in respect to the main task. The estimates provided a statistic which could be considered representative in some way of the subject's reaction to his experience with the scores he was making. The interrelations of these estimates with the scores made appeared to provide a basis for investigating several questions concerning the subject's responses to the experimental situation.

In the following discussion, one resultant measure--the relation between the test score and the just-succeeding estimate--is used to discover a relation concerning card-calling scores. In order to keep the formulation of this problem clear, the actual insight leading to the relation noted is left out of the discussion of the analysis to follow but is pertinent here.

This writer was interested in the hypothesis that estimate-changes might show characteristic emotional responses by the subject to his experience with high or low scores. It was desired to see if a measure of "affectability" could be found. The first measure, tried on one series of the data, was the shift of estimate with a preceding change of card-calling score. It was noted that if this shift were directly correlated, the subject seemed to score lower than if the shifts were apparently unrelated. A second measure, a score-to-estimate correlation, was instituted because it included the extent as well as the direction of changes. But when these were compared with the card-calling scores which contributed part of the correlation, it was considered that there might be a spurious relation. A high scoring subject might be more "affectable" just because he was scoring high or because the experimenters evidenced greater interest.

It was then conceived as necessary to set up two independent tests such as those used in the following analysis. The first part of the test period with a subject could be limited to providing data for known psychological variables; the second part would constitute the independent card-calling test with which they were to be compared.

(Footnote continued) "distraction" or separate focus for the attention important to, but readily distracted from, the main task of card calling.

The Data

The data of this report are the observations of a number of different ESP investigators within experimental work primarily intended to study relationships other than those to be examined in this study. The investigations may be grouped into two classes in terms of the stimulus-response relation. The first class includes pre-shuffle card-calling in which the subject responses were compared with stimuli orders determined randomly after the subject responses were made. The second class includes subject responses to stimuli existent at the time of the response. The data from experiments utilizing pre-shuffle calling will be denoted as Series I; the data from experiments utilizing post-shuffle calling will be denoted as Series II. The stimuli were ESP symbols on regular ESP cards.

The total data were included in order to avoid questions of sampling. Since experimenter-subject relations are thought to be of importance in card-calling experiments, it seemed feasible to include the work of various investigators. The pooling of all observations likewise permitted the greater security of larger numbers.

Series I is made up of subseries from six experimental projects. These subseries will be identified as A, B, C, D, Hp, and P. Series II is made up of four subseries: O, E, S, and Hc, constituting observations from four different experimental projects.

In Series A, eleven college student subjects were tested in 18 subject sessions. The experimental project was the first of four planned to study the effect of fixed rates of movement in the subject's matching response in card-calling tests in which the responses were to be made before the card order was randomly determined (pre-shuffle calling).

On a table provided with a screened-touch-matching screen[3] were five blank cards in a row. The subject was instructed to point with a small pointer (a plastic penholder) to the blank cards, the pointing to indicate that the experimenter should place the top card of a shuffled deck of 25 ESP cards opposite the blank card. The pointing was to signal the subject's intention that the unseen card designated would match the symbol to appear in the blank space at the end of the run. A run

3. This screen is illustrated in Pratt and Woodruff (11).

consisted of 25 such designations, after which the cards in the experimenter's hand would be distributed in piles opposite the blank "key cards." The rate of response was fixed by the beat of a metronome.

After completion of the run, the experimenter asked the subject to guess how many hits would be tallied in that run. The subject's estimate was recorded. The second experimenter, stationed in a partitioned cubicle wholly screened from the experiment table, started a Merritt shuffler (18) immediately after the end of the run. After three revolutions of the shuffler, the cards were removed and the first five different symbols were placed in order in five boxes of a tray, and the serial number of the run was fixed to the end of the tray. During this procedure, the first experimenter had removed the screen and had placed the five piles of cards face up on the table, each symbol isolated and in line with the designated blank card. After the layout was complete, the first experimenter placed the tray over the space occupied by the five blanks in such a way that the tray cards lay just over the blanks. The first experimenter then photographed the table lay-out. In the photograph would appear the five key cards in the tray, and below them the 25 cards of the run arranged as the subject had matched them. The hit score was counted individually by the subject and the two experimenters, and, when all agreed, this score was recorded by both experimenters. A deck of cards shuffled by the second experimenter was exchanged for the gathered-up cards. After the screen was replaced, this deck was cut with a paper knife and the subject instructed to be ready for the next run. Experimental engagements were made for one hour, but in a few cases the subject worked for a shorter period. Customarily the session involved 24 runs, although one consisted of as few as 12, and others as many as 28. The number of runs was necessarily a multiple of four.

A schedule of rewards enabled the subject to win 50 cents if he scored 10 hits, $1.00 if he scored 11 or 12. (Higher rewards were offered for higher scores but no subject scored above 12 in these experiments.) He was also to be rewarded 25 cents if, in the counting procedure, both experimenters announced an incorrect count.

The ritual of the experiment was designed to objectify precaution against a number of conceivable possibilities of error or misinterpretation and thus to render any results of the experiment equal in validity to the most secure ESP research heretofore published.

The STM screen blocked subject vision of the matching cards, and the cubicle blocked his vision of the randomizing procedure with the key cards. His only experience with these cards was to see their faces during the check-up. All shuffling occurred out of sight of the experimenter or subject. The photographic record precluded error of observation in recording the distribution of cards; the three independent counts, with the subject motivated by promise of reward, assured a high degree of accuracy in duplicated data. Each experimenter was individually responsible for all records, yet the outcome of no run could be influenced by an act of either one.

Series B consisted of 7 subject sessions with 5 subjects. The subjects were all individuals associated with this laboratory and were thoroughly familiar with ESP test procedures. Otherwise, the procedure was identical to that of Series A. (Although the experiment was designed to include 400 runs, it remains incomplete.)

Series C consisted of 16 subject sessions with 9 of the same subjects as Series A. The procedure was practically identical with that of Series A, the only difference being in the variation of rates of matching.

Series D consists of 25 subject sessions with 8 of the same subjects as in Series A. The procedure was similar to Series A except that each session was strictly limited to 16 runs, and further variation of rate of matching was employed.

In all of these subseries, the experimenters were C. E. Stuart and J. G. Pratt. Stuart worked with the subject at the experimental table and Pratt handled the shuffling and key card determination in the cubicle. When either acted as subject in Series B, Stuart's duties were carried out by Mrs. D. H. Clark, the Laboratory secretary.

Series P consists of 10 subject sessions with 8 women college student subjects. The experimenters were J. G. Pratt and J. L. Woodruff. The research was intended to repeat with pre-shuffle card calling a study of ESP response to variation in size of stimulus symbol noted in previous research (11). The subject made his responses in a screened-touch-matching set-up similar to that described for Series A. There was, however, no limitation upon his rate of movement. Also, the key cards were selected after the run by a throw of three dice which determined the selection of an envelope among a large number of such envelopes containing random

arrangements of key cards. Independent records of the card lay-out and of the keys were written on numbered record sheets and placed in a locked box for later checking by the record librarian. Upon request of this writer, the experimenters introduced the recording of subject estimates during a short period of the experiment. The subject sessions which included estimates were all that could be used here.

Series H consists of 24 subject sessions with 15 subjects, all seniors or graduate students. In this research by Miss Lois Hutchinson and J. G. Pratt, scores in pre-shuffle card-calling were to be compared with scores in post-shuffle calling. The experiment was quite different in technique from that of Series A but entirely comparable in terms of the high degree of precaution introduced against experimental error. The subject was instructed to call verbally the symbols of a previously shuffled and recorded deck of 25 ESP cards. The deck remained intact in its box, however, during the calling. (This is the "DT" procedure.) The pre-shuffle calling runs were achieved by instructing the subject to call the order of the deck as it would be after it had been cut at random, the random cut being determined by calculations based upon the fall of three special dice. Before each run, the subject was asked for his guess as to how many hits he would make. After the calls were made but before they were checked, he was again asked for an estimate of his score.[4] The subject sessions of Series H utilized here consist of all sessions in which as many as 20 runs were made, 10 of DT and 10 of PDT[5]. These provide two subseries, one in which the PDT runs were given last in the session and another in which the DT runs occur last in the session. The first we have denoted Series Hp and the second, Series Hc.

Series O consists of 15 subject sessions with 10 subjects. These 15 sessions were part of a previously reported study of the rate of matching upon ESP performance (20). During this work, the present writer first considered the notion of getting subject estimates of scores before the scores were checked. The usefulness of these estimates as constituting other than a type of ESP test was not, at the time, considered; so in only a relatively few subject sessions were estimates consistently recorded. The experimental procedure was screened-

4. This estimate, more comparable to those of the other series, was the one utilized here.
5. DT symbolizes the technique of calling an unbroken deck of cards. PDT symbolizes the DT procedure when the order of cards is not determined until after the calls have been made.

touch-matching with open key cards visible to both subject and experimenter. The pointing responses by the subject were limited by a metronome beat as in Series A. The number of runs in the subject session was limited only by convenience.

The experimental conditions for objective precaution against error were not as elaborate as those of later experiments. One experimenter conducted the research and recorded the results. The data consisted of run scores, and did not include the individual card distributions. Since this writer conducted the experiment, he can at least record the subjective impression that great care for accurate observation was exercised.

Series E consists of 10 subject sessions from 4 subjects. The experimental conditions were the same as in Series O except that for 9 of the 10 sessions, the subject was limited strictly to 20 runs.

Series S consists of 4 sessions from exploratory tests conducted by C. E. Stuart and B. M. Smith for an experiment to study the effect of distraction upon ESP performance. The procedure was screened-touch-matching in which the key cards were enclosed in envelopes in such a way that the symbols were not visible to the subject nor to the experimenter. The two experimenters kept independent records of scores and estimates.

The results of the experimental subseries above are presented in Table I. It will be noted that no subseries is significant in total deviation except Series O. The latter is to be expected since the work of which it is a part was totally significant (20: p. 179). The fact that the total of all work selected yields a significant critical ratio of 2.71 implies merely that we are justified in concluding that extra-chance scoring occurred in the experiments, a fact already established as highly probable in the case of Series O.

For the present study, each experimental subject session is separated into two parts, the first part to consist of the first 10 runs and the first 11 estimates, the second part to consist of all remaining runs. This separation is necessary to provide a card-calling test which will be wholly independent of the statistics used to measure the variable to be compared to the card-calling score.

In the separation of the subject session into two parts, the data of the first part will be utilized to

TABLE I

Various Statistics of Subseries

Series	No. Subs.	Subject Sessions	Total Runs	Dev.	Ave.	C. R.	Experimenters	Dates
A	11	18	400	+41	5.103	1.03	Pratt-Stuart	2/3/39 - 3/1/39
B	5	7	168	+24	5.143	.93	"	1/28/39 - 2/20/39
C	9	16	400	-8	4.980	.20	"	3/1/39 - 3/14/39
D	8	25	400	+33	5.083	.83	"	3/15/39 - 5/3/39
Hp	11	12	240	-7	4.971	.23	Hutchinson-Pratt	2/6/40 - 3/1/40
P	10	10	200	+34	5.170	1.20	Woodruff-Pratt	3/8/40 - 3/20/40
Total: I	38*	88	1808	+117	5.064	1.38		
O	8	15	364	+96	5.264	2.52	Stuart	11/3/36 - 9/6/37
E	4	10	228	+19	5.083	.63	"	4/3/39 - 1/22/40
S	4	4	76	+20	5.263	1.15	Smith-Stuart	3/15/40 - 3/19/40
Hc	10	12	240	+30	5.125	.97	Hutchinson-Pratt	2/6/40 - 3/1/40
Total: II	28*	41	908	+165	5.182	2.74		
Total	55*	129	2716	+282	5.104	2.71		

*These are not totals of columns since a subject might serve in more than one subseries.

TABLE II

Example of Subject Session Data Used for T-Test and Card-Calling Test

Series A Subj.: Dil. Feb. 20, 1939

	Run	Run Score	Estimate
Data for T-test	1	5	8
	2	4	7
	3	7	5
	4	4	6
	5	1	8
	6	6	5
	7	5	4
	8	7	6
	9	5	6
	10	6	7
Data for CC test	11	5	5
	12	2	6
	13	5	7
	14	7	5
	15	7	6
	16	4	6
	17	4	5
	18	4	8
	19	5	8
	20	7	7
	21	8	3
	22	9	2
	23	5	1
	24	5	4

The first column lists the number of the run in the session. The second column lists the number of correct correspondences or "hits" between the subject's call responses and the randomized key cards or card order. The third column lists the subject's estimate, before the checking of the run, of the number of hits he has made on that run. The dotted lines show items later to be correlated to provide a T-score.

measure a non-card-calling variable and will be referred to as the T-test. The remaining runs of the subject session will constitute the card-calling test data for that session. Table II exemplifies a subject session so divided into T-test and CC test.

The Specific Problem

This analysis will be concerned with the question: Is there a test, T, applicable in the same subject session as a card-calling test, which permits the classification of the card-calling scores into two classes: A, a chance distribution of scores, and B, a distribution of scores with the mean significantly above chance expectation?

If this question may be answered affirmatively and if class B is a numerically adequate and statistically unbiased sample of card-calling test scores, then those operations of test T which produce the B classification are predictive correlates of variation in card-calling scores and therefore constitute an experimental condition highly favorable to extra-chance scoring as tested by card-calling techniques.

For the T test, the hit scores 1 to 10 are paired respectively with estimates 2 to 11. A product-moment correlation between the two orders is computed. The r so computed is the T-score for that subject session. A T-score may have a value between -1.000 and +1.000. Since the T-scores are correlation coefficients based upon 10 pairs of items, their serial position in the range might be largely accidental. We may group them roughly and arbitrarily into five subranges, -1.000 to -.601, -.600 to -.201, -.200 to +.200, +.201 to +.600, and +.601 to +1.000.

The card-calling score for each subject session is the total number of hits made on the remaining test runs of the session. This will usually be expressed as an average per run or a deviation from a chance expectation of one-fifth the number of runs.

Results

T-scores computed for Series I, consisting of pre-shuffle card-calling tests, ranged in value from -.68 to +.87. These scores were grouped in the five equal sub-ranges from -1 to +1. The subject sessions were thus grouped into five classes which could be characterized

respectively as those in which there had been a high positive correlation between scores and estimates, those in which there had been a moderate positive correlation, those in which there had been a low correlation, those in which there had been a moderate negative correlation, and those in which there had been a high negative correlation.

Thus, each class may characterize a "condition" specific to the session in which the individual CC score[6] was observed. The individual CC scores, when pooled for each class, give the results presented in Table III. All T-score groups except the mid-range group yield a deviation negligibly different from chance. The mid-range is significant.

Since the mid-range T-scores preceded significantly extra-chance CC scores and other ranges preceded CC scores which were not significantly deviant from chance, the first question of this analysis has been provisionally answered. The data from Series I are presented in Table IV, as grouped into two classes, A and B, A having a chance mean of 5.02, and B having a significantly extra-chance mean of 5.34. It is observed that the difference is consistent in each of the six subseries, as is shown in Table V.

Table III

CC Scores grouped in 5 T-Score Ranges, Series I

T-Score range	Subject Sessions	Runs	Dev.	Aver.	C. R.
+1.0 to +.60	10	112	+20	5.179	.95
+.60 to +.20	32	284	+10	5.035	.30
+.20 to -.20	29	310	+105	5.339	2.98
-.20 to -.60	14	176	- 25	4.858	.94
-.60 to -1.0	3	38	- 4	4.895	.32
Total	88	920	+106	5.115	1.75

The question has been only provisionally answered because there is no *a priori* expectation that the extra-chance scoring is contingent upon the range in which it occurred. Without such *a priori* expectation of extra-chance scoring in any particular sub-range, it may be considered equally likely to occur in any one of the five.

6. "CC score" will indicate the card-calling score.

TABLE IV

A and B Class T-Scores and Concomitant CC Scores, Series I

T-Score class	Subject Sessions	Runs	Dev.	Aver.	C. R.
A	59	610	+1	5.002	.02
B	29	310	+105	5.339	2.98
Total	88	920	+106	5.115	1.75
Diff.				.337	2.48

TABLE V

A and B Classification of Subseries CC Scores, Series I

	A				B			
	Runs	Dev.	Aver.	C.R.	Runs	Dev.	Aver.	C.R.
Series A	112	+15	5.13	.71	108	+36	5.33	1.73
" B	74	+8	5.11	.47	24	+22	5.92	2.25
" C	176	-21	4.88	.79	56	+2	5.04	.13
" D	108	+9	5.08	.43	42	+15	5.36	1.16
" Hp	70	-37	4.47	2.21	50	+13	5.26	.92
" P	70	+27	5.39	1.61	30	+17	5.57	1.55
Total: I	610	+1	5.00	.02	310	+105	5.34	2.98

Thus the probability that the CC scores in the mid-range should average 5.33 with a C.R. of 2.98 is not 0012 but about 5 times that value, or .006. This latter figure is the probability that one of the five arbitrary subranges should include the scoring rate noted in the mid-range.

While this probability is in itself barely significant, the contingency may be validated by a repetition of the test upon other research data. Since the results of Series I would at the very least support the *a priori* expectation that a similar contingency might hold among similar data, then the classification of other data into classes A and B should provide significantly extra-chance scores in class B and chance scores in class A.

When the data of Series II are classified into A and B classes, it is found that class A yields scores negligibly different from chance, and class B yields

scores above chance to a significant degree. In Table VI, this classification is summarized. The probability that the class B scores of Series II are an accidental sample of a binomial population with a success probability of 1/5 is .00013. If we ignore the chance hypothesis for the moment and consider Series II as a sample of CC scores nearly normal in distribution but with the *a priori* mean and standard deviation unknown, the probability that classes A and B are accidental divisions of these ESP scores is .00097. Further, it was noted from Table I that the total average score of Series II was significantly above chance with an average of 5.2. The A and B classes are statistically proper, yet not only does B remain significant with a higher average, but A is negligibly different from a chance mean of 5.0.

These observations on Series II clearly verify the hypothesis that the A and B classification of T-scores effects a real distinction between a highly favorable and a negligible expectation of extra-chance scoring in the contingent card-calling tests.

TABLE VI

A and B Class T-Scores and Concomitant CC Scores, Series II

T-Score class	Subject Sessions	Runs	Dev.	Aver.	C. R.
A	29	360	-13	4.964	.34
B	12	138	+86	5.623	3.66
Total	41	498	+73	5.147	1.64
Difference				.659	3.10

TABLE VII

A and B Classification of Subseries CC Scores, Series II; and Totals.

	A				B			
	Runs	Dev.	Aver.	C.R.	Runs	Dev.	Aver.	C.R.
Series O	150	-7	4.95	.29	64	+55	5.86	3.44
" E	100	-11	4.89	.55	28	+17	5.61	1.61
" S	30	+4	5.13	.37	6	+ 4	5.67	.82
" Hc	80	+1	5.01	.06	40	+10	5.25	.79
Total: II	360	-13	4.96	.34	138	+86	5.62	3.66
Total: I	610	+ 1	5.00	.02	310	+105	5.34	2.98
Total: I and II	970	-12	4.99	.19	448	+191	5.43	4.51

Further Observation Regarding Validity

The T-test-CC-test comparison has been approached so far in a somewhat arbitrary and "operational" way. This approach has seemed feasible in order to keep in their simplest terms the points clearly established, and has been necessary for the demonstration of extra-chance card-calling within a certain class of the experimental observations. Now that the relation has been established, certain hypotheses may be considered in regard to the items of the comparison.

The subject-session T-score is a product moment correlation between the first 10 test scores and the succeeding estimates. The experimental situation in which the estimates are made is very similar to situations used for the study of subject level of aspiration, and similar reactions of these subjects were observed. The meaning, importance, and implications of the estimate may be very different for different subjects. The instruction was simply to "guess how many you got right this time." Some subjects tried seriously to make a true independent judgment each time. Other estimates were frankly "desired" scores. One subject made estimates in one session from zero to 25. Some utilized the estimate as "magic": "If I guess a seven, I know it will be a three, so I'm going to guess three."

The average level of estimate in the T-tests of Series I was +.862, a value significantly above the average scoring level. The estimate average was above the score average in 64 sessions, equal and below in 24 sessions. This indicates that the social situation of the experiment was such as to modify intelligent judgments in the direction of a desire to "score above chance."

But it is equally notable that the level of estimate was, for all but one subject, always within reasonable limits of attainment. In Series I, only 3 of the 88 subjects ever exceeded an estimate average of 7 hits per run.[7]

7. One of these, subject Jer, was characteristically different from any other subject in the wildness of his estimates. The two others were the present writer (aver. est.: 7.1) and his collaborator in these experiments, Dr. Pratt (aver. est.: 8.1). To the latter two "sophisticated subjects," the maintenance of confident expression was more important than exactness of estimate.

Furthermore, the variation in estimate was consistently below the variation in test scores. That is, the central tendency of the estimate in a given subject was greater than that of the test scores in 76 of the 88 cases. The average of the test variances was 3.487; the average of the estimate variances was 2.270. The difference is statistically significant.

It is evident that the subject did not make his estimates with the independent randomness of a chance machine; neither did he make them strictly in accordance with his own experience of past scores. The fact that they were neither independent responses nor strictly determined responses suggests their variability as a meaningful characteristic of the individual subject response.

Isolation of all the factors that go into the statement of an estimate by the subject would be as difficult as the evaluation of experiences leading to judgments in everyday life. A few of the more obvious ones may be listed:

a. The subject's experience with the general average of past scores and with their variability;

b. His experience with or conception of what accidental coincidences should produce;

c. His response to social pressures in the experiment, such as his desire to perform in a way pleasing to his psychology instructor, or his notion of what others have done;

d. His interpretation of experience of effectiveness or frustration in the act of calling or matching the particular run of cards;

e. Any systematic or patterned tendencies of response, etc.

Each of these, insofar as they are measurable, would provide a variable in some relation to the ultimate judgment leading to an expression of estimate. Certainly the basis of one possible variable is the experience of the subject with the run score just preceding the estimate.

The T-score is a measure of the degree to which the subject is affected by his just previous score in making an estimate. If the subject were to use his immediately preceding test score as the best estimate of his expected accomplishment, the T-score would most likely be high and positive. If the subject were to use his immediately preceding test score to determine a

"compensating" estimate for the next run (as might occur if the subject makes the assumption that his scores will average five in a cyclic fashion), then the T-score would likely be high and negative. If he were to try to make each estimate an independent judgment, his T-score would likely be close to zero.

The first two hypothetical situations would be examples of what ESP investigators have characterized as "score-consciousness." Although never experimentally established, it has been frequently noted that subjects who are overly rational in their attitude, who tend to evaluate each run score in terms of increment or decrement to a total or expected total, are rarely as successful as those whose reaction within the test situation is free and spontaneous.

The T-score is, however, not a very good measure of specific score affectability; in fact, it is a very crude measure. The crudeness follows from two possible sources of spurious variation:

a. The correlation coefficient as a statistical constant based upon 10 paired items has a large probable error. For example, the probable error of a zero correlation is about ±25; which means that from a population of pairs having no correlation about half the samples of 10 pairs drawn will have a correlation coefficient of .25 or greater in absolute value.

b. Variation in the subject's mode of response may alter the accuracy of a T-score. If the subject responds directly to attained scores part of the time, and then responds negatively for a time, the two may cancel to indicate no correlation even though estimates have been almost wholly determined by the experience with the past score.

The first source of error can be corrected only by increasing the number of items in the T-test. But as the reliability of the correlation coefficient increases only as the square root of the number of pairs of items, it is doubtful whether the correction would be practicable. A subject-session of 20-24 runs may occupy a 50-minute experimental period. The extension of the experimental period to allow an adequately long test would penalize the subsequent card-calling test period with the likelihood of fatigue and boredom.

Variation attributable to an actual change in the subject's mode of response cannot be isolated when the correlation is itself assumably a measure of that variation.

Regardless of the argumentative case for a low *a priori* statistical reliability and validity of the T-score, however, instances suggest that it does measure some consistent type of behavior. The fact that it provided an effective criterion for the classification of the data of each of 10 subseries suggests that it must represent consistently some factor correlated to the performance. Other observations of consistency are available from the comparison of subjects each used for several sessions. For example, subject Pec served in 11 of the subject sessions. The mean of his T-scores was -.109, their standard deviation, ±.271. Subject Wst served in 7 sessions with a mean T-score of +.488 distributed with a standard deviation of ±.199. The difference between these mean T-scores is highly significant with $t = 4.75$.[8] The fact that a significant difference is demonstrable supports the hypothesis that the T-score measures some characteristic of the individual subject's behavior which is persistent from session to session.

Up to this point, the card-calling test has been accepted as that part of the subject session following the T-test. This general inclusion is adequate if the factor measured by the T-test is assumed to be constant throughout the experimental period. It is necessary, however, to take into consideration the possibility that the T-test measures a transitory state characteristic of the first part of the experimental session and that its relation to the card-calling test runs which follow may itself be limited to that transitory condition. Insight into this possibility may be effected by limiting the number of card-calling runs in the subject session. One limitation that might be considered is to fix the number of card-calling test runs as equal to the T-test runs. The A and B classification of the data thus sampled is arrayed in Table VIII. The average scores and critical ratios, when compared to Tables V and VII, show negligible differences.

Since this strict limitation necessarily excludes Series D, another sampling is suggested in order to include Series D and certain other sessions. This sampling would define the card-calling test as the runs following the T-test up to and including the 10th run. This sampling also is given in Table IX. This sample is also only slightly different from the all-run sample. Thus,

8. A t-value for the difference is computed here instead of the critical ratio of the difference because of the small number of items in the comparison. The t-value corrects for "degrees of freedom" when small numbers are used (17: p. 92).

it is certain that variation in session length, within these small limits, at least, does not alter the validity of the A and B classification.

TABLE VIII

Data Sampled to Include Exactly 10 Runs as CC Test

	A				B			
	Runs	Dev.	Aver.	C.R.	Runs	Dev.	Aver.	C.R.
Series A	40	-10	4.75	.79	70	+33	5.47	1.97
" B	50	+ 7	5.14	.50	20	+14	5.70	1.57
" C	120	- 9	4.93	.41	40	+ 4	5.10	.32
" Hp	70	-37	4.47	2.21	50	+13	5.26	.92
" P	70	+27	5.39	1.61	30	+17	5.57	1.55
Total: I	350	-22	4.94	.59	210	+81	5.39	2.80
Series O	90	- 4	4.96	.21	30	+33	6.10	3.01
" E	70	+ 6	5.09	.36	20	+11	5.55	1.23
" S	30	+ 4	5.13	.37				
" Hc	80	+ 1	5.01	.06	40	+10	5.25	.79
Total: II	270	+ 7	5.03	.21	90	+54	5.60	2.85
Total	620	- 5	4.99	.10	300	+135	5.45	3.90

TABLE IX

Data Sampled to Include in CC Test 10 or Less Runs Following T-Test

	A				B			
	Runs	Dev.	Aver.	C.R.	Runs	Dev.	Aver.	C.R.
Series A	84	+ 1	5.01	.05	76	+30	5.39	1.72
" B	50	+ 7	5.14	.50	20	+14	5.70	1.57
" C	120	- 9	4.93	.41	40	+ 4	5.10	.32
" D	108	+ 9	5.08	.43	42	+15	5.36	1.16
" Hp	70	-37	4.47	2.21	50	+13	5.26	.92
" P	70	+27	5.39	1.61	30	+17	5.57	1.55
Total: I	502	- 2	5.00	.04	258	+103	5.40	3.21
Series O	96	- 6	4.94	.31	45	+38	5.84	2.83
" E	72	+ 8	5.11	.47	20	+11	5.55	1.23
" S	30	+ 4	5.13	.37	6	+ 4	5.67	.82
" Hc	80	+ 1	5.01	.06	40	+10	5.25	.79
Total: II	278	+ 7	5.03	.21	111	+63	5.57	2.99
Total	780	+ 5	5.01	.09	369	+166	5.45	4.32

Analysis of Further Implications

Success and Failure

The T-test data provide statistics amenable to interpretation with respect to other behavioral characteristics of the subject. It has been observed by investigators (but not experimentally established) that the success or failure of the subject in his performance seems to influence further scoring. In many respects, the T-test situation is comparable to customary experimental situations used for tests of level of aspiration (1; 3) particularly if the subject feels that he is effective in influencing the test scores.[7] The estimates partake of the same ambiguous character as bids in level of aspiration experiments. The estimate may be at times an intelligent judgment, at other times, the expression of a desire. The subject knows that a score of 5 is a "chance" score and soon learns that the experimenters prefer scores above 5 and seem disappointed at scores below 5. As noted, previously, the average level of estimate was significantly above the scoring level.

If the estimate is interpreted as a bid, it may be that the subject experiences success when he equals or exceeds his estimate and experiences failure if his score falls below the estimate. If we so assume, the estimates and scores for the 10 runs of the T-test may be classified, then, into "successes" and "failures."

When the 129 T-tests are classified according to the relative number of "successes" and "failures," it is observed that all had at least one failure and that, as would be expected from the average level of estimate, the most of the subjects "failed" slightly more often than they "succeeded." The distribution of the T-tests according to the number of "failures" is given in Table X.

Comparison of "success-failure" difference with the ESP tests must involve care with regard to sampling. Some of the subject-sessions were longer than others; and while success in performance was never made a criterion for continuing a test series, the fact that this might partially motivate a subject (or experimenter) who had other reasons for alteration of the customary number of

7. Some subjects are quite willing to engage in the card-calling experiment while insisting in the belief that the correspondences recorded are wholly accidental. The experimenter does not attempt to change this belief but tries to get the subject to take a "play attitude," to act as if he were producing the correspondences.

runs suggests strict sampling to avoid this possibility. The 10-or-less-runs sample or the exactly-10-runs sample are useful for this study (although the latter excludes Series D, an entirely valid sample).

Scores on the exactly-10-run ESP tests, when classified according to "failure or success" difference, show no linear relation, but instead, a tendency to positive scoring in the cases when the "failures" and "successes" were about equal. The 6 + 4, 5 + 5, and 4 + 6 classes include 650 runs with a positive deviation of 163, significantly greater than the remaining 350 runs with a negative deviation of 32. The same observation holds for the more inclusive 10-or-less-runs sample. (See Table X.)

An interesting result appears when this classification is further broken down to A and B classes according to the T-test. (The small numbers make it feasible to group the end classes.) The peaked distribution of Table X separates into two distributions. The A class varies about its chance mean. The B class is below and close to the chance mean for from 10 to 7 "failures," and then rises to a fairly consistent level of about 5.50 with an increasing number of "successes." (See Graph 1.) The indication is clear that the T-score classification was not effective when the subject "failed" unduly is his estimates.

The interpretation that undue failure is instrumental in inhibiting scoring, rather than the supposition that success enhances scoring, is borne out by a study of the actual occurrence of success and failure. Previous studies have affirmed that success of striving in an experimental task is usually followed by increase of level of aspiration, and failure by lowering of level of aspiration (3: pp. 67-68; 7: pp. 17ff.).

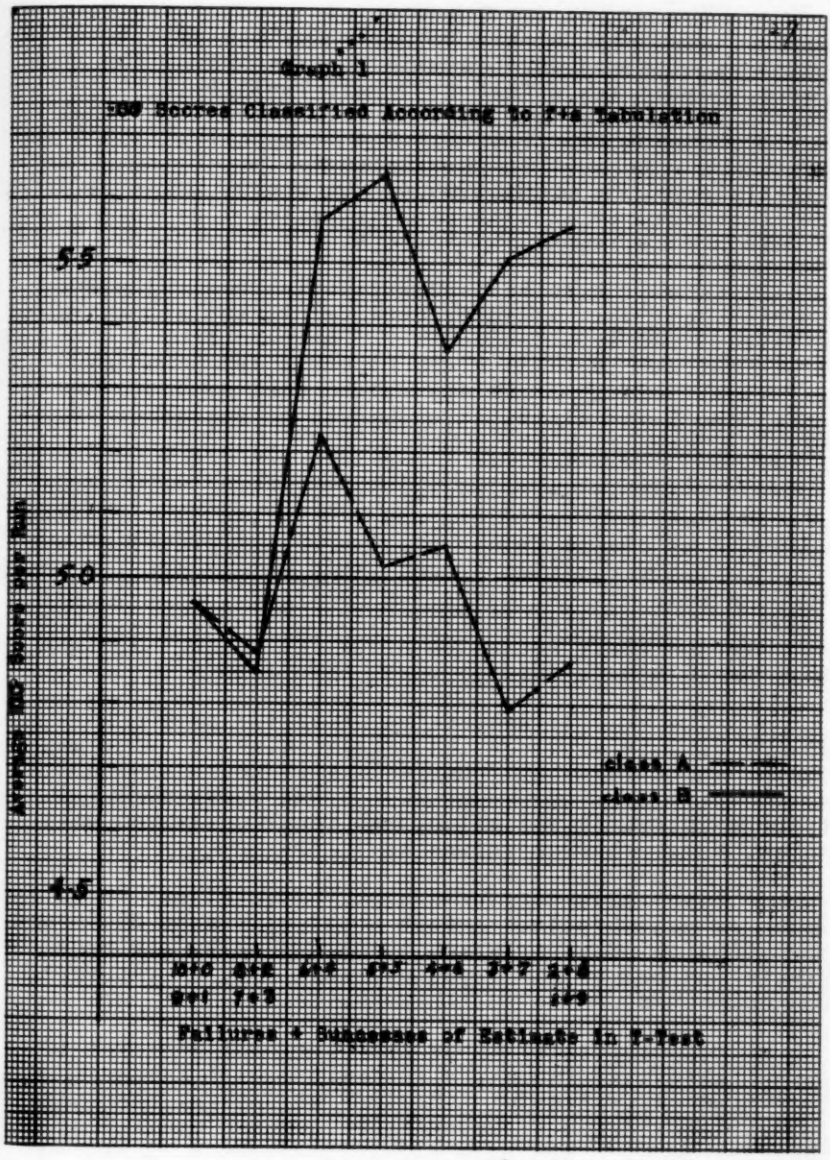
This observation has been made with tasks in which the subject could readily perceive the relation of his skill to the score achieved. In the card-calling test, the manner of performance and the resulting score can only have such quasi-magical relation as the subject is willing to imagine for himself. It is evident that if the subject took all his scores wholly seriously as evidence of a personal skill, the card-calling test situation would soon become one of extreme frustration. Some compromise in acceptance of responsibility for the scores is necessary in order to make the situation endurable.[8]

8. For example, a frequent result was the subject's playful protest that the second experimenter "brought out the wrong cards," a protest that sometimes covered real annoyance.

TABLE X

CC Scores Classified According to Failure + Success Tabulation of T-Tests: Series I and II (CC Score from 10-or-less Runs)

	10+0	9+1	8+2	7+3	6+4	5+5	4+6	3+7	2+8	1+9	0+10
Subj. Sess's (total)	5	4	9	15	29	28	22	13	3	1	0
Total CC Scores											
Runs	34	40	36	126	246	256	238	121	22	10	0
Dev.	-12	+ 9	-40	+14	+86	+64	+33	0	+12	- 5	
Ave.	4.65	5.23	4.53	5.11	5.35	5.25	5.14	5.00	5.55	4.50	
Class A											
S.S.	2	3	6	14	19	18	15	9	1	1	0
Runs	16	30	56	116	158	160	172	86	6	10	
Dev.	-10	+ 8	-24	+ 4	+36	+ 3	+ 9	-18	+ 3	- 5	
Ave.	4.37	5.27	4.57	5.03	5.23	5.02	5.05	4.79	5.50	4.50	
Class B											
S.S.	3	1	3	1	10	10	7	4	2	0	0
Runs	18	10	30	10	88	96	66	35	16		
Dev.	- 2	+ 1	-16	+10	+50	+61	+24	+18	+ 9		
Ave.	4.89	5.10	4.47	6.00	5.57	5.64	5.36	5.51	5.56		


Graph 1
Raw Scores Classified According to F+S Tabulation
5.5
5.0
4.5
class A
class B
Failures + Successes of Subjects in F-Test

The principle of increased level of aspiration with success may be used to determine the extent to which success and failure were experienced by the subjects. In Series A through D, *all* runs, both of T-test and CC test, were tabulated in nine classes: S+, S-, SO, S±, S=, SC, F+, F-, FO. The designation indicates whether a "success" in scoring above the estimate, an exact successful prediction of the score, or a "failure," was succeeded by an increased estimate, a decreased estimate, or no change.

A null hypothesis distribution was calculated to represent what would be expected if the estimates were replaced by a random set of numbers from a binomial distribution. But since the observed level of estimates was significantly above the chance mean of 5.0, the null hypothesis distribution had to be corrected for that difference.

A chi-square goodness of fit test of the observed distribution of S's and F's to the distribution expected if change in estimate were independent of success or failure yields a negligible chi-square for the successes in scoring above the estimate. A probability value of .60 affirms the fact that the subjects raised and lowered their estimates after "successes" in the same way a chance machine might have done.

The exactly successful calls (S) gave a chi-square of 7.19 with $p = .03$. This value is not significant and affords only suggestive evidence that estimates which exactly predicted the run score were followed by more increased estimates, and fewer decreased estimates, than a chance series would provide.

The failures give a highly significant variation from the null hypothesis with a chi-square of 27.06 and $p = .000001$. The variation was provided by the much greater number of F+ responses, and fewer F- responses than expected on the null hypothesis.

The implications of this analysis are that if the principle of increased aspiration with a success and decreased aspiration with failure hold true, then there is little evidence of experienced success and failure as measured here.[9] The only significant reaction was the

9. That success and failure are experienced in the card-calling test situation is obvious to anyone who has tried seriously to be a subject. So the failure of the measures used here to indicate that experience is no doubt in part the fault of those

frequent response to a "failure" by an increased estimate, a response which may indicate a compensation reaction to failure, but may as well indicate a rationally calculated compensatory reaction to a low score. In any case, the absence of evidence of reaction to "success" and the presence of some response characteristic of "failure" affirms the interpretation that the "failure" situation may be inhibitory to scoring ability.

Level of Estimate and Extra-Chance Scoring

The T-test provides data for computation of a level of estimate for the beginning of each subject session. As noted previously the average level of estimate was 5.862 for Series I, and 5.893 for Series II, a total average of 5.873.

No direct relation between the individual level of estimate and his subsequent CC scores was observed. But one observation was of suggestive interest. Some subjects adopted a level of estimate equal to or below the chance expectation of five hits. This was the case in 24 subject-sessions of Series I and 9 subject-sessions of Series II. The contingent card-calling tests included 366 runs with a significant positive deviation of 141 hits (See Table XI.) It is evident that these subject-sessions yielded significantly better CC scores than those in which the subjects signified expectation of higher-than-chance scores.

TABLE XI

Level of Estimate and Card Calling Scores

	Estimate Level 5 or less				Estimate Level above 5			
	Runs	Dev.	Ave.	C.R.	Runs	Dev.	Ave.	C.R.
Series I	244	+77	5.32	2.46	676	+29	5.04	.56
Series II	122	+64	5.52	2.90	366	+ 9	5.02	.24
Total	366	+144	5.39	3.69	1042	+38	5.04	.59

Ave. Diff. = +.3489 ± .1217

C. R. of Diff. = 2.87.

(Footnote continued) measures. As Gould notes, the magnitude of differences which may constitute "success" or "failure" varies among subjects and among tasks (3: p. 68). So far there has been, in studies of level of aspiration, no statistical treatment of the principle of increase with success and decrease with failure. McGehee tabulates his findings in the classifications used here but avoids statistical treatment or discussion (10: pp. 9-10). Examination of his results suggests that the control of a null hypothesis might reveal results very similar to those noted here.

The result is interesting in that a subject estimating at the chance level or below must resist the normal social pressure of the experimental situation. He knows that the experimenters want him to score above chance. Estimates averaging below that level would seem to indicate socially independent judgments and even negativism. Neither of these traits have heretofore been suggested as correlates of extra-chance scoring ability. A possible explanation may lie in the observations of experimenters of level of aspiration. It was noted that subjects admitted to setting low estimates in order to avoid failure, (3: pp. 93-94). The below-chance estimates might indicate, therefore, an interest in successful performance so great that the subject is forced to use his estimate as a protective device.

Effect of Repeated Sessions:

The unit for this analysis has been the subject session. As remarked previously, however, many subjects served for more than one experimental session. The actual distribution of the 55 subjects among the 129 sessions was:

Experimental sessions	1	2	3	4	5	6	7	9	11
Number of subjects	31	11	3	1	1	4	2	1	1

The 31 subjects who served for only one session may or may not have had previous experience with card-calling tests. But since the feature of making estimates of their scores was new to each subject in his first session there is some opportunity to trace the influence of repeated sessions with this particular experimental condition.

Pratt and Woodruff observed that when new stimulus sizes were introduced in their experiment there was an apparent increase in scoring coincident with the novel materials, and a later decline of scoring with succeeding experimental sessions (11, pp. 147-152). While the number of observations of the present analysis is too small to make a strictly comparable study there are enough to permit examination of scoring decline, as might be evidenced with the A and B classification of card-calling scores.

The results of classifying subject sessions as "first," "second" and "later" are quite striking, as may be noted in Table XII. The only significant scoring of the six groups occurs among those subjects who had B class T-scores in their first experimental session.[10]

10. The highly significant CR of 4.68 does not require correction for selection of one of six cases, since in this study we have

TABLE XII

A and B Classification of CC Scores in Successive Subject Sessions

	SS	Runs	Dev.	Ave.	C.R.	SS	Runs	Dev.	Ave.	C.R.
First Session	36	423	+14	5.03	.34	19	232	+142	5.61	4.66
Second Session	14	178	-23	4.87	.86	10	112	+ 38	5.34	1.80
Later Sessions	38	369	- 2	5.00	.05	12	104	+ 11	5.11	.54

First Sessions	Series I	Series II	Total
Ave. Diff.	.4425	.8941	.6207
C.R. of Diff.	2.27	3.27	3.66

The roughly linear decline indicated by the averages 5.61, 5.34, and 5.11 for the B-class subjects in first, second, and later sessions is suggestive of a real decline in scoring, although an analysis of variance among the groups indicates that the averages do not vary significantly. No such regularity occurs among the A class scores.

The implications of this observation are not unambiguous. At first it appears to support a general belief among ESP investigators that the first experience by the subject of a new experimental procedure is critically important, and that the subject is more likely to do well the first time than with repeated testing. But this holds true here only for B class subject sessions. It may well be that only the inferred effect of the T-score upon the subsequent card-calling scores declines with repeated testing. Whatever the proper interpretation it is methodologically interesting that, in this investigation of the T-test to card-calling test relation, repeated sessions with a given subject served only to dilute results which were highly significant when based upon a proper sample of first sessions.

To give a general picture of individual subject performance Table XIII lists the basic data of this report by subjects. The first two columns designate the subject and the series. When subjects were used for more than one session the chronological order of their

(Footnote continued) already established the expectation of extra-chance scoring in the B class, and the Pratt-Woodruff observation permits *a priori* expectation of highest scoring in the first session.

Table XIII

LIST OF INDIVIDUAL T-SCORES AND CC SCORES BY SUBJECT SESSIONS

Series and No.	Sub.	T-Score	CC-Score Runs	Dev.	Series and No.	Sub.	T-Score	CC-Score Runs	Dev.	Series and No.	Sub.	T-Score	CC-Score Runs	Dev.
S4	Ann	+.775	10	+ 2	A9	Has	-.177	14	+11	D18	Pec	+.224	6	- 9
Hp1	Ber	+.763	10	0	A14	"	-.485	14	+ 8	D23	"	-.152	6	+ 1
Hc1	"	-.215	10	- 5	C4	"	-.235	14	+10	S1	Por	+.224	10	+ 2
P7	Bre	-.146	10	+ 7	C12	"	+.384	14	- 6	B4	Pra	+.346	14	- 7
Hp2	Bin	-.287	10	-13	D3	"	+.079	6	+12	01	Rhi	+.351	15	0
A7	Blo	+.379	14	+11	A8	Hor	+.581	14	+ 2	P10	Rog	+.180	10	+16
S2	Bla	+.304	10	0	A12	"	+.139	14	- 2	Hc9	Ron	+.422	10	+ 5
014	Brz	+.746	18	+ 9	A18	"	+.424	6	+ 6	Hp7	"	+.062	10	+13
02	Cha	-.217	21	- 6	C10	"	+.872	14	- 4	P9	Rou	+.532	10	+22
03	"	-.242	12	+10	D1	"	+.698	6	+ 1	010	Ryn	+.120	20	+29
05	"	-.454	18	- 8	D9	"	+.287	6	0	011	"	-.323	10	- 8
06	"	+.254	15	+ 3	D12	"	+.473	6	+ 5	Hp8	Sch	-.096	10	+ 5
09	"	+.492	6	- 2	D19	"	+.808	6	+ 7	C5	Smi	-.009	14	0
B3	Cla	-.495	14	- 3	D24	"	-.537	6	+ 2	C13	"	-.380	14	-15
B5	"	-.282	14	+ 3	Hp6	Hou	-.637	10	- 8	D4	"	-.211	6	+ 1
Hp3	Clo	+.132	10	- 1	08	Hut	-.296	16	- 4	D21	"	+.091	6	+ 5
04	Col	+.035	9	+ 3	C2	Jer	+.702	14	- 4	Hp9	Sou	+.223	10	+ 1
Hc2	Cru	+.542	10	0	C11	"	+.422	14	0	P3	Ste	+.550	10	+ 3
P2	Dei	+.837	10	- 2	D2	"	+.258	6	- 9	P8	"	+.511	10	- 2
A10	Dil	.000	14	+ 7	D10	"	-.092	6	+ 1	B6	Stu	+.659	22	+21
C6	"	+.183	14	- 2	D20	"	-.025	6	- 4	B7	"	+.244	10	- 5
C14	"	-.335	14	- 9	D25	"	-.112	6	+ 1	E10	"	+.381	2	+ 2

D5	"	+.510	6	- 2	P4	Jol	+.413	10	+ 4	A13	Tra	+.318	14	+ 4
D11	"	+.293	6	+ 3	Hp10	Jst	+.289	10	- 2	C3	"	+.156	14	+15
D17	"	+.395	6	- 9	Hp11	"	+.079	10	- 5	Hp12	Tru	-.102	10	+ 1
D22	"	+.437	6	+ 9	Hc10	"	.000	10	+ 7	Hc11	"	-.241	10	- 4
A5	Don	-.048	6	- 3	A6	Kep	+.343	2	+ 3	Hc12	"	-0.14	10	- 7
Hp4	Ede	+.408	10	-11	07	Lee	-.160	10	+ 6	A17	Whi	-.078	14	+14
Hc3	"	+.492	10	+ 6	A4	Map	+.324	6	+ 2	C9	"	-.219	14	- 2
Hp5	Ell	-.382	10	- 4	P5	May	-.648	10	+ 4	C17	"	+.643	14	+ 2
Hc4	"	-.034	10	+ 7	013	Mdo	+.063	19	+15	D8	"	+.656	6	+ 1
P1	Erm	+.764	10	- 2	Hc7	Met	+.378	10	- 7	D16	"	+.383	6	- 9
P6	"	-.068	10	- 6	012	Mfa	-.768	19	- 1	E2	"	+.846	14	+ 5
A15	Eye	-.121	14	+ 7	Hc8	Mon	-.251	10	- 2	015	Wie	+.011	6	+ 2
C7	"	+.518	18	+ 1	A2	Pec	-.021	18	- 2	B2	Woo	-.149	10	+ 4
C15	"	-.040	14	-11	A3	"	-.676	18	0	S3	"	+.184	6	+ 4
D6	"	+.513	6	+11	A11	"	+.095	14	+ 4	E3	Wst	+.138	14	+14
D12	"	+.586	6	+ 2	A16	"	-.210	14	- 8	E4	"	+.592	14	- 2
D14	"	+.349	6	- 6	C8	"	-.311	18	0	E5	"	+.351	14	0
Hc5	Foo	+.304	10	+ 8	C16	"	-.370	14	+ 6	E6	"	+.687	14	+ 1
Hc6	"	+.173	10	+ 3	D7	"	+.246	6	+11	E7	"	+.698	14	- 8
B1	Gre	-.161	14	+ 8	E1	"	-.179	14	+ 3	E8	"	+.333	14	- 2
A1	Has	+.298	10	-13	D15	"	+.158	6	- 8	E9	"	+.667	14	- 7

sessions is retained in the table. The third column represents the correlation between scores and subsequent estimates for the first ten runs of the series. The last two columns represent the remaining number of runs and the deviation of the observed number of hits from a mean expectation of five hits per run.

The Precognition Hypothesis

By an accident of procedure this analysis, although carried out upon data from pre-shuffle card-calling tests, does not provide unequivocal evidence that extra-chance scoring occurred among those pre-shuffle tests in Series I. If the relation of A and B classification of T-scores to subsequent CC scores had been determined from Series II, then it would have been possible to designate the B classification of Series I as a proper sample which, being significant, would have established the occurrence of extra-chance scoring in Series I. A common-sense conclusion would suggest that the order of examination of the two series might be ignored and that each supports the other to an equal degree. But this conclusion would overlook the *a priori* status of the two series. The occurrence of extra-chance scoring in Series II, which consists of post-shuffle card-calling tests, would not be considered highly unusual since there is wide support for such an observation under these experimental conditions (16: pp. 95ff). The observation of extra-chance scoring in pre-shuffle calling conditions has no such body of verifying research and has only very recently been amenable to demonstration (8; 14; 15). It is desirable, therefore, to look for some extra-chance characteristic of Series I which is not dependent upon the expectation that extra-chance scores will occur specifically in class B.

An analysis of variance of the CC scores grouped into four T-score classes, 0 to $\pm.20$, $\pm.20$ to $\pm.40$, $\pm.40$ to $\pm.60$, and greater in absolute value than $\pm.60$, shows a mean variance within classes of 3.77 and a mean variance between classes of 15.79. An F statistic test of the significance of the difference (19: pp. 173-178) shows that the observed F of 4.19 is well beyond the 3.34 required for a significant difference.

Strictly interpreted, this difference indicates that variation in scoring rates in the card-calling tests of Series I was not independent of their classification according to T-score groups. This extra-chance characteristic of the data is thus adequate to establish the fact that extra-chance scoring occurred within

Series 1, a series consisting of pre-shuffle card-calling tests. The findings, therefore, provide relevant and supporting evidence for the hypothesis of precognition as proposed by Rhine (14: pp. 39ff.)

Counter-Hypotheses

No report of research offering findings in support of the hypothesis that there exists a cognitive process of an extra-sensory nature may well ignore the wide range of counter-hypotheses that have been proposed, and criticisms that have been directed to ESP research.[11] But at this date in the history of such investigation, it is possible to adopt a standpoint considerably more liberal than that of the majority of critics. These critics have implicitly assumed that the proposal of any reasonable hypothesis counter to the inference of a cognitive process was so self-evidently preferable from a scientific standpoint that the counter-hypothesis was valid without test. Today, however, the ESP hypothesis has adequately wide support to permit the questioning of counter-hypotheses directly in terms of evidence for and against their validity.[12]

The present research is fortunate in having in Series I sets of precautionary experimental conditions that are equal to the best devised for ESP investigators. They permit exclusion or control of all the counter-hypotheses so far proposed. A further circumstance removes from consideration many possibilities of motivationally determined error or mal-observation; namely, the fact that with the exception of Series P, the observations were made without possible knowledge by anyone concerned of the use to be made of them here. And even in Series P, the classification could not be determined until the conclusion of the experimental session.

Many of the counter-hypotheses listed in *Extra-Sensory Perception After Sixty Years* can be grouped into more general classes for consideration here.

The first group of counter-hypotheses may be phrased in the question: "Might not these results be a

11. These are fully summarized in *Extra-Sensory Perception After Sixty Years*, Part II.
12. For example, it cannot be proved that any given significant deviation is not an accidental occurrence with a high degree of improbability of occurrence. But this likelihood must itself be weighed in the light of the fact that no such accidents have occurred in exactly comparable data with the effect of intended coincidences removed (5: p. 303).

spurious outcome of inadequate or improper methematical methods?" The answer to this must lie in the justification of the methods used. The assumption is customary and it has been empirically justified that chance data from a card-calling test situation will closely approximate the binomial distribution (4). Variation from this is customarily used to detect the extra-chance nature of the test scores (6). This method has been utilized here in listing deviations and critical ratios. The assumption of binomial distribution has been controlled, however, by establishing critical ratios of the difference between classes A and B. In finding the critical ratio of these differences, the empirical variance of the data, and not the variance of a binomial assumption, was used. These are all standard and widely used statistical techniques.

A second hypothesis is the question whether these results might have been spuriously produced by improper selection of data in some fashion. As noted above, effort was made to get every relevant observation from all studies for this report. This report includes all data which permitted the analysis used. No data could have been list or misplaced, since all records were numbered. An often raised alternative hypothesis has been the possibility of improper selection due to the experimenter's exercising his option to conclude the experiment at a point where the results might be favorable to his hypotheses. The major part of this work was concluded before the relations studied were known to exist. The only option available to this investigator was the stopping point o March 20 for utilizing data from research in progress. This date had no relation to the results of any research. Further, it may be noted that Tables VIII and IX are samples which exclude possible optional stopping in the subject-session.

The next question concerns the validity of the CC tests themselves in respect to their freedom from the possibility of sensory cues, rational inference, or accidental similarity of patterning between subject calls and card arrangements. We must here differentiate the individual series in terms of their experimental conditions. All research in Series I was conducted under the best possible precautionary conditions. Sensory cues are impossible from card orders not yet established. The possibility of rational inference was avoided by the fact that the subject never handled the cards used for matching. Decks were constantly changed and the key cards were selected by a reliably random mechanical criterion. In all matching procedures, the cards in the

table lay-out are gathered by sweeping them together so that pile orders are completely disarranged. The pack is then shuffled with a dove-tail shuffle at least three times and is cut with a knife-cut before it is next used. A proper sample of 200 key card orders taken from the photographic records of Series A through D, to see if any consistent patterning were evident was made. A contingency table showed only chance variation.

The same care in avoiding possibility of sensory cues or shuffling inadequacies was observed in the experiments of Series II, although only in Series Hc were the precautions objectively demonstrable.

Of persistent expression in the literature is the hypothesis that errors in observation might account for certain card-calling test results. In most of the research considered here, this possibility has been completely excluded. In Series A, B, C, and D three persons counted their scores and the two experimenters kept independent records. A third independent record of the itemized hits was the photograph of the table lay-out. In order to check the recording accuracy, a sample of 200 runs was arbitrarily selected from the photographic records and the hits counted independently of the record of either experimenter. No errors were found. The experience of Pratt and Woodruff in finding only three errors in 60,000 recorded items indicates that with the exercise of even reasonable experimental caution, recording errors in ESP tests are bound to be negligible.[13]

Some Suggestions Toward Further Research

Since the field of specific psychological correlates with extra-chance scoring in card-calling tests remains almost totally unexplored the implications for further research from this study are so numerous that any catalogue of them would reveal more the interests of the investigator than the limitations of the problem. The direction of a few of these further steps deserve notice here, however, if only to emphasize the need for such exploration.

13. The implications of the recording error hypothesis, the main proponent of which has been J. L. Kennedy, have never been fully developed. In insisting that recording errors may occur, and that they do occur in experiments controlled to detect them, Kennedy has persistently overlooked their magnitude. In his original paper, he concluded that the motivational direction of recording errors had been established, a conclusion unsupported by his data, and ignored the obvious interpretation that, considering the many possibilities of error introduced, his student observers evidenced a high degree of accuracy (9).

Concomitant T-tests and card-calling tests have been defined in this analysis as consisting of a 10 run series followed by an undefined number of runs. Justification for this limitation lies simply in its practical outcome: that a relation could be observed. If, however, it is assumed that the T-test measures some characteristic of the subject's performance which persists through the experimental session, then it would be interesting to note whether the T-test and CC test relation would be noted if both scores were calculated from some other serial section of the runs in the session. Furthermore, although the relation of a T-score calculated from the total runs of the session may correlate spuriously with CC scores based upon the same data it might be possible to set up some measure of this correlation which would permit, by partial correlation methods, determination of the "score affectability" relation to extra-chance scoring within the total session.

Within the same region of score-affectability as a predictive correlate of extra-chance scoring would come experiments in which a T-score would be computed from a task varying from the identity of condition considered here, to a marked difference in type of task, such as dart-throwing, puzzle-solving, etc.

An obvious implication from the standpoint of card-calling test method is the practical usage of the T-test classification in the collection of data. If efficient usage of laboratory time is not an essential element in a given experiment the outcome of a preliminary test would permit only B class subjects to be utilized in the experiment. Further, a useful corollary of such a method of subject selection would be to develop skills of presentation of the experimental tasks directed to predisposing the subject to respond with low score-affectability.

Conclusion

The aim of this investigation was to determine whether an objective test of some behavioral characteristic of the experimental situation might be utilized to separate subject-sessions into two classes: one in which it would be highly probable that extra-chance scoring in card-calling tests might occur, and the other in which the scores would show only chance variation. The type of card-calling test situation that suggested the most likely source of measures of variation other than in success of card-calling were those subject-sessions in which the subject made estimates of his scores

before each run. All such subject-sessions were collected for this study. Series I consisted of all pre-shuffle card-calling tests. Series II consisted of all post-shuffle card-calling tests.

In order to insure complete independence of the variables the subject session was divided into a T-test and a card-calling test. The first 10 runs constituted the T-test. The remaining runs constituted the card-calling test.

It was found that a product-moment correlation of each T-test card-matching score with its succeeding estimate yielded T-scores for each subject-session which represented the degree to which the subject's estimates were influenced by experience with the just-preceding score. The T-scores permitted grouping of subject-sessions of Series I into two classes: class B, those which showed little or no influence of previous scores on their estimates (T-scores of -.20 to +.20); and class A, those which indicated a somewhat larger influence of scores upon estimates (T-scores greater in absolute value than .20). The total of CC scores of the class B sessions was significantly greater than chance. The class A group CC scores were scores wholly comparable to a chance distribution.

The observation in Series I was repeated upon the subject-sessions of Series II. The same division of the card-calling test data into a chance class A and an extra-chance class B ensured. This second observation supported the validity of the T-score grouping in Series I.

Although the *a priori* reliability of the T-score for a given subject-session is low, examples of several T-scores for single subjects show that the average T-score for a given subject may be significantly different from the average of another subject. This observation indicates that the T-score may measure some mode of behavior characteristic of the subject from session to session.

An interpretation of the estimates as scores aspired to with concomitant experiences of success or failure leads to the observation that if the subject is unduly unsuccessful in estimating his scores (over-estimates in 7 or more of the 10 runs), the A and B class separation of subject-sessions shows no difference in CC scores, with both at a chance level. Since it is found that subjects do not respond significantly to "successful"

estimates but do respond significantly to "failure" of estimates, it is suggested that the "failure" situation inhibits extra-chance scoring.

When the subject's level of estimate is calculated from the T-test, it is noted that levels of estimate below or equal to the expected chance scoring average preceded higher card-calling scores than estimate levels greater than the expected chance average.

The fact that an analysis of variance in Series I showed a significantly extra-chance distribution suggesting an underlying cognitive process in pre-shuffle card-calling tests supports the precognition hypothesis.

References

1. Frank, J. D. Some psychological determinants of the level of aspiration. *Amer. J. Psychol.*, 1935, 47, 285-283.
2. Gibson, E. P. Meteorological Effects in the ESP Test. (In manuscript.)
3. Gould, R. An experimental analysis of "level of aspiration." *Genet. Psychol. Monog.*, 1939, 21, 3-115.
4. Greenwood, J. A. Analysis of a large chance control series of ESP data. *J. Parapsychol.*, 1938, 2, 138-146.
5. Greenwood, J. A. & Stuart, C. E. A review of Dr. Feller's critique. *J. Parapsychol.*, 1940, 4, 299-319.
6. Greenwood, J. A. & Stuart, C. E. Mathematical techniques used in ESP research. *J. Parapsychol.*, 1937, 1, 206-225.
7. Hoppe, F. Erfolg und Misserfolg. *Psychol. Forsch.*, 1930, 14, 1-62.
8. Hutchinson, L. Variations of time intervals in pre-shuffle card-calling tests. *J. Parapsychol.*, 1940, 4, 249-270.
9. Kennedy, J. L. Suggestions concerning the nature and production of "extra-sensory perception." (In manuscript).
10. McGehee, W. Judgment and the level of aspiration. *J. Gen. Psychol.*, 1940, 22, 5-16.
11. Pratt, J. G. & Woodruff, J. L. Size of stimulus symbols in extra-sensory perception. *J. Parapsychol.*, 1939, 3, 121-158.

12. Rhine, J. B. The effect of distance in ESP tests. *J. Parapsychol.*, 1937, 1, 172-184.

13. Rhine, J. B. Extra-Sensory Perception. Boston: Bruce Humphries, 1935.

14. Rhine, J. B. Experiments bearing on the precognition hypothesis: 1. Pre-shuffle card-calling. *J. Parapsychol.*, 1938, 2, 38-54.

15. Rhine, J. B. Experiments bearing on the precognition hypothesis: 3. Mechanically selected cards. *J. Parapsychol.*, 1941 (in press)

16. Rhine, J. B., Pratt, J. G., Stuart, C. E., Smith, B. M., & Greenwood, J. A. Extra-Sensory Perception After Sixty Years. New York: Henry Holt, 1940.

17. Rider, P. R. An Introduction to Modern Statistical Methods. New York: John Wiley & Sons, 1939.

18. Smith, B. M. Note on a shuffling machine. *J. Parapsychol.*, 1938, 2, 231-232.

19. Snedecor, G. W. Statistical Methods. Ames, Iowa: Collegiate Press, 1937.

20. Stuart, C. E. The effect of rate of movement in card-matching tests of extra-sensory perception. *J. Parapsychol.*, 1938, 2, 171-183.

21. Warner, L. & Clark, C. C. A survey of psychological opinion on ESP. *J. Parapsychol.*, 1938, 2, 296-301.

THE ULTRA-PERCEPTIVE FACULTY

J. Hettinger Ph. D. (London)

Introduction

1. In this and a second article to follow I propose to give a comprehensive account of a piece of research which I carried out at King's College, London University, Professor F.A. Aveling, Head of the Psychological Department, being my supervisor.

I became interested in psychical research as far back as 1928 but, although I gained a good deal of experience on the mental side of the subject by many years of frequent attendance at private, group and public demonstrations given by well, and less well, known Sensitives, it was not until 1933 that I realized that my continuing to collect so-called "experiences" was not enough and that, to be useful, I ought to concentrate on experiments of a more constructive nature. It may be that the training in Physics and Electrical Engineering and the continuous contact with inventors in the exercise of my profession as a patent attorney are to a certain extent responsible for my tendency towards a more pragmatic approach to psychical research.

I joined King's College in the autumn of 1933 with a view to attending the psychological classes in my spare time and making myself fully acquainted with psychological methods of investigation. At the same time I started, so to say, to feel my way as regards the post graduate research which I contemplated undertaking[1]; some of the students serving as Subjects of experimentation, and a Sensitive I had known for years to give satisfactory results with the particular "psychic power" I had decided to concentrate upon acting as percipient. The power in question was that generally referred to in psychical research under the misnomer of "psychometry", according to which the Sensitive mentally "concentrates" on any article submitted by the experimenter, and during the "concentration" receives a number of impressions, visual, emotional etc., which the owner of the article verifies as having a particular meaning for, or as being applicable to, him or her.

---

1. Actually commenced in May 1934.

The research comprises the following distinct steps:

a. Statistical experiments with a view to ascertaining the probability of the existence of the so-called "psychometric" power;

b. Time-factor experiments;

c. A new method of psychic investigation;

d. Miscellaneous experiments.

The statistical and time factor experiments formed the subject of my Pd. D. Thesis published[2] under the title: *The Ultra-perceptive Faculty*, with the kind support of a grant from the Publication Fund of the London University. The present article deals solely with these experiments. The new method of investigation and the miscellaneous experiments form the subject of a second volume: *Exploring the Ultra-perceptive Faculty*, published by the same firm[2]. A second article will deal with these further explorations.

2. Dr. Rhine's work in America became known to me after my own research had already begun and, as will be appreciated from the above introductory remarks, I followed totally different lines from those adopted in his investigation. It is true that, in the first instance, we had a common object in view, viz., to test statistically the probable existence of some form or other of so-called "psychic power," the reality of which has been in dispute for many decades, but this ends the line of contact between the two pieces of research. The main differences are:

a. My experiments are concentrated on "psychometry", as above defined, in contradistinction to clairvoyance and telepathy with cards bearing certain designs;

b. I use professional Sensitives alleged definitely, more or less permanently, to possess the ability the existence of which is to be tested, as against: a) a more or less large number of persons who are not professionals and who may or may not possess the alleged faculty or b) persons, again non-professionals, who seem temporarily to be endowed with the ability and then lose it, so that they are unable to display it again and again for the purpose of further demonstrations.

c. The first difference referred to above necessarily includes another important one, namely, that the number of different items perceived in the case of

2. Rider & Co., London.

"psychometry" is unlimited, as against the limitation to five or more different designs on cards. On the face of it this third difference may seem to introduce great complications from a statistical point of view, as compared with the statistics in the case of a well defined number of items to be perceived, such as used by Dr. Rhine. How, in spite of the apparently unavoidable complications, the quantitative results have been made available to very simple statistical calculations is, as will be realised from the account which follows below, one of the important features of this research.

3. At the time my own experiments were started the expression Extra-sensory perception or ESP had not yet gained its present popularity and, considering the double character (sensory and purely intellectual) of the perception involved in these experiments, I could not possibly accept that expression as being scientifically applicable thereto. Let me explain.

In experiments with cards in which the percipient gets the image of a wavy line, circle, etc., one may have no compunction in accepting the expression ESP, since the image is something sensory and, in so far as it is not due to ordinary vision, may be referred to as "extra-sensory".

Suppose, however, a "psychometrist" says that a person, miles away, on whose silk handerchief she is mentally concentrating, is "exceedingly religious" - which proves to be correct - and then adds that the impression was purely *mental*, unaccompanied by any image. In the absence of any *sensory* perception, in the recognised psychological sense, I fail to see how "extra-sensory" is applicable to such a case. On the other hand, if the expression is intended to include mental impressions of the hitherto called "psychic" character and at the same time exclude purely mental perceptions recognised as normal, "extra-sensory" does not convey any differentiation between the two, quite apart from the fact that the qualification "sensory" rules out its applicability to purely intellectual perception.

I therefore submit, with all due respect, in spite of the popularity the expression "extra-sensory perception" has gained in the meantime that, from a psychological point of view, so far as *certain kinds of perception* involved in psychical research are concerned, the use of the expression in connection therewith is not appropriate, and to a certain extent confusing.

It will readily be granted that our ordinary perception includes both sensory perception and purely intellectual perception without sensory experience, and that the alleged perception which comes into question in psychical research lies beyond this normal double range. I therefore suggested the use of the expression "ultra-perceptive faculty" to denote the ability of perceiving beyond the normal range of the scale of perception, in the same sense as "ultra" is always used as a prefix to denote something beyond the normal; as, for instance, in "ultra-violet", which means the invisible rays of the spectrum following immediately beyond the ordinary visible violet.

The following diagram reproduced from page 19 of my book[3] will be found self-explanatory.

ULTRA-PERCEPTIVE FACULTY	ORDINARY PERCEPTION				INFRA-PERCEPTIVE FACULTY
PSYCHIC	Intellectual		Sensory		INSTINCTIVE
	Intuition	Contemplation	Localized	Non-Localized	

Since, according to the leading dictionaries, the prefix "ultra" is not necessarily limited to its association with an adjective, but may also be applied to a noun, I propose using in future, in addition to: "ultra-perceptive faculty" and "ultra-perceptive cognition", the simplified form of: "ultra-perception", for which preference has already been shown in certain quarters.. I am even greatly tempted to suggest "ultra-percipient" as a substitute for "Medium" and "Sensitive", but such further substitutions will better wait until the concepts we are dealing with are more clarified and settled.

4. May I, in conclusion of this introductory part, express the hope that the work reported in these two articles will appeal to other psychological researchers and prompt them to an attack on the problem from the various angles expounded therein.

It is not for me to discuss the relative merits of the researches respectively initiated by Dr. Rhine in America and by myself in England, nor does it appear necessary to do so; but, in view of the experience I have gained of the subject, I cannot abstain from highly recommending my own lines of experimentation, even if it be

3. J. Hettinger: "The Ultra-perceptive Faculty".

only by way of an alternative to the ESP card experiments. I do so, confident that once the reliability of the methods propounded is fully realised, as a result of personal experimentation, the impetus the experimental side of this youngest branch of psychology will receive, as regards the exploration of both normal and ultra-perceptive cognition, will gradually lead to psychological achievements comparable in theoretical and practical importance with some of those which hitherto crowned the progress of the more concrete sciences.

STATISTICAL EXPERIMENTS

PRELIMINARY TESTS

(1st Series)

The first series of experiments had for its main object to ascertain a few facts which might serve as a guide for the selection of some useful method of scientific investigation.

The series extended over 172 tests, with a total number of items exceeding 2000; the average being about twelve separate items per test. The number of articles submitted at each sitting varied between four and six, and each sitting lasted for about one hour, so that each article was "psychometrised" for about ten to fifteen minutes. The intervals between the impressions received by the Sensitive with respect to the separate items given with each article were of about one minute duration, which included the time taken by the verbal statement which the Sensitive made and a period of silence, which seemed slightly to vary from object to object, as well as with one and the same object. The statements made were written down as accurately as possible.

The Sensitive in this, and in all subsequent series up to the 8th, was Mrs. F. Kingstone, who is a well-known professional of good reputation; the sittings took place at her home, between about 6.30 and 7.30 p.m. This series extended over thirty-five sittings, three of which I did not attend; for the latter, the articles were sent to the Sensitive by post.

The objects submitted belonged to sixty-three different Subjects of most varied occupations and temperaments, including a number of King's College students.

The records of the statements made were submitted to the respective Subjects, who stated against each item whether it was correct or not. Only very few of these records were not duly filled out and returned.

The nature and character of the items. When the records of this series were analyzed as to their contents, it was found that the items given by the Sensitive varied considerably in their character and that those which seemed to be mostly acknowledged as correct were those which concerned the affective life of the Subject. Here is an example: 'Trouble with the right ear.' Events, actions and thoughts, which seem to be connected with some feeling or emotion on the part of the Subject are also well represented. Here is an example of each: 'A presentation of a walking-stick, silver mounted'; 'a lot to do with books in paper covers'; 'She thinks a lot of a little child - I think it is a little girl - not her own'.

A large number of items concerned the interests and behaviour of the Subject, things he has seen or heard and people with whom he is or has been associated. Some of the items referred to the personal characteristics of the Subjects and it was really impressive to find how well they applied to them.

A large number of items were too vague or too general. Some of the Subjects admitted them as correct, whilst others left them unanswered. The following example will suffice to show the nature of these items: 'A good prospect around this person, all favourable, but there is a lot of hard mental work.'

Although the items were often differentiated as to whether they applied to the owner of the object "psychometrised" or to someone else associated with him, this was not always the case. The same remark applies to the differentiation as between the present, past and future.

Trivialities. A few of the Subjects remarked about the triviality of a large proportion of the items; this was mainly due to a misconception of the object of these experiments and to the wrong assumption of a theory, which the writer did not set out to test. Perception due to the ultra-perceptive faculty must necessarily embrace the whole of the field that is accessible to ordinary perception, and, generally, the bulk of it may be said to consist of trivialities: the sight of a car, the noise of a typewriter, the scent of a rose, etc. The main question is: is there such a thing as an ultra-perceptive faculty? It may be that the simpler the character of the things perceived by the Sensitive, the less difficult it will be to prove the existence of such a faculty.

The nature of the objects. The only stipulation made with respect to the objects submitted was that they should have been in contact with, or handled by, the respective Subjects. The character of the objects varied considerably. Use was made of wallets and purses; pencils and fountain-pens; watches and chains; rings and necklaces; brooches, bangles and cuff-links; written notes and diaries; mirrors, scissors and penknives; locks of hair, combs and hair-slides; gloves and silk handkerchiefs; keys and magnets; a cigar piercer, a clothes peg, a slide rule, etc. etc. A very large number, however, consisted of private letters in sealed envelopes (55) and of blank sheets of paper (28), which the Subjects were asked to carry on themselves and handle from time to time.

Judging from the records, this series did not supply any definite information as to whether any particular *material*, e.g. leather, paper, silk, metal, wood, etc., gave better results than the others, although, at first, it appeared as if metal was not suitable.

The Sensitive was under the impression that she had to handle the object and be in *direct* contact with it, by grasping it in one hand, holding it between the palms of both hands, or pressing it against the forehead. Noticing a certain tendency on her part to refer to matters one could have inferred from the objects, the latter were enclosed in *sealed envelopes*. The results in no way suffered by this alteration; on the contrary, the writer had the impression as if an improvement had thereby been brought about, which he attributed to the Sensitive's own imagination being less active.

At one sitting, the Sensitive was asked not to touch the sealed envelopes, but just 'concentrate' on each one of them as they were placed one after the other, next to her, on a table. Again the results were not impaired by the *Sensitive not touching the object.*

Relation between Subject, Experimenter and Sensitive. Results were obtained, whether the experimenter did or did not know the Subjects; also, when he did know the Subjects, whether he did or did not know whose object he submitted each time in the sealed envelope.

This may seem to point to the conclusion that the experimenter's mind does not play any part in the *modus operandi* of the ultra-perceptive faculty. Such a conclusion, however, would be premature, not only because the experiments in this series were not at all of a systematic

character, but also because in the case of some records a few of the items were fully applicable to the experimenter, and were not accepted by the Subjects as applicable to them. This might point to telepathy from the experimenter, but still another inference would be that this was due, not to his presence at the sitting but to his having handled the articles. It will thus be seen that no valid conclusion can be arrived at from general observation alone.

One interesting experiment consisted in submitting, on two occasions, sealed envelopes containing the Sensitive's own letters, addressed to the experimenter. When the Sensitive was subsequently asked which of the items would be applicable to her and which not, without, of course, her knowing that the records were her own, the results were found to be rather poor; the best item was: 'A lady who puts on a better dress in the evenings (not an evening dress) for a special purpose, which I cannot quite make out'. This was quite correct and was connected with her occupation as a medium.

Tentative methods. In this preliminary series no attempt was made to obtain results which could be considered statistically, although the percentage method was applied to all the 111 records which were filled in, and in which all answers which were not definite, such as, 'probably', 'may be', and the like, were disregarded.

It was clear from the very outset of the experiments that the percentage method was not suitable for a 'scientific proof' since, in some tests, the percentage of the recognized items was very low: 1 or 2 out of 12, yet highly significant, and in others it was very high: 8 or 10 out of 12, yet too general to be individually significant.

Tentatively the following control methods were tried:

1. Multiple control. Copies of all the four records of one sitting were given to all the four Subjects, and each one was asked to select the one he or she thought would be the most applicable to him or her. Knowing from some of the previous experiments, not recorded herein, that the task was by no means an easy one, the Subjects were given two chances, and they were asked to mark the most preferred record as: 'Choice No. 1' and the next most applicable one as: 'Choice No. 2', unless they were satisfied with one single choice.

Some selected rightly under 'Choice No. 1' and some under 'Choice No. 2', and some were right, although

they selected one record only. On the other hand, some went wrong altogether, and a few stated that it was impossible to select, as practically all the records seemed to be more or less applicable. Owing to the difficulty encountered by some of the Subjects, this method was abandoned after a few tests.

2. Made-up control. Each of the records was associated with another record which was made up, but in this case the following practical difficulty was encountered: I did not consider myself to be the proper person to supply the control items with respect to the Subjects I personally knew, and I did not find it convenient to use another person who would supply the thousands of made-up control items which were to follow with the contemplated experiments. Besides, I further came to the conclusion that in the event of the ultimate statistical verdict being in favour of the existence of the psychometric faculty, such a made-up control might be severely criticised on the ground that a medium is more clever than others in selecting the appropriate items.

3. Combined control and evaluating method. It was thought that by combining the control with the evaluating method the difficulty of selection could be overcome.

Each record was paired with a control and, according to their significance, all the items were given one of the following values: 3, 2+, 2, 1+, 1, 0. All the values of the items found to be applicable were treated as positive, and those of the items not accepted as applicable were treated as negative quantities.

The total result of three such tests showed a slight difference in favour of the records of the items given by the Sensitive.

It was clear that this method could not be used extensively with a large number of Subjects, since such evaluation could only be made by the Subjects themselves, and any statistics would have been open to very serious criticism, on the main ground that but very few could have been depended upon to evaluate properly. Besides, even with the best judgment, such a method would be open to the further criticism of being largely arbitrary, and that it would consequently lend itself to mathematical manipulation which could easily be controlled so as to bias the results one way or the other.

4. A checking method. A letter about six months old was enclosed in a sealed envelope and submitted to

the Sensitive. The record was handed over to the writer of the letter with the intimation that it was hers and she was asked to fill it in and, in addition thereto, to submit it to five of her colleagues for the same purpose. The idea was to check how many of the items would be claimed by the others as being applicable to them. Here is the result:

Writer of the letter	A	B	C	D	E
10	8	6	4	4	3
10	10	10	10	10	10

Of all the methods tentatively tried in this series, the latter was the first one which seemed to lend itself satisfactorily to an investigation of a statistical nature. This method was applied to the second series of tests.

THE CHECKING METHOD
(2nd and 3rd Series)

This method was used in the second series, and with a certain modification in a third series of tests.

In the second series the articles of five different Subjects were submitted to the Sensitive, in turn, and five separate records were obtained. The Subjects were then asked to fill in the first column in their own record and additionally a column allocated to them in each one of the other four records which were not theirs.

Forty tests to which the method was applied yielded 584 items, out of which 237 were admitted as being applicable to the Subjects to whom the records actually belonged. Thus, the average per Subject was 40% as compared with nearly 50% in the first series. The total of the items admitted by the checking Subjects was 321, the average per checking Subject thus being less than 14%. On page 148 is a copy of one of these records.

The main object of this series was not to find out the relative percentages per actual and checking Subject respectively, but ascertain how many of a number of records will show a maximum for the actual Subject as compared with all the four checking Subjects associated therewith.

Comparing the entries of all five Subjects on the record of one Subject, it was found that in 29 out of the 40 records, the Subject's score surpassed each one of the

26th. Nov., 1935. 2nd. Series. Test No. 17.

Pocket Book

	Subject	Controls			
	17	16	18	19	20
1. I hear a man singing after a dinner in a concert - "Rocked in the cradle of the deep."	-	-	-	-	-
2. This person has been working under high pressure, now relaxed a bit - but will soon start again.	-	-	-	yes	yes
3. I get a lot of thoughts about business matters; uppermost in the mind; time is important; such a lot to do.	yes	-	-	yes	-
4. A nature that does not mind how much he does for humanity. I get a human element here.	-	yes	-	-	-
5. In lots of ways this person is an idealist.	yes	yes	-	-	-
6. Five years ago he lost someone by death, a very near relative; left a deep impression.	-	-	-	Yes S* my wife.	-
7. I see sandwichmen advertising something; must have a symbolic meaning, as regards publicity of something.	-	-	-	-	-
8. An elderly lady in the room and as soon as this person comes home, he goes upstairs to see her. She is rather deaf.	yes	-	-	-	-
9. I see a gentleman using the left hand. I think it means he uses the left hand just as sure as the right one.	yes	-	-	-	-
10. Waiting for something important in March.	yes	-	-	-	-
11. I get an irritation about humming a tune and keeping time.	-	yes about whistling.	-	-	-
12. I see a lot of new bandages rolled up.	-	-	-	yes, in the foot	-
13. A recent change of diet. (reluctantly)	yes	yes Breakfast.	-	-	-
14. Worrying about hair getting thin.	-	-	-	-	-

* S denotes that the Subject considers the item to be especially significant.

checking scores, and a comparison of the entries of one and the same Subject in all five records showed that in 29 out of the 40 records, the Subject's score in his own record was greater than his checking scores in the other four records of the group.

Statistical meaning of the results. Throughout this work we are concerned solely with statistics applicable to *simple sampling*, as in the case of coin-tossing or dice-throwing.

In the case of a coin which has no tendency to fall on one side more often than on the other, we may expect that in a long series of throws 'heads' will come uppermost approximately the same number of times as 'tails'; similarly in the case of a die, that it will fall with each of its faces uppermost an equal number of times. In the first case the chance of throwing 'heads' is $\frac{1}{2}$, and in the second case the chance of throwing any specified face is 1/6. Generally such chances are referred to as the chances of *success* as compared with the chance of *failure*, which in the above examples are 1/2 and 5/6 in the case of the coin and die respectively.

If in a statistical sample a proportion p of successes is different from the proportion *theoretically* expected, the question arises whether this difference is solely due to fluctuations of simple sampling or to some subsisting cause. This question is answered by comparing the difference observed with the *standard deviation*, which in the case of simple sampling is also called the *standard error*. If the difference is found to be so great as to exceed a certain arbitrarily fixed amount of fluctuations of a simple sampling, which are $\pm$ three times the standard deviation (standard error), then the same may be considered to be *significant* and to be due to some subsisting cause other than chance fluctuations.

The standard deviation or standard error of a proportion is given by the formula:

$$\sigma_p = \sqrt{\frac{pq}{N}}$$

in which p and q are the proportions of success and failure respectively, and N is the total number of trials. The *probable error* PE_p is $\frac{2}{3}\sigma_p$.

A difference of $2\sigma_p$ or 3 PE_p is often taken as significant. However, for the sake of safety, in view of the fact that a number of textbooks dealing with

statistical methods seem to differ slightly as to the actual arbitrary value, this value, following G. U. Yule (An Introduction to the Theory of Statistics, p. 311), has been taken as 4 PE_p., for which the odds are 142 to 1. Values from slightly below 3 × PE_p, for which the odds are 16 to 1, to slightly below 4 × PE_p are considered as being on the border line of significance.

Now the quantitative results may be statistically utilised as follows:

The number of successes S, viz. the number of records or items accepted from among the records or items given by the 'Sensitive' to the Subjects, and the number of failures C, viz. the number of records or items accepted from among the 'control' records or items, are added together, and the total is considered to constitute the number of trials N in the above formula with

$$p = \frac{S}{N} \text{ and } q = \frac{C}{N}, \; p + q \text{ thus being equal to 1.}$$

Both the 'Sensitive' and the 'control' are assumed to have the same chance, the theoretical probability thus being 1/2, and the proportion p of successes obtained is compared with this theoretical probability.

Referring now to the above tests we get:

$$p = \frac{29}{40} = 0.725; \quad q = \frac{11}{40} = 0.275; \quad \text{and } N = 40$$

The standard deviation therefore is:

$$\sigma_p = \frac{0.725 \times 0.275}{40} = 0.0685$$

and the probable error: $PE_p = \frac{2\sigma}{3}p = 0.0475$.

The ratio obtained was: $\frac{29}{40}$ or 0.725

The significance of this ratio, (in view of the theoretical probability of the above tests, viz. $p = \frac{1}{5} = 0.2$) is therefore tested:

$$0.725 - 0.2 = 0.525, \text{ or } \frac{0.525}{(PE_p = 0.0475} \text{ viz}$$

$$\underline{\underline{11.48 \times PE_p}},$$

The result is *highly significant.*

A reliability test. The highly significant figure just referred to seemed too good to be acceptable without further analysis.

The tests clearly showed that a few Subjects scored very high on their own records and very low on the records of the others. The question naturally arose whether this was not entirely due to an unconscious attitude, which inclined those Subjects to accept as many items as possible on their own records, because they knew those records were theirs, without their being equally liberal as regards the records which they knew did not belong to them. If it could be proved that this was the case, the results would then lose the significance they seemed to convey, and the method employed would have to be discarded as unreliable.

A reliability test was therefore carried out which consisted in setting out the five records with the five columns as per sample above given and handing them over to the five Subjects, with the difference, as compared with the previous tests, that each Subject was purposely wrongly informed which record was her own.

The analysis of the five records showed that, whilst it is true that some Subjects tend to return a maximum for the record they *think* is theirs, it cannot be inferred therefrom that this applies to all Subjects. However, the fact that one of them entered a high maximum in the fictitious record, substantially surpassing her entries in all the other records, including the one which was her own, is a clear indication that when working with a number of different types of Subjects, more especially when some of them are entirely unknown to the experimenter, the checking method used in this series is not a reliable one.

A variation of the checking method. It was thought that the drawback of the checking method used in the second series of tests could be overcome by following the same procedure and using the same form of record, but not disclosing to the Subjects which one of the records belonged to them; the idea being to ascertain how many will score a maximum on their own record, if they assumed that any of the records might be their own.

This third series was a very short one, for the comments of the Subjects showed that, this time, they found the task of filling in the records far more difficult.

Fully realizing that not only the Sensitive but also the Subject played an important part in this form of investigation, the one perceiving the item and the other recognizing it, and that, like the Sensitive, the Subject too ought to be psychologically at ease, the method was immediately abandoned for a more simplified form.

THE METHOD OF ADMIXED CONTROL ITEMS
(4th, 5th and 6th Series)

In view of the further findings, viz. (a) that reliability imposes the necessity of not informing the Subject which items have been given by the Sensitive in connection with his article, and (b) that the method employed ought not to be such as to render the task of the Subject unnecessarily more difficult, a method was evolved which, it was thought, would fulfill these requirements, and which, in addition thereto, would also take into consideration the psychological predisposition of the Subjects, some of whom are inclined to claim a comparatively large number of items, whilst others only a very few.

This method, which may be referred to as the *'method of admixing control items'*, consists in mixing twelve items given by the Sensitive with twelve 'control items on one and the same record, and asking the Subject to state which of the items are applicable to him and which are not. By comparing the number (R) claimed from among the items given by the Sensitive with the number (W) claimed from among the control items, we obtain a measure which can be used statistically. If the ratio $\frac{R}{W} = 1$, this means that the Sensitive and the 'control' have scored the same number; $\frac{R}{W} > 1$ means that the Sensitive has scored more than the control; and $\frac{R}{W} < 1$ means that the control has scored more than the Sensitive.

By obtaining a sufficiently large number of records and counting on the one hand the number of all $\frac{R}{W} > 1$, and on the other hand the number of all $\frac{R}{W} < 1$, plus of all $\frac{R}{W} = 1$, we get two values, to which we can apply the theory of probabilities. Similarly, if we add together all the R's on the one hand, and all the W's on the other hand, we obtain again two values, which can be treated likewise.

The 'control' items were obtained as follows: all the items previously given by the Sensitive were written down on slips of paper, which were placed in a box, and the typing assistant was instructed, when typing each record, to draw twelve items from the said box, referred to as the 'guess-box', and intermix them with the twelve items given by the Sensitive in connection with the respective article.

It will be seen that in this way the two conditions above stipulated are fulfilled; further, that whether the Subjects are inclined to accept a large number of items, or hardly disposed to claim more than a very few, a kind of compensation is automatically effected in the fraction $\frac{R}{W}$, for, since the tendency to claim a large or a small number necessarily applies to both numerator and denominator of the fraction, theoretically speaking, the character of the latter as to its being = 1, >1, or <1 is retained throughout all the tests.

A sample copy of one of the records of this 4th series, in which 63 different Subjects participated, is shown on page 154. The control items from the 'guess-box' are shown cancelled. This was done after the return of the record by the Subject, who then received his copy showing the control items already cancelled therein. The exchange of the two copies was always done in such a way that the Subject was not able to make any alterations after he had received his own copy, nor could the experimenter alter the latter after he had received the actual record duly filled in by the Subject.

This series comprises 79 tests, but three records were not returned. The remainder of 76 records yielded 958 items, which were intermixed with the same number of control items in the manner above explained. Out of the 948 items given by the Sensitive 328 were claimed by the Subjects, and out of the same number of control items obtained from the 'guess-box' 222 were claimed by them. The scores thus were 34.5% and 23.5% respectively, as compared with the corresponding figures of 40% and less than 14% of the tests in the second series, when the Subjects knew which of the items were given by the Sensitive in connection with their own articles.

As regards the number of cases in which the Subjects claimed a greater number from among the items given by the Sensitive as compared with those belonging to the 'guess-box', the figures were as follows:-

$\frac{R}{W} > 1$ (50); $\frac{R}{W} < 1$ (19); $\frac{R}{W} = 1$ (7).

5th. March, 1936. 4th. Series. Test No. 13.

Small leather purse.

1. A restless mind; cannot stand still even if they get the opportunity.	Yes	R
2. I get some irritation.	Yes	W
3. Mind a good deal on domestic matters.	Yes	R
4. This person has a friend in the police force.	No	
5. I get a girl in the surrounding on this person, going to an art school.	No	
6. Arrangements being made by letter for someone living far away to come on a visit.	No	
7. I get neuritis in the arm.	Yes-my Mother.	R
8. Prizes received for something, but not sport.	No	
9. A big disappointment, just getting over it.	No	
10. I see such a lot of pencils (plain brown) in a bundle.	No	
11. A lot of small articles, all wrapped carefully in tissue paper.	No	
12. I get a picture painted and dissatisfied with the effort.	No	
13. I get dealing with and balancing out a lot of figures.	Yes	R
14. I think these people follow the catholic faith.	No	
15. I get a stout gentleman connected with this person who has a difficulty in getting up and down the stairs.	Yes-18 months ago.	R
16. I get something about moving into a new flat.	Yes-18 months ago.	R
17. This person is fond of home-made lemonade.	Yes	W
18. Some joke at someone's expense; not a very pleasant sort of impression.	No	
19. I get two children (girls) in the surrounding of this person, and I get the impression they have lost their father	No	
20. I see a class of children in some school.	No	
21. Fond of drinking cold water	No	
22. I see a small fire but not much damage done.	No	
23. Been somewhere recently, where there was heavy snow.	No	
24. I see something on the ground covered with a net.	No	

This means that, in 50 out of 76 cases, the 'Sensitive' scored a higher number than the 'guess-box', as compared with 26 cases in which the 'guess-box' was not surpassed by the Sensitive.

The Statistical Results. 1. Referring first of all to the *frequency of the cases* in which the Sensitive proved to be superior to the 'guess-box', and to that of the cases in which the 'guess-box' was not surpassed by the Sensitive, but either surpassed, or was at least equal to, her, we get as a result of the above tests:

$$\sigma_p = \sqrt{\frac{p \times q}{N}} = \sqrt{\frac{\frac{50}{76} \quad \frac{26}{76}}{76}} = 0.054$$

and $$PE_p = \frac{2}{3}\sigma_p = 0.036$$

The ratio obtained was: $\frac{50}{76}$ or 0.657.

Comparing this ratio with the theoretical probability of the above tests, (viz. $p = \frac{1}{2} = 0.5$), we have:

$0.657 - 0.5 = 0.157$, or $\frac{0.157}{(PE_p = 0.036)}$, viz.

$$\underline{4.3 \times PE_p}.$$

which is *significant*.

2. Referring to the *frequency of the items* claimed from among those given by the Sensitive, and to that of the items claimed from among those due to the 'guess-box', we get:

$$\sigma_p = \frac{p \times q}{N} = \sqrt{\frac{\frac{328}{550} \times \frac{222}{550}}{550}} = 0.02$$

and $PE_p = \frac{2}{3}\sigma_p = 0.0133$.

The ratio obtained was: $\frac{328}{550} = 0.596$.

This ratio compared with the theoretical probability of the tests, (viz. $p = \frac{1}{2} = 0.5$), gives us:

$0.596 - 0.5 = 0.096$, or $\frac{0.096}{(PE_p = 0.0136)}$, viz

$$\underline{7.2 \times PE_p}.$$

which is *significant*.

Modifications of the method. Two modifications of this method were tried in the hope of possibly improving the results.

The first modification, referred to as: *'Selection by mutual consultation'* (5th Series), consisted in intermixing the items given to two Subjects on one record and asking each of them to co-operate and decide between them as to which items were more applicable to the one and which to the other.

The second modification referred to as *'Selection by Graded Significance'* (6th Series) consisted in intermixing the items (8, 10, or 12) given to one Subject with the items (8, 10, or 12) given to another Subject and asking each one of them to mark the records as follows: yes 1 against each item that strikes the Subject at once as being correct and significant, and yes 2 against any item subsequently selected, but so that yes 1 + yes 2 should be 8 (10 or 12) or less, but not more.

Both series were very short as it was obvious practically from the outset that the expected improvement would not be realised.

THE METHOD OF SELECTION OF ONE OF TWO PAIRED RECORDS

(7th, 8th and 9th Series)

The Subjects having reported that sometimes some of the items impressed them as being rather insignificant, it was thought that a greater deviation from the chance probability might be obtained if, instead of the items being intermixed, each Subject received two records, one being entirely his own, and the other entirely a 'control' record, and he decided by *general impression* which of the two, *considered as a whole*, was his.

In fact, this is only a simplification of the method of multiple control, referred to under *'Tentative Methods'*, which was discarded, since the Subjects found it very difficult to select their own record when the same was associated with as many as three 'control' record. It was hoped that, by using one control only, the difficulty of selection would be greatly reduced, and that the selection of a record as a whole, taking into consideration the significance of the items rather than their number, would yield improved results.

In addition to making a selection by general impression the Subjects were also asked to annotate both

records, in order that a comparison might at the same time be made between the number of items accepted on the proper record and the number of those accepted on the 'control' record.

In all the tests of this series, with the exception of the last four, the records of two Subjects were paired together, the one acting as a 'control' of the other. In the four exceptions, the 'control' records were compiled by taking one item from each one of the previous records, according to a pre-determined order.

Each Subject received two such records, and he had to state which one he selected and make his annotations in one of two columns provided for him in each record.

A sample copy of two paired records is reproduced below. In this particular case, both Subjects selected rightly and, in addition thereto, both scored a maximum on their own record as compared with the 'control' record.

The scoring of Subject No. 3 was: on own record (No. 3), 8 out of 11; on 'control' record (No. 4), 4 out of 10; and of Subject No. 4: on own record (No. 4), 9 out of 10; on 'control' record (No. 3), 1 out of 11.

The 7th Series extended over 54 tests. 28 Subjects selected the right record, 16 the wrong one and 10 made no selection. The aggregate number of items on the 54 records was 648, of which 207 were claimed on the right records and 162 on the 'control' records, the score thus being: nearly 32% and 25% respectively. These figures show a slight *decrease* on the *right* records and a slight *increase* on the 'control' records as compared with the percentage scores in the 4th series, which were 34.5% and 23.5% respectively. Such slight variations do not justify any inference being drawn therefrom.

The Statistical Results. 1. In this series the number of cases in which no selection was made ought not to be added to the number of cases wrongly selected, as was done in the previous series with regard to $\frac{R}{W} < 1$ and $\frac{R}{W} = 1$, the respective numbers of cases of which were added together in favour of the 'control', as against the number of cases of $\frac{R}{W} > 1$ in favour of the 'Sensitive'. The aim of the previous series was to find to what extent the 'Sensitive' was superior to the 'guess-box', so that the said addition was fully justified; in this series,

11th August 1936.

Combined Tests No. 3 & No. 4. 7th Series. Test No.3.

Selected by owner of wallet.	Entry by owner of wallet	Entry by owner of ring
1. I get the impression that this person has been cross-questioned in a Court of law.	Yes. 1934.S.	No.
2. This person has an attractive way of speaking.	Yes.	No.
3. This person has been warned not to exert themselves.	No.	Yes. S.
4. This person is often in the company of someone who has a slight impediment in speech.	Yes.	No.
5. Making new residential arrangements for some fresh accommodation.	Yes.	No.
6. This person is filled with other peoples' difficulties.	Yes. S.	No.
7. I see an unusual broad ring, plain and slightly raised - I see no stone.	No.	No.
8. Concerned about a doctor who is retiring and is giving up his practice.	Yes. Friend	No.
9. An elderly lady closely connected; worried about leaving her after a short stay; reluctant to leave her; she is not laid up.	Yes. Relative	No.
10. I get a lot of women choir voices and having an interest in it. Not casual, some connection with it.	No.	No.
11. I see a mole on this person's right hand.	Yes.	No.

the main object was to find how often the right record would be selected as compared with the wrong (control) one, and the Subject was directed not to make a selection if he found it practically impossible to do so.

So far as the *frequencies of the record selection* are concerned we get, as a result of the tests, the following:

11th August 1936.

Combined Tests No. 3 & No. 4. 7th Series. Test No. 4.

Selected by owner of ring.	Entry by owner of ring	Entry by owner of wallet
1. Feeling very happy about a newly formed friendship of the opposite sex.	Yes. S.	Yes
2. Some slight anoyance about going up a dark staircase.	Yes.	No.
3. In temporary charge of a department.	Yes. S.	No.
4. I am on a cruise for some time.	Yes.1934	No.
5. Something slightly wrong with the instep of the foot.	Yes.	No.
6. This person would love to have her fortune told.	Yes.	Yes.
7. Helping another person with some domestic work in the spare time in a friendly way.	Yes. Relative	No.
8. This person is thinking a lot about 18 months ahead.	No.	Yes.
9. I see a new blue dress, long; fitting beautifully and tightly.	Yes. S.	No.
10. Some annoyance about gossip and wishing other people to mind their own business.	Yes.	Yes.

$$\sigma_p = \sqrt{\frac{p \times q}{N}} = \sqrt{\frac{\frac{28}{44} \times \frac{16}{44}}{44}} = 0.074$$

and $$PE_p = 0.0493.$$

The ratio obtained was:

$\frac{28}{44}$ or 0.06363 and the

deviation from the chance probability:

$$\frac{2.74 \times PE_p}{}$$

which is *very near* 3 x PE_p, sometimes taken as *significant*, as compared with 4.3 x PE obtained in the 4th series.

2. Referring to the *total scoring* of the items respectively claimed on the right and on the control record, irrespective of the fact whether the right or the wrong record was selected, or whether no selection was made, we get for the aggregate 369 of accepted items, of which 207 are on the right records and 162 on the control records:

$$\sigma_p = \sqrt{\frac{\frac{207}{369} \times \frac{162}{369}}{369}} = 0.0258,$$

and $$PE_p = 0.0172.$$

The ratio obtained was: $\frac{207}{369}$ or 0.56 and the deviation from the chance probability

$$\underline{\underline{3.48 \times PE_p}}$$

which is on the border line of significance.

This figure is not so high as that obtained in the 4th series ($7.2 \times PE_p$).

3. Since the results of this series are lower than those obtained in the 4th series, we may conclude that, from the point of view of testing the existence of an ultra-perceptive faculty, the method of selection of a whole record is less satisfactory than the method of 'admixing control items' used in that earlier series.

In the 8th and the 9th series which followed and, in fact, in all the subsequent series, use was made, in addition to Mrs. Kingstone, of a second Sensitive, viz. Miss F. Fallows, recommended to the writer as a good 'psychometrist'.

A modification. As regards the 8th series I decided to mark the records of all the Subjects who previously gave a fair number of good results as belonging to the 'A' group, and those of the Subjects who previously gave poor results, together with those of all the new Subjects, as belonging to a mixed group 'M'.

This series comprises 58 tests with Sensitive 'K' and 130 with the Sensitive 'F'. The great difference in the number of tests is due to the fact that, whereas 'F' was able to give me two sittings a week, 'K', owing to numerous other engagements, could not spare more than one evening a week, now and then even missing a week.

The figures obtained showed that the expected improvement of obtaining a substantially better result for

the group A, as compared with the group M, had not been realised. This, however, may have been due merely to wrong grouping and to a too hasty transfer from Group M to A, namely, when some of the Subjects were found to give good results at the outset.

The Statistical Results. The tests with 'K' showed, as regards frequencies of record selection, a deviation from chance probability which was practically nil; and as regards *frequencies of items* a deviation of $2.42 \times PE_p$ which is below significance.

The corresponding figures for the tests with 'F' were: $4.4\ PE_p$ and $8.18\ PE_p$, both of which are *significant*.

A comparison between the two Sensitives. 1. This series was the first one in which I was able to make a comparison between two Sensitives.

Although the statistical figures given show a marked difference between the results obtained with the one and with the other Sensitive respectively, it was noticed with both Sensitives that at some sittings they were not 'quite up to the mark', whilst at other sittings they seemed to be 'in good form' throughout. Most of the disturbances in one case were due to singing in the street and telephone bells and, in the other case, to a neighboring wireless set, in addition to worry.

Since 'K's' sittings *sometimes* gave good results almost *throughout one and the same sitting*, and 'F's' sittings *sometimes* gave poor results also practically *throughout one and the same sitting*, the logical inference therefrom is, that, although the statistical method supplies us with a means of quantitatively determining the ultra-perceptive ability of different Sensitives, such determination is meaningless, unless the psychological state of the Sensitive at the time of the sitting is also taken into account.

2. Judging from a few instances it was first thought that the one Sensitive gave better results with some Subjects, and the other one with the others, but this was not confirmed as the sittings continued. However, this seems to be a point well worth investigating.

3. A marked difference between the two Sensitives is that 'K' gives quick expression to what she perceives and accompanies it by her interpretation, whilst 'F' has often difficulty in expressing herself. Another important

difference is that 'F' is far less conscious of her surroundings, although not actually in a trance, and that she yields completely to feeling, actually changing her facial expression according to such feeling.

A further modification was introduced in the 9th series, which was a short one. The method used in this series differs from the preceding one in that each of the items, in both the 'right' and the 'control' columns, is given by the experimenter a value of 1, 2, or 3, according to its apparent significance, which values are entered on the records before they are dispatched to the Subjects for selection and annotation. The guiding principle of this *evaluation in advance* is the likelihood that the items are, or are not, common to a large number of subjects, this being judged, to a certain extent, in the light of the annotations in the records of the preceding series. Evaluation in advance by one single person, viz. the experimenter, is to be preferred to evaluation by the Subjects themselves, since the latter are bound to differ greatly in their respective judgments. But quite apart from this I myself somewhat distrust evaluation, because of its necessarily arbitrary character. In fact, it was introduced into the method of 'paired records selection' only tentatively, by way of an experiment.

All items of a *general* character were marked I; those which were specific, II; and those which seemed exceptional, III.

The results were such that I decided to carry out a longer series of tests with *evaluation of the items in advance* and compare the results with *the scoring without evaluation.*

THE METHOD OF PAIRED ITEMS
10th, 11th and 12th Series

Most of the Subjects reported that sometimes they found it exceedingly difficult to decide on their selection, since both columns contained items which seemed to be especially significant. It occurred to the writer that this difficulty might be overcome by arranging that the selection be made between individual items instead of between whole records. This led to the adoption of the method used in these series, according to which each item given by the Sensitive is paired with a 'control' item and the Subject is asked to select one of the items in each pair, knowing that only one of them is his.

However, the Subjects were further instructed not to accept any of them, if neither was applicable, or accept both, if each one was equally applicable.

As regards the 'control' items, the 'guess-box' used in the 4th series was replaced by three 'guess-lists'. In the 10th series these were prepared by tabulating some of the items previously given and valued I, II, and III respectively, further items being added thereto as they were accepted in the course of this series.

The table on page 164 is an example of a record thus obtained. 50 records were obtained with Sensitive 'K' with an aggregate number of items of 500 (Sensitive) and 500 (control), of which 157 and 134 were respectively claimed, the percentage scorings thus being 31.4% and 26.8% for the Sensitive and control respectively. In the case of Sensitive 'F' 53 records were obtained with an aggregate number of items of 530 (Sensitive) and 530 (control), of which 215 and 158 were respectively claimed, the percentage scorings thus being 40.5% and 30% respectively.

For full details of the tests of this series I must refer the reader to pages 87-98 of my book[4].

It will suffice to state here that so far as the comparison between the scores with and without evaluation in advance is concerned, it was found that evaluation did not appreciably change the ratio between the score of the Sensitive and that of the control as compared with the results without evaluation.

The statistical calculations with respect to the total scorings without evaluation resulted in the following figures for the deviation from chance probability:

Tests 'K': $\underline{2.06 \times PE_p}$ which is not significant;

Tests 'F': $\underline{\underline{2.98 \times PE_p}}$ which is on the border line of significance.

Modification. As the results obtained in the 10th series showed a decline as compared with the preceding ones, the writer thought that this might be due to the fact that its 'control' items were taken from among the 'right' items which had been accepted by previous subjects, and a good many of which were, therefore, presumably of a character having a greater chance of being accepted also by many of the other Subjects.

4. J. Hettinger: "The Ultra-perceptive Faculty".

Sensitive K.

27.4.37. WALLET 10th Series. Test No.35.

1	Owner sitting in a place where there is a big draught and cannot get away from it		No
1a	I feel a *need* for this person for more rest and *sleep.*	I	Yes R
2	Owner - house hunting		Yes R
2a	Owner has to catch two trains, not sufficient time allowed for the one to catch the other	II	No
3	A lady's hand holding up a *small, unusual* mirror with handle		Yes W
3a	This person wears a ring that is uncomfortably thing	II	No
4	Owner is connected with an elderly lady with whom they have to exercise a lot of patience		Yes R
4a	Owner does not want to have about them anything but what is useful	I	No
5	This person is a fair man and drives a car		Yes R
5a	A lot of happiness out of a *new* friendship	II	No
6	I get a piece of broken glass		No
6a	Owner has recently been bothered with pains in the back	I	No
7	Plans being made by the owner and carried out about business and home		Yes R
7a	Owner not able to express their own life and must live in agreement with others	I	Yes W
8	This person uses sealing wax *quite a lot*		No
8a	Owner looking at a clock on the wall that replaced another, not yet accustomed to it	II	No
9	Owner is putting in a lot of energy in something that has been wasted		Yes W
9a	Owner has a special liking for grape fruit	I	Yes R
10	Owner used to like dancing, but not now	II	Yes R
10a	Owner stayed for a time in a place not their own and then changed by going to their own home		Yes W

Series 11 and 12 according to the method of paired items differ from the preceding one, in that the items are not paired with control items taken from any previously prepared 'guess-lists', but with the items given to another Subject at the same sitting.

Such method of pairing the 'right' with the 'control' items might cause the results to be vitiated to the disadvantage of the Sensitive, if the Sensitive reacted not only to the article on which she concentrated but also to the articles of the other Subjects brought by the experimenter to the sitting, and one of which supplies the control items. However, this is only a supposition, and not a proved fact, and for the sake of convenience the possibility of such interference was disregarded and the items were paired as stated.

Another innovation introduced in these series was that the Sensitives no longer touched the envelopes containing the articles; these were placed on a table or chair and the Sensitives concentrated on them at a distance of about two feet.

Owing to the inability of Sensitive 'K' to give me more frequent sittings, the numbers of 'K' tests in these series are substantially smaller than those of the 'F' tests.

In the 11th series there were 26 'K' records with an aggregate number of items of 260 given by the Sensitive paired with 260 control items. Of these, 87 and 68 were respectively claimed by the Subjects, the respective percentage scorings thus being: 33% for the Sensitive and 26.15% for the control. The number of 'F' records was 70, with an aggregate number of items of 700 which were paired with 700 control items. Of these, 293 and 223 were claimed in favour of the Sensitive and the control respectively, the respective percentage scorings thus being 41.85% for the Sensitive and 31.85% for the control.

The statistical calculations with respect to the total scorings resulted in the following figures for the deviation from chance probability:

Tests 'K': with evaluation in advance: $\underline{3.9 \times PE_p}$, which is on the border line of significance, and without evaluation: $\underline{2.26 \times PE_p}$ which is not significant.

It is interesting to note that the introduction of evaluation increased the total scoring of sensitive 'K' sufficiently to convert an insignificant into a

significant result. However, it is to be borne in mind that both figures lie in the proximity of the threshold of significance, so that we are not justified in drawing any conclusion from such a result, which may be considered to constitute a special case.

Tests 'F': with evaluation in advance: $\underline{\underline{6.35}} \times \underline{\underline{PE_p}}$ which is significant, and without evaluation: $\underline{\underline{4.62}} \times \underline{\underline{PE_p}}$ which is also significant.

Note:- The corresponding figure in the preceding series was $2.98 \times PE_p$.

It will be seen that also in the case of the Sensitive 'F', the scoring with evaluation gives an improved result in favour of the Sensitive. This improvement is only apparent since the ratio $\frac{\text{Sensitive}}{\text{Control}}$ for the scoring with evaluation was found not to be substantially different from that for the scoring without evaluation and, therefore, the higher value ($6.35 \times PE_p$ as compared with $4.62 \times PE_p$) can only be attributed to the fact that the doubled values of a substantial number of items (II) substantially reduced the standard deviation and consequently the probable error.

The 12th series was a short extension of the 11th one. The method of control by means of paired items, which was found suitable, was retained, but evaluation was discarded since, apart from its being to a certain extent arbitrary, it was found to be an unnecessary addition.

In 22 'K' records, the Sensitive scored 72 and the control 62; in 49 'F' records, the corresponding figures were 204 and 132 respectively. Owing to the very small number of 'K' tests in this series, they have not been considered separately statistically. The 'F' tests showed a deviation of $\underline{\underline{6 \times PE_p}}$, which is significant.

SUMMARY OF THE TWELVE SERIES OF TESTS

The table on page 167 is a summary of the main statistical results obtained in the twelve series of tests.

Although the various methods differed from one another in some particular respect, they had one factor in common, namely, that the Subject had always to decide on a number of items, of which exactly one half was given by the Sensitive whilst the other half consisted of

	Number of Tests	Total number of items	Total number of items accepted N+S+C	Total number of right items accepted S	Total number of control items accepted C	% of right items accepted	% of control items accepted	Probability calculated on total scoring $A \times PE_p$.
"K"								
TENTATIVE METHODS								
1st Series	111	1266		605		47.8		
CHECKING METHODS								
2nd Series	40	584		237		40		
3rd Series	9	135		40		30		
	160	1985						
CONTROL METHODS								
ADMIXED CONTROL ITEMS								
4th Series	76	948	550	328	222	34.5	23.5	$7.2 \times PE_p$
5th Series	16	176	110	63	47	35.8	26.7	—
6th Series	16	140	102	58	44	41.4	31.4	—
PAIRED RECORDS SELECTION								
7th Series	54	648	369	207	162	32	25	$3.48 \times PE_p$
8th Series	53	498	403	218	185	43	36	$2.42 \times PE_p$
9th Series	10	118	89	48	41	40.6	34.6	—
PAIRED ITEMS SELECTION								
10th Series	50	500	291	157	134	31.4	26.8	$2.06 \times PE_p$
11th Series	26	260	155	87	68	33	26.15	$2.26 \times PE_p$
12th Series	22	220	134	72	62	32.7	28.2	—
	323	3508	2203	1238	965			
"P"								
PAIRED RECORDS SELECTION								
8th Series	116	1261	963	565	398	45	32	$8.18 \times PE_p$
9th Series	12	142	92	55	37	38.7	33.1	—
PAIRED ITEMS SELECTION								
10th Series	53	530	373	215	158	40.5	30	$2.98 \times PE_p$
11th Series	70	700	516	298	223	41.85	31.85	$5.62 \times PE_p$
12th Series	49	490	336	204	132	41.6	27	$6 \times PE_p$
"P"	300	3123	2280	1332	948	Total %		$12.22 \times PE_p$
+ "K"	323	3508	2203	1238	965	calculated		$8.52 \times PE_p$
	623	6631	4483	2570	1913	38.75	28.85	$14.6 \times PE_p$

'control' items. In view of this fact, we are entitled to add all the results in one collective group and consider them statistically with regard to the number of items the Subjects accepted from among those given by the Sensitives and from the 'control' items respectively.

Referring to the general results obtained from the 4th series onwards, in which 'control' items were used, we get:

Total number of tests: 623.

Total number of items given by the Sensitive = total number of 'control' items: 6631.

Total number of items accepted (S) from among the items given by the Sensitive: 2570, viz. a percentage of: 38.75%.

Total number of 'control' items accepted (C): 1913, viz. a percentage of: 28.85%.

Total number of 'right' + 'control' items accepted (N=S+C): 4483.

From the above percentage figure we find that the Sensitive scored 34% higher than the 'control'.

Applying the usual formulae:

$$\sigma_p = \sqrt{\frac{p \times q}{N}} \quad \text{and} \quad PE_p = \frac{2}{3}\,\sigma_p,$$

we obtain the following cumulative results for the deviation from chance probability:

For Sensitive 'K': $\underline{\underline{8.52 \times PE_p}}$
For Sensitive 'F': $\underline{\underline{12.22 \times PE_p}}$
For 'K' + 'F': $\underline{\underline{14.6 \times PE_p}}$

These figures are substantially higher than those obtained with the individual series. Such result, however, was to be expected, since the values of the standard deviation σ_p and the probable error PE_p necessarily decrease with the increase in the frequency N, viz. the total of the items accepted in the various series.

We see from these cumulative results that both Sensitives 'K' and 'F' have given statistically significant figures, and their joint result:

$$\underline{\underline{14.6 \times PE_p}},$$

has therefore to be interpreted to signify that, whatever the explanation of the Sensitives' ability may be to

perceive the thoughts, feelings, actions, characteristics, etc., of other persons, in the absence of such persons, but in the presence of articles belonging to them, *from a statistical point of view, this result is not attributable to mere chance.* Hence we arrive at the conclusion that *the general statistical result obtained in this experimental investigation is probably due to the existence of an ultra-perceptive faculty possessed by the Sensitives.*

This conclusion is a very important one since, so far as I am aware, it is the first time that such proof has been established along strictly scientific lines in connection with professional mediums giving expression to that which they perceive, exactly in the manner they usually display their powers in public or private demonstrations, without any predetermination as to the things they are expected to perceive.

Scoring relationships in control methods. It has been found throughout *all* the series in which a control method has been used, viz. from the 4th series onwards, that, if we consider the total respective scorings of the 'Sensitive' and 'control' in each one of the records, the 'control' scores more often than the 'Sensitive' in the lower range of scoring, and the 'Sensitive' more often than the 'control' in the higher range.

The distribution of these total figures is shown in the following table, in which the results of all the series (4 to 12) of both Sensitives have been added together:

SCORES	0	1	2	3	4	5	6	7	8	9	10	11	12	
FREQUENCIES	32	61	83	93	104	79	59	48	36	10	14	3	1	623 (Sensitives)
	46	102	120	111	111	67	28	20	13	3	2	0	0	623 (Controls)

It will be seen from this table that, up to the score of 4 inclusive, the number of records in which both the Sensitives and the controls scored 0, 1, 2, 3, and 4 respectively is greater in the case of the control as compared with the Sensitive for each one of these scores, whilst from score 5 to 12 the opposite holds good.

It is also interesting to note how, with the exception of the scores 9 and 10 of the Sensitives, the frequencies of the scores are gradually increasing up to a line of demarcation, between 4 and 5, and are then gradually decreasing.

By multiplying the scores 1 to 12 with the corresponding frequencies, we obtain the following distribution of the total scorings:

SCORES	1	2	3	4	5	6	7	8	9	10	11	12	
TOTAL SCORINGS	61	166	279	416	395	354	336	288	90	140	33 (Sensitives)	12	(2570)
	102	240	333	444	335	168	140	104	27	20	0 (Controls)	0	(1913)

These relationships will be more easily followed from the diagrams, Figures 4 and 5, in which the distributions above tabulated have been graphically represented.

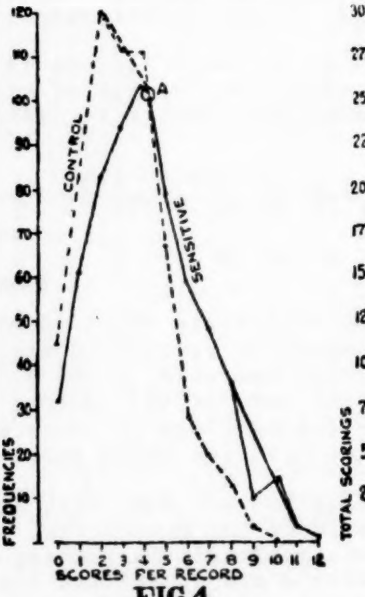


FIG 4.
SCORE FREQUENCIES DISTRIBUTION

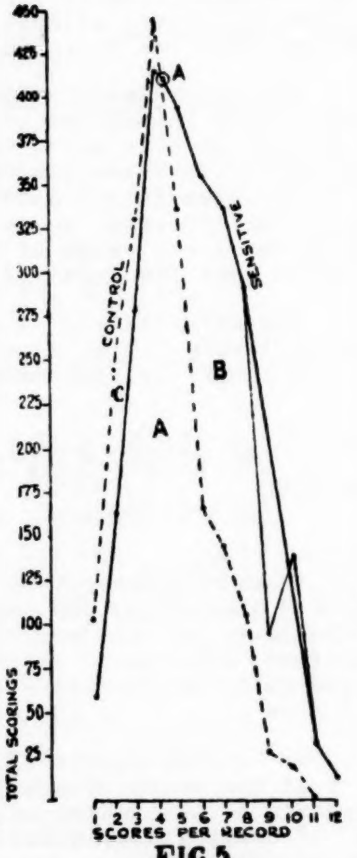


FIG. 5.
TOTAL SCORINGS DISTRIBUTION

Differentiation of the results according to Subjects. As the tests proceeded, it was found that some Subjects gave more or less consistently higher figures in favour of the Sensitives (R) as compared with the control (W), whilst other Subjects alternated from one side to the other to such an extent that there could be no question of consistent results.

Altogether 112 different Subjects participated in the control series (4th to 12th) and the individual results of those most frequently used were analysed. It was found that the ultra-perceptive ability of a Sensitive displayed through the intermediary of objects, in the absence of the Subjects to whom they belong, differs from Subject to Subject, attaining a high degree in the case of some Subjects and being almost negligible or nil in the case of others.

TIME FACTOR EXPERIMENTS

From the very beginning of this research it was intimated by the Subjects that some of the items given by the Sensitive occurred *after* they had parted with the article, and *before* or a short time *after* the Sensitive 'psychometrised' it. This means that, whilst some of the items concerned the *past*, some of them referred to the *present* and some to the *future*.

Although fully aware that J. W. Dunne's *An Experiment with Time* received a good deal of attention, not only in the general press but also in scientific publications, I did not consider it a practical proposition to extend the statistics to future events, the task undertaken being already sufficiently complicated. For this reason no account was taken of the intimations in question, although the facts are mentioned in some of the records or in their accompanying notes or letters.

It was only after the termination of the statistical tests, hereinbefore reported on, that I realised the importance of testing the *time factor*, provided it could be done in a simple manner without introducing complications. I also thought that once the proof of the probable existence of an ultra-perceptive faculty has been established, from a logical point of view, a scientific investigator is fully justified in proceeding with his experiments on the assumption that such faculty is a fact, with a view to ascertaining other facts which will assist in advancing the solution of the problem. However, when such an assumption is made, one is no longer bound to use a handicapping method of testing, such as

one of the various 'control' methods hereinbefore described which, however, as it has been shown, was absolutely necessary for proving the probability of the existence of such a faculty.

A further series of experiments was therefore carried out without any 'control', with the main object in view, not of ascertaining the prediction of future events, but of determining whether any of the items perceived by the Sensitive *coincided* with the *time of the test*, or whether they were of *recent* occurrence, viz. *on the day of, or one, two or more days before, the test.* My intention was thereby to ascertain whether there was any relationship between the actual perception by the Sensitive and the conscious or subconscious or subconscious activity of the Subject at *the actual time of the test.*

This series of tests, directed to the investigation of the *time factor*, which were not intended to provide a further proof of the existence of an ultra-perceptive faculty, being mainly exploratory, differed from the preceding series in two respects:

(a) The use of a control method was dispensed with, and

(b) The Subjects were instructed to accept only such items which were applicable within two weeks before the time of the test.

Altogether 147 records were obtained with a total number of items of 1658 given by the two Sensitives 'K' and 'F'.

Of this total 667 were accepted as applicable within the period stated. The items accepted include also such items which the Subjects considered to be permanent characteristics constantly applicable to them.

Of the accepted number, 167 were applicable within *five days* before the time of the test, 217 *on the day* of the test, and 66 at the *actual* time of the test.

These figures give the following results:

1. The total number of items accepted is 40% of the total number of items given by the Sensitives as compared with 38.75% of the total given in the summary of the statistical results.

One would have expected that, since in this series the items holding good beyond the fourteen days were

excluded, the percentage would be smaller than in the statistical series which were carried out without such a limitation of time. The slightly higher percentage is most likely due to the fact that, in this series, there was no 'control' and the Subjects *knew* that all the items had been given to them by the Sensitives without their being intermixed with any other extraneous items. We have already seen from the preliminary experiments that in such cases some Subjects tend to score higher.

2. Calculated as percentages of the total number of items accepted, we get for the items applicable within five days: 25%; for those applicable on the day: 32%; and those applicable at the actual time of the test: 10%; the remainder of 33% representing the items applicable in the time between five and fourteen days, plus the items accepted as permanent characteristics which are constantly applicable.

It is important to add that the artictes submitted to be 'psychometrised' almost always left the Subjects one or more days before the day of the sitting, and in many instances even weeks before, I having retained them for repeated tests. That in this series the events mentioned by the Sensitives were in most cases of quite recent occurrence, and that they happened after the Subjects had parted with their articles, was a most illuminating confirmation of the observations previously made by many Subjects in this respect.

These general results as regards *recency* and the *time factor* constitute an important finding in the present investigation, since they seem to give us a clue as to one direction in which we may possibly find an explanation of the *modus operandi* of the ultra-perceptive faculty, and organize further research of a promising character.

In this particular series special endeavours were made that the Subjects should receive the records the day after the test, and I am extremely obliged for the promptitude with which they filled them in while the events of the preceding days were still fresh in their memory.

The items perceived were more or less of a trivial character, although quite specific. Here is one example from each of the above groups.

Between five and fourteen days.

SENSITIVE: 'Not more than a week ago, some arrangement of a party to meet and did not come off, someone let them down.'

SUBJECT: 'Yes. Tried to arrange a party for next Saturday, but all I asked had a prior engagement and could not come.'

Within five days.

SENSITIVE: 'Very peculiar feeling about a bullet, does not touch me, cannot understand what it means.'

SUBJECT: 'Yes. About Monday or Tuesday I dreamt I was shot at' (test on Wednesday).

On the day of the test.

SENSITIVE: 'Two bunches of violets.'

SUBJECT: 'My mother and I were each presented with a bunch of violets at two o'clock that day.'

At the actual time of the test.

SENSITIVE: 'Owner read some papers I want to call proofs.'

SUBJECT: '*Yes. Yes.* On this date at *six o'clock*' (time of the test: 5.58-6.8 p.m.).

ANALYSIS AND THEORY

Statistical experiments. Assuming the alleged ability of 'psychometric' perception had been a generally accepted fact, by the law of 'cause and effect' we should have further assumed that, unless all the parts involved in the general process responsible for such perception happened to be in the required relationships, the perception would not occur. Just as an electric bell would remain silent, however hard the push-button be pressed, if the battery voltage is too low or the bell armature does not make contact or if there is otherwise something wrong with the electric circuit, so would a Sensitive fail to perceive, if the chain of links involved in the general phenomenon referred to was faulty in some part or other; and above all there would be no 'psychometric' perception whatsoever if the person claiming to have the ability mentioned is not at all a 'psychometrist', viz. a person capable of reacting in the chain of links.

It will thus be seen that, knowing nothing of the *modus operandi* of the phenomenon, the whole of the investigation might have completely failed owing to some failure in the chain, which would have been impossible to detect. Had we obtained entirely insignificant figures we could not have inferred therefrom that the alleged 'psychometric' ability did not exist, for these figures might have been entirely due to some failure in the chain. But having been rewarded by results pointing to 'significance' and therefore to the probability of

the existence of an 'ultra-perceptive faculty', we must further inquire into the manner as to how the experiments were conducted to obtain such results, so that it might be properly appreciated that this has been a completely unbiased finding.

The tangible links which come into question in the statistical experiments are: (1) the *Subject* whose article is to be psychometrised; (2) the *article* itself; (3) the *Sensitive*. To these links there must be added: (4) the *experimenter* himself; and (5) the *items* perceived. These are the five factors involved in the conduct of the experiments. The invisible links between the Subject (who is absent) and the article, and between the latter and the Sensitive, of which links we know nothing, do not concern the present discussion.

(1) As regards the *Subjects* used in the experiments, they were never present at the sittings and the Sensitives never knew who they were.

(2) As regards the *articles*, with very few exceptions, they were always enclosed in sealed envelopes, and, in the case of one of the Sensitives, most of the tests were carried out without her touching the envelopes.

(3) Therefore, as regards the *Sensitives*, the question of 'fraud', sometimes alleged against some mediums, does not arise in this investigation.

(4) The *experimenter* too was completely ignorant with respect to the bulk of the items given, viz. as to whether they were applicable or not, and in many cases he did not even know the Subjects. Moreover, he had no say whatever in the acceptance or non-acceptance of the items, as this always rested with the Subjects.

(5) Finally, as regards the *items*, they could not possibly have been perceived from the articles, even if the latter had been properly inspected by the Sensitives, since, as has been pointed out, *the items did not concern particulars thereof, but a wide range of the Subjects' experience.*

The heterogenous character of the items might make it appear that they could not be quantitatively analysed. However, a definite '*yes*' for items applicable to and accepted by the Subject, and a definite '*no*' for items not applicable to and not accepted by him, completely disposed of this difficulty and rendered the results of the experiments suitable for statistical investigation.

The tendency of some of the Subjects to accept copiously, and of others to accept niggardly, was also impartially dealt with by the use of *control methods having a balancing effect*. Each record contained an equal number of 'control' items and of items given by the Sensitive, and, whatever the tendency of the Subject, that tendency applied equally to both kinds of items, the Subject not knowing which was which.

It is sometimes alleged that the statistical method does not afford certainty of a proof. No useful purpose will be served by arguing this question one way or the other. Suffice it to point out that it is used extensively for the determination of the *probability* of all kinds of events, and it is in this sense, and in this sense only, that I claim this research to have led to a positive answer to the question as to whether there is such a thing as an 'ultra-perceptive faculty'.

If a coin is tossed 1000 times and we get only a small deviation from 500 'heads' and 500 'tails', this is a result we should have expected; but if the deviation is comparatively large, say we get 600 'heads' and 400 'tails', this is a result pointing to the fact that very likely the coin has a bias with respect to 'heads', viz. that its constitution is different from that of a properly balanced coin, in the sense that it is caused to fall one way more often than the other.

Similarly in the present investigation, assuming the Sensitive had been merely guessing, the chance of the Subjects accepting 'control' items and items given by the Sensitives would have been the same, so that out of a total of 4483 accepted by the Subjects, there should have been, within certain deviation limits denoting lack of significance, approximately 2240 'control' items, and approximately the same number of items given by the Sensitives.

However, the figures obtained by the statistics were: 1913 'control' items (accepted) and 2570 'Sensitive' items (accepted), which latter figure was well outside the range of insignificance. *We are thus led to the conclusion, analogous to the second example in the coin tossing, that there must be something characteristic of the Sensitives that enables them to score higher than they would by mere guessing.*

Although this is only an indirect proof of the probable existence of an ultra-perceptive faculty, as against the assumption of 'coincidence' or 'guessing',

I submit that it is a scientific proof of a nature which would be readily acceptable in any branch of some of the more positive sciences as establishing a definite differentiation.

Time factor experiments. The expression 'time factor' has been used to indicate either the approximate coincidence or the lag between the time of the test and the time when the Subject experienced the event mentioned by the Sensitive.

We have seen that the time lag varied considerably and that the limitation to fourteen days produced a large number of events recognized as applicable within that period. The experiments also showed that, out of that number, a large percentage (32%) applied to events on the day of the test; in these cases we are concerned with a *lag of a few hours*. The experiments further showed that the items found to be applicable at the actual time of the test were also quite numerous (10%, not included in the 32%); these cases may be considered as *approximately coincident in time*.

The general trend of the 'time factor' tests seems to disprove the theory according to which the article submitted is supposed to have recorded all kinds of events while it was in contact with the owner, which events are subsequently sensed by the Sensitives through some form of vibrations coming direct from the article. Such a theory would not hold good for a very large proportion of items given in these tests in connection with articles which were no longer in possession of the Subjects at the time when the events mentioned by the Sensitives occurred.

The analysis of the results shows that the *recency* of the items perceived is an important factor in this form of perception.

In conjunction therewith it is, however, to be pointed out that, although the range of *recency* proved to be comparatively wide - from apparent coincidence in time to the arbitrary limit of fourteen days - the various series of tests previous to the 'time factor' series produced quite a number of exceptional items of much older standing than that limit, which were accepted by the Subjects as significant to them. On the face of it one might, therefore, be inclined to draw the opposite conclusion, namely, that this form of perception is independent of time as regards the actual advent of the item perceived.

To say, on the one hand, that in the form of perception we are dealing with the *recency* of the item perceived is an important factor, and, on the other hand, that the perception seems to be *independent of time*, may appear to be mutually contradictory. The solution of this paradox lies in the following psychological facts:

(a) Every personal experience, including any form of acquired knowledge, is stored in the subconscious mind, where it has the potentiality of emerging into the conscious at any subsequent time.

(b) Whether the experience is in its stored form in the subconscious, of which storage we know very little, or whether it is emerging or has already emerged into the conscious in the form of a thought, accompanied or not by some feeling or action, all these transmuted forms of the original experience constitute a more or less accurate mental copy of that original experience, which may thus be considered to be *always* associated in some form or other with the individual who had the experience.

(c) Recent experiences have a tendency to emerge into consciousness more readily and more often than old experiences.

A joint consideration of these facts, and of the results obtained in the 'time factor' experiments, is self-suggestive of a very plausible explanation as to what is the main fact in the exercise of the form of ultra-perceptive faculty we have been considering, as apart from any theories regarding its actual *modus operandi*. This main fact may be postulated as follows:

When a Sensitive 'psychometrises' an article belonging to a Subject, a mental 'rapport' seems to be established between the Sensitive and that Subject, and whatever is perceived by the Sensitive in virtue of the ultra-perceptive faculty seems to be present in, or accessible to, the psychological make-up of the Subject at the time of the test.

It is important to note that this postulate is not in the nature of a theory, but merely a statement of fact which has been derived from other well established psychological facts and numerous experimental observations.

I am fully aware that, taken absolutely literally, the above postulate is open to one important criticism, namely, that it disregards the perception of *future events*.

A few instances of this character were revealed by the experiments. Some of them were traced by the Subjects to have been in their minds at the time of the tests; however, there were others which could not be thus traced, nor be arrived at by reasoning. Such latter items are admittedly puzzling.

The present research was not extended to the problem of precognition of future events, and all I can say in this respect at present is that, once we have admitted the probable existence of an 'ultra-perceptive' faculty with respect to present and past events, not because of any satisfactory explanation but because of the significance of statistical figures, we ought not to shrink back from a possible existence of 'ultra-perceptive precognition' of future events because we cannot grasp it and supply an explanation. If and when its probability is proved, as in the case of past and present events, we should have no compunction in admitting it as one of the further facts ascertained in connection with the 'ultra-perceptive' faculty.

Hypotheses and a 'Theory of generalized facts'. As regards the hypotheses hitherto put forward, I must refer the reader to the comments in my book herein already quoted, where he will also find a 'theory of generalized facts' propounded as a result of these experiments. Here is a summary of that theory:

Man's potentialities of cognition extend beyond the normal sensory and intellectual perception, and this *ultra-perceptive* faculty is associated with a *state of consciousness* which is more or less different from the ordinary normal state. The cognition is dependent upon a *rapport* being established between the percipient and a source from which the items are perceived, the *rapport* being a *mental* one when the source is a person. In the latter case the actual source of the items is the *psychological make-up* of that person at the time of the *ultra-perceptive cognition*.

An *object* on which a Sensitive concentrates mentally *enables a mental rapport to be established* between the Sensitive and the owner of the object, the psychological make-up of the owner at the time when the Sensitive concentrates on the object being the source which supplies the items of the ultra-perceptive congition.

I have endeavoured to embody in this article the main points of my research as reported in the first

volume: "The Ultra-perceptive Faculty", but the latter contains further material which students of this branch of scientific research will find of interest and useful as regards further experiments.

In a second volume entitled: "Exploring the Ultra-perceptive Faculty", I give an account of a further piece of research which I carried out at King's College, London University, as a sequel to the statistical and time factor experiments dealt with in this article. It mainly concerns a new method of investigating this particular faculty, consisting in the Sensitive 'psychometrising' an object belonging to a Subject while the latter is perusing an illustrated paper normally, without any effort at concentration on the experiment. The results obtained indicate that the method may be destined to play an important part, not only in investigations such as referred to but broadly in connection with other psychological problems. In a second article I will describe this method together with miscellaneous experiments to which the same was applied.