

The Journal of Parapsychology

Volume 1

DECEMBER, 1937

Number 4

EDITORIAL COMMENT

THIS NUMBER concludes the first volume of the JOURNAL OF PARAPSYCHOLOGY. The total value of the JOURNAL cannot be fairly estimated at this time. Its usefulness, however, can scarcely be questioned. In these pages have been reported over 10 times the amount of research published prior to the March number. Practically all the confirmatory work was done at institutions other than Duke University.

New experimental methods reported have included the various matching techniques, the "blind matching" and "screened touch matching" being especially adapted to careful experimental conditions.

The range of subjects used has been widened to include children, both in a home and in a classroom situation; twins; and institutionalized blind people.

First step advances in understanding of the nature of ESP have included studies of variation in form and size of the stimulus patterns, comparison of motor and verbal responses. The effect of distance, of volitional direction, and of groups has been discussed.

When to these are added the presentation of studies of mathematical techniques and consideration of criticisms, the first volume has allowed both the student and the interested layman direct contact with experimental work in the field of parapsychology from which must arise the important future research.

* * *

In the "Editorial Introduction" to this JOURNAL, criticism of experimental work in Parapsychology was invited as a necessary aid to enlightened research progress. Several critical articles have appeared in other journals, although they have not been submitted here. However, since the wide public interest in parapsychology has given news importance even to criticism of the subject, most readers of this JOURNAL are doubtless aware of the general nature of the criticisms.

Recognizing the value of controversy, but recognizing equally the need to keep issues sharply drawn and at a level pertinent to the evidence, we have devoted much of this number to articles which deal directly with points criticised.

Probably the most frequently raised question of procedure in extra-sensory perception has concerned the mathematical methods used in measuring the scores of the experiments. The scoring table for small numbers of runs presented here by Professor E. V. Huntington makes a valuable contribution to the research, and deals effectively with the statement of a critic that "there is no method available for evaluating merely the total scores of small numbers of runs."¹ A formal consideration of the criticisms directed at mathematical methods is reported by Stuart and Greenwood, who find, after the most thorough analysis of this question to date, that whereas minor changes in method are advisable, the mass of criticism has been directed at points which are experimentally trivial.

An interesting, and for many the most convincing reply to critical attack is that supplied by Dr. Lucien Warner's study. In an experiment designed to meet all criticisms, valid and invalid alike, he finds an unquestionably significant result which again verifies the adequacy of the concept of extra-sensory perception.

Another controversial question, which will doubtless become more important as mathematical challenges recede, is the adequacy of experimental conditions to exclude sensory cues to the symbols on the cards. In this number Dr. J. B. Rhine summarizes the published evidence from experiments wherein these conditions are unquestionably adequate.

Two further studies conducted outside the University but in active contact with the Parapsychology Laboratory at Duke University are here reported. Mr. E. P. Gibson finds results which substantiate those reported by Woodruff and George in the study of the relative success of various calling and matching techniques. The death of Dr. C. Hilton Rice brought to a close an investigation which might well have been of major importance to the field of parapsychology. Experiments bearing upon his particular interest, the role of imagery in extra-sensory perception, remain incomplete. Dr. J. G. Pratt had the opportunity of working with Dr. Rice and studying his results a few weeks before the latter's unexpected death. Dr. Pratt reports here a summary of this work.

A part of Dr. Rice's project consisted of the extra-sensory percep-

¹C. E. Kellogg, "New Evidence (?) for Extra-Sensory Perception," *The Scientific Monthly*, 1937, XLV, p. 336.

tion of free drawings. This kind of experiment is not new to parapsychology, but results have always been difficult to evaluate in an objective manner. Dr. Pratt applies the "method of correct matchings" to this kind of material for the first time.

The scores reported by Dr. Riess as resulting from a distance experiment are the highest scores ever published in the history of Parapsychology. The author's apparent incredulity is therefore not difficult to understand, although the reasons given are. The paper as submitted gives no ground for an explanatory hypothesis other than ESP.

We reprint under "Notes" a press release signed by Dr. Burton H. Camp, President of the Institute of Mathematical Statistics. After study of the mathematical techniques applied to ESP investigations he finds that "the statistical analysis is essentially valid."

* * *

It was the intention to publish in this number the first of a series of reports of experiments made to bear on the question of precognition. But the need of space for criticism and inquiry regarding already published work has necessitated postponing this paper until the next issue.

A TEST CASE

LUCIEN WARNER

ABSTRACT: In reply to criticism directed at various points of previous research, an experiment was set up to avoid all the debated points, regardless of their applicability. A single subject made 250 calls of cards chosen in a strictly random fashion. The subject was in a closed room one flight below and at some distance from the room in which the experimenters handled the cards. The only communication was one-way; the subject signalling by a light as each call was made. In 250 trials there were 93 hits, or 43 more than chance expectation. The result is highly significant.

Dr. Warner has written two previous articles in this Journal, "Telepathy in the Psycho-physical Laboratory" and "The Role of Luck in ESP Data."

THE EXTENSION of research in extra-sensory perception by many colleges and institutions here and abroad has been accompanied by a similar extension of criticism. In general, there are two ways of answering such criticisms, if answered they may be. The first of these would be to point out their inapplicability, their logical weaknesses, or to cite research overlooked by the critics which has already answered their questions.

The second method, and for many people the most satisfactory, is that of subjecting the criticisms to new and modified experiment, designed to avoid whatever faults have been pointed to, whether they are actual or are entirely superficial.

The research reported in this article is a direct response to recent criticisms, and is submitted in reply. It is hoped that these data will encourage the critics to further comments, for progress is made through the interaction of conscientious experimentation and sincere criticism.

Specifically, this test was planned to meet two points which have recently been raised in an effort summarily to dispose of all ESP work.¹ In the first place, it has been said that evidence for ESP has been obtained spuriously through the *post facto* selection of data. This is a thought which had occurred to me two or three years ago

¹Kellogg, C. E., "New Evidence (?) for Extra-Sensory Perception," *The Scientific Monthly*, 1937, XLV, 331-341.

and which I had advanced recently.² However, I expressed it as a warning rather than as a device for explaining away the evidence for ESP. As such a device the *post facto* selection article is feeble for when, without any selection whatever, we combine the results of all ESP work done at Duke and elsewhere we still arrive at a deviation from expectation significantly greater than chance.

The second criticism which the conditions of the present test meet concerns the method used in measuring the deviation of obtained results from chance expectation. It has been said that the methods previously used underestimate the deviation which may be obtained by chance. Even granting the point without argument, the magnitude of the error involved is trivial when compared with the extreme deviations obtained.³

However obvious it may be that these two criticisms have already been answered in published work, the issue at stake is not a small one, and there are some scientists who have considered the ESP evidence (superficially at least) and believe that nothing has been proven. So I decided to disregard, for the moment, all previous work and to say to myself (for the Nth time): "There must be something wrong with this work somewhere, and I mean to make one more crucial test."

As a subject, I selected M. H. of Greenfield, Massachusetts, and on October eighth, after telephoning for an appointment, I drove there and obtained the record reported below. Thanks are due Mrs. Mildred Raible for assistance in the experiment.

The selection of M. H. was not fortuitous. But the skeptics are not concerned with what subjects are selected so long as selection is made *before* and not after the data are obtained. The believers in ESP, on the other hand, would consider me entirely unjustified in selecting for a test case any subject who had not proven the ability which they call ESP. Thus by selecting M.H., who had in the past made scores which deviated rather widely from chance, I have satisfied both camps.

The selection of the method to be used was also made with care, but in this case I erred on the side of the critics. I altered the conditions under which the subject worked, a procedure which the ESP protagonists will at once tell you is likely to interfere with a subject's supposed psychic ability. M. H. had previously been tested while in

² See last paragraph of my article: The Role of Luck in ESP Data, *JOURNAL OF PARAPSYCHOLOGY*, I, 84-92.

³ C. E. Stuart and J. A. Greenwood, "A Review of Criticism of the Mathematical Evaluation of ESP Data," *JOURNAL OF PARAPSYCHOLOGY*, 1937, I, 295-304.

the same room with the investigator, chatting informally between runs. I told her that this evening she was to be locked in a room by herself while the experimenters and, of course, the cards she was to read clairvoyantly, were to be in a room on another floor. The doors to both the experimenters' room and the subject's room were closed throughout the experiment. The only communication existing between the room during the work was *one-way*. The subject could press a button which flashed a light in the experimenters' room. This she did after writing down her guess for a given card as an indication that she was ready for the next. It is highly improbable that the subject could have heard the experimenters' voices if they had shouted. As a matter of fact, there was not even whispering in the experimenters' room during the work. Visual communication between the two rooms was a manifest impossibility. Not only were the two closed doors separating the rooms on different floors, but they were oriented so that even had they been on the same floor one door could not have been seen by a person standing in the doorway of the other. If the old hyperaesthesia label is to be hung on these results I shall find myself in a wilderness with a different name, but still a wilderness.

Dr. Kellogg⁴ objects to Rhine's statistical treatment as inapplicable to a fixed deck of cards. He states "for Rhine's pack of five sets of five, there are over 623 trillions of different arrangements, so that a solution by tabulation and comparison is beyond human capacity." Hoping, therefore, to simplify the task of whoever may care to analyze the few data presented in this article, I abandoned the use of a fixed deck. Each card to be guessed was obtained by making a fresh cut in a freshly shuffled deck. This is, of course, a far slower method and reduced the number of data we could obtain in the single evening our subject, M.H., could give us. Furthermore, it introduced into the situation still another unaccustomed element, for this subject had been accustomed to calling cards very rapidly, so rapidly, in fact, that it had been all a recorder could do to keep up with her verbal guesses. Perhaps these changes in the situation account for the fact that her scoring on this test evening was not up to her highest standard.

To summarize, then, before giving the results: the experimenters, having told the subject that work was to begin, retired to the upstairs room, shut the door, cut a shuffled deck of cards and placed the cut card face down on the table by itself. It was not seen by the experimenters until after the guess had been made and recorded by the subject in her room downstairs. The subject signalled when this had

⁴ *Op. cit.*

been done by pressing a button which flashed a light in the experimenters' room. When this signal was given, the card cut was exposed and recorded, and a different deck, newly shuffled, was cut to obtain the next card to be guessed. This procedure was repeated without interruption until the end of the work. At the conclusion, the subject's record of guesses was compared with the experimenters' record of cards cut.

It has been complained that ESP data are never published in full detail. In a short series like the present one such publication is possible.

In the following record, s represents star; c, circle; p, plus; r, rectangle; and w, waves. The upper lines of letters indicate the calls of the subject; the lower lines, the cards actually cut. Correct calls are capitalized. The grouping into fives is merely for convenience of inspection, as the experiment was continuous.

TABLE I
THE DATA OF THE EXPERIMENT

SCpw, pPrrS, rcpwc, rrCpw, sswcc, PWcCS, swcsp, pcwsw, pRprr, WCrrr
 scrrc, wpcws, pprcs, wwcwc, cwrpr, pwscs, ccscw, wsrpc, wrswr, wccsc
 pwCrw, cSswc, srCsC, Rwcps, Spcww, sPCSS, pscpR, WsRsp, Cwcsc, Cpsww
 rscsp, psrrw, rprc, rcwcs, scacc, rpsc, cpps, wrcrp, ccpws, cccsr
 WpcWw, pPwR s, Sp rWw, PCscP, pcwrC, CPSwW, ScrCP, CW rpp, wrCcw, ppCpw
 wwswa, wprrw, srcwp, pccpp, srcpc, cpsrw, srscp, cwccw, rccpr, rscwc
 wrrsc, prwrc, swcRR, WwWPC, CsSpp, PWcWP, RWSsP, pSfcw, PWCSs, cprrs
 ccwwr, rcccs, cprrr, wcpcc, ccssc, pwrwp, rwsrp, rscsc, pwcsw, rcpw
 wrwsC, PWRR c, WsSwP, wPppS, spRrs, Rswwp, cRSW s, pRCWw, RwPPr, pcwrs
 rcpcc, pwrpp, wsrsp, cpsss, wsrsp, rcprr, wrswr, srcwc, rcppe, cprcw

Examination of the record shows that the subject scored 93 hits in the 250 trials, or 43 more than expectation. The binomial hypothesis is theoretically proper for this experiment. The critical ratio is 6.8. With the use of Gram-Charlier curve⁵ the probability that this might occur by chance, corrected for the moments of the binomial curve, is found to be about .00000000064.

The data are further summarized in a contingency table.

The figures in parenthesis are the expected frequencies. Before any mathematical treatment it can be noted that the diagonal cells, representing hits, contain more correspondences than any other cell of their respective row and column. An evaluation of the entire table, with fluctuations in the card series and subject's call series ruled out, gives a chi-square of 61.34. The probability of such a value being a

⁵ This number, JOURNAL OF PARAPSYCHOLOGY, p. 302.

TABLE 2.
CORRESPONDENCE OF SYMBOL CALLED WITH CARD OCCURRENCE

Calls	CARDS					Totals
	C	R	W	P	S	
C.....	22 (14.20)	6 (10.00)	3 (8.60)	10 (8.60)	9 (8.60)	50
R.....	12 (11.93)	16 (8.40)	5 (7.22)	5 (7.22)	4 (7.22)	42
W.....	17 (15.62)	12 (11.00)	18 (9.46)	5 (9.46)	3 (9.46)	55
P.....	8 (14.05)	8 (10.60)	9 (9.12)	19 (9.12)	9 (9.12)	53
S.....	12 (14.20)	8 (10.00)	8 (8.60)	4 (8.60)	18 (8.60)	50
Totals.....	71	50	43	43	43	

chance occurrence is less than one in a million. The five diagonal cells contribute 39.84, or over half of the total chi-square. Independently evaluated, the probability of such a chi-square from the diagonal cells is again less than one in a million. It is obvious that the large number of hits was due neither to chance fluctuations nor to sampling errors.

DISCUSSION AND SUMMARY

In an experiment of this nature, the conclusion is limited but definite. The subject attempted to choose the correct one of five symbols denoted at random. She succeeded 93 times out of 250 trials, a success ratio of 37.2 percent. Analysis shows that this is much higher than could be explained upon chance hypotheses. No sensory contact or cues were possible under the circumstances. There was no basis for rational judgment as to the order of the card symbols. Yet there was a measureable relation between the turn of the cards and the subject's responses. This would constitute an adequate definition of extra-sensory perception.

THE WORK OF DR. C. HILTON RICE¹ IN EXTRA-SENSORY PERCEPTION

J. G. PRATT

ABSTRACT: Dr. C. Hilton Rice made studies of extra-sensory perception with ESP cards and with freely chosen rough drawings. He was particularly interested in the role of imagery in the subject's response. Dr. Pratt, in a review of Dr. Rice's experimentation, finds that good evidence for extra-sensory perception was obtained in the card tests with two subjects. The reproduction of drawings was clearly above chance expectation but the conditions were not adequate to rule out all hypotheses other than ESP. The experimental results were not such as to permit definite conclusions as to the role of imagery.

In the presentation of the results with drawings, a method for strict quantitative evaluation applicable to material of this character is used for the first time in a parapsychological study.

Dr. Rice was a practicing physician in Montgomery, Alabama. Dr. Pratt is a member of the staff of the Parapsychology Laboratory at Duke University.

THE EXPERIMENTER whose work is reported in this study began investigating ESP while he was still very dubious of its occurrence. He had read the report of experiments in this subject done at Duke Uni-

¹ Deceased May 26, 1937. Under the circumstances it is necessary to introduce the man who intended to write this report. Dr. Rice was a physician whose specialty was pediatrics. But in his later years he turned more and more to psychotherapy—a field for which he was particularly fitted by his capacity for quick insight into problems of human personality. He gained the respect and admiration of his colleagues for his work in this field in spite of the fact that he was a pioneer in the South. By way of avocation, Dr. Rice was greatly interested in the philosophy of science. He had held for twenty-five years a working hypothesis of the nature of life as a unitary whole. He spent much time reading scientific literature to see how well his hypothesis stood the test of facts. He looked upon all life and its environment as an essential unity—an organism with close interdependence among the various forms of life and between all life and the physical environment. His published writings consist of a brief statement of this hypothesis ("An Hypothesis," *Psyche*, Jan., 1929), a discussion of the bearing of the Duke experiments in ESP upon the body-mind question ("The Matter-Mind Relationship", *The South Atlantic Quarterly*, 1936, pp. 161-166), and a short paper on the use of suggestion in the treatment of habit spasms ("The Treatment of Habit Spasms by Psychotherapy," *The Jour. of the Medical Association of the State of Alabama*, Sept., 1933). A book-length manuscript on his philosophical views is as yet unpublished.

versity,² but remained skeptical of the conclusions there reached. However, he accepted the challenge which has been generally made to test the matter for himself, and began making exploratory tests after the manner of those previously reported.

When the preliminary tests (included in this report) convinced Dr. Rice that ESP was occurring, he became interested in conducting experiments to help to explain this phenomenon. The particular problem with which he was most concerned was the role of imagery in ESP. His experiments were cut short by his unexpected death. The present study shows how his results contribute to the general evidence for the occurrence of ESP and also how they bear, as preliminary experiments, upon the problem of the role of imagery in this mental function.

I. INVESTIGATIONS USING ESP CARDS

EXPERIMENTAL PROCEDURES

The card tests were all made with the usual packs of 25 ESP cards. The cards were usually laid down (either by the subject or the experimenter, as the case might be) into 5 piles, each pile being intended for one of the 5 symbols. The record of the score for the pack was made after all 25 cards had been placed.

The various conditions used in the card tests were as follows:

(1) *Unobserved*. That done by a subject when entirely unobserved; the subject shuffled the pack, placed the cards out into the 5 piles, made the check-up, and recorded the score.

(2) *Observed work with subject handling cards*. These were trials in which an experimenter was present and shuffled the cards, and the subject handled them himself in making his choices; in this case, both people observed the checking and recording.

(3) *Observed work with experimenter handling cards*. In this condition the experimenter alone handled the cards. The subject called out where they were to be placed.

(4) *Observed work with screened cards*. Like procedure (3), with the additional feature of screening the cards from the subject's sight. The experimenter sat behind and close to a folding screen of opaque cloth 6 feet high. One of the twenty-inch panels was directly between the subject and the cards. The experimenter could see the subject, who sat about 10 feet from the screen, by placing his eye at a narrow crack between the two panels.

² Rhine, J. B., *Extra-Sensory Perception*, Bruce Humphries, Boston, 1934.

(5) *Tests with sealed cards.* Cards sealed in heavy manila envelopes were used, sometimes with an observer and sometimes not. Card-board shields were placed over the symbols to insure against the possibility that the symbols might be seen through the envelope. There was never any evidence that the sealings had been tampered with.

(6) *Tests for GESP.* The above procedures were used for studying clairvoyant perception, but they could be adapted to testing general extra-sensory perception simply by having the experimenter look at the card. In addition to this method for studying GESP, some trials were made without cards with the agent behind a screen drawing the outline of the symbols in the air with his finger-tip.

RESULTS OF CARD TESTS

All of the work done with ESP cards is summarized in Table I. In all, 1,569 packs were called, making 39,225 single trials. The total number of cards correctly called was 10,762 which is 2,917 more hits than the mean chance expectation. The S.D. for this number of trials is approximately 80.8, giving an X-value for the observed deviation of 36. This means that the results would happen by chance an average of about one time in 10^{280} such series. From this general summary, it is evident that some knowledge of the order of the cards in the shuffled pack was being obtained in some way.

Whether this knowledge was extra-sensory in origin must be decided on the basis of the adequacy of the conditions to exclude all other possibilities. The procedures used in No. 4, 5, 6, 9 and 11 of Table I were such that the cards were either completely screened from the subject or were sealed in opaque envelopes. Of these the results of experiments No. 4 and 5 are clearly significant, those of No. 9 are almost unquestionably so (odds of 150 to 1 against chance), and experiments No. 6 and 11 gave inconclusive results. Combining this work done under the very strictest conditions we find that the total number of runs is 565, the hits are 358 more than the average chance expectancy, and the average chance occurrence of so high a result would be about one per trillion such series. These tests are those which can be rejected only on the basis of some kind of incompetence in the experimenters involved. The most significant results were obtained by Dr. Rice, but experiment No. 9, which Dr. Rhine observed, yielded striking results, and the first day of experiment No. 6 in which the writer was the observer gave results with odds against chance of 700 to 1.

These experiments, selected for the rigor of the condition of testing and not for the results, show that ESP was demonstrated. The

TABLE I.
SUMMARY OF WORK WITH ESP CARDS

Subject	Exp. No.	Condition	Date	Runs	Dev.	Av./25	X-Value*	Remarks
Miss T..	1	Unobserved.....	July 1935.....	136	671	9.9	28.2	Beginning to get images. First day, 20:+27, P = 1/700. His highest scores.
			Nov. 1935.....	24	243	15.1	24.3	
	2	Observed, sub. holding cards....	Feb. 1936.....	19	56	8.0	6.3	
			May 1936.....	220	321	6.5	10.6	
	3	Observed, exp. holding cards....	June 1936.....	206	434	7.1	14.8	
			July 1936.....	90	113	6.3	5.8	
4	Same + screening.	July-Dec. 1936....	158	164	6.0	6.4		
5	Sealed pack.....	Sept. 1936-May 1937	209	153	5.7	5.2		
6	Screened-J.G.P....	May 1937.....	184	25	5.1	.09		
C.H.R....	7	Clair., alone.....	Dec. 1936.....	9	46	10.1	7.5	
	8	Clair., Miss T. observer and holding cards....	Dec. 1936.....	77	144	6.9	8.0	
	9	Sealed cards.....	J.B.R. witness....	4	10	7.5	2.45	
	10	5 agents in turn....	Jan. 1937.....	30	104	8.5	9.3	
A. J....	11	Sealed cards.....		10	6	5.6	1.0	Agents drawing in air. C.H.R.'s daughter
	12	P.T., C.H.R. agent.		27	48	6.8	4.5	
Rem....	13	Clair., unobserved.		83	268	8.2	14.4	Getting images. Stopped, not seeing images.
	14	Clair., observed....		2	-2	4.0	-----	
Mrs. W..	15	Observed by various people....		33	60	6.8	5.1	
Others..	16	Observed by various people....		48	53	6.1	3.8	
TOTAL..				1,569	2,917	6.9	36.1	

C.H.R. = Dr. Rice. J.B.R. = Dr. Rhine. J.G.P. = The writer.

*Using the corrected S.D. for maximum variance suggested by Greenwood & Stuart in this *Journal* 1, 3, 212.

present paper, therefore, adds to the general body of evidence proving the occurrence of the phenomenon. No other definite conclusions can be reached in this study. However, a discussion of the results in Table I brings out a suggestion which may be of great importance of how ESP functions.

DISCUSSION OF RESULTS

Dr. Rice was interested in finding out as much as possible about how the subjects experienced their ESP impressions. After he felt that the ESP ability of his subjects had been established, efforts simply to accumulate evidence became secondary to this primary issue. In trying to find out how the genuine ESP impressions differed from ordinary imagination Dr. Rice frequently asked his subjects for introspections. It is important to discuss the results of each subject to see whether the introspections help to explain the results.

Miss T.'s results. Miss T. was Dr. Rice's secretary. She began ESP tests in July, 1936, first working by herself at her desk. At this time she was unable to score in the presence of an observer; her own work for July, however, averaged 9.9 for 136 packs. This was clearly significant of something. But when she was asked to work in Dr. Rice's presence, she completely failed to exceed chance. In the face of continual failure, the work was stopped for a time.

In November, Miss T. again began working alone with the cards, and at once began reporting phenomenally high scores. She attributed her new success to a new mode of experience: she was "seeing" the cards in a manner which was for her particularly vivid imagery. From Table I it may be seen that 24 consecutive packs done at this period average 15.1 per 25. An attempt to demonstrate to Dr. Rice her new-found power again resulted in failure. The subject became ashamed of her reports of high scores, as she professed to realize fully that unobserved work could have no scientific value. At Dr. Rice's suggestion she stopped keeping records of work done alone.

The rule of not keeping records of unwitnessed work was generally followed, but for some reason 19 runs dated February, 1936, appear in the records. In these she made a considerably lower average than in those of the previous November. This is probably an exception to the general average for working alone, as she reported in answer to questioning that she continued to get phenomenal scores when by herself. She attributed her high scores when alone to her ability to get clear, vivid images only under those conditions.

Occasional trials were made in Dr. Rice's presence, and in May, 1936, she began to score consistently above chance. The table shows that she made in his presence 220 runs in May, and 206 in June which averaged respectively 6.5 and 7.1. In June the experimenter was able to handle the cards himself without reducing the score entirely to chance; 90 runs averaged 6.3. This is a highly significant result. For these tests, the subject was about 4 feet from the cards and worked most of the time with her eyes closed. There was not the slightest indication that she was trying to look at the cards or even wished to do so. During all this time the subject blamed herself for failing; this attitude is not hard to understand if she could, as she claimed, average around 15 per run when alone.

Between July and December, 1936, the subject succeeded under conditions which leave no escape from the conclusion that ESP was demonstrated, unless one wishes to accuse Dr. Rice of gross incompetence

or self-deception. The experimenter handled the cards and kept them screened from the subject at all times. She did 158 runs with an average of 6. The average chance expectancy for a result as high as this is less than 10^{-9} . During this same period, Miss T. worked with sealed cards. Sometimes she took the envelopes which Dr. Rice had marked in code and worked alone; at other times she worked as usual in his presence and he checked the results at once from his knowledge of the code. Some of the cards were sealed by himself, others in the Parapsychology Laboratory of Duke University. This work extended from September, 1936, until May, 1937. A total of 209 runs were made, giving an average of 5.7. Even this is highly significant, with odds of about 2,000,000 to 1 against chance.

I had an opportunity to work with Miss T. for a few days in May, 1937, doing the last card work to enter into this report. In view of the fact that she had been so long in succeeding for him, Dr. Rice expected her to fail to score with me. However, she began rather well with me on the first day. The cards were handled by me and screened from her sight; all known possibilities of sensory cues were thus eliminated. The first day gave results that would have happened by chance an average of only once in every 700 such days. Thereafter the scores fell, and all the tests given by me taken together were clearly within the range of chance occurrences.

The conclusions indicated are that Miss T. demonstrated ESP under excellent conditions of testing. And if ESP was concerned in the work done alone, it was of a variety which showed a much higher level of scoring than that shown by her witnessed trials.

In order to render this report as true as possible to the spirit in which the research was done, I quote here a report found among Dr. Rice's records. In this he describes at greater length than has already been done Miss T.'s development and experience as a subject:

"More than half of her runs were observed by me. Miss T. had never played cards and cannot yet handle them deftly. The bottom card was drawn and placed opposite a marker and the count was made after all the cards had been placed. After the first thousand I played a chance series against her. On 21 series of 250 each, she beat me 19 times. At this point I became convinced something more than chance was at work . . . The only explanation she could give at this time was 'I just seem to follow my hand.' . . . Being extremely self-conscious, her scores began to drop as soon as she was aware of being watched. 'I've always been painfully self-conscious in the presence of others,' she

said 'and to feel that I am being observed is agony. I could score well before you when it seemed like a game we were playing, but when I know you are watching every move I just can't do anything. But I can go right on beating chance when I'm alone.'

"Miss T. has been my secretary for 19 years. She is middle-aged and unmarried. I have never known a more honest person and have discovered no other subject who has worked with the card experiments as patiently as she. She understood perfectly that high scores made when alone proved nothing, and we stopped keeping a record. But time and again she would say, earnestly and insistently: 'Doctor, I tell you there is something to this thing. I can't explain it, but I keep making good scores when I'm alone.' Then she would show me again how she shuffled and placed the cards. . . .

"A number of times I had asked Miss T. about the nature of the 'hunch' or feeling of certainty about particular cards. I suggested that she work slowly with closed eyes to discover, if possible, how the thought of the symbols got into consciousness. That is, did she 'hear' the name 'circle' or 'see' an image of it? One night at home she tried the slow method and when I came to the office the next morning she was in a state of awed excitement. What she reported seemed unbelievable.

"'Last night I tried the cards as you suggested. I closed my eyes and concentrated on the bottom card and I began to see images of the symbols. I never knew what you meant by images until then; but they would form and fade out and the right images would persist or keep coming back clearer than the others. I would call the clearest image and then look at the card to see if it was right [a deliberate departure from her usual procedure]. On one run I got a 20.'

"Miss T. is neither subtle nor inventive, and it would have been wholly out of character for her to have acted such a role of deception . . . I was certain no fraud had been practiced in the first ten thousand trials because I had witnessed over half of the runs, but it was obvious now that she either had discovered an important clue that enabled her to make the most of an unknown mode of perception or she had resorted to conscious or unconscious faking . . .

"By working alone when she was in the 'mood' and the images were vivid, Miss T. has reported a high series of 600 trials with an average of 15 hits in 25. She has never been able to beat chance by telepathy, but by patient effort through many trials to relieve her of her self-consciousness she has climbed to an average of 6 in nearly 3,000 trials under all manner of rigid test conditions with the cards in my hands.

This low average compared to the high score claimed when alone is enough to arouse suspicion of deception, but I have not yet detected any evidence of fraud. I have spied on her through a crack in the door and from a mirror in my back office and have never seen her make a suspicious move. Her technique when alone is the same as when under observation. She shuffles the cards well, then sits with eyes closed and concentrates on the bottom card. When the card is drawn she never glances at the back of it, but places it face down before the marker and promptly closes her eyes.

"Once I walked in unexpectedly when she was making a run. 'I'm really seeing them this time,' she said in an intent way that warned me not to break the spell. I slipped quietly into a chair opposite her desk and watched every move. She was staring straight ahead through closed eyelids, and from time to time moved her eyes toward the window, then away from it. From my position I could see the symbol as she raised the card to place it. Of the seven cards left in her hand she placed five correctly. Then I made the count and found she had made 18 hits out of the 25.

"Not less than 20 times in the past year I have got a flash of what Miss T. calls 'genuine images' and not once have they been wrong. . . . I seldom try the cards because I do not get clear images frequently enough to make high scores. Furthermore, staring at the dark background of the closed eyelids produces an eyestrain that soon becomes annoying. However, I have obtained the best evidence an investigator could wish to observe in the card experiments. To a limited degree I have experienced a mode of perception that is non-sensory, but which is registered in consciousness as an hallucinatory image of a card symbol.

"It is this phenomenon of image projection that has interested me most, for I am quite certain that here is an important clue that opens the way for a better understanding of ESP. If I am working along the right lines, it should become possible for good subjects to improve their scores with the cards."

Dr. Rice's card work with himself as subject. Dr. Rice refers to some work of his own in which he verified Miss T.'s experience of a vivid image of the symbol in connection with her highest success. His own data shows that he began working alone in December, 1936, just after Miss T. began doing phenomenal work, and got his best results when alone. He had an average of 10.1 for 9 consecutive packs, a highly significant series. When he began serving as subject with Miss

T. holding the cards he was not able to keep up this rate of scoring. However, he did quite significant work under the conditions in which he did not see or touch the cards at all (cf. experiments 8 and 9 in Table I).

Dr. Rice was also able to work successfully under the GESP condition, in which either telepathy or clairvoyance is theoretically possible. We have seen that Miss T. never succeeded at this, due probably to her excessive shyness. In Dr. Rice's case, he did better under witnessed conditions when someone was thinking of the card or drawing the symbol in the air (both are GESP conditions), though even these scores were strikingly lower than those of his clairvoyant tests done alone. The fact that Dr. Rice was capable of scoring as high as is usually expected of only the best ESP subject, under conditions which he called "making the most of an unreliable mode of perception," and that he might fall to a much lower level of scoring or fail altogether when he did not get "images" will be of interest in connection with a discussion of his work with drawings which appears in a later section.

Work of other subjects. Aside from the general observation made by Dr. Rice that several other subjects had been able to verify Miss T.'s experience of the vivid flash or image in connection with ESP success, there is little to see in these results which has not already become apparent. In the case of subject Rem, work done alone was reported as giving a score of 8.2 for 83 runs when the subject was seeing the symbols in imagination, but a test observed by Dr. Rice failed completely, and Rem. asked to have the test stopped because he could not get into the mood.

During December, 1936, Dr. Rice attempted to have his subjects specify which of their witnessed calls they felt certain were correct. This has been attempted by other investigators, with poor results, as it seems to have a disturbing effect upon the subject. Woodruff and George³ reported that they had subjects estimate scores on individual runs and found no success in so doing. By making his tests incidental, having a subject call out during the course of the run when he had a vivid impression of the symbol, Dr. Rice got 61 such indications from 6 subjects, and 35 of them were correct. This represents an average of about 15 per pack, and though the number of trials is small, it is interesting that this about equals Miss T.'s results reported when working alone.

³"Experiments in Extra-Sensory Perception", J. OF PARAPSYCHOLOGY, 1937, I, 18-30.

From these card tests a strong suggestion arises that ESP scoring ability might be improved if one can get control of a peculiar mode of experience associated with success. It must be remembered, however, that Dr. Rice emphasized the feature of vivid imagery for all his subjects. It is not at all beyond the range of possibility that what the subject said about imagery was due to suggestion. Even if the images were genuine indices of success, we cannot assume that they were a clue to the nature of the ESP percept. They may simply have been the way in which his subjects experienced their impression just as other subjects have said they commonly used visual imagery, and as still others have said that they use no imagery at all.

SUMMARY OF CARD RESULTS

A total of 1,569 runs of 25 trials each with ESP cards was recorded with 2,917 more successes than the most probable chance number. This is an average of 6.9 per run and is clearly beyond the ordinary range of chance occurrence. The most rigorous tests number 565 runs with a deviation of 358 hits above chance; this is highly significant, with odds of more than a trillion to 1 againsts its being a chance occurrence. On the strength of these rigorous tests, the conclusion is reached that ESP was demonstrated in Dr. Rice's experiments.

The other work is interesting for the suggestion which it contains that imagery may play a peculiar and important part in ESP. No definite conclusions on this point can be reached as a result of the card experiments.

II. INVESTIGATIONS USING DRAWINGS

The question of the role of imagery was raised by the card work, but no definite conclusions were reached. In this section of the report will be presented and discussed the results of another investigation by Dr. Rice designed to bear further upon this problem. In these experiments he attempted to give imagery a more obvious role in producing the results by testing his ability to perceive simple drawings by ESP.

PROCEDURE

These tests consisted simply of having someone (usually one of the people who had acted as subjects in the card tests) make a drawing of any kind on a blank piece of paper which Dr. Rice then attempted to reproduce. The person making the drawing, whom we may call the

agent, was instructed to put down anything he chose without any qualifying remarks to Dr. Rice. Dr. Rice wanted no suggestion about the drawings, as he felt these started rational thoughts which interfered with the mental function to be studied. While the agent's drawing was being made, Dr. Rice sat either across the room with his back turned or in the next room out of sight of the agent. When the agent was ready he would say "All right," and would then continue to look in silence at what he had drawn until Dr. Rice finished marking down any images or impressions he received. The agent's original (O) and the subject's reproduction (R) would be compared at once by both people. Usually, several tests were made in this manner in the course of a single session.

These drawing tests were started early in April, 1937. Dr. Rice experimented for the next 8 weeks (until his death) with this method. This study will be concerned only with those drawing tests in which Dr. Rice himself was the subject; these include all that were made under his direct observation and are about 80% of all the drawings found among his records. During the first 7 weeks, he was the subject in 49 tests. Several people took part as agent. In the last week, I was able to conduct 15 tests, 7 of them being with drawings prepared in advance by copying on cards some of the originals of the Sinclair tests as reported by W. F. Prince.⁴

RESULTS WITH DRAWINGS

The results (O's on the left and R's on the right) of 20 of the 64 tests are shown in the accompanying illustrations. The numbers show the chronological order of these tests among the entire series of 64. These 20 tests include the 8 results in which the subject's R's showed the most striking similarities to the agent's O's, as well as the 12 with the least similarity, as shown by the objective method of evaluation later to be explained. The reader may be interested to judge for himself at this time which of these tests will appear, from an objective approach, to be successful results.

Two questions must be answered before attempting to say how results of this sort bear upon any special problem in ESP. First, do the O and R of each test show on the whole more than a chance similarity? Secondly, do the conditions of testing justify attributing a significant correspondence, if it is found, to ESP?

⁴*The Sinclair Experiments Demonstrating Telepathy*, Bulletin XVI, Boston Society for Psychic Research, 1932.

	<p>T >>X</p>
<p>1 AGENT-REM</p>	
<p>4 AGT- MISS T</p>	
<p>5 AGT- MISS T</p>	
<p>7 AGT- MISS T</p>	
	<p>→ 1 1</p>
<p>9 AGT- MISS T</p>	
<p>99</p>	<p>«</p>
<p>11 AGT- MRS. M</p>	
<p>Λ</p>	<p>Λ CW T φ</p>
<p>13 AGT- MRS. M</p>	
<p>20 AGT- A. J.</p>	
<p>24 AGT- REM</p>	
<p>25 AGT- REM</p>	
	<p>Found, checked signs XX / BO before drawing.</p>
<p>30 AGT- MISS T</p>	

	<p>life</p>
<p>32 AGT- MISS T</p>	
<p>33 AGT- MISS T</p>	
<p>34 AGT- MISS T</p>	
<p>43 AGT- A. J.</p>	
	<p>circle inside of circle</p>
<p>47 AGT- MR. R</p>	
<p>54 AGT- J.G.P.</p>	
<p>61 AGT- J.G.P.</p>	
	<p>circle inside of circle</p>
<p>62 AGT- J.G.P.</p>	
	<p>N M H = </p>
<p>64 AGT- J.G.P.</p>	

EVALUATION OF THE RESULTS

Freehand drawings are not new as a means of testing ESP. Many of the studies of this nature made before 1929 have been reviewed by Dr. Prince,⁵ who was the first to develop a quantitative method of evaluating such results. By using his method, he was able to conclude quite soundly that many experiments with drawings have yielded significant results.

The use of drawings in parapsychological research has been more or less in disrepute for a long time because it has appeared to be practically impossible to evaluate the results. Dr. Prince's method was a step in the right direction, but in addition to being quite laborious it is based upon the assumption that ESP is a very rare ability. His method consists essentially of repeating the experiment to be evaluated, using the O's of the original experiment for pseudo-subjects, chosen at random, to "guess"; in this way several sets of R's are obtained for the one set of O's. Then judges decide for each O how the real subject's R compares with those of the pseudo-subjects. This method depends for its validity upon having a fairly wide difference in ESP ability between the true subject and the pseudo-subjects. It would fail in case there were either only a slight gift on the part of the true subject or an equally marked ability shown by half or more of the pseudo-subjects.

There is a more direct method of evaluating such work. This is a procedure called the method of correct matchings, which has an established statistical basis.⁶ The method is one which has been widely

⁵ *Ibid.*

⁶ Chapman, D. W., "The Statistics of the Method of Correct Matchings," *The American Jour. of Psychol.*, 1934, XLVI, pp. 287-298.

FIG. 1—The results of 20 tests are shown, the 8 most successful and the 12 most unsuccessful as indicated by the objective method of evaluation used. The originals, drawn in each case by the agent indicated, are on the left in the pairs and Dr. Rice's reproductions are on the right. The numbers indicate the chronological order of these tests with respect to the whole series of 64. Notes: No. 30: Dr. Rice wrote "Faint, doubtful images" before comparing his reproduction with the agent's original. The words "before drawing" indicate that this impression was received telepathically, i.e., after the agent had made her choice in imagination but before she had made any marks on paper. No. 32: The subject made one reproduction "before" the agent had started her drawing and another "after" she had completed it. No. 47: After Dr. Rice had completed his drawing and before he saw the original, he wrote "Circle inside of circle." No. 62: The note, "Resembles object looked at by J. G. P." was made after the two drawings were compared and may not properly be considered as contributing to the significance of the result; this note was deleted in this drawing as used in the evaluation of the results.

used in psychological work, but this is its first application in a parapsychological study. The method itself can best be explained by showing its application to the material in hand. The question to which it makes possible an answer is: Do the O's and R's of the 64 pairs of drawings show on the whole more than a chance degree of similarity within the pairs? This evaluative method has nothing to do with the actual interpretations to be made, but it may show that the interpretations are based upon a sound result rather than mere impression or ill-defined judgment.

All of the O's and R's were traced on two sets of 64 slips of paper each (small scratch-pad size). Each set was then shuffled thoroughly and the random order of slips in each set numbered from 1 to 64.

As the method of correct matchings had never before been used for results of this type, it seemed desirable to find the results of matching different numbers of slips in a group. This was done in order to get a basis for estimating the reliability of the method and to see how different sizes of matching units support further analyses of the results. Three sizes of matching units were tried with the following results:

(1) Four judges were asked to match all 64 of the R's to the 64 O's. The successes were 2, 4, 4, and 5, giving an average of 3.75 where a value close to 1 would be most expected by chance. From the tables which Chapman gives, we find that the probability that this would occur by chance is less than 1 in 1,000 (this is as far as the tables go), which justifies assuming some principle tending for correct matchings.⁷

(2) The above 4 judges and 2 additional ones were asked to match 16 groups of 4 R's each to 16 such groups of O's. In each matching unit, the O's and R's belonged together, but the judges had no clues to the correct matchings except those intrinsic to the drawings themselves.⁸

⁷ On each of 3 correct matchings there was agreement between 2 judges; these are numbers 13, 47 and 54 among the test pairs reproduced in this report. However, there was agreement among 3 judges in 3 cases and between 2 judges in 15 cases that certain incorrect matchings belonged together. This does not invalidate the evaluative procedure or the significance of the results for the average number of correct matchings, but only shows that similar drawings might occur in two series as long as this without any direct causal connection. A control check was made by comparing the order of the O's for each judge with the numerical order of the R's, beginning with R no. 2 instead of no. 1, and going on through no. 64; the last O in each judge's order was checked against R no. 1. The correct matchings with the judgment factor here eliminated were 0, 1, 0 and 2. The average (.75) is not significantly different from 1.

⁸ In forming these sub-groups, the R's were taken by groups of 4 as they appeared in the shuffled set of slips, and the corresponding O's were selected from

Each of the 6 judges had 16 separate matching groups, making 96 separate matching groups in all. The number correctly matched was 174, or the average number of correct matchings per group of 4 pairs was 1.81. By consulting the same tables (Chapman, *ibid.*) as before, we find that the probability that this would occur by chance is again less than 1 in 1,000 and is therefore highly significant.

(3) The 6 judges who acted in the second matching condition and 3 additional ones were asked to judge in 32 matching units consisting of 2 O's and 2 R's each. The 4 drawings making up each group had been obtained together in two tests. The object of having them matched by the judges was to see whether an R tended to be more similar to the O of the same test than to one other O chosen at random. In each of the 32 matching units, a judge had to be right for both pairs or wrong for both, with a chance probability of 1/2 for either result. The results may be considered, therefore, on the basis of 32 trials for each judge with an average chance expectation of 16 successes. The actual number of units or trials in which correct judgments were given were 21, 20, 19, 18, 24, 21, 21, 23, and 22. This gives a deviation of 45 beyond mean chance expectancy, which would on the average happen by chance less than once in a million times.⁹

The three applications here made of the method of correct matchings (64 with 64, 4 with 4, and 2 with 2) all give highly significant results. Whereas it might appear that the two methods using sub-groups might have been favorable to correct matchings because of the particular groupings into smaller units, this could have no effect upon the number of correct matchings when the whole series was the unit. The consistency of these three results leaves no doubt that there is on the whole a greater than chance similarity between the O's and R's which were drawn together in the tests.

The results of the 20 tests, 8 successful and 12 unsuccessful, actually reproduced in this study are those showing the least and the most similarity as determined on the basis of the matching condition in

the other shuffled set of slips and presented in the order in which they were found. For example, the first matching unit was presented as follows:

O's	1	2	3	4
R's	2	43	49	54.

For a perfect score, the R's would need to be rearranged as follows:

O's	1	2	3	4
R's	2	54	49	43.

⁹The formula used is $C.R. = dev. / \sqrt{npq}$ where n (trials) = 288, p (the chance of a success on each trial) = 1/2 and q (the chance of a failure on each) = 1/2. The C. R. is 5.3 which gives the odds against chance stated.

which the unit was 4 pairs. It will be remembered that 6 judges served in that condition. When all 6 judges correctly matched a particular R with a particular O, this was arbitrarily taken as showing a striking similarity. When no judge matched an O and an R which really belonged together, this was arbitrarily taken to indicate a striking lack of similarity. By these criteria, pairs No. 1, 4, 13, 20, 32, 34, 47 and 54 include the best results of the whole series, and the remaining pairs include those drawings showing the least similarity.¹⁰ This procedure for limiting the number of pairs to those which should contain the best and worst results (as far as similarity goes) is not infallible, but it is a method by which the similar and dissimilar pairs can be segregated with relative safety for further study.

INTERPRETATION OF THE RESULTS

Thus far we have been concerned only with finding and applying a suitable method for determining whether there is anything more than a chance similarity between agent's and subject's drawings in the 64 tests. The method applied shows that there is a significant degree of resemblance, and this fact raises the second question of how to account for the results. Was the more-than-chance relation shown in these drawings due to ESP?

Two features of the conditions of testing stand in the way of giving an unqualified affirmative answer. One is the fact that the agent made his drawings in the same room with the subject or one room away with the door open. This leaves the alternative possibility that the subject was unconsciously influenced by the sounds of the pencil as the agent prepared an O. If the subject were accustomed to draw complete pictures and if the O's were fairly complex, it would appear to be impossible to detect what drawing was made by the sound of the pencil. But in the present experiments the agent frequently prepared very simple drawings and the subject deliberately avoided trying to form his impressions into completed pictures. It is not unreasonable to suppose, therefore, that the subject might unconsciously have detected the characteristics of some parts of the drawing of the agent as it was being made; the sound of drawing an angle and a curve, for instance, are quite distinct. It is impossible to tell whether sensory cues of hearing entered into these results.

¹⁰ The fact that there was perfect agreement among the judges in these 8 cases shows that they were guided by the objective similarities in the drawings rather than by ESP in making their correct matchings.

The conditions may also be criticized for the fact that the O's were prepared by the agent just before each test, usually just after the results of the last test had been examined and discussed. There is a possibility that the agent and the subject were influenced in this manner to make the same or similar choices. The agent might tend, for example, to alternate in his O's between figures with predominantly curved and predominantly straight lines, and the subject might unconsciously anticipate these changes. Or they may both have noticed some incidental object or event in the environment and have been influenced by this fact to produce similar drawings.

Because of these aspects of the testing conditions, it cannot be concluded that these results independently prove the occurrence of ESP. However, independent proof that ESP occurs was not the primary purpose with which these tests were made. For those who accept perception beyond the known senses as already proved, it may seem more or less likely that the ESP hypothesis has some bearing upon these results. If ESP be accepted, it is no more proper to conclude that it was *not* demonstrated by the results under consideration than to conclude that it *was*. We must leave the matter just as it stands: the results may signify the unconscious use of sensory cues or suggestions arising from the common environment; or they may signify the occurrence of ESP; or both factors may have entered to some degree.

DISCUSSION OF THE RESULTS

These tests were designed as a further study of the role of imagery in ESP. Although Dr. Rice considered this work only preliminary in character, we may consider tentatively what light is thrown upon this particular problem. The reason for considering the results in this regard is to see whether they give any leads for further investigations along this particular line. For our present purposes we may proceed on the assumption that the results of the drawing tests indicate that ESP was occurring and see whether there is any relation between the kind of images recorded by the subject and the success of the tests.

In this matter each one will have to judge for himself. It seems to me that an examination of the 20 pairs of drawings presented in this study show no essential difference in the images (subject's R's) of the successful and the unsuccessful tests. There was a tendency for the subject to mark down things without knowing their meaning, but he did this as much when he failed as when he succeeded. Also, he tended to put down incomplete diagrams or fragments. In the successful tests,

the fragments have a striking resemblance to some part of the O. In the failures, fragments are frequently present in great numbers but they have no obvious relation to the O. In only one case among the 20 reproduced did the subject stop with a single drawing (no. 43), and that was a failure. In another case, the agent drew a footrule, and the subject drew a heavy straight line with the words: "Saw only straight line"; 5 of the 6 judges in the second matching condition correctly matched these two. There appears, therefore, to be no relation between the number of images recorded by the subject and his success.

As far as the actual results show there is no detectable difference in the characteristics of the R's when the subject was successful and when he was unsuccessful in copying what the agent had drawn. Only in one case did the subject record his introspections; in test no. 30 he wrote "Faint, doubtful images," and in that test he was unsuccessful. Perhaps further introspections might have revealed a more definite role of imagery in ESP than the present results indicate.

CONCERNING TESTS WITH DRAWINGS IN GENERAL

This study offers an excellent opportunity to compare the method of drawings with the one based upon ESP cards, as Dr. Rice used both methods. From his results with cards, the definite conclusion was reached that ESP had been demonstrated by two subjects, Miss T. and himself. When the results with drawings were considered, however, we were not able to reach that conclusion on the basis of those tests alone. The drawing results were evaluated by an objective, quantitative method and the similarities shown were clearly beyond the range of what we might reasonably expect to find by chance. The inability to conclude that this was due to ESP emphasizes the greater difficulty of conducting an adequate test with drawings and leads to the question whether such a test might be conducted under conditions adequate to demonstrating the occurrence of ESP.

A drawing test for ESP may be conducted with clear-cut results by meeting the following requirements: (1) All the original drawings should be prepared in advance of making any actual tests. They should be made upon cards or pieces of paper so that they can be shuffled easily. (2) When a test is to be made, the precautions necessary to insure against the subject's obtaining sensory cues from the agent must be observed. In making a test, the agent concentrates upon one of the drawings taken at random from the prepared set while the subject tries to reproduce it. The test can be conducted clairvoyantly

simply by not looking at the drawing selected. If it is not inconsistent with the purpose of the experiment, the subject may know his success immediately after each test in case each O is returned to the set and they are all shuffled again before the next drawing is made. (3) The results must be evaluated in an objective manner, using the method described in this study unless a better one be discovered. When the test is planned from the outset with the evaluative procedure in mind, the task may be greatly simplified by the manner in which the whole procedure can be organized. For example, when O's prepared in advance are used, it is permissible to group the results into matching groups of the desired size simply by following the chronological order of the tests. If the method of correct matchings is used, it is recommended that the matching unit should contain from 4 to 10 pairs; smaller and larger numbers do not appear, on the basis of the present study, to help as much toward further analyses of the results. The number of judges needed will vary with the quality of the results; the less the amount of resemblance, the more judgments necessary to show its significance. The numbers used in the present study were adequate for this case.

Investigations with drawings have had a prominent place in the history of parapsychological investigations. Two of the most important investigations using this method are the work of Sinclair¹¹ whose material was further studied by Prince¹² and the experiments of René Warcollier and his collaborators.¹³ However, investigators using the drawing method have never put their investigations on a strictly measurable basis, and it is probably for this reason that they have not been more successful in getting their results considered as evidence of ESP. The simpler, quantitative procedures, chiefly those based upon ESP cards, have produced evidence which has been convincing to a much larger number of people. This appears to indicate that, as far as establishing the fact of ESP is concerned, the card tests have yielded better results. We cannot know what the case would have been, however, if the tests with drawings had been conducted with as much attention to the conditions of testing and the evaluation of the results as the card tests received.

¹¹ Sinclair, U., *Mental Radio*, New York: Albert & Charles Boni, 1930.

¹² *Op. cit.*

¹³ Warcollier, R., *La Télépathie*, Paris, 1921, and subsequent articles in *Revue Metapsychique*. An English translation by Dr. Gardner Murphy and Mrs. Sidney Gridley will be published soon by The Boston Society for Psychic Research.

But of more immediate concern is the question of the relative importance of these two methods in future parapsychological investigations. Should both procedures be continued? Are there problems of importance which can only be answered by some variation in the method of drawings? The drawing method when used carefully is much more difficult and time-consuming than the simpler card methods. Are there limits to the usefulness of the simpler methods of card tests which justify a greater emphasis upon drawings and similar approaches than they have been receiving?

The questions raised in the last paragraph will probably bring out strong arguments, *pro* and *con*. It is not possible in this paper to reach definite answers nor desirable even to take sides. As far as this matter is concerned, the aim of the present study has been achieved in emphasizing some of the difficulties of the drawing method and in showing how some of them may be overcome.

However, some pointed suggestions arise which in themselves sufficiently justify the experimental use of both methods until the usefulness of each shall no longer be a matter of conjecture.

(1) A wider range of people may show the ability of ESP if both methods are used. Some may work successfully only with the cards, others only with drawings, and some may be able to adapt to both.

(2) The use of drawings may give a more direct approach to the study of how ESP impressions get into consciousness. There may be something about the process in question which can best be studied in a situation permitting the subject to show as much freedom and spontaneity as he can in the drawing method.

(3) It is possible that ESP ability will not become inhibited after a time with drawings as has appeared to be the rule with the card tests.

(4) By the drawing method, we may better be able to isolate the individual successes due to ESP and make a special study of the conditions under which they occurred. In the card method, this is approached by analyzing the conditions under which good *runs* or *series* occur.

(5) By the card method it is possible to study negativism in regard to ESP, but this appears to be relatively impossible with drawings. The card tests give a nicer distinction between results obtained under two different conditions.

Enough has been said, without attempting to exhaust the list of possibilities, to indicate that these two methods are not mutually ex-

clusive, but that they may supplement each other. The investigator must choose his methods to suit his experimental needs.

SUMMARY OF DRAWING RESULTS

Dr. Rice was the subject for 64 tests in which he attempted to reproduce freehand drawings made by a number of agents. An objective evaluation of the results shows that he was more successful than might reasonably be expected by chance. However, the conditions of conducting the tests make it impossible to conclude definitely that the results demonstrate ESP.

The results are discussed for what they may be worth as they bear upon the question of the role of imagery in ESP. There is no evidence that imagery played a different role in the successful and unsuccessful tests. While the present experiments do not lead to definite conclusions, the results suggest that the role of imagery in ESP may depend upon individual differences in subjects. The question of whether vividness of imagery and ESP success are correlated cannot be answered in the present study.

Some suggestions are made concerning the requirements for an adequate test of ESP by means of drawings.

A CASE OF HIGH SCORES IN CARD GUESSING AT A DISTANCE

BERNARD F. RIESS

ABSTRACT: A single subject, chosen on the basis of reputed "psychic" ability, called two series of tests with ESP cards, totalling 84 runs through the pack. Two runs were made each night at irregular intervals, and upon a pre-arranged time scheme. The subject was one-quarter mile from the room in which the experimenter handled the cards. The latter looked at each card following an agreed-upon time schedule. In the first series of 74 runs the average number of hits per 25 calls was 18.24. After an intervening illness on the part of the subject, 10 runs under similar conditions yielded an average of 5.30.

Dr. Riess is Assistant Professor of Psychology at Hunter College of the City of New York.

THE FOLLOWING article presents the results of an experiment which is unfortunately incomplete and which will remain so for an indeterminate time. The data, however, are in themselves so unusual that the author feels justified in publishing them despite the risk of adding to the supply of unverifiables in scientific literature.

In view of the nature of the results it is necessary to present the history of the investigation in some detail. During the Fall of 1936, Dr. J. B. Rhine lectured to the Psychology and Education Club of Hunter College on the general topic of ESP. The author attended and, after the lecture, had some discussion with Dr. Rhine on the psychological factors involved in the Duke experiments. Training and interest in comparative psychology had made the author skeptical of the whole problem. However, during discussion of the lecture in a general psychology class, one of the students volunteered to have a friend, who claimed to possess high ESP, act as subject. Believing that such demonstration with ESP cards under fairly well controlled conditions would illustrate the topic with proper emphasis, the instructor initiated the present experiment. A pack of ESP cards was obtained from Dr. Rhine and mimeographed record sheets were set up in accordance with instructions. After much urging, the young lady was persuaded to serve as subject. Because of physical limitations it

was impractical to conduct the investigation in the College laboratory. When the results were found to conflict with the preexistent expectations of the instructor, he immediately sought for control groups to determine, under much better conditions, whether others would yield the same high scores as the single individual reported on here.

METHOD

The usual pack of ESP cards was employed. Since both the subject and the experimenter lived in the same suburb of New York City, the experiment was conducted in their respective homes. At 9:00 P.M. the experimenter would expose a card from the freshly shuffled deck lying on his desk. He looked at the card and noted the symbol on the mimeographed record sheet. At 9:01 P.M. the subject, seated at a table in her home guessed at the symbol on the first card exposed by the experimenter in his home. The experimenter, meanwhile, had placed the first card to one side and at 9:01 had exposed the second card. This procedure was repeated until the deck was exhausted. It was then reshuffled manually and, after an interval of 10 minutes, the whole experiment was done again. In this way 50 trials were run off during each experimental session. The distance between the two homes was approximately one-quarter mile and the respective rooms in the houses were so situated that they faced in opposite directions to each other.

The experiment was begun on December 27, 1936 and continued at irregular intervals until April, 1937. By that time 1850 trials had been run. The scores on this part constitute Series A in the statement of results. During the months of April, May, and June the subject was unable to continue because of poor health and the experiment was not undertaken again until July 18. During this month and the succeeding month 250 trials were run, constituting Series B of the results. In the latter part of August the young lady moved to the Middle West and has since refused to participate further in the experiment. It is hoped that she may again be induced to serve in a new series in which more adequate statistical controls will be employed.

SUBJECT

Miss S. had attended a woman's college for three years as a music major. She had been compelled to leave her academic environment because of ill health. On her return home, she occupied herself with music, giving lessons and attending concerts. At the time of the experiment she was 26 years of age. Her medical record is of some

interest in view of the fact that she gave, as excuse for the lapse of time between the two series of trials, the statement that she had had a "nervous breakdown." Her physician was approached by the experimenter and in view of the possible importance of physiological facts in explaining the difference in scores on the two series gave the following report on her condition. He said that she had had a condition of hyperthyroidism of long standing which he was treating and for which he had prescribed treatment during the three-month interval between series. He had recommended a change of scene and interest which brought about the change in residence. From the limited observation of the author, Miss S. seemed to be a young lady of high intelligence and limited interests. She was obviously not well adjusted to her life in the community and on several occasions stated that she was "sick of living at home." Just prior to the discontinuance of the experiment she was asked whether she thought her scores were different from what they had been during Series A. She said that she did not believe that there was any difference in scores although she was much less curious about the scores than in the former series. At no time in either series was she informed of the nature of her scores.

RESULTS

The results of the two series are presented in Tables I and II. In Table I is stated the summary of the complete experiment. In Table II, the scores per 25 cards are given for both series in terms of successive runs of 25 cards.

DISCUSSION

The author has sought merely to present the raw data of an investigation which is both incomplete and not in accordance with the preferred scientific methodology. In order to satisfy all the mathematical criteria it would be necessary to present the cards either shuffled after each guess or taken one at a time from a series of 25 decks of cards. From the psychological point of view it is unfortunate that

TABLE I.
SUMMARY OF SCORES ON THE TWO SERIES

	Number of trials	Number of hits	Average hits per 25	Deviation from np	Deviation /S.D.
Series A.....	1850	1349	18.24	979	25.6
Series B.....	250	53	5.30	3	0.5

TABLE II.
TOTAL HITS ON EACH RUN OF 25 CARDS

Run	Hits A	Hits B	Run	Hits A	Run	Hits A	Run	Hits A
1	5	2	21	21	41	16	61	18
2	7	4	22	20	42	21	62	18
3	10	7	23	19	43	22	63	19
4	12	12	24	20	44	17	64	18
5	15	7	25	18	45	16	65	17
6	8	5	26	14	46	18	66	18
7	16	4	27	15	47	19	67	19
8	13	3	28	15	48	21	68	20
9	18	5	29	16	49	20	69	20
10	21	4	30	12	50	19	70	20
11	11	..	31	19	51	16	71	19
12	15	..	32	21	52	21	72	20
13	19	..	33	22	53	23	73	21
14	24	..	34	24	54	19	74	21
15	21	..	35	20	55	17
16	21	..	36	18	56	22
17	22	..	37	22	57	21
18	24	..	38	21	58	18
19	25	..	39	19	59	15
20	24	..	40	19	60	14

this subject like so many others who have been tested has not remained available for long-continued research. In view of the many uncontrolled factors, the data as presented are to be thought of as suggestive only. It is to be hoped that any other experimenter finding a subject equivalent to Miss S. will immediately start with correct methods and controls.

The results of the tests made upon 67 college girls who were run for a comparable number of trials under carefully controlled conditions have yielded no significant data. The scores are, however, still being studied and will be presented in a later paper.

A STUDY OF COMPARATIVE PERFORMANCE IN SEVERAL *ESP* PROCEDURES

EDMOND P. GIBSON

ABSTRACT: A comparison of scoring rate in card-calling vs. card-matching procedures in *ESP* was made with 67 subjects. The total results of the preliminary tests were significantly above chance and eleven of these subjects were then selected for the more extensive investigation involving the comparison of procedures. Of these, nine obtained scores that exceeded the criterion of significance for extra-chance results. No significant differences were obtained, however, between calling and matching. Some striking individual differences in general performance were found. Further confirmation of the U-shaped curve of successes in the DT method was found.

Mr. Gibson is City Engineer of East Grand Rapids, Michigan, and conducted these experiments mainly in his home, with a short series at Duke University.

INTRODUCTION

FROM preceding reports it appears that different methods of testing for *ESP* have given different averages. In the Rhine report¹ a constant difference was found in the same subjects in the results obtained through the use of the BT and DT procedures.² Woodruff and George³ obtained similar differences from the use of various procedures. OM was found to yield higher averages with a given subject than BM. Even with a screen interposed, SOM_i still yielded twice the deviation from chance found in SBM. Woodruff and George also found that BT procedures usually gave averages comparable to those of the same subjects working with the OM method. In their experimentation, the BM procedure gave lower averages than other techniques.

It thus appears that, in the first mentioned work, there was a marked difference between two card-calling methods (DT and BT) and in the second, a great difference between two matching methods (OM and BM). But the general comparison of calling and matching procedures needs further investigation, and toward that end the research reported here has been directed. The investigation involved the use of the vari-

¹ Rhine, J. B., *Extra-Sensory Perception*, pp. 85 and 89-91.

² For the meaning of abbreviations used in this study, see the Glossary, p. 306.

³ *JOURNAL OF PARAPSYCHOLOGY*, 1937, I, 18-30.

ous techniques mentioned above, but so distributed as to give also a general comparison of the relative merits of verbal and manual responses in the ESP tests.

In making such a comparison, two general methods of procedure are possible: the parallel method, in which an investigation of two or more test procedures is made during the same experimental period; and the serial method of comparison, in which the investigation of one test procedure is effected at a time. Each method has its own peculiar advantages and handicaps. The parallel method was used by Woodruff and George and was followed to some extent at first in this study; but it was dropped as the experimentation progressed when it was found that the subjects tended to develop a preference for one of the several test procedures under simultaneous investigation. This preference itself became a factor which seemed to influence the scoring rate. In view of this situation, the successive method was used exclusively after the study got well under way. This method permitted the subject to concentrate his attention wholly upon the particular procedure being tested and to attempt to do his best at it. Arbitrary and often strong preferences for particular methods did not develop so readily, and such preferences as existed were less active. In this sense a fairer comparison of procedures is believed to have been made, though comparing different procedures in succession makes it impossible to be sure of having throughout the series identical physical conditions, general health, and experimental interest on the part of each subject. To compensate in a measure for these possible differences, large numbers of trials were made in all cases in which it was possible.

GENERAL CONDITIONS

The experiments, falling within the period from December, 1934, to March, 1937, were made mainly at the experimenter's home in Grand Rapids, Michigan. A few were made at the homes of various subjects, and one short series in the Parapsychology Laboratory at Duke University. All the data of each subject in each technique is presented intact, free from omissions.

*Subjects.*⁴ Sixty-seven subjects were tested for ESP ability in exploratory tests, in which the GESP and BM procedures were generally used. Eleven of the subjects succeeded in scoring high enough to warrant their use in the comparative tests, and were retained for that purpose. These eleven are referred to as the major subjects of the

⁴All of the subjects have given liberally of their time to the present inquiry and have cooperated splendidly throughout. Their generous aid is gratefully acknowledged.

investigation. Of the remaining 56 subjects, several made significantly high scores in the one or two procedures they tried but were unable for various reasons to work in enough different procedures to make possible the desired comparison. They do not therefore concern the issue of this report.

Of the eleven major subjects, chosen, as stated above, because they qualified for and worked in most of the procedures under comparison, a brief description may be of interest.

No. 1 is the principal subject, having made one-third of the total number of trials. She is the wife of the experimenter, a college graduate, formerly a teacher, and in her early thirties.

No. 2, daughter of the experimenter, was five years old when the investigation began.

No. 3, the experimenter's son, was nine.

No. 4 is a journalist, under thirty years of age, who has done some automatic writing.

No. 5 is a Czech. He holds a degree in metallurgical engineering, has considerable ability as an artist, and is exceptionally versatile. He also has done automatic writing. He is in the forties.

No. 6, the wife of No. 5, was formerly an interior decorator. She is in the thirties.

No. 7 is a college graduate, now about fifty years old. He is now engaged in public service work; he was formerly director of a Y.M.C.A.

No. 8 is a registered nurse and college graduate. She is about forty-five.

No. 9, the wife of No. 4, is a university graduate, formerly engaged in public welfare work. She is twenty-five years old.

No. 10 is the wife of a prominent local engineer. She does automatic writing. She is in her late forties.

No. 11, the author, is the City Engineer of East Grand Rapids, Michigan. He has been engaged avocationally in parapsychological experimentation since 1932.

Materials. The tests were made with ESP cards. A shuffling machine was used instead of hand shuffling during certain periods of experimentation. A screen similar to that used by MacFarland and George⁵ was used in the STM procedure.

Procedure. The following procedures, described in the glossary, were used by the eleven major subjects:

Calling	Matching
GESP	OM
Distance GESP	BM
DT	STM

⁵ JOURNAL OF PARAPSYCHOLOGY, June, 1937; see photograph opposite page 95.

Not all eleven subjects worked with each procedure, but each worked with a sufficient number of them to make possible a general comparison.

Precautions. The usual safeguards to prevent the operation of sensory perception and inference were taken. Hand shuffling usually consisted of at least two fan or dovetail shuffles plus one or more cuts by the observer. Mechanical shuffling was used at times without any lowering of the scores.

In the GESP method, the possibility of the subject's being guided by sensory cues was minimized if not completely eliminated by placing him with his back to the observer or behind a partition in the next room.

In the GESP (long distance) tests, calls were made and recorded by the subjects to indicate cards looked at successively at specified intervals by an agent who was at a distance of from seven hundred to two thousand miles. During part of the experimental period the agent was at Durham, North Carolina, and during the rest of the time in San Diego, California.

The checking was done at the Parapsychology Laboratory of Duke University, to which the records both from the agent and from the subjects were sent.

In the DT procedure the danger of sensory cues is limited entirely to the top and bottom cards. The 23 cards inside these are effectively screened.

The OM work was done on non-reflecting table surfaces. New packs of cards were used as soon as old ones became slightly worn or soiled. In the BM method new key cards were used even more frequently than new decks. In addition, a number of sets of five cards were prepared and used successively.

In the STM test for clairvoyant perception the screen was quite adequate to remove the possibility of the subject's seeing the cards of the deck held by the experimenter. The observer's aperture was shielded from the subject's vision by a second screen.⁶

In general, except where prohibited by the experimental conditions, the checking was done by the observer with the subject as a witness. Whenever possible, additional witnesses were employed. In most of the experiments involving the other subjects, Subject No. 1 acted as additional observer. The possibility of a subject's improving his score by keeping track of the cards left in the deck was completely ruled out by the fact that no checking up was done before 25 trials had been made.

Methods of Evaluation. The statistical reliability of the results is

⁶ JOURNAL OF PARAPSYCHOLOGY, 1937, I, opposite p. 95.

determined by the established formula applied to such material; that is, by dividing the total *deviation* from the *mean chance expectation* by the *standard deviation*, to obtain the *critical ratio*.⁷

Results. A total of 93,875 trials was made with 67 subjects, giving an average of 5.54 hits per run of 25. These tests yielded a positive deviation of 1,954 hits. This deviation gives a critical ratio of 15.9, a highly significant figure which indicates clearly the operation of an extra-chance factor in the experiments as a whole.

Of the total 67 subjects the eleven major subjects used in the comparison obtained an average of 5.65 per run for 69,400 trials. This figure gives a critical ratio of 17.1.⁸ The results of the 56 minor subjects (Group I) are not significant; 24,475 trials averaged 5.15 hits per run, giving a critical ratio of 2.37, which falls short of the criterion of significance (2.5). The figures for these two groups appear in Table I.

It is clear from the above totals that whatever the explanation may be for the results of these tests as a whole, they cannot be regarded as due wholly to chance. The odds against obtaining a deviation so large as that obtained for these 67 subjects are of the order of 10^{57} .

In the remainder of this paper attention will be given exclusively to the group of eleven major subjects (Group II) and to the comparison of the results with the two general procedures, matching and calling.

One block of data consisting of 10,400 trials will be left out of the general comparison because it was produced by Subject No. 1 in unobserved tests. The average scores of this group depart but slightly from the observed results for the same subject, but the comparative score averages submitted below rest on the basis of observed results alone.

Calling Tests. The tests using calling responses, GESP and DT procedures totalled for the observed series 17,735 and averaged 5.55 per 25. (This includes the 4,750 trials with distance GESP, which averaged only 5.25 and considerably lowered the general average.)

⁷ See glossary for further information on the italicized terms.

⁸ These eleven subjects who took about 75% of the tests reported were used more extensively because of their greater success in scoring. As a safeguard against unsafe selection of results, therefore, all results of both high- and low-scoring subjects are reported in the grand total of 93,875 trials. The work of the eleven subjects who made comparative tests on several methods may safely be analyzed apart from the non-comparative tests given the other 56 subjects because, first, the work as a whole is significant; and, second, the comparison is in itself a separate experiment, justifiably treated apart, in which the objective is not extra-chance deviation but a comparison of score averages.

Matching Tests. On the matching tests (the OM, BM, and STM methods) the total observed series of 41,265 trials averaged 5.60 hits per 25, or only 0.05 above the calling methods. This is a very small difference, not statistically significant, and clearly affords no ground for discrimination between the two general types of test based on manual and on verbal modes of response by the subjects.

The totals and averages of these comparative tests are given in Table II.

Comparison of Procedures. The results of OM were higher in general average than those of any other procedure. The 6,875 trials yielded an average of 6.67 per run. GESP was next with an average of 5.97 per 25 for 4,535 trials. BM, with an average of 5.51 per 25 for the large number of 20,690 trials, was next lower than GESP and slightly higher than DT, which gave an average of 5.49 per 25 for 8,450 trials. GESP (long distance) and STM were lowest. Distance GESP, which averaged 5.25 per 25 for 4,750 trials, was higher than STM, which yielded only 5.19 per run for 13,700 trials. The total score for each procedure except distance GESP and STM is independently significant.

The scores of individual subjects in the calling and matching techniques, as given in Table III, showed six subjects with higher averages in calling than in matching, as against four who did better when matching than when calling. There was no general preference among the subjects for any of the six types of procedure. Five subjects obtained their highest scores when using the GESP procedure, two when using OM, two with BM, and one with DT. The remaining subject did not work in the calling techniques; hence, his scores are not compared here.

If the averages of all of the eleven subjects in each technique are ranked from highest to lowest, they are as follows: OM, GESP, BM, DT, GESP (long distance), STM. If they are ranked according to number of individuals scoring highest in each, GESP is first with five subjects, OM and BM next with two each, and DT last with only one subject who secured his highest average in it.

The outstanding point of this work is the obvious fact that neither matching as such nor the calling method in general stands out above the other type of procedure. Evidently two test procedures, one of each type, OM and GESP, were more successful generally than the other procedures tried, though which of these two is better would seem to be a matter of individual variation. It further appears from these comparisons that BM ranked about third, with DT still lower.

These results seem in line with those reported from other investigators. The marked difference in the scoring between the GESP and DT techniques noted above may be compared with the similar results obtained by Rhine⁹ experimenting with Pearce and Linzmayer. Rhine compared BT (a calling condition similar to GESP) with DT. Both Pearce's and Linzmayer's averages were higher in BT than in DT.

Woodruff and George¹⁰ also report experiments involving BT, OM, and BM. A comparison of their work with the data obtained in this inquiry is interesting. It appears in Table IV.

That this failure to score as successfully in BM as in OM is due to a psychological barrier rather than one inherent in the nature of the procedure is suggested by the large individual difference in BM in the present report. See in Graph I that four subjects did their best matching work in BM. The work of Price and Pegram,¹¹ who in working with blind subjects found that the BM procedure succeeded as well as OM, supports this view. In this case, where there was probably no psychological prejudice in favor of OM as against BM, BM yielded slightly higher results. Further, this work with blind subjects yielded practically identical averages for calling and for matching procedures.

Individual Comparisons. The averages of individual subjects for the various test procedures are given in Table III.

In the case of the principal subject, No. 1, there is a remarkable evenness in response to the various experimental procedures and a regularity and consistency which is sustained through a long series of experiments running over two years. Starting in December, 1934, she scored very evenly during the ensuing year in three methods involving both calling and matching. The whole of the first year's work was significant. Following a period of illness her work dropped to chance. It quickly rose again in the long distance experimentation and continued at a significant rate until the fall of 1936, when it again slipped to chance with the introduction of the STM procedure.

This protracted series of trials with significant deviations is unusual in the light of the work of the other subjects, who generally scored significantly over much shorter periods. It is noteworthy that this subject exhibits in general personality an unusual degree of calmness, steadiness, and persistence of interest. Table V reviews the stages of her "career in ESP testing."

⁹ Rhine, *op. cit.*

¹⁰ Woodruff and George, *op. cit.*

¹¹ Price and Pegram, "Extra-Sensory Perception Among the Blind," JOURNAL OF PARAPSYCHOLOGY, 1937, I, 143-155.

No. 4 brought a more intense interest into his experimentation as time went on and he was inclined to be either enthusiastic or discouraged by the tendencies of his scoring. His performance throughout is brilliant but is characterized by pronounced ups and downs, the downs being often so low as to spoil what otherwise would have been excellent scoring. This subject contrasts with No. 1 also in personality. He is nervous and less steady. He easily dissociates to the degree of doing automatic writing.

No. 5, working in the OM procedure, made the highest average of any subject. In fact, his brilliant scoring on this procedure gave it a great part of its margin over GESP. His 350 trials made in one evening using this technique yielded 192 hits, or an average of 13.72 hits per 25. During this experiment he directed the placing of many of the cards from across the room, where he walked slowly back and forth in what seemed a semi-dissociated state. This work was done in the presence of two observers and the deck was invariably shuffled and cut by one of these. In the distance GESP technique he no longer scored significantly; his deviations were then fairly consistently subchance. A chronological examination of his data tends to show that his ESP ability is like that of No. 4, somewhat fitful. As stated, No. 5 is artistic, gifted in several arts. In other activities he works characteristically by spells, brilliant periods being followed by depressions of performance. What stands out in the work with Nos. 4 and 5 is their peculiar success with OM and, in the case of No. 4, with BM, as compared with No. 1, who went placidly on through the different methods with an average between 5.5 and 7.0 until she tried the STM procedure.

Similar peculiarities in the ESP performance of subjects noted by other observers have seemed to be due to personal idiosyncrasies. Rhine reported that at first Zirkle could work successfully at pure telepathy while for a time he failed consistently with pure clairvoyance. With Pearce the opposite occurred. Woodruff and George reported a subject, M.S., whose ESP abilities improved with the use of a screen, while others were less successful with the same change. This peculiarity shown by M.S. was noted also in the work here reported in the case of No. 9. Carpenter and Phalen in "An Experiment in Card Guessing"¹² report that their subject L. could not score high with the DT technique because he could not "imagine" the individual cards so arranged. He supported this contention by producing a chance score in 1,500 trials after he had previously worked significantly using the

¹² JOURNAL OF PARAPSYCHOLOGY, 1937, I, 31-43.

BT procedure. Later on, then, L. did significant work using the DT method.

Peculiar differences such as these occur in the work of several investigators. They fall outside the scoring tendencies of the groups tested and may be due to the attitude or personality of the individual rather than to any inherent differences in difficulty in the conditions of the tests.

Work with the two children, No. 2 and No. 3, was not sustained due to a gradual loss of interest and a consequent effect upon the scoring. It was carried far enough to indicate that ESP would operate as long as interest could be maintained, as noted by L. E. Rhine in her work with child subjects.¹³

The DT Curve. In *Extra-Sensory Perception* Rhine reported that when the frequency of hits in DT work was plotted in groups of five cards down through the pack, the results were likely to follow a U-shaped curve, hitting occurring at a higher frequency in the upper five and bottom five cards of the ESP deck. The work of No. 1, who had made the largest number of trials in this procedure, was selected for a special study in the scoring order in DT. This study, over a series of 4,100 consecutive trials, partly observed, partly unobserved, shows a very interesting curve which when plotted in groups of fives is definitely U-shaped and similar to the curve first reported by Rhine and subsequently confirmed in the work of Sharp and Clark¹⁴ and by Pegram.¹⁵ These DT curves are shown in Graph I. The top and bottom cards were hit at a slightly higher rate than other high points on the curve, but not enough higher to justify the hypothesis of sensory cues.

SUMMARY

In the total of 93,875 trials in ESP tests given to 67 subjects in this research, an average of 5.54 per 25 trials was obtained; and the critical ratio is 15.9. The results are evidently not to be explained as due to chance factors. Significant work under conditions such as BM and DT offer special safeguards against sensory cues. Also the highest scores occurred in OM tests with the subject not touching or looking at the cards, nor shuffling them between runs, and with two observers present.

The principal group of trials (69,400) was made by eleven subjects who averaged 5.65 hits per 25. These subjects were given comparative tests in calling and matching methods of response, with GESP and DT

¹³ JOURNAL OF PARAPSYCHOLOGY, 1937, I, 102-113.

¹⁴ J. PARAPSYCHOL., I, 141.

¹⁵ J. PARAPSYCHOL., I, 191.

techniques of calling and with OM, BM, and STM in matching. The most important point encountered is that the subjects tested do not as a group differ widely on the matching and calling methods used.

There are, however, marked individual differences shown. The largest number of subjects did their best on GESP, but the highest scores were made at OM. Two did their highest scoring at BM, and one at DT.

A check was made of the frequency of hits made for each of the 25 trials in the run in the entire series of DT runs made by Subject No. 1. A curve was plotted to show this frequency distribution of hits throughout the run, and again these were grouped by five trials, making five groups of frequencies. These curves show that there is much more success in calling during the first and last five calls and that this is not due entirely to the first and last cards, which conceivably might be more exposed to sensory cues if careless handling of the testing were supposed. The U-curve shown by this analysis is similar to those found by other investigators for the DT technique.

Within the limits of the range of subjects tested and the variations of the conditions used, the feature that stands out is the large role of individual difference under the different conditions and, for some subjects especially, the large variation occurring from one period of testing to another.

TABLE I.
TOTAL RESULTS FROM ALL SUBJECTS

	Number of Subjects	Trials	Deviation	C.R.	Average per 25
Group I — Minor Subjects.....	56	24,475	148	2.37	5.15
Group II — Major Subjects.....	11	69,400	1,817	17.1	5.65
Total.....	67	93,875	1,965	15.9	5.52

TABLE II.
CALLING VS. MATCHING TESTS

CALLING				MATCHING			
Proc.	Trials	C.R.	Average	Proc.	Trials	C.R.	Average
GESP.....	4,535	6.52	5.97	OM.....	6,875	13.82	6.67
DT.....	8,450	4.54	5.49	BM.....	20,690	7.40	5.51
GESP (l. d.)....	4,750	1.73	5.25	STM.....	13,700	2.24	5.19
Total.....	17,735	5.55	Total.....	41,265	5.60

TABLE III.
AVERAGES OF INDIVIDUAL SUBJECTS FOR VARIOUS TEST PROCEDURES

CALLING										
Subject	GESP		DT		DT Unobs.**		GESP (I.d.)		Total	
1.....	1,250	<i>6.12*</i>	1,150	5.63	6,425	<i>5.87</i>	1,800	<i>5.58</i>	10,625	<i>5.82</i>
2.....	400	5.90	1,475	<i>5.76</i>	1,875	<i>5.80</i>
3.....	350	6.20	3,475	<i>5.41</i>	3,825	<i>5.49</i>
4.....	675	5.19	675	5.19
5.....	635	5.80	375	4.40	1,050	<i>4.74</i>	2,060	<i>4.97</i>
6.....
7.....	100	7.50	200	5.38	300	6.08
8.....	100	5.25	100	5.25
9.....	425	6.00	125	6.80	550	6.19
10.....	200	5.75	200	5.75
11.....	600	<i>6.46</i>	1,650	5.48	1,700	5.20	3,950	<i>5.51</i>

MATCHING										
Subject	OM		BM		BM Unobs.**		STM		Total	
1.....	2,400	<i>6.96</i>	1,400	<i>7.21</i>	4,375	<i>6.06</i>	2,575	<i>6.03</i>	5,575	4.97
2.....	900	5.17	150	4.83
3.....	1,600	4.92	100	5.25
4.....	400	<i>9.12</i>	1,025	<i>7.66</i>	2,000	5.16
5.....	350	<i>13.72</i>	750	5.13	825	5.00
6.....	600	6.00	300	5.42	400	5.63
7.....	200	5.62	800	6.20	100	3.50
8.....	1,600	5.59	500	4.80
9.....	325	3.25	450	6.17	2,475	5.72
10.....	315	<i>7.06</i>	325	5.69
11.....	2,600	<i>5.44</i>	8,575	4.96	1,250	5.36

*Italicized averages are significant.

**Unobserved results not included in averages and totals given in text.

TABLE IV.
COMPARISON OF INDEPENDENT RESULTS

WOODRUFF AND GEORGE			PRESENT EXPERIMENT		
Technique	Trials	Av. per 25	Technique	Trials	Av. per 25
BT.....	6,825	6.92	GESP.....	4,535	5.97
OM.....	10,975	6.97	OM.....	6,875	6.67
BM.....	8,700	5.43	BM.....	20,690	5.51

TABLE V.
PERFORMANCE OF SUBJECT NO. 1 BY STAGES

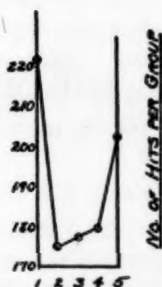
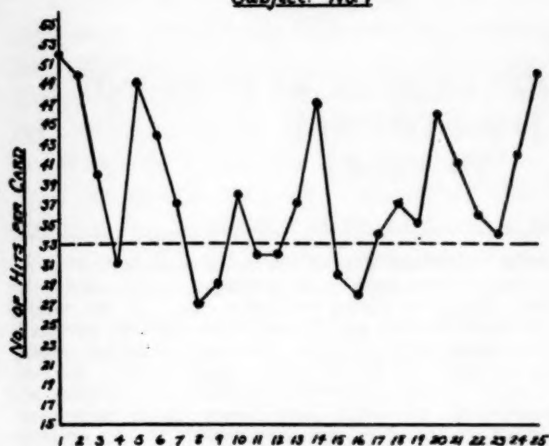
Test Procedure	Period	No. Trials	Average
GESP.....	12-11-34 to 3-17-35	1,250	6.12
DT.....	1-14-35 to 11-7-35	1,150	5.63
BM.....	6-3-35 to 10-15-35	4,375	6.06
BM (Unobserved).....	6-4-35 to 8-19-35	2,575	6.03
DT (Unobserved).....	10-24-35 to 12-20-35	6,425	5.87
GESP (I.d.).....	12-4-35 to 4-15-36	1,800	5.58
OM.....	6-17-36 to 9-2-36	2,400	6.96
OM (Unobserved).....	8-17-36 to 8-19-36	1,400	7.21
STM.....	8-9-36 to 1-3-37	5,575	4.97
Total of all procedures.....	12-11-34 to 1-3-37	26,950	5.88

STUDY OF THE RANK ORDER OF HITS

D.T. TECHNIQUE

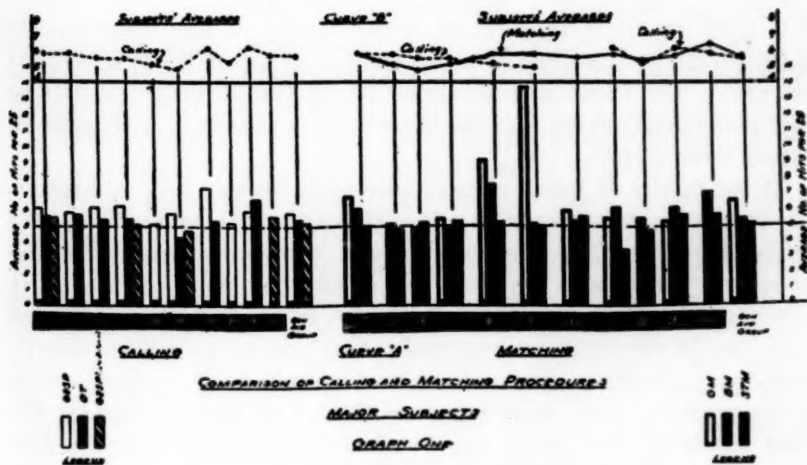
100 TRIALS

Subject: No. 1



ORDER OF CARDS BY RANK
IN GROUPS OF FIVE

ORDER OF CARDS BY RANK



THE QUESTION OF SENSORY CUES AND THE EVIDENCE

J. B. RHINE

ABSTRACT: Experimental conditions are chosen to exclude definitely all possible sensory cues: In clairvoyance tests with packs of 5-suit cards the subjects are not to see or touch the cards during the entire run. All the results from tests made under this narrowly defined set of conditions are then surveyed, in three general groups: I, opaque envelopes; II, screened untouched cards; III, distance tests with walls between. In the 13 series found to meet the conditions the total deviation is large enough to be very significant. The odds are 10^{100} to one that so high a score would not result from chance. Both the hypothesis of sensory cues and that of chance, then, are inapplicable to the explanation of work done under these conditions.

THERE has recently been some little discussion as to whether the results of tests of ESP ability reported are significantly different from what would be expected by chance alone. It is a question only of the sufficient soundness and applicability of the methods of evaluation used. The indications are rather strong that there will shortly be quite general agreement that the hypothesis of chance is inadequate to explain these results.

If the results of tests for extra-sensory perception are not explainable by chance, it is proper to ask next whether the senses have been adequately excluded by the experimental conditions from influencing the scores. Those who have followed closely the published reports are familiar with the now extensive literature in which various advanced conditions in the tests are reported which appear to render sensory perception impossible. Some of the most outspoken critics commenting on the research seem, however, generally not to have considered the better test conditions, and it may be of advantage to assemble the data for the focus of their attention.

Still better justification for this survey is the evident need to take stock, for the benefit of those engaged in the research as well as those otherwise interested in it, of the most conclusive evidence against the hypothesis that the results published are explainable by sensory cues.

Such a treatment may well serve both as a target for helpful criticism and as a register of the state of progress made.

In order to deal effectively and with reasonable brevity with the question of sensory cues it is necessary, first, to limit the discussion to that hypothesis, making all other considerations secondary. Second, it must be clear at the start just what criteria of adequately safeguarding conditions will be followed in assembling the results of the experiments which bear upon the hypothesis.

If the criteria are made high enough and there are still significant results eligible, agreement on the exclusion of sensory cues is practically assured. To lift the question, then, to a level where agreement will be general it is necessary to use clear-cut standards of evidentiality which leave no ground for argument. In choosing the conditions which may be represented as barring out sensory perception, the error, if any, must be on the side of over-caution, even though large masses of valuable data will thereby be dropped from consideration on this immediate issue.¹ The lines drawn will seem unnecessarily exclusive and perhaps highly arbitrary to many experimenters and critics. For naturally all experimenters have been convinced that in their test conditions sensory cues were excluded. As a matter of fact, there has been no demonstrated grounds for doubting that this has been the case. But the question at this point is not, "How much data is there that may be reasonably regarded as extra-sensory?" but rather, "Is there *any* series of tests in which even the most drastic critic can see that the senses could not possibly have been factors?"

The criteria adopted in this survey for the classification of the evidence bearing on the question of sensory cues are as follows:

1. Only the tests of the extra-sensory perception of objects (clairvoyance) will be considered. All pure telepathy and GESP tests (in which both telepathy and clairvoyance are possible) will be rejected, no matter how well the screening may appear to have been or how far the distance between the subjects. This extreme discrimination is made partly to focus the issue and simplify judgment of the case for and against sensory cues. It is done in great part, too, because otherwise there would be difficulty and possible dispute as to what distance is adequate to exclude sensory cues of an auditory nature.

At the same time it leaves out not only the contributions of Estabrooks and those of Brugmanns in which sender and receiver were in different rooms with closed doors, but the work of Sinclair involving

¹ Valuable as such data may be on other questions concerning ESP. This theme will be resumed in a later paragraph.

thirty miles between subjects, that of Usher and Burt at still greater distances, and a considerable range of Duke work in telepathy and GESP conducted with short distances (subjects two rooms apart), the Junaluska series at 250 miles, and considerable work at still greater distances. This wholesale rejection of all tests permitting telepathy is done with full recognition that much of this work could be included without raising any question of sensory cues, but it would be at the expense of simplicity and clarity on the particular problem being now considered.²

2. Test results will be rejected unless the report shows that there was no vision or touching of the cards by the subject between the preparatory cut (following shuffling) by the experimenter and the end of the run of at least 25 calls or trials. The senses of vision and touch are those against which barriers are needed in clairvoyance tests. Since no one knows the card order (if cards are the basis of the tests), auditory cues are not possible.

Thereby is discarded again a huge block of data for which some investigators will properly maintain a stout defense: For instance, the DT tests in which the back of only the top card and the edges of the others are visible (unless careless handling exposes the bottom card to view), but in which the evidence shows that the hits significantly above the mean chance expectation are not confined to the top and bottom cards, as would be expected were sensory cues the explanation. Even those screened tests are left out in which the subjects handle the cards under or behind an opaque screen even though the subject only takes the card by thumb and forefinger by its corner or edge. Also excluded are the screened open matching performances by children reported by L. E. Rhine, the screened open and screened blind matching by Woodruff and George. In the latter the key cards were inverted and unknown; and the cards of the target pack (to be matched against the key cards) were held under the screen by the subject. However improbable it seems, a combination of visual cues from the backs of the key cards and of tactual ones from the cards handled might conceivably be supposed to account for the scoring of the subject.

* Rejection of any work from this narrow classification does not imply doubt of the actual exclusion of sensory cues in the work considered. Rather is it highly probable that sensory cues could have had nothing to do with most of these results under the conditions of testing. It is doubtful if any qualified critic would suppose they did. There *might*, however, be a dispute about some of this work because sensory cues are remotely conceivable. The aim back of these narrow criteria is to make sensory cues under the conditions admittedly inconceivable.

Our criteria are most rigorous in ruling out screen tests of BT in which subjects remove the cards *after the call* is made. No sensory cues can be had after the call, and at most, short of actual trickery, with which we are not concerned here, only an accidental touch of the edge or back could be had of the next (uncalled) card. But we can well afford in this survey to be even unnecessarily narrow and strict. The work rejected by this ruling has many valuable features, and like practically all the work referred to as excluded from this survey, it is statistically significant. Carpenter and Phalen conducted two series of screened BT, 8,750 trials, critical ratio 3.9. Also, Woodruff and George's 6,825 trials with a critical ratio of 5.9 is in this discard. In both instances there was tactual contact although it ordinarily came *after the call*.

So far, then, as tactual and visual sensation go, if the cards are simply not in any way or degree seen or touched by the subject during the run, the role of sensory cues is obviously impossible. We have only to read the reports, select all the data meeting the conditions, and the case for and against sensory cues is determined.

3. Only the results of tests made with the ESP cards (or a modification thereof, leaving the probability of success of each trial $1/5$) will be included. The ground is again largely that of avoiding debatable issues and keeping the question within easy understanding. That is, the results of tests made with drawings, playing cards, and other objects are, although measurable, not so easily judged; and the reports of the experiments made with other than ESP cards are not so readily available to readers of this JOURNAL.

This provision will appear especially ruthless to those investigators who favor other means of testing than that of using ESP cards. It is admittedly ruthless and is laid down here only to restrict the issue to an area of easy settlement of the single question of sensory cues.

4. Once the above-mentioned criteria are defined, *all* the reported test results that meet them are included in the totals under consideration. Selection *within* the defined class and without the guidance of an acceptable principle is statistically improper and unsafe. But the selection and separate treatment of an *a priori* clearly defined class of data *taken as a whole* is not only usual in science but is essential.

The class of data, then, included here and taken in its entirety, regardless of score values, is defined as the score-totals of all of the clairvoyance tests conducted with ESP symbols under conditions in which the subject could have had, according to the reports, no tactual

or visual sensation of the backs, edges, or faces of the cards to be called until after all the calls of the run were made.

* * * * *

The results to be surveyed fall easily into three general groups, marked by differences in test conditions: Those in which the cards were enclosed individually in sealed, opaque envelopes; those in which portable opaque screens were used to bar vision of the cards; and those in which the cards were in a different room or a different building from the subject and screened from vision by a wall between.

In spite of the narrow criteria defining the evidence acceptable, there are thirteen experimental series eligible, based on the work of 118 subjects investigated by ten experimenters, giving the huge total of 142,825 trials. A number of trials many times as large as this has been rejected as failing in some respect to meet the criteria defined above.

I. CARDS SEALED IN OPAQUE ENVELOPES

The condition described by this heading offers the smallest number of reported tests of the three groupings. Only three researches are involved, but there is a total of 25,225 trials for the three series.

1. For the first series of tests with opaque envelopes we draw from a report published in this number, Dr. Pratt's article on Dr. C. Hilton Rice's work. "Cards sealed in heavy manila envelopes were used, sometimes with an observer and sometimes not. Cardboard shields were placed over the symbols to insure against the possibility that the symbols might be seen through the envelope. There was never any evidence that the sealings had been tampered with."³

The following quotation will describe the procedure briefly and give a summary of the results: ". . . Miss T. worked with sealed cards. Sometimes she took the envelopes which Dr. Rice had marked in code and worked alone; at other times she worked as usual in his presence and he checked the results at once from his knowledge of the code. Some of the cards were sealed by himself [Dr. Rice], others in the Parapsychology Laboratory of Duke University. This work extended from September, 1936, until May, 1937. A total of 209 runs were made, giving an average of 5.7."⁴ Dr. Pratt's table shows a critical ratio for this series of 5.2.

In similar fashion Dr. Rice tested his subject, A. J., to the extent of 10 runs, obtaining an average of 5.6 per run.

³ Page 241, this number.

⁴ Page 244, this number.

2. In the other and larger opaque-envelope investigation belonging in this classification, that of Price and Pegram with blind subjects, ". . . each card had a blank card glued to the face of it, so that the symbol was inside, and this was sealed in an opaque brown envelope. Each envelope was taped around the outside with gummed paper to make the sealing secure. These cards were identified by a numeral printed on the back of the envelope in a code unknown to the subjects."⁵ As to procedure, ". . . (a) the subjects were never told what symbol card was in any envelope; (b) the code was never revealed to any subject; (c) the code numbers were so small as to be indistinguishable by any subject; (d) three different packs of sealed cards with three different codes were used."⁶

Price and Pegram⁷ report among other tests the results of 778 runs⁸ with the cards sealed in opaque envelopes given to (partially or totally) blind subjects. They obtained an average of 6.0 hits per run. This gives a critical ratio of 13.1.

In 133 of these runs made by blind subjects the method used was blind matching, BM.⁹ In this the key cards were face down and were arranged in an order unknown to the subjects. The target pack was sealed in opaque envelopes. The average reported was 5.95 and the critical ratio was 5.4.

In 329 runs¹⁰ made by the blind subjects with the opaque envelopes the BT method was used with a 9" x 12" notebook before the pack as a screen. The average in this series is 6.06, critical ratio of 9.4.

For completeness there must be included 300 trials or 12 runs of 25 which were given my subject, A. J. L., in 1932, using sealed opaque envelopes. The results were exactly what would be most expected by chance—an average of 5.0.¹¹

The total of 1,009 runs of both these researches made with cards sealed in opaque envelopes gave an average per run of 5.9 hits. The deviation from the mean chance expectation is 905 and is 14 times the standard deviation.

⁵"Extra-Sensory Perception Among the Blind," *JOURNAL OF PARAPSYCHOLOGY*, I, 145.

⁶*Ibid.*, p. 148.

⁷*Ibid.*, p. 145.

⁸In order to delete the GESP tests from the pooled totals as published, I have had to consult the authors for some of the figures given here.

⁹See glossary.

¹⁰Cf. Footnote No. 8, above.

¹¹Rhine, J. B., *Extra-Sensory Perception*. Boston: Bruce Humphries, 1934. P. 66.

II. TESTS WITH SCREENED CARDS

Under this subheading are three series totalling 91,700 trials or 3,668 runs of 25. These represent the work of 81 subjects and of six different experimenters. The tests were all made with the cards screened from vision by the subject throughout the run and with no tactual contact with the cards.

1. The first series of this group is likewise taken from Pratt's report of Dr. Rice's work, and consists of 10,475 trials in clairvoyant card calling by two subjects made in part under the observation of Dr. Rice himself, partly by Dr. Pratt, and partly (when Dr. Rice was himself subject) by his secretary, Miss Maude Thweatt. All vision and touching of the cards was eliminated, as the account of the conditions given by Pratt will show: "Like procedure (3), ["In this condition the experimenter alone handled the cards. The subject called out where they were to be placed."] with the additional feature of screening the cards from the subject's sight. The experimenter sat behind and close to a folding screen of opaque cloth 6 feet high. One of the twenty-inch panels was directly between the subject and the cards. The experimenter could see the subject, who sat about 10 feet from the screen, by placing his eye at a narrow crack between the two panels."¹²

The total of 419 runs of 25 average 5.8 hits per 25 and give a critical ratio of 8.0.

2. The second series is made up of 21,500 trials given a single subject by Pratt at Columbia. These trials used a special combination of the BM or blind matching and the STM or screened touch matching procedures. Pratt describes the conditions thus: "As the first step in this direction, the experimenter prepared 28 sets of key-symbol cards in advance. Each set of 5 symbols was shuffled, a blank card was used to cover the face of the bottom symbol to prevent its being accidentally seen at any time, and the 6 cards were fastened together by an elastic band. These 28 sets of key cards were used in rotation, one set being re-shuffled and placed back with the others after being turned face up for checking the results of a series of trials.

"A set of receptacles for the place cards was made by fastening five open shallow boxes to a strip of cardboard. These boxes were just large enough to hold the key-cards, but were deep enough for 50 cards. Five blank cards were provided to cover the face-down key cards after they had been placed in the boxes.

"In arranging the key cards, the following routine procedure was strictly observed. The experimenter first placed the row of boxes

¹²Page 240, this number.

on his knees *under* the table. (Due to the fact that the table was covered with a blanket which hung nearly to the floor, the subject could not have seen the boxes in this position if she had tried.) Then the experimenter selected one of the sets of key cards. Holding them face down under the table, he slid the 5 symbol cards one after another into the 5 boxes. Then the blank cards were placed over the key cards while the boxes were still under the table. Having completed this, the experimenter placed the row of key cards under the edge of the screen¹³ before the subject, picked up and cut a deck of shuffled cards, and gave the signal for the subject to begin.

"Her responses consisted in pointing, as previously described, to one after another of the key cards, indicating to the experimenter how to place the cards of the deck to match those before her."¹⁴

The first 7,800 trials gave a critical ratio of 5.3. Then the subject ceased to score above chance in all her work. She made 13,700 trials more at this level and thus reduced her critical ratio to 3.1; still it is significant.

3. In the third series the 66 subjects were either partially or totally blind. Sixty-one were students in a school for the blind. The investigators, Price and Pegram of the Duke Parapsychology Laboratory, say of their conditions: "(1) No person was allowed to stand beside or behind the experimenter. (2) For all GESP and BT work a stiff-backed notebook (9" x 12") was used as a screen behind and close to which the experimenter held the cards so that the cards were invisible to the subjects and spectators while the calls were being made. (3) The subject was kept at least four feet from the experimenter, usually seated at the opposite side of the table. . . . Two further checks on the results were attempted. (1) Two experimenters worked independently with some of the same selected subjects. (2) A considerable part of the work was witnessed by teachers at the school or was observed and checked by a second experimenter."¹⁵

Only screened work with the BT method is accepted for this series, and with this method no subject touched the cards either before or after calling. There were all together 39,275 trials made under SBT (screened BT) conditions by these blind subjects, averaging 5.5 per 25 and giving a critical ratio of 10.6.

Of these, 8,225 trials were made with the additional protection

¹³ "In this procedure a wooden screen, two feet square, stood upright on the table, separating the subject and experimenter. The lower edge of this screen was raised three inches above the table." JOURNAL OF PARAPSYCHOLOGY, I, 12.

¹⁴ *Ibid.*, 14.

¹⁵ *Op. cit.*, 148.

of coded sealed opaque envelopes enclosing the cards (as described above) with the same screening and other safeguards maintained as with open cards. With this summation of blindness (many subjects totally so), screening, and opaque envelopes, the average was in fact higher, 6.06, and the CR 9.5.

4. MacFarland and George combined the precautions of DT and STM techniques in a series of 20,450 trials by 13 subjects which leave the highest average yet encountered in this review. Their conditions are given in considerable detail:

"In the experimental situation, the subject and the experimenter sat on opposite sides of a table with an upright screen between them. The screen which was used was rectangular in shape, 18 inches high and 24 inches long. It had an opening at the bottom which was $2\frac{1}{4}$ inches high and 12 inches long. As an additional precaution, a small shield ($2\frac{1}{2}$ " x 16") was built in front of this opening in such a way that the subject could not possibly see beyond the screen even with eyes brought to the level of the table, and yet both subject and observer were able to see key cards placed in the opening in the screen. Two sets of key cards were used, one normal and the other distorted. Each set consisted of one card of each of the five designs mentioned above. The procedure which was followed consisted of a combination of the STM (screened touch matching) and the DT (down through) techniques. In this procedure the experimenter placed the deck of cards which was to be used face-down on his own side of the table and out of the subject's sight. The key cards were then placed in the opening in the screen. The subject pointed to the key card which he believed matched the first card in the deck, then to the key card which he believed matched the second card in the deck, and so on 'down through' the entire deck of 25 cards. The cards in the deck *remained in position* until all 25 responses were completed. This constituted what we refer to hereafter as a 'run.' As the subject pointed to each key card, the experimenter recorded the response on a data sheet prepared for the purpose. The responses were checked and the score recorded at the end of each run. Two decks of normal cards and two of distorted were used throughout the experiment. Normal decks were alternated with distorted in the order of presentation. When the normal decks were used, the key cards were also normal; when the distorted decks were used, the key cards were distorted. Each deck was carefully shuffled after the run in which it was used. At the beginning of the day's work with each subject the decks were again shuffled, except

in those instances referred to later in this report. Although the screen would seem to minimize the importance of the caution, old decks were always exchanged for new ones whenever the old ones became soiled. . . . In all but a very few cases, at least one other observer was present. There was therefore double witnessing of each check-up, as well as of the behavior of the subject throughout the test."¹⁶

The astonishing critical ratio of 29.2 results from the evaluation of these results. The 20,450 trials averaged 7.1 per run.

The three series just reviewed total 91,700 trials by 81 subjects averaging 5.8 and producing a deviation from mean chance expectation of 3,082, which is 25.0 times the standard deviation.

III. TESTS WITH DISTANCE AND WALLS SEPARATING CARDS AND SUBJECTS

The separation of subject and cards by walls and by distances greater than that which the ordinary across-table ESP test involves is a general condition for the next group of six series of investigations. These were conducted by seven different experimenters on a total of 34 subjects. The six series total 34,225 trials, with an average of 5.7 hits per 25, approximately that of the other two groups of data surveyed. The deviation of the total of the scores from the mean chance expectation of this group of 1,369 runs is + 987, which is 13.1 times the standard deviation.

For lack of a more discriminative principle, that of degree of distance will be used as a basis of order of presentation.

1. The 750 trials reported by Martin, of the University of Colorado, were made with the subject in the room adjoining that in which the experimenter handled the cards. These 30 runs, which averaged 7.8 hits per 25 and give a critical ratio of 7.5, were conducted under the following conditions: "The experimenter shuffled the cards thoroughly, cut them and placed them on the table before her. Exceeding care was taken that the bottom card was visible to no one. The subject then recorded twenty-five guesses on a record blank. The experimenter then read the actual order of the cards to the subject who recorded them. The experimenter carefully watched the recording of each card and was often checked by a third person."¹⁷

"*Series 3.* This series of 30 runs or 750 trials was conducted with the experimenter (the author), seated at a table in one room, and the subject seated at a table in an adjoining room. The subject recorded

¹⁶ *Op. cit.*, 94-95.

¹⁷ *Ibid.*, 186.

her guesses after receiving a 'start' signal from the experimenter. A door was ajar between the rooms to permit of signals. The order of the pack was recorded by the subject in the presence of the experimenter."¹⁸

2. A short series of 61 runs or 1,525 trials using DT and BT procedures was conducted by Stuart with H. P. as subject, and a report was included (without much detail because the results did not warrant it) in my semi-popular volume, *Extra-Sensory Perception*. An average of 5.3 was obtained with the non-significant ratio of 1.1.

The distances were from 8 to 30 feet with either one or two walls between cards and subject. The pack was cut by the experimenter before the test and the recording was done by him also, although always in the presence of the subject and commonly with the latter turning the cards over (always close under the experimenter's eye, however).

3. With the help of Miss Ella Phillips Crandall as a second observer, I conducted with the subject E. J. G. 3,725 trials in clairvoyance tests (both DT and BT) at short distances, obtaining an average of 5.7 and a critical ratio of 4.4. I quote from my report as to the conditions: "The calls were recorded by an observer as made, and either at the end of the run of 25 calls or after several packs were called, the check-up would be made. Both subject and observer would check both card and call together, both observing both. No opportunity for error or deception was allowed. During most of the tests, including the best, more than one observer was present. . . . But a still further advance in conditions consists in the separation of subject and cards by distance and a wall. This was done first with a short distance of about 12 to 15 feet, and with one wall, cutting off vision; and second, with 30 feet and two walls. An open door permitted communication, but in some experiments the doors were closed; signalling with a telegraph key was used for the BT condition to indicate when the card had been removed and the call recorded."¹⁹

4. The report by Warner in this issue comes well within the scope of our criteria. It is hardly necessary to quote at length from his account of conditions since it is easily available. The distance, the double observation and double checking, and other features of the research make it, as its author entitles his report, "A Test Case." His average per 25 trials of 9.3 hits gives a critical ratio of 6.8. This is the highest average of the series here surveyed.

5. The series next in turn has become well known already as the Pearce-Pratt series. It consists of 1,850 trials by H. P. with Pratt

¹⁸ *Ibid.*, 188.

¹⁹ *Character and Personality*, III, 95-96.

as main experimenter and with myself as a second observer in 150 trials. The general average was 7.5. (During the double observation series it was 9.3.) The critical ratio is 10.7.²⁰ The conditions are given: "The observer and subject synchronized their watches, and arranged to work at a stated time and distance. At the specified time the observer would take the top card from a shuffled pack of ESP cards in the room agreed on and lay it face down on a book in the center of the table without looking at its face. Thirty seconds later the subject in his cubicle in the Duke Library would record a call for the card. At the end of the minute, the observer would remove the card and take the next one. The cards as removed would be kept in order for later recording. Two runs were made per day.

"Groups A, C, and D were made with the observer in the Physics Building and the subject in the Library. The distance was about 100 yards. In group B, the observer was farther away with the cards, approximately 250 yards. The subject was in the same place in the library as in A, C, and D. . . .

"In Groups A-D, the records were sealed up after each sitting and delivered to me before subject and observer got together. In Group D, I was present with Dr. Pratt as a second observer.

"The cards were shuffled between runs, and this shuffling occurred just before using, with H.P. absent from the room. . . . Two packs were used at a sitting."²¹

6. The last and largest series of this group has also the largest distance as a condition. MacFarland, at Tarkio College, has through the medium of the mails supervised 26,125 trials by 30 subjects with distances extending up to 1,400 miles separating subjects from the cards. Reference to the table given in my preliminary report of this work²² will show that the highest average (5.64) was at a distance of 100 to 300 miles. The general average was 5.5 and the critical ratio 8.3. The following paragraph gives further information: "The tests were made with the DT technique, the cards being kept intact in packs by the experimenter throughout the test period and removed only when checking up. The subjects filled out five columns of a record sheet, one for each pack in the experimenter's desk, at any time they wished on a given day, and sent the sheet to the experimenter to be checked. Double checking was carried out, and the general sponsorship of Dr.

²⁰ This series was used as the basis of the examination of methods of evaluation made by Greenwood and Stuart (*JOURNAL OF PARAPSYCHOLOGY*, I, 206). By consulting this as well as the original report further information may be had.

²¹ *JOURNAL OF PARAPSYCHOLOGY*, I, 75. ²² *Ibid.*, 181.

R. W. George, head of the Department of Psychology, was exercised over this series. The cards were well shuffled and kept under careful observation by the experimenter." 23

The total of the individual trials in Groups I, II, and III, with envelopes, screens, and distance, respectively, is 142,825. The average per run of these 5,713 runs is 5.8. The positive deviation of 4,624 is 30.0 times the standard deviation. The table presents the figures of all series in summary.

TABLE I.
CLAIRVOYANCE TESTS MEETING SPECIAL CRITERIA FOR EXCLUSION OF SENSORY CUES

I. CARDS SEALED IN OPAQUE ENVELOPES							
Experimenters	No. Sub-jects	Av. per 25	Trials	Runs	Deviation	S. D.††	C. R.
P & P.....	66	6.0	19,450	778	746	56.9	13.1
CHR.....	2	5.7	5,475	219	159	30.2	5.3
JBR.....	1	5.0	300	12	0	7.1	0
Total.....		5.9	25,225	1,009	905	64.8	14.0
Total sealed less screened sealed ‡ ...		5.8	16,900	676	555	53.0	10.5
II. TESTS WITH SCREENED CARDS							
CHR & JGP.....	2	5.8	10,475	419	333	41.8	8.0
MacF & G.....	13	7.1	20,450	818	1,704	58.3	29.2
JGP.....	1	5.2	21,500	860†	188	59.8	3.1
P & P.....	66	5.5	39,275	1,571	857	80.9	10.6
Total.....	81	5.8	91,700	3,668	3,082	123.5	25.0
III. TESTS WITH DISTANCE AND WALLS SEPARATING CARDS AND SUBJECTS							
LW.....	1	9.3	250	10†	43	6.3	6.8
DLM.....	1	7.8	750	30	84	11.2	7.5
CES.....	1	5.3	1,525	61	17	15.9	1.1
JGP & JBR.....	1	7.5	1,850	74	187	17.5	10.7
JBR.....	1	5.7	3,725	149	110	24.9	4.4
MacF & G.....	30	5.5	26,125	1,045	546	65.9	8.3
Total.....	34**	5.7	34,225	1,369	987	75.5	13.1
GRAND TOTAL.....	118**	5.8	142,825*	5,713	4,624	±154.2	30.0

* There are 8,325 trials appearing twice in this table. These are, however, not duplicated in this total. They averaged 6.06 hits per 25.

** These totals do not check with numbers of subjects given because of duplications in some series.

† Equivalent to this number of runs of 25, though actually not so broken up in the experiment.

†† Corrected for Matching Hypothesis; i. e., about 2 per cent larger than for binomial.

‡ Repeated in Group II below.

DISCUSSION

The aim of this survey was to assemble the results of all tests in which there were exceptional assurances that sensory cues were com-

■ *Ibid.*, 180.

pletely barred by the conditions—barred so effectively that there can be no dispute of the fact. And to this end, many borderline series were rejected for which a case probably convincing to most scientific readers might be made.

The criteria were made so exclusive that the 142,825 trials are not to be suspected of explanation by the hypothesis of sensory cues. Nor are the scores any more readily explainable by chance. The critical ratio of 30.0 represents (if we use the normal probability integral tables as an approximately correct basis of estimate) a probability of only one in 10^{195} that so large a total deviation (as 4,624) would occur by chance alone.

Something has evidently operated in an extra-chance manner in this series of tests reviewed, and under the conditions reported it could not have been a sensory process. But the critical reader must, of course, consider other alternatives as well. Since this treatment is limited to the sensory cue question, it is necessary to be very brief in considering further steps in the weighing of the evidence. Perhaps the most important alternative left for judgment is that of the general reliability and competence of the investigators. For it would appear that if the reports are to be taken with full reliance—even if one of them be so regarded—there is hardly any alternative to the hypothesis that some extra-sensory mode of perception is occurring.

But there is no statistical method for handling reliability of investigators since there is nothing comparable to "mortality tables" for predicting incompetent or unscrupulous experimenters. Nor do we know how to determine how many experimental confirmations of ESP are required for "significance" of investigators' reliability. The next step, then, with these uncertain and non-measurable factors, must therefore be left to the general judgment and inclination of the reader. They are inherent in all research but become prominent when results do not fit in with pre-existent views.

The question of possible dishonesty or of unintentional error on the part of the subjects themselves does not arise when the cards are adequately protected from the senses. The subjects' active participation in the checking was also eliminated in the majority of the series of this survey.²⁴

The limited goal of this survey has, I believe, fully justified the discrimination made between data which does and that which does not meet the criteria laid down for the sure exclusion of sensory cues. That

²⁴ Passive witnessing by the subject is advantageous, but the actual checking and scoring should be wholly in the experimenter's hands.

the criteria are unfairly high for a just decision on some of the rejected data must be clear, but the goal was that of obtaining an undebatable set of criteria rather than a fair one.

With another objective in view, the rating given to various series of test results in this survey would naturally be very different. In fact, some of the best work from the point of view of bearing on the nature of ESP is not quite eligible to this group. A considerable and important series of comparative tests bearing on the character of ESP have been set aside, comparisons (1) of different size and number of stimuli, by L. E. Rhine; (2) of various methods, by Woodruff and George and by Gibson; (3) of screened cards versus unscreened, by Woodruff and George; (4) of intelligence relationships, by Bond; (5) of various drug effects, by myself; (6) of color versus form stimuli, by Carpenter and Phalen; (7) of high-aim versus low-aim calling, by Pegram; (8) of groups versus individual tests, by Sharp and Clark; (9) of telepathy with clairvoyance, in my earlier work; and indeed many other contributions to the growing knowledge of what ESP is in relation to other more familiar processes.

It must be granted, too, that research in telepathy and GESP has not been given fair representation here. Furthermore, it has in general been neglected in late years in actual research for the more simply controlled tests of clairvoyance. But the problems involved are too important to permit this phase of ESP to continue uninvestigated. Telepathy tests can of course be controlled for sensory cues, and I am confident such cues have been excluded in a great part of the research on that branch. The ground for needing to exclude telepathy research in clarifying the issue in this paper does not, fortunately, apply to the research itself.

The standards enforced in this survey are not to be adhered to strictly in judging that work in which the primary aim has been to inquire into the nature of ESP. One may accept with relative confidence, all the more so against the background of the results surveyed here, the exclusion of sensory cues in experiments such as the screened matching tests (with carefully examined cards), screened card calling without the subject touching the cards until after the call, GESP with subjects in separate rooms, and many other conditions in which significantly high total deviations have been found by capable experimenters.

However, it will doubtless follow that in future research in ESP, both in telepathy and in clairvoyance, better safeguarding against "remotely conceivable" sensory cues will be maintained. The means and

methods are more easily available today than ever before. It may appear unusual that all the precautionary conditions now part of the routine were not applied from the start. But in all research the obvious of today was obscure but yesterday.

The primary advance in ESP research, however, has consisted not simply in laying down heavier barriers to sensory cues, but in the development of such ways and means of approaching subjects with a view to measuring the ability in question that they may perform under conditions favorable at the same time both to the ability itself and to the avoidance of experimental weakness. It has been this double objective of the research which has made its progress difficult and slow. But its severest critic must admit that in spite of difficulties it has made some progress.

SUMMARY

In order to remove the question of the adequate exclusion of sensory cues in the ESP research from the range of all dispute, there was made a survey of all results which meet the following very exclusive criteria, designed to admit only data from tests made under conditions in which, as described, sensory cues are undeniably impossible: (1) Clairvoyance tests made with (2) packs of cards with five suits in which the cards were both (3) invisible to and (4) untouched by the subject (5) throughout the run.

There were found 13 series of trials totalling 142,825, made by 118 subjects, with 10 experimenters, and averaging for the 5,713 runs 5.8 hits per 25, giving a critical ratio of 30.0. The odds against so high an average for this length of series resulting by chance alone would be about 10^{195} to one.

It appears logical to conclude that sensory cues were not responsible for the fact that the subjects as a group scored significantly above the mean chance expectation in the ESP tests falling within the scope of this survey.

A RATING TABLE FOR CARD-MATCHING EXPERIMENTS*

EDWARD V. HUNTINGTON

ABSTRACT: If a subject restricts his calls so that he names just five of each symbol when a deck of cards is being called, then the probability of a particular average score for a number of runs may be calculated very exactly. A formula which gives these probabilities and tables of probabilities derived from it are presented here. A rating table for average scores is given.

Dr. Huntington is Professor of Mechanics in the Department of Mathematics of Harvard University.

WITH a deck of 25 cards composed of 5 suits of 5 cards each, the number of possible "runs" is known to be $(25!)/(5!)^5$. Calling this number T , we have $T = 623,360,743,125,120$. If each of these runs is "matched" against any fixed run, taken from a duplicate pack, the number of "hits" found is called the "score" for that run. The score may have any value from 0 to 25, inclusive (except 24).

The mean of the distribution of the scores has long been known to be $M = 5$. The second, third, and fourth moments about the mean (as recently computed by T. E. Sterne in *Science* for November 26, 1937) are $\mu_2 = 25/6$, $\mu_3 = 125/46$, $\mu_4 = 3625/69$.¹ The standard deviation is therefore $\sigma = \sqrt{\mu_2} = 2.041,241,45 +$. Hence, if we define skewness $= \gamma = \mu_3/\sigma^3 [= \sqrt{\beta_1}]$ and kurtosis $= \delta = (\mu_4/\sigma^4) - 3$, we have $\gamma = 0.319,498,66$ and $\delta = 0.026,086,96$. These values of M , σ , γ , and δ for the basic distribution are exact.

Now consider all possible samples of n drawn from the parent population of T . By known formulas, which are exact for an infinite distribution of any shape, the mean, the standard deviation, skewness,

* Presented to the American Mathematical Society, December 30, 1937, at its meeting in Indianapolis. A longer account of the subject, in non-technical language, will appear in the spring number of *The American Scholar*.

¹ A general formula has been worked out independently by E. G. Olds; see abstract No. 428, *Bull. Amer. Math. Soc.*, vol. 43, p. 779, Nov., 1937.

and kurtosis for the distribution of the means of the samples of n are given by $M_n = M$, $\sigma_n = \frac{\sigma}{\sqrt{n}}$, $\gamma_n = \frac{\gamma}{\sqrt{n}}$, $\delta_n = \frac{\delta}{n}$

Let S_n = the average score in a sequence of n runs (that is, the mean of a sample of n). To obtain a good approximation to the distribution of S_n , we may let $t_n = (S_n - 5)/\sigma_n$ and fit a Charlier curve having the four constants just given. The area under this curve from any value of S_n out to infinity will indicate the relative frequency of average scores of S_n or better. This area is given by

$$\int_t^{\infty} F(t)dt = \int_t^{\infty} \phi(t)dt + \frac{\gamma}{6} \phi''(t) - \frac{\delta}{24} \phi'''(t).$$

Remembering to start each area at the beginning of the panel of which S_n is the midpoint, these areas can be readily computed by the aid of the tables for $\phi(t)$, $\phi''(t)$, etc., given by Glover or Rietz.

In the following tables, the fraction P opposite any average score S shows the proportion of cases in a complete cycle of T_n events in which an average score of S or better will occur.²

n = 4		n = 8		n = 16		n = 100	
S	P	S	P	S	P	S	P
5.00	.538	5.00	.527	5.00	.519	5.00	.508
5.50	.345	5.50	.268	5.25	.330	5.10	.319
6.00	.193	6.00	.099	5.50	.179	5.20	.170
6.50	.092	6.50	.026	5.75	.081	5.30	.075
7.00	.038	6.75	.012	6.00	.031	5.40	.027
7.50	.013	7.00	.005	6.18	.013	5.46	.014
8.00	.004	7.25	.002	6.25	.010	5.50	.008
8.25	.002	7.50	.001	6.50	.003	5.60	.002
8.50	.001	7.87	.0001	6.56	.002	5.70	.0004
9.00	.0002	6.75	.006	5.77	.0001
9.25	.0001	7.00	.0001

From these tables, and similar ones for $n=12$, $n=20$, $n=50$, the following "rating table" can be written down. (This table corresponds to the rating table written by Rhine in the instructions accompanying the ESP cards. The first row and the last column are added by the present writer for convenience of reference.)

* The author is indebted to Dr. J. A. Greenwood for assistance in verifying the entries in these tables.

RATING TABLE FOR AVERAGE SCORES OBTAINED IN N RUNS

Number of runs: n	Average score obtained in n runs		
	Good (1 in 75)	Very good (1 in 500)	Extremely good (1 in 10,000)
1.....	10.00	12.00	14.00
4.....	7.50	8.25	9.25
8.....	6.75	7.25	7.87
12.....	6.41	6.83	7.33
16.....	6.18	6.56	7.00
20.....	6.05	6.40	6.80
50.....	5.66	5.86	6.10
100.....	5.46	5.60	5.77

In this table, "good" corresponds to relative frequency of about .013, "very good" to about .002, and "extremely good" to about .0001. In other words "good" means about 1 in 75, "very good" about 1 in 500, and "extremely good" about 1 in 10,000.

The differences between this table and the rating table given on the ESP cards are too small to be of any practical importance, provided the experiments are of the "pure matching" type. (In the "free calling" type the standard deviation is known to be less than in the pure matching case, so that the table may be used with some confidence even in the free calling case, since the error will be in general on the safe side.)

A REVIEW OF CRITICISMS OF THE MATHEMATICAL EVALUATION OF *ESP* DATA

C. E. STUART and J. A. GREENWOOD

ABSTRACT: Reports of ESP test results usually evaluate the data mathematically by giving a deviation from mean chance expectation, a critical ratio, and a statement of probability relating to the critical ratio. Critics have suggested errors both in the theory and in the application of these measures. It is shown herein that the deviation from mean chance expectation is a valid measure both theoretically and practically. The accuracy of the critical ratio depends upon the closeness of approximation of the theory used to the results of the tests. The theory in use, the Binomial hypothesis, and a proposed theory, the Matching hypothesis, are both good approximations; and experimental use of either may be a matter of choice. Probability statements previously reported by investigators of ESP are found to involve slight underestimations of the true probability. Calculation of the error shows it to be negligible except for small numbers of runs.

Mr. Stuart is a member of the staff of the Parapsychology Laboratory and Dr. Greenwood is Instructor in Mathematics at Duke University.

IN A RECENT article (2) the writers discussed the techniques available for the evaluation of data in tests for extra-sensory perception. The limits of such a project did not allow full attention to an equally important consideration, namely, the validity of certain criticisms directed at the techniques used in already published reports.

These criticisms have appeared in several articles by various writers. They have been proposed frequently in correspondence. No critical point has been the suggestion of a single individual. Although reference can properly be made to published sources only, it can be generally assumed that the critic mentioned is but one of a number. This understanding justifies our grouping the critical points in some general fashion. The possible unfairness of separating a point from its background of comment is compensated in large part by the fact that these criticisms do fall into particular classes regardless of the arguments from which they arise.

Let us state specifically the standpoint of this discussion. It is granted that the logical examination of assumptions within the theory

of probability is of concern to any research which uses statistical methods. It is recognized that the conclusions of any experiment using those methods are based upon the type of method used. But the judgment of the experimenter must determine his choice of method, and this judgment will be based upon the type of measurement he wishes to make. The usefulness as well as the accuracy of a method must be considered. The criterion of such a judgment cannot be stated with any objective exactness. We shall attempt, however, to evaluate critical comment with respect to two possibilities: (1) its statistical or mathematical validity, and (2) whether, in our opinion, its effect upon conclusions might be of such a magnitude that an experimenter should adjust either his experiment or his evaluative method to correct for it.

The Mean Chance Expectation:

The statistic measured in card tests of extra-sensory perception is a "deviation from mean chance expectation." Upon the assumption that the probability of guessing each card as it is presented is $\frac{1}{5}$, the mean chance expectation is $\frac{1}{5}$ the number of guesses made. The first point of critical attack has been to question whether the observed deviation has any meaning at all. If the mean chance expectation is not the value assumed, then the differences themselves are in doubt. The challenge may be stated: *The probability of a chance success on a given card is not 1/5; therefore the deviations are not true measures of variation from chance.*

This challenge has been countered by proof that the probability of chance success on a given card is $1/5$. Wright (9) has demonstrated this in a short note in answer to a statement by Weisberg (6). The writers have given a demonstration and proof in their previous article (2, p. 209 ff.). The particular criticism must be rated *irrelevant*.

The reasoning upon which this criticism has been based is not easily recognized as fallacious even after it has been demonstrated that it leads to a false conclusion. The following remarks by Mr. Louis Weisberg illustrate clearly the plausible nature of the argument.

"When a single card is drawn at random from the pack of 25, the chance of its belonging to a specified kind is 1 in 5. Professor Rhine apparently assumes that if a second card, or a third card, or any number of cards, is drawn from the same deck, the probability for each card is 1 in 5. But this is clearly not so. When the second card is drawn, there remain 24 cards in the deck, any one of which is equally likely to be drawn. Of these 24, 4 are the same kind as the first one drawn and there are 5 each of the other 4 kinds. The chance that the second card

drawn is the same kind as the first is 4 in 24, while the chance that the second card is any specified one of the other 4 kinds is 5 in 24. As each successive card is *drawn*, the chances vary according to a pattern that is not difficult to follow." (6)

We have italicized the word *drawn* to indicate an easily overlooked point of inadequate description. Whenever a card is known, the probabilities of occurrences in the rest of the deck may be stated (or "changed") with respect to the known card. But when the drawn card is not known, the pattern of varying conditional chances (this pattern was given in our previous article, p. 210) lead directly back to a statement of $1/5$ probability on each draw.

The contention that mean chance expectation is not $1/5$ is of historical interest only. All its proponents have withdrawn the criticism upon further discussion.

The Standard Deviation:

The second point of critical attack has been directed to the statistical evaluation of the difference between the observed number of successes and the expected number. Rarely in an experiment is the exact expected number of hits found. These chance variations from the average expected number are themselves averaged in the expected *variance*. The customary method in ESP research has been to assume that the theoretical expected variance was that of Binomial theory. The observed deviation was measured by dividing it by the standard deviation (the square root of the variance), the quotient being a critical ratio. The criticisms aimed at this procedure may be summarized to state: *The critical ratio is not accurate because the standard deviation used is not properly applicable. The correct standard deviation is probably much larger than the one used.*

Typical of this challenge is Kellogg's statement: ". . . So while the average for all series may be about $1/5$ the deviations will be greater than would be the case in truly random sampling . . . All these factors work in the same direction, to increase variability over that of simple sampling." (3, p. 182).

Since the true variance of the correspondences between a series of subject's calls and a series of shuffled cards is a function of the various orders which both series may take, the true variance will change somewhat with any limitation placed upon *both* series. In the case of experiments with ESP cards the card series is commonly limited by being made up of 25 symbols equally divided into five suits. There is no such limitation on the subject's calls. The question whether a

subject's calls can be truly random, i.e., each call independent of every other, is not one which can be answered with any finality. If the calls may be considered random, then the theoretical variance of the binomial is exactly applicable, and the criticisms would be irrelevant. But the critic may point out justly that the calls *may* not be random, that indeed the subject may call any systematic order he pleases.

As we pointed out in our previous article, the systematic order which would produce the greatest variance is that in which the subject's calls are limited to five calls of each symbol. With this extreme limitation the standard deviation of a run score distribution is 2.041. The standard deviation based upon binomial theory is 2.000. (The restricted order we shall hereafter refer to as the "Matching hypothesis," which means that it is equivalent to matching one deck of ESP cards against the shuffled order of another deck.) Any other call distribution, wherein more of one symbol is called than another, decreases this standard deviation. As such variation is the rule rather than exceptional, it is safe to state that the properly applicable standard deviation lies between 2.000 and 2.041.¹

The criticism being considered is, therefore, valid, but as critical ratios in ESP research are usually so high that such a small change in their absolute value would change their significance not at all, it must be rated *relevant* but not important. The experimenter must choose whether he will use the Binomial hypothesis, which *may* slightly overvalue his deviations, or the Matching hypothesis, which will slightly undervalue his deviations. For any but borderline critical ratios the effect upon experimental conclusions is, of course, trivial.

Skewness:

A third point of critical attack is directed to the third statistic used in reports of ESP research. Given a deviation from mean chance expectation, and a critical ratio based upon a standard deviation, a statement of the meaning of such a critical ratio in terms of a fraction of probability is usually reported. This probability is the ratio of the area under the Normal curve to the right of an ordinate equal to the observed critical ratio, to the total area under the curve. Such an evaluation assumes that the deviations will be normally distributed about their expected means. The challenges of this method may be summarized: *The expected distribution of scores is not normal; and there-*

¹For example, the true theoretical standard deviation for the distance series discussed in our previous paper is equivalent to 2.029 for each run. A formal discussion of this point will be presented by one of the present writers in a future issue of this JOURNAL.

fore the areas under the normal curve will not give true probability values. The distribution must be evaluated by finding the areas under the distribution curve which best fits the data.

This argument is typified by Willoughby's statement: "Since, in the case under discussion, p is $1/5$ and q is $4/5$ this distribution is necessarily skewed, with the longer tail toward the right . . . It seems clear that normal probability tables are not applicable to the evaluation of differences between values pertaining to skewed curves, and that, since the curves are theoretically and necessarily skewed, the multiplication of cases will not remedy the situation." (8, p. 79).

This criticism involves a number of considerations. (a) What is the best theoretical curve to use to approximate the run scores of the experiments? (b) What is the nature and magnitude of the difference between these curves and the normal curve? (c) Is the difference constant or if not, how does it change when run scores are added to give a total score?

The first question is not easy to answer definitely. Willoughby (7) suggests the Binomial series, and Kellogg (4) insists upon the Matching hypothesis. Since we may consider both of these valid approximations we may limit the discussion to answering the other questions in respect to these hypotheses.

The differences between these hypotheses and the normal curve may be stated in terms of constants based upon the moments of distribution. These constants, for a distribution of scores of runs of 25 cards, are presented in Table I.

TABLE I.
CONSTANTS OF DISTRIBUTION CURVES OF VARIOUS HYPOTHESES

	Mean m	Standard Deviation σ	Skewness γ	Kurtosis ² δ
Matching Hypothesis	5	2.04	.319	.026
Binomial Hypothesis	5	2.00	.300	.010
Normal Hypothesis	5	(2.00 or 2.04)	.000	.000

While the display of these constants answers specifically the second question, the differences are neither small enough nor large enough to

²In Pearson's notation $\gamma = \sqrt{\beta_1}$ and $\delta = \beta_2 - 3$. We are indebted to Prof. E. V. Huntington for these constants in the matching hypothesis. The others may be easily computed from a table of curve constants given by Thornton C. Fry (1).

present without discussion. How would they affect the probabilities reported from the data of an experiment?

When distribution of run scores in ESP tests are reported, they are evaluated by a Chi-square fit to the Binomial hypothesis. If the Matching hypothesis is the valid test, however, the probability would be erroneous to the degree that the Binomial is a bad fit to the Matching case. A recent paper by T. E. Sterne (5) gives the results of a consideration of this exact point.³ Sterne's distributions are presented in Table II. Some judgment of the magnitude of the error is possible from a study of this table. It will be noted that the hypotheses for 100 runs would be practically identical, and that they differ in only a few units in the thousandths place. Sterne finds the fit good, but offers the significant summary: "To distinguish observationally between the true values and the Binomial values . . . would require some 30,000 trials [runs]." As the number of runs reported in any given ESP experiment is usually only a small fraction of this quantity, it may be safely concluded that for practical experimental evaluations the Binomial and Matching hypotheses are interchangeable.

But data in ESP research is customarily evaluated not in terms of the distribution of hit frequencies per run but in terms of total scores

TABLE II.
COMPARISON OF MATCHING AND BINOMIAL HIT FREQUENCY EXPECTATION

Hit Frequency	Matching	Binomial
0.....	.0043	.0038
1.....	.0247	.0236
2.....	.0730	.0708
3.....	.1374	.1358
4.....	.1852	.1867
5.....	.1919	.1960
6.....	.1597	.1634
7.....	.1096	.1108
8.....	.0633	.0624
9.....	.0311	.0294
10.....	.0131	.0118
11.....	.0047	.0040
12.....	.0014	.0012
13.....	.0004	.0003
14.....	.0001	.0001
15.....	.0000	.0000

³Dr. Sterne approximates the values for the Matching hypothesis by fitting a Type I curve to the exact moments. In correspondence, Dr. T. N. E. Greville has given us the exact values. The two distributions agree to the degree of accuracy claimed by Sterne. The incomplete table presented by Dr. C. E. Kellogg (4, p. 336) contains errors in the third place, and should not be used.

found by adding the hits of a number of runs. The important area of the critical attack under consideration is then the third question proposed above. The view expressed by Willoughby is shared by Kellogg who says "When several runs are combined to give a total score, the discrepancy [between hypotheses] increases the longer the number of runs in question."

That this statement is not true can be readily demonstrated. A total score may be expressed as an average score per 25 calls by dividing the total score by the number of runs. The mean chance expectation of this value is seen from Table I to be 5 for all hypotheses. The distribution of this mean value is not that of the original observations, but is a function of the number of runs. The constants for the distribution of the mean are :

$$\text{Standard Deviation} = \frac{\sigma}{\sqrt{N}} ; \text{Skewness} = \frac{\gamma}{\sqrt{N}} ; \text{Kurtosis} = \frac{\delta}{N}$$

TABLE III.
DIFFERENCE BETWEEN THE VARIOUS DISTRIBUTIONS

Difference between:	m	σ	γ	δ
Matching and Binomial.....	0	.04	.319	.016
Matching and Normal.....	0	(.00)	.319	.026
Binomial and Normal.....	0	(.00)	.300	.010

The actual discrepancies between the three hypotheses are recorded in Table III. The discrepancies between the constants of the mean of a number of runs may be found by dividing each of these values by \sqrt{N} or N. (For 100 runs the discrepancies between the constants of the Matching and Binomial means are reduced to .004, .0319, and .00016 respectively.) The relative difference between the standard deviations of the Matching and Binomial does not increase with the number of runs but remains at a constant two per cent. The other statistics obviously decrease in difference as the number of runs increase.

There remains, however, an important part of the criticism unanswered, namely, what is the magnitude of the error introduced by using the normal curve tables as an expression of the area under the above slightly skewed curves? It has been shown that these errors may be made as small as we please by taking a sufficiently large number of runs. But it may validly be asked how large the errors are for small numbers of runs.

We have shown that for ordinarily reported numbers of runs the Matching hypothesis and the Binomial are practically interchangeable. To find how much the Normal approximation underestimates the Binomial, it is necessary to fit a Gram-Charlier approximation to the Binomial series.⁴ The formula for this case states the Binomial series in terms of the Normal Probability Integral and its derivatives.

Exact statements of the probabilities involve two corrections. The first is an adjustment for the fact that, while the normal curve is continuous, hit scores can have only integral values. The best approximation for this correction is to subtract .5 from the observed deviation. The second correction is that of the actual difference in area under the normal and binomial curves. Table IV gives the value read directly from the Probability Integral Tables, this value corrected for continuity, the exact probability of such a critical ratio upon a Binomial hypothesis, and the exact probability of such a critical ratio upon a Matching hypothesis.⁵ The last two columns list the errors of understatement when the Normal curve tables are used without correction as statements of the probability of Binomial and Matching hypotheses.

The error of understatement is greatest for small numbers of runs with a small critical ratio. With a given critical ratio the relative error decreases with increase in number of runs. The absolute value of the errors is so small that in the main experimental conclusions should not be affected. But the errors are important in the range close to the criterion of significance. If one sets as a criterion the probability value of 1/150 the table shows that for the customarily used "significant" critical ratio of 2.5 the probability criterion is attained only when the critical ratio is that of a large number of runs. This point of criticism is, therefore, important; and the experimenter should adjust his methods to compensate for it.

The 1/150 criterion may be met by a slight change in the point of significance of critical ratios. For scores based upon less than 25 runs the criterion should be increased to 2.80; for scores of more than 25 runs a critical ratio of 2.65 meets the probability criterion.

The experimenter should further choose whether he will report probability values as those of the normal approximation or those computed exactly for the Binomial. The difficulty of computation for large critical ratios and the fact that the errors are too small to affect an

⁴ See T. C. Fry *Probability and its Engineering Uses*, p. 255 ff. for the statement and development of this formula.

⁵ Values for the Matching Hypothesis were calculated by the formula presented by E. V. Huntington in this number of the JOURNAL, p. 293.

TABLE IV.

COMPARISON OF PROBABILITY STATEMENTS BASED UPON THE NORMAL CURVE WITH THEIR CORRECTED VALUES

	Normal	Normal with Correction for Continuity	Binomial	Matching	Error B-N	Error M-N
X = 2.5						
4 runs.....	.0062	.0088	.0114	.0131	.0052	.0069
25 runs.....	.0062	.0071	.0081	.0093	.0019	.0031
100 runs.....	.0062	.0067	.0075	.0082	.0013	.0020
X = 3.0						
4 runs.....	.0014	.0020	.0034	.0038	.0020	.0024
25 runs.....	.0014	.0016	.0020	.0024	.0006	.0010
100 runs.....	.0014	.0015	.0018	.0020	.0004	.0006
X = 4.0						
4 runs.....	.000032	.000053	.000148	.000180	.000116	.000158
25 runs.....	.000032	.000039	.000063	.000083	.000031	.000051
100 runs.....	.000032	.000035	.000053	.000065	.000021	.000033
X = 10.0						
100 runs.....	5.9x10 ⁻²⁴	7.6x10 ⁻²⁴	14.8x10 ⁻²⁴	443.9x10 ⁻²⁴	8.9x10 ⁻²⁴	438x10 ⁻²⁴

experimental conclusion will probably weigh against use of the exact values.⁶

The third point of critical attack has led to a relevant contribution, the Matching hypothesis. Any experimenter displaying his data in a distribution of hit-frequencies per run should choose whether he will use the Matching or the Binomial hypothesis as an approximation to the expected chance distribution. It has led to an important re-interpretation of the probability statements for small critical ratios. Although their development is interesting as a mathematical exercise, the rest of the critical concerns under this heading must be rated experimentally trivial.

⁶If he wishes, the experimenter may avoid use of the normal approximations entirely. The criterion of significance is an arbitrary constant. It may be defined not as a probability value to be exceeded, but directly in terms of the critical ratio. Say a critical ratio of 2.7 is accepted as the criterion. Observed critical ratios may then be judged "not significant" when below 2.7, "significant" when 2.7 or above, and "highly significant" when above, say, twice the criterion or 5.4. These are the conclusions commonly deduced from the critical ratio method. Probability statements may be calculated exactly whenever they are needed for further analysis.

Summary:

This discussion has covered the ground of specific mathematical criticisms to date. Many unsupported statements and diffuse arguments which have been offered in publication may be avoided in future discussion if the above "vulnerable" areas are duly considered.

Critical attacks upon the mathematic evaluation of ESP data have centered upon the three statistics commonly reported: (1) the mean chance expectation from which a deviation is measured; (2) the standard deviation by which the critical ratio is found; and (3) the statement of the probability of the computed critical ratio.

It is shown that these attacks have been irrelevant in regard to the mean chance expectation, mathematically valid but experimentally trivial in regard to the standard deviations, and mathematically and experimentally valid in regard to the probability statements. The error of the probability statements is discussed, and it is shown that these may be corrected by a small change in the criterion of significance. These changes would in no way affect any of the conclusions heretofore established in ESP research.

REFERENCES

1. FRY, T. C., *Probability and its Engineering Uses*, New York: D. Van Nostrand Co. 1928. Pp. xiv + 476.
2. GREENWOOD, J. A. and STUART, C. E. *Mathematical Techniques Used in ESP Research*, *Journal of Parapsychology*, 1937, I, 206-225.
3. KELLOGG, C. E. "Dr. J. B. Rhine and Extra-Sensory Perception." *Journal of Abnormal and Social Psychology*, 1936, XXXI, 190-193.
4. KELLOGG, C. E. "New Evidence (?) for Extra-Sensory Perception," *The Scientific Monthly*, 1937, XLV, 331-341.
5. STERNE, T. E., "The Solution of a Problem in Probability," *Science*, 1937, LXXXVI, 500-501.
6. WEISBERG, LOUIS, [Note in "Personal and Otherwise" section] *Harper's*, February, 1937.
7. WILLOUGHBY, R. R. "A Critique of Rhine's 'Extra-Sensory Perception'," *Journal of Abnormal and Social Psychology*, 1935, XXX, 199-207.
8. WILLOUGHBY, R. R. "The Use of the Probable Error in Evaluating Clairvoyance," *Character and Personality*, 1935, IV, 79-80.
9. WRIGHT, E. H., [Note in "Personal and Otherwise" section] *Harper's*, February, 1937.

NOTES

The following press release has been brought to the attention of the editors of the JOURNAL:

December 30, 1937.

Statement from the president of the Institute of Mathematical Statistics.

Dr. Rhine's investigations have two aspects; experimental and statistical. On the experimental side mathematicians of course have nothing to say. On the statistical side, however, recent mathematical work has established the fact that assuming that the experiments have been properly performed, the statistical analysis is essentially valid. If the Rhine investigation is to be fairly attacked it must be on other than mathematical grounds.

[Signed] BURTON H. CAMP,
President, Institute of Mathematical Statistics.

A Handbook for Testing Extra-Sensory Perception edited by C. E. Stuart and J. G. Pratt has been published by Farrar and Rinehart of New York. This *Handbook* has been written in simple, non-technical language specifically for those who wish to carry out ESP tests and experiments.

Imperfections in the commercial reproductions of the ESP cards preclude their unscreened use for experimental purposes. As reports of screened work have indicated that subjects do practically as well and in some cases better when they have no sensory contact at all with the cards, the screen should be so uniform a condition that card imperfections are not a matter for experimental concern. However, the publishers of the cards have been able to overcome the principal difficulties and future printings will therefore be much improved.

Much interest and inquiry has centered about the broadcast programs of the Zenith Radio Corporation and the results of their tests of extra-sensory perception. Frequent reference to Duke University in the program has created the not unreasonable impression that the program is under the direction of the Parapsychology Laboratory. In the organization of early programs a member of the staff of this laboratory was a member of an advisory council composed of psychologists. Neither this nor other connection exists at the present time. The program has been observed with interest, and with the hope that the most would be made of an unusual research opportunity. The responsibility for statements about extra-sensory perception, experimental methods used, tabulation of results, and statistical evaluation of the results, is of course entirely that of the Zenith Foundation.

GLOSSARY

In order to avoid constant redefining of commonly recurring terms in papers appearing in this journal, the following definitions are submitted for convenient reference. In case of any discrepancy between Glossary and usage in the text of an article the latter should be followed. Words defined elsewhere in the Glossary are italicized in the text of the definitions.

AGENT: In tests for *telepathy*, the person whose mental states are to be apprehended by the *percipient*.

AVERAGE SCORE = AVERAGE: Average number of *hits* per *run*.

BT (BEFORE TOUCHING): The technique in which the top card of the face-down *deck* is called and, after being called, is laid aside for *checking* at the end of the *run*. Each card in the *deck* is treated in the same way.

BM (BLIND MATCHING): The technique in which the *subject matches a deck of ESP cards* to five *key cards* which are laid out face down before him in an unknown order. Unless otherwise stated, the order is also unknown to the experimenter.

CALL v.: To designate a card by naming one suit either orally or by indicating it manually (as pointing or writing).

CALL n.: The response described above; also, the resulting selection.

CHANCE: The complex of undefined causal factors irrelevant to the purpose at hand.

CHANCE EXPECTATION = MEAN CHANCE EXPECTATION: The most likely *score* if only *chance* obtains.

CHANCE AVERAGE: *Mean chance expectation* in terms of *average per run*.

CHECK: To determine a *score* after the completion of a *run* by comparing the order of the *subject's calls* with the order of cards in the *deck*.

CLAIRVOYANCE: *Extra-sensory perception* of objective events as distinguished from *telepathic perception* (of the mental or subjective events of another person).

CRITICAL RATIO = X VALUE: The observed *deviation* divided by the *standard deviation*.

DECK: Twenty-five *ESP cards*, five of each suit.

DEVIATION: The amount an observed number of *hits* or an *average score* varies from the *mean chance expectation* or *chance average*. A *deviation* may be told (for a series of *runs*) or *average (per run)*.

DT (DOWN THROUGH): The technique in which the cards are called down through the *deck* before any are removed or *checked*.

EMPIRICAL CONTROL: An experiment which wholly or partially follows the main experiment with the exception that the conditions are designed to exclude the possibility of *ESP*.

ESP (EXTRA-SENSORY PERCEPTION): Response to an external event (perception) not presented to any known sense.

ESP CARDS: *Cards*, each bearing one of the following five symbols: Star, circle, three parallel wavy lines (called "waves"), rectangle, plus.

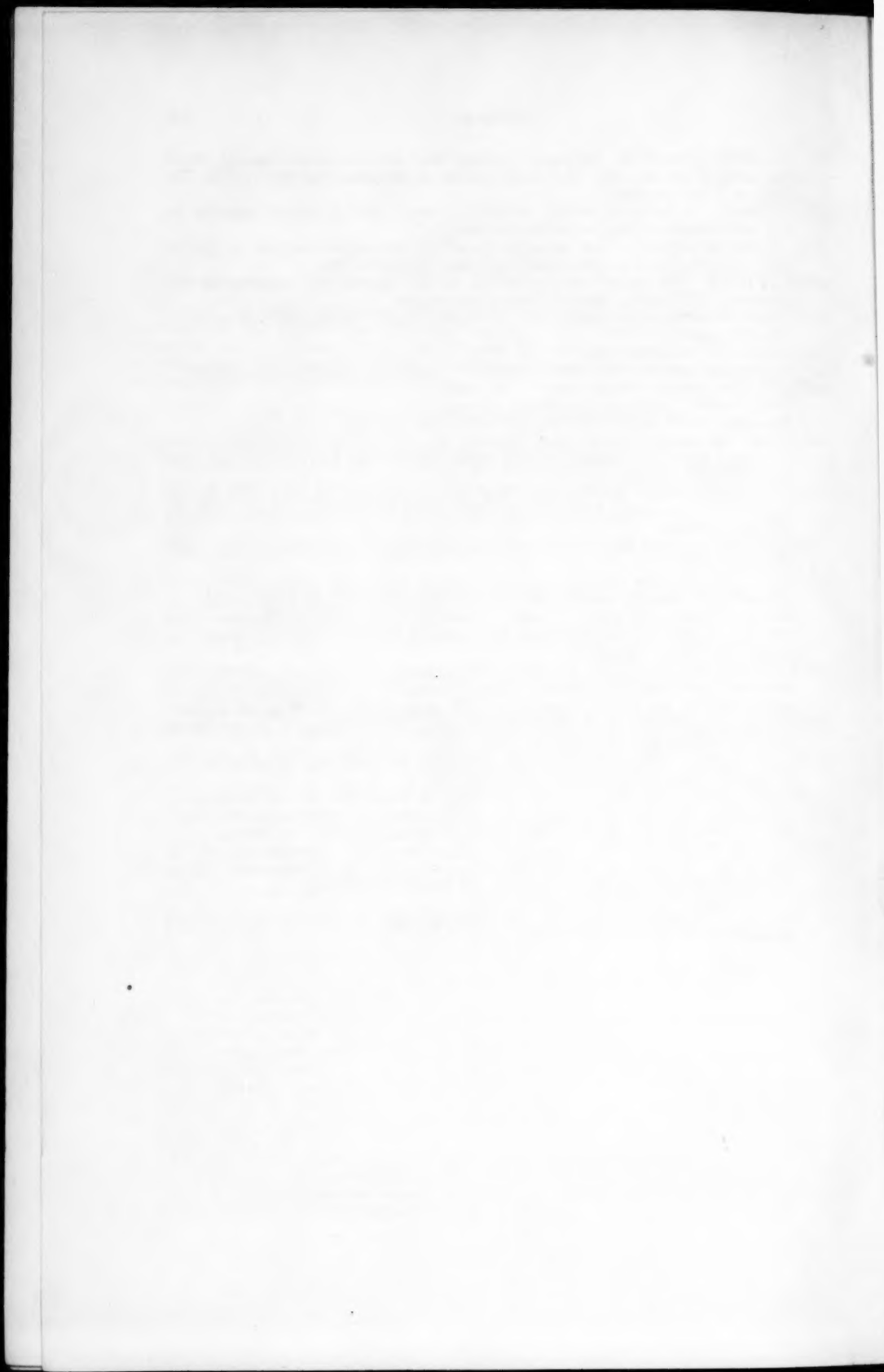
ESP SYMBOLS: See plate opposite page 1, this journal Vol. I, No. 1, March 1937.

ESP TESTS: A considerable number of techniques come under this heading, conveniently represented by initials, the principal ones being: *BT, DT, PT, GESP, BM, OM, STM*.

EXTRA-CHANCE: *Significantly different from chance expectation*.

GESP = GENERAL (OR UNDIFFERENTIATED) EXTRA-SENSORY PERCEPTION: A technique designed to test the occurrence of *extra-sensory perception*, permitting either *telepathy* or *clairvoyance* or both to operate.

- KEY CARD:** One of the five cards (where there are five suits) against which the cards of the test *deck* (i.e. *target cards*) in the *matching tests* (*OM*, *BM*, *TM*, etc.) are *matched*.
- MATCHING:** A form of *calling* in which a *target card* is placed opposite the *key card* which the *subject* selects to identify it.
- OM (OPEN MATCHING):** The technique in which the *subject matches a deck of ESP cards* to five *key cards* which are face up before him.
- PERCIPIENT:** The person who perceives. In this journal the common use will pertain to *ESP* rather than to sensory perception.
- PT (PURE TELEPATHY) = TELEPATHY:** The word "pure" emphasizes the exclusion of *clairvoyance*.
- RELIABILITY:** *Significance*.
- RUN:** The presentation and *calling* of 25 *ESP cards* or *symbols* in succession.
- SCORE:** The number of *hits* made in one *run*.
TOTAL SCORE: *Score* of any number of *runs*.
AVERAGE SCORE: *Total score* divided by number of *runs*.
- SCREEN:** An opaque barrier used between the subject and the card or agent. The main types of screens will be illustrated in this journal on their first introduction in print.
- STM (SCREENED TOUCH MATCHING):** The *touch matching* technique with an upright *screen* preventing vision by the *subject* of the handling of the cards by the experimenter.
- SERIES:** Several *runs* that are grouped consecutively or according to some other principle.
- SIGNIFICANCE:** The unlikelihood that a given event (or series) will not on the average occur by chance more often than once in 150 such events (or series). [Arbitrarily taken to mean a *deviation* in the expected direction such that the *critical ratio* is 2.5 times the *standard deviation* (or four times the probable error) or greater.]
- STANDARD DEVIATION:** A unit of measurement in statistical method; that *deviation* above and below *mean chance expectation* which is expected to include about $\frac{2}{3}$ the *chance scores*. For *ESP cards*, $S. D. = 2.04 \sqrt{\text{no. of runs}}$.
- SUBJECT:** The person who is experimented upon. Most commonly the *percipient* in *ESP*, though also the *agent* in *telepathy*.
- TARGET CARD:** The card which the *percipient* is attempting to perceive (i.e. to *call*, or otherwise indicate a knowledge of).
- TARGET DECK:** The *deck* of cards the order of which the *subject* is *calling*.
- TELEPATHY:** *Extra-sensory perception* of the mental activities of another person. It does not include the *clairvoyant* perception of objective events.
- TOUCH MATCHING:** The technique in which the *subject* indicates his *call* by tapping or touching one of the five *key cards* while the *experimenter* places the *target card* thus called in front of the *key card* indicated.
- TRIAL:** A single *call* of a presented *symbol*.
- X VALUE = CRITICAL RATIO:** The observed *deviation* divided by the *standard deviation* or by the *probable error*.



INDEX

- Adams, D. K., 52-62, 71
 Age range, 32, 82, 102, 105, 114, 144, 149
 American Mathematical Society, 292
 Arkin, H., 222, 224
 Attention curves, 200
 Austin, Mary, 1
 Automatic writing, 271
- BM, *see* blind matching
 BT procedure, 19-30, 127, 145, 150, 175, 193, 264, 270, 272, 279, 281, 283, 286; screened, 19, 283
 Bard College, 31, 32
 Barrett, Wm., 173
 Bechterew, W., 70
 Bender, Hans, 52-62, 71
 Bigham, 203
 Binet, A., 203
 Binomial hypothesis, 89, 116, 214-216, 237, 294, 299-303
 Blind matching, 10-17, 18-30, 127, 145, 147, 150-151, 231, 264, 267-275, 281-282; screened, 18-30, 264, 278
 Blind subjects, 82, 83, 143-155, 231, 281, 283-284
 Bond, E. M., 82, 102, 114-122, 290
 Bond, Frederick Bligh, 1
 Bonn, Univ. of, 52
 Boston Society for Psychic Research, the, 6, 8, 249, 257
 Brugmanns, H. J. F. W., 70, 277
- Calkins, M. W., 203
 Camp, B. H., 233, 305
 Campus Distance Series, *see* Distance Series, Pearce-Pratt
 Carpenter, C. R., 31-43, 186, 188, 271, 279, 290
 Chapman, D. W., 251-253
 Chi-square, contingency test, 218-221, 237-238; Fisher technique, 179; fit to binomial, 89, 116, 214-216, 300; formula, 215; grouping, 216-217; method discussed, 215-233; sign test, 217-218
 Children, 82, 102-113, 153, 231, 272; retarded, 82, 114-122
 Clark, C. C., 82, 123-142, 153, 155, 272, 290
- Clark University, 1
 Color cards, 32, 36-42
 Collusion hypothesis, 79
 Colorado, Univ. of, 185, 285
 Colton, R. R., 222, 224
 Columbia University, 10, 31-32, 48, 282
 Combination, methods of, 222-223
 Comparison, 290; calling vs. matching, 273; of distances, 182; of groups with individuals, 137-138; of methods, 22-23, 150-151, 264-275; of normal and distorted symbols, 93-101; of single and multiple symbols, 102-113; of small and large symbols, 102-113; of successive experimental periods, 151-152
 Competition, 106
 Conan Doyle, Sir Arthur, 1
 Constant Stimuli, Method of, 45
 Contingency test, *see* Chi-square method, contingency test
 Control series, 16, 28-30, 33, 56, 77, 84-92, 97-98, 185-186, 209, 212, 213, 215-217, 220-221, 244
 Cooper, C. C., 175-178
 Coover, John E., 1, 70, 72
 Cox, W. S., 84, 159
 Crandall, E. P., 286
 Crandon, L. R. G., 1
 Creery sisters, the, 102
 Critical Ratio method, 207-213
 Criticisms of ESP Research, 9, 81, 161, 206, 231-233, 234, 295-304
- DT, 38, 74, 94, 126, 127, 138-140, 175, 185, 187-188, 193-205, 264, 267-273, 278, 284, 286-287;
 DT₇₅, 139-140, 193, 200-205; DT₁₂₅, 140
 Davenport, C. B., 213, 222, 224
 DT curve, 100-101, 138-140, 160, 199-205, 272-275
 Deland, Margaret, 1
 Difference, evaluation of a, 221-222
 Discrimination, weight, 44-51
 Distance, 78, 104, 161, 172-184, 187-188, 260-263, 266-269, 280, 282-288; series, Bailey-Ownbey, 176; Cooper-Ownbey, 175; Duke-San Diego, 180-182;

- Garrett, 176; Junaluska, *see* Turner-Ownbey; Miles-Ramsden, 173; Pearce-Pratt, 74-76, 79, 177, 183-184, 209-222, 286; Pearce-Stuart, 175; Tarkio, 180, 183; trans-Atlantic, 173; Turner-Ownbey, 78-79, 178, 183-184, 278; Usher and Burt, 173, 174, 278; Wales-Somers, 173; Zirkle-Ownbey, 76-79, 176.
- Distorted Symbols, compared with normal, 93-101, 284
- Drawings, 54, 57-59, 104, 174, 233, 248-259, 279
- Driesch, Hans, 1
- Dufay, Dr., 173
- Duke University, 6, 8-10, 52, 63, 71, 91, 103, 124, 143, 160, 163, 185, 206, 231-232, 234, 239, 244, 264-267, 280, 295, 305.
- Durham Workshop for the Blind, 146
- Dusart, Dr., 173
- ESP cards, *illustr.* opp. p. 1, 159, imperfections in, 305
- Ebbinghaus, H., 203
- Ekas, M. P., 213, 222, 224
- Elderton's tables, 214-215, 217, 219
- Error, 153, 289
- Estabrooks, G. H., 70, 72, 277
- Estimation by subject, 27-28, 50, 247
- Experience, effect of, 132-133, 152
- Extra-Sensory Perception*, 6, 9, 18, 63, 71, 80, 103, 125, 155, 174, 192, 225, 240, 264, 272, 281, 286
- Fatigue, 160, 171, 196-199
- Fisher, R. A., 179, 213-214, 218, 221, 223-224
- Fisk, G. W., 66-67, 226-227
- Fordham Univ., 163
- Frobes, J., 203
- Fry, T. C., 299, 302, 304
- GESP procedure, 145, 147, 150, 153, 187, 241, 266-275, 278, 283
- GSTM procedure, 127
- Galileo, 2
- Garrett, Mrs. E., 176, 286
- Gault, R. H., 173
- George, R. W., 15, 18-30, 69, 82-83, 93-101, 151, 155, 180, 188, 232, 247, 264-266, 270-271, 278-279, 284, 288, 290
- Gestalt relation, 160, 204
- Gibert, Dr., 173
- Gibson, E. P., 232, 264-275, 290
- Glossary, 83, 156-157, 228-229, 306-307
- Glover, J. W., 293
- Gram-Charlier curve, 237, 293, 302
- Grand Rapids, Mich., 265
- Greenwood, J. A., 161, 179, 206-225, 232, 235, 242, 287, 293, 295-304
- Greville, T. N. E., 300
- Gridley, Mrs. S., 257
- Groningen Univ., 6
- Group Tests, 123-142, 231
- Guilford College, 191-192
- Guthrie, M., 102
- HP as subject, *see* Pearce, H.
- Habitual calling order, 71, 76-77, 79, 119-120
- Hall, G. Stanley, 1
- Hallucinated images, 58
- Handbook for Testing ESP, A*, 162, 305
- Hansen, 71
- Harvard University, 5, 292
- Hericourt, J., 173
- High-aim, *see* volitional direction
- Houdini, Harry, 1
- Hunter College, 260
- Huntington, E. V., 232, 292-294, 299, 302
- Hyperaesthesia, 74, 78-79, 236
- Hyperthyroidism, 262
- Hypnotic tests, 172-173
- Ilga K., 102
- Illinois, Univ. of., 159
- Imagery, 232, 243, 245-249, 255-256, 271
- Institute of Mathematical Statistics, 233, 305
- Institut Métapsychique International, 4
- Intelligence, 82, 122, 262
- Introspective accounts, 242-247
- Intuition, 143
- Jaensch, E. R., 53
- Janet, P., 70-71, 173
- Jastrow, J., 1
- Johnson, Miss G. M., 63-69, 226-227
- Junaluska Series, *see* Distance series, Turner-Ownbey
- Kelley, T. I., 208, 218, 224
- Kellogg, C. E., 209, 218, 224, 232, 234, 236, 297, 299-301, 304
- Kubis, J. F., 161, 163-171
- Kurtosis, 292
- Learning, 42-43
- Lehmann, A., 71

- Lemmon, V. W., 81, 212, 224
 Linzmayer, A. J., 13, 270, 281
 Lodge, Sir Oliver, 1, 173
 Low-aim, *see* volitional direction
 "Luck" hypothesis, 83, 84-92, 161
- MP procedure, 145, 147, 150-151
 Manual response, 24, 63, 231, 265
 Martin, D. R., 161, 185-190, 285, 288
 Matching hypothesis, 288, 292-295, 298-303
 Mathematical evaluations, 11, 20, 34-35, 49, 72, 92, 128, 161, 163, 189, 206-225, 231-233, 237-238, 253, 267-268, 295-305
 Mean chance expectation, 209-211, 296-297, 299, 301, 304
 Mechanical selector, 66
 Method of Correct Matchings, 233, 251-254
 Metropolitan Achievement Test, 114
 Milner, Dr. and Mrs. C. A., 191
 Minnesota, Univ. of, 159
 Motor response, *see* manual response
 Murchison, C., 1
 Murphy, Gardner, 1, 10, 45, 173, 257
 Music accompanying test, 56
 Myers, F. W. H., 173
 McDougall, William, 1
 MacFarland, J. D., 82, 83, 93-101, 180, 266, 284, 287-288
 McKinley, W., 95
- Negative results, 84, 188-189, 213-214
 Neureiter, F. von, 103
 New Card Series, 72-74
 New York City, 260-261
 New York University, 123
 Normal distribution curve, 208, 212, 216, 298-303
 Norton, H. W., 223, 225
 Numbers as stimuli, 114
- OM, *see* Open Matching
 Observer, effect of, 133-136, 142
 Olds, E. G., 292
 Omer, F., 95
 Open Matching, 11, 18-30, 106, 127, 145, 147, 150-151, 193, 264, 267-275;
 Screened, 18-30, 106, 264, 278, *illust.* opp. p. 18; *illust.* opp. p. 107
 Ownbey, Miss Sara, 76-79, 138, 175, 178
- PT procedure, 175
- Parapsychology, compared with psychological research, 7
 Pearce, H., 72-76, 140-141, 175-177, 192, 208, 270-271, 286-287
 Pearson, K., 208, 214, 225, 299
 Pearson's tables, 209, 214
 Pegram, M. H., 82-83, 143-155, 160-161, 191-205, 270, 272, 281, 283, 288, 290
 Pennsylvania, Univ. of, 5
 Peters, C. C., 213, 225
 Phalen, H. R., 31-43, 186, 188, 271, 279, 290
 Podmore, F., 173
 Pohlmann, 203
 Pratt, J. G., 10-17, 75-76, 130, 162, 177, 194, 208, 225, 232-233, 239-259, 280, 282, 286-288, 305
 Precognition, 65-66, 184, 233
 Preference, 94, 97, 108, 265
 Price, M. M., 82, 83, 143-155, 270, 281, 283, 288
 Prince, W. F., 1, 129, 249, 251, 257, Princeton Univ., 159
 Probability, concept of, 207
 Probability integral tables, 208, 302
 Probable error, definition and formula, 208
 Psychological research, 1-9
 Psychophysical experiments, 32, 44-51
- Radio broadcast experiments, 173-174, 305
 Raible, Mildred, 44-51, 235
 Rapport, 33; between brothers, 121
 Recording, method of, 15, 34, 95, 126, 180, 186, 240, 284-287
 Reliability of investigators, 289
 Rewards, effect of, 26-27, 112, 147
 Rhine, J. B., 6, 10, 20, 31-33, 45, 52, 63, 70-80, 84, 96, 138, 150-151, 155, 161, 163, 172-184, 185, 191-192, 199, 206-208, 213, 222-223, 225, 232, 236, 240-242, 260, 264, 270-272, 276-291, 293, 296, 305
 Rhine, L. E., 82, 272, 278, 290
 Rice, C. H., 232, 239-259, 280, 282, 288
 Richet, C., 70, 102, 173
 Riess, B. F., 233, 260-263
 Rietz, H. L., 293
 Rouke, F. L., 161, 163-171
- SBM, *see* Blind matching, screened
 SBT, *see* BT, screened
 SOM, *see* Open matching, screened

- Schiller, F. C. S., 1
- Scoring table, 232, 292-294
- Screens, description of, 12, 20, 39, 40, 94, 105-106, 126, 187, 240, 267, 280, 282-284; effect of, discussed, 25-26
- Screened touch matching, 12, 15, 94, 126-127, 231, 266-275, 282, 284; *illustr.* opp. p. 18, and opp. p. 95
- Sealed cards, 13, 53-62, 83, 144-145, 147-150, 153-154, 241, 280-281, 284
- Selection of data, 71, 224, 234-235, 241-242, 268, 279; of subjects, 21, 33, 86-87, 124, 148, 185, 265, 283
- Sensory cues, possibility of, discussed, 12-13, 21, 34, 42-43, 46, 50-51, 56, 66, 71, 73-74, 79, 97, 152-153, 175, 186-187, 194, 232, 238, 244, 254-256, 267, 276-291
- Sharp, V., 82, 123-142, 153, 155, 272, 290
- Shuffling method of, 14-15, 20, 71-72, 74, 79, 95, 180, 194, 236-237, 266-267, 284, 287-288; comparison, 98-100
- Significance, criteria of, 188, 213-214, 302-304
- Sinclair, Upton, 174, 249, 257, 277
- Sixth sense, 143
- Skewness, 212, 292, 298-303
- Smith, Burke M., 19, 63-69, 226-227
- Soal, S. G., 159, 174
- Social factors, 82
- Society for Psychical Research, The, 3, 8, 59, 63, 102-103, 173
- Standard deviation, described, 208, 297-298
- Stanford University, 5
- State School for the Blind, 144
- Sterne, T. E., 292, 300, 304
- Stuart, C. E., 13, 161, 162, 175, 179, 191-192, 206-225, 232, 235, 242, 286-288, 295-305
- Suggestion, 105, 248-249, 255
- Sullivan, J. W. N., 82
- TM procedure, 126
- Tarkio College, 18, 93, 180-181, 183, 287
- Telepathy, 44-51, 76-80, 173, 290; in twins, 163-171
- Thweatt, M., 242-247, 282
- Turner, M. F., 78-79, 178
- Twins, 161, 163-171, 231
- Tyrrell, G. N. M., 19, 63-69, 226-227
- Tyrrell, Machine, the, 64-66
- University College, London, 159
- Van Voorhis, W. R., 213, 225
- Variance of ESP deck, 211-212
- Variation of stimulus, 93, 103
- Verbal response, 144, 150, 227, 231, 255, 268
- Volitional direction, 104, 191-205, 231
- Walden, G. H., 48
- Walker, R., 95
- Warcollier, R., 173, 257
- Warner, L. H., 44-51, 83, 84-92, 161, 178, 232, 234-238, 286, 288
- Wave hypothesis, 78
- Weber-Fechner law, 111
- Weisberg, L., 209, 225, 296, 304
- Whispering hypothesis, 46, 76, 79
- Willoughby, R. R., 84, 209, 212, 214, 218, 225, 299, 301, 304
- Witnessing, 95, 97, 148, 153, 190, 194, 242, 267, 283-285, 289
- Woodruff, J. L., 15, 18-30, 69, 151, 155, 188, 232, 247, 264-265, 270-271, 278-279, 290
- Wooley, V. J., 173
- Wright, E. H., 225, 296, 304
- Wundt, W., 44
- Yule, G. U., 211, 225
- Zener cards, 226; *illustr.* opp. p. 1, opp. p. 159
- Zenith Radio Corporation, 305
- Zirkle, G. A., 176, 178, 271

183,

0; in

227

, 255,

231

161,

214,

194,

, 155,
278-

opp.