

# The Journal of Parapsychology

---

Volume 1

SEPTEMBER, 1937

Number 3

---

## EDITORIAL COMMENT

WHEREVER the topic of extra-sensory perception is discussed, the question is sooner or later likely to be asked: Have other universities or colleges besides Duke been investigating this subject? And, if so, what have they found? There are several questions regarding any research that are answered most satisfactorily by repetition of experiments in other laboratories. Confirmation is usually very reassuring and failure to confirm very damaging.

The readers of this journal need not be told that there have been many successful repetitions of the ESP research. The experiments that have not confirmed the Duke work are not so readily available. However, the ratio of favorable to unfavorable investigations seems to be about four to one, either on the basis of published or unpublished work.

So far as published work goes, to our knowledge, there have been two repetitions which have failed. (Although in one of these the investigator writes that his results are "encouraging," it is not conclusively confirmatory.) One of these was by Mr. William S. Cox, a senior undergraduate student at Princeton University,<sup>1</sup> and the other was by Mr. S. G. Soal, now working at the psychological laboratory of University College, London.<sup>2</sup> Two others have been mentioned in the public press as having been conducted, one at the University of Illinois and the other at the University of Minnesota. There is information on hand of three other attempts that have been made which failed to confirm the case for ESP but which have not yet been published. These seven cases represent the total to date of the non-confirmatory repetitions of the ESP work.

<sup>1</sup>"An Experiment on Extra-Sensory Perception," *Journal of Experimental Psychology*, 1936, XIX, 429-437.

<sup>2</sup>"A Repetition of Dr. J. B. Rhine's Work in Extra-Sensory Perception," *Journal of the Society for Psychical Research*, 1937, XXX, 55-58.

It is important to the students of this field to know of these results and to study the conditions under which they were obtained. These results should be as readily available to the research worker as those favorable to the occurrence of ESP. It should be made clear that the pages of this *JOURNAL* are fully as open to articles that are unfavorable as to those that are favorable to the claims for ESP. The fact that none have been submitted thus far which failed to support the ESP hypothesis may have been due to some misapprehension on this point.

So far, then, as answering the common need for repetition goes, the case for ESP is obviously greatly strengthened by the confirmatory series. But a new problem for study is created by every non-confirmatory report; we shall need to account for these as well as the successes. From the study of their causes and conditions may come the discovery of keys to better understanding and control of whatever principles are at work.

\* \* \* \* \*

The present number of the *JOURNAL* continues the confirmatory side of the repetitional picture, with contributions from three institutions besides Duke. Two of the reports are based on theses submitted for graduate degrees. A much hastened solution of many of the problems of this varied field is assured when graduate students find in them interesting and available research projects.

This issue of the *JOURNAL* brings the total number of individual trials made in ESP tests reported during the year to a point well above 1,000,000. This is over twelve times the number reported in the monograph dealing with the original Duke experiments.

\* \* \* \* \*

Further steps toward relating ESP to other mental processes have been taken in research reported in this number. Pegram finds marked U-shaped curves in her DT tests and suggests a gestalt relation analogous to memory curves. Of special interest is the inversion of these curves when the subject tried to miss the symbols; i.e., to aim for low scores.

Another point of considerable importance is the relatively very much greater average deviation on the negative side in the low-aim tests than occurred on the positive side in the usual high-aim runs. This would at least superficially appear to be connected with the peculiar motivation of the subject, but psychology has as yet little method for the objectively reliable study of motivation by which this could be followed up.

The question of the possible effect of fatigue on ESP scoring would

apparently be tested adequately in the 5,000 trials in a single day Miss Pegram gave herself.

A problem of widespread interest is the possible telepathy between twins. This oft-suggested problem has been attacked by Kubis and Rouke, with at least interesting secondary results. The fact that in only one pair were the twins identical, and the fact that in the tests the twins *did not know* that they were being tested for similarity (or community) of ESP response, leave the question still open, as the authors recognize. Nevertheless, an important research project has been opened up.

The investigation of distance as a condition enters slightly into the Martin research, and becomes the main topic of the Rhine report. This is a survey of earlier work, with some new evidence added. Spatial influence upon ESP scoring is apparently not in evidence in any instance.

While the case for ESP does not rest upon any single way of evaluating the results, or upon any special formula or mathematical method, the success of the research into the many problems of its nature does greatly depend upon statistical method and the mathematics of probability. From the beginning of the ESP testing at Duke and even before, it has been customary to rely upon the existing authorities for methods of analysis, as does psychology proper, and other sciences in general. The character of the research in extra-sensory perception has, however, led those who are not familiar with the very specialized methods of evaluation to fear that, if something is wrong with the ESP reports, it is probably in the mathematics. It is important, then, to make these methods as fully explicit as possible, not only for those who are frankly unfamiliar with statistical methods, but for those who are quite at home in some branch of the varied mathematical field but who find these specialized techniques obscure.

Warner's article, "The Role of Luck in ESP Data," in the last number of the JOURNAL, went far to clarify some of the obscure aspects of the methods of measurement of the ESP results. In this number Greenwood and Stuart deal directly with the evaluation of ESP scores and the methods available.

These columns are open to criticism of this phase of the ESP research, as they are to any other phase. The editors will welcome any new light that may be thrown upon any aspect of this complicated field of problems.

\* \* \* \* \*

The interest in the tests for ESP which has grown up both among the lay and the scientific public has led to a considerable demand for test cards, instructions, and record sheets. For some time the Parapsychology Laboratory at Duke has been furnishing these to those who have written for them.

At last, however, it became necessary to put these materials on the market and to shift the burden of distribution to commercial houses. It has been possible to arrange to have the ESP test cards printed in two types of pack, both to be on sale at book and novelty stores at very modest prices. In Pack No. 1 the symbols will be in black and in No. 2 they will be in five colors, a different color for each type of symbol. Record sheets will be available in pad form.

The instructions for the conduct of the tests have been expanded into book form, with directions for finding and testing subjects, evaluating results, safeguarding the conditions, and carrying out experiments. This *Handbook for ESP Tests* is edited and arranged by Mr. C. E. Stuart and Dr. J. G. Pratt, of the Parapsychology Laboratory at Duke, and is to be published shortly by Farrar and Rinehart. It is intended to be of use both by laymen and in academic scientific research.

The provision of these materials facilitates the participation of the general public in these tests. It is hoped that this participation will advance the scientific inquiry: first, by the discovery of unusually gifted subjects who will be brought to the attention of college investigators; and second, through the development of such a degree of interest as will induce a number of capable experimenters to begin to explore this field for themselves—if only in the endeavor to discover some essential error which, as they may believe, must have escaped the notice of those who claim to have found positive evidence of ESP.

## AN EXPERIMENTAL INVESTIGATION OF TELEPATHIC PHENOMENA IN TWINS

JOSEPH F. KUBIS and FABIAN L. ROUKE

---

**ABSTRACT:** Six sets of twins, one identical and five non-identical, were studied to see if spontaneous telepathic perception might occur between them when both of a pair called ESP cards simultaneously. The twins were placed in separate rooms and instructed to call at a signal the ESP cards looked at by the experimenter. Their calls were compared both with the experimenter's cards and the corresponding calls of the other twin. The subjects did not know that the latter check was to be made. No significant number of correspondences between subject's calls resulted. Two of the subjects did, however, score significantly high on the cards looked at by the experimenter.

The research was carried out as partial fulfillment of research requirements for a master's degree at Fordham University, where Dr. Kubis is Instructor in Psychology and Mr. Rouke is a graduate student.

---

THE SUBJECTS in this series of experiments were six sets of twins. The common opinion that twins are alike mentally as well as physically was the idea that prompted this investigation. The problem was to try to discover if this supposed concurrence of thought could be demonstrated in a laboratory test for ESP.

The procedure used was that instituted by Dr. Rhine at Duke University. The agent, going through a pack of ESP cards would concentrate on each design as it was turned up, and the subject would try to name the design.

The statistical procedure used to calculate the significance of the responses was also that used by Dr. Rhine. The critical ratio or  $X$  value was obtained from the formula  $X = \frac{d}{\sqrt{npq}}$

where  $n$  = total number of guesses made  
 $p$  = probability of successful guesses (1/5)  
 $q$  = probability of unsuccessful guesses (4/5)  
 $d$  = deviation of score from chance expectancy.

Both twins of a set acted simultaneously as subjects with one agent. Their responses were to be checked against the correct symbols as they

were in the pack the agent used, and also checked against each other for correspondence. This situation seems to furnish an opportunity to compare the telepathic communication between the twins and the outside agent and between the twins themselves. The twins were not told that this check for correspondence was being made. The situation in this experiment, where they were not actively trying to send impressions to each other, will show how much alike their thoughts are, and whether or not they independently get the same ideas at the same time. It is to be observed that the conditions of this test were such as to show only "spontaneous" telepathy between the twins. Telling them to try to give the same responses might give a totally different result.

A signalling system of a switch and two lights was used. The agent operated the switch which flashed a signal for both subjects. When the agent turned over a card and concentrated on it the switch would be closed momentarily. The subject was directed to record the first symbol that came into his or her mind after each flash. The time interval between the calling of successive cards was decided by the subjects and was regulated to suit them. After recording the correct symbols from each pack the cards were shuffled thoroughly by hand before being used in another trial. In most cases five packs were turned up in succession and then the five packs were scored. The subjects knew of their scores as the experiment progressed. This was an important factor in sustaining their interest. If the subject desired it an observer was present to record the responses.

The connection between the agent and the subjects is not one of pure telepathy; the experiment is in undifferentiated extra-sensory perception because of the use of the cards. The connection between twins, if any exists, would be one of pure telepathy.

The experiments had to be performed at the homes of the various subjects. Slight differences in the procedure, necessitated by varied conditions available in the separate homes, will be fully explained below in the account of the actual experimentation. If separate rooms for each of the participants could not be used, the radio was turned on softly to obviate any effects of possible subconscious whispering. The authors acted as agents in all the simultaneous reception trials.

#### EXPERIMENTATION

Altogether records of over twenty-thousand trials were made with the six sets of subjects. Fifteen hundred trials were made with each set in the simultaneous perception experiment described above. In addition, a few added trials were made with some of the twins with one

acting as percipient and the other as agent. All of the subjects showed an encouraging interest in the investigations. The possibility of further trials with set E and set F was excluded by their late introduction to the experiment. And the illness of  $B_2$  interrupted work with that set for almost one month.

Twins  $A_1$  and  $A_2$  were age 21, female and non-identical. In their home the first floor, consisting of three rooms, front, back and center, was used. The agent turned the cards and worked the switch in the center room which was shut off from the rear room by a closed door and from the front room by heavy portieres which were kept drawn together during the experiment. Each subject was in a separate room with the agent in a third between them. This set of subjects regularly worked with an observer to record their responses. All experiments had to be done at night, usually between the hours of eight and ten o'clock. Besides 1,500 trials each in the simultaneous reception experiment, 825 trials between the twins themselves were recorded. Of these  $A_1$  was the subject in 450 and the agent in 375.

Twins  $B_1$  and  $B_2$  were age 20, female and non-identical. During the earlier series of trials these twins visited the home of Twins  $A_1$  and  $A_2$  and the experimentation was done there. A little less than half of the trials were finished this way. The remainder of the experiments were performed in their own home. The agent was in the foyer and the subjects were in separate rooms which opened off the foyer. These twins worked with observers to record their responses in the beginning, but as they became more accustomed to the experiment they preferred to record their own. This set of subjects did not make any trials other than the fifteen-hundred done in simultaneous reception. With this set too, all experimenting had to be done after eight o'clock in the evening.

Twins  $C_1$  and  $C_2$  were age 17, female and non-identical. Experimentation was done in a large basement playroom. The agent sat at a table in the center of the room. The subjects sat at opposite ends of the room with their backs to the agent. The agent was about twenty feet distant from each subject. All the responses were recorded by the subjects themselves. As in the two previous cases all experimentation was done at night. More trials were made with these subjects than with any of the others. In addition to the fifteen-hundred they did in the simultaneous experiment, two thousand trials were performed by the twins themselves. Twin  $C_1$  acted as agent for one thousand trials with her sister as subject, and then the functions of agent and percipient were reversed for another thousand trials.

The fourth set of subjects,  $D_1$  and  $D_2$  were age 17, female and identical. These twins chose to record their own responses from the beginning of the experiment. At their home the agent was in the living-room, one subject was in the dining-room directly behind it, and the other sat in the reception hall which was to the side of the living-room. As before, the experimentation was started in the evening. Later on it was possible to see the twins in the afternoon, and this was much more agreeable to them, as they were accustomed to retiring very early and were fatigued when they had to stay up for the experiment.

$E_1$  and  $E_2$ , the fifth set of twins used, were age 16, male and non-identical. Once again experimentation was started in the evening, but only a few trials were run off this way when it was made possible for the experimenter to be free in the afternoon. This new time was eagerly taken by the subjects who said that they would much rather work in the afternoon; that they were not so tired. The three participants were located in three rooms one behind the other, and each subject was shut off from the agent (in the middle room) by a closed door. They, too, recorded their own responses from the start of the experiment. Because the investigator was not able to contact these subjects until late in the period of experimentation, only the fifteen-hundred trials in the simultaneous situation were performed. \*

Twins  $F_1$  and  $F_2$  were age 10 and non-identical.  $F_1$  was male,  $F_2$  female. They are the youngest of three sets of twins in the one family. However, the other two sets were not sufficiently interested to continue in the experiment after the first few trials. Here, again, experimentation had to be performed in the evening. For the first two series of trials some other member of the family acted as observer to record responses, but for the rest of the experimentation the twins recorded their own. In this house the agent was in a central reception hall with one subject in a room on each side, both rooms being separated from the hall by closed doors. As in the case of  $E_1$  and  $E_2$  the experimentation was limited to the fifteen-hundred trials in simultaneous reception.

#### RESULTS AND CONCLUSIONS

On the whole, the results of these experiments are not significant. In particular, they do not show that as the test was made, any telepathic connection exists between the twins. The average deviation of each set for the whole series of the experiments was well within chance expectancy. The results are recorded in the tables below,<sup>1</sup> and on Fig. I.

<sup>1</sup> Cf. Tables I to VI.

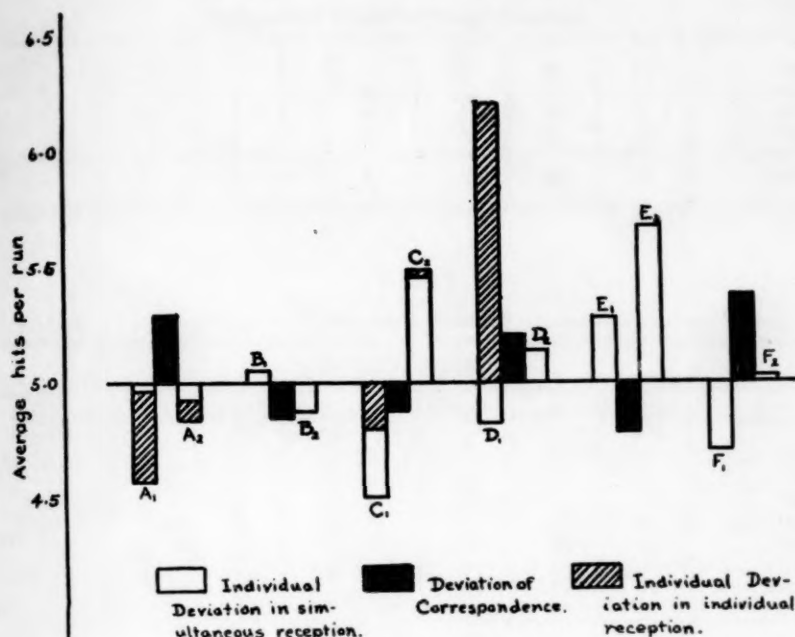


FIGURE I. Comparison of Average Scores for Individual Subjects.

TABLE I  
RESULTS WITH SET A.

SIMULTANEOUS RECEPTION							
Date	Runs of 25	Correct Calls		Deviation		Correspondence	Deviation of Correspondence
		A1	A2	A1	A2		
2/9 .....	10	52	48	2	- 2	50	0
2/16 .....	20	100	109	0	9	115	15
3/4 .....	20	97	102	- 3	2	111	11
3/9 .....	10	49	36	- 1	-14	41	- 9
Totals .....	60	298	295	- 2	- 5	317	17

INDIVIDUAL RECEPTION—SUBJECT A2—AGENT A1					
2/4 .....	4	..	22	..	2
3/4 .....	5	..	21	..	- 4
3/15 .....	10	..	44	..	- 6
Totals .....	19	..	87	..	- 8

INDIVIDUAL RECEPTION—SUBJECT A<sub>1</sub>—AGENT A<sub>2</sub>

2/4 .....	6	32	..	2	..	
2/9 .....	3	11	..	-4	..	
3/4 .....	5	23	..	-2	..	
3/15 .....	10	50	..	0	..	
Totals .....	24	116	..	-4	..	

TABLE II

## RESULTS WITH SET B.

## SIMULTANEOUS RECEPTION

Date	Runs of 25	Correct Calls		Deviation		Correspondence	Deviation of Correspondence
		B <sub>1</sub>	B <sub>2</sub>	B <sub>1</sub>	B <sub>2</sub>		
2/9 .....	6	29	24	-1	-6	33	3
3/9 .....	10	57	52	7	2	50	0
3/14 .....	20	112	95	12	-5	103	3
3/21 .....	24	105	122	-15	2	105	-15
Totals .....	60	303	293	3	-7	291	-9

TABLE III

## RESULTS WITH SET C.

## SIMULTANEOUS RECEPTION

Date	Runs of 25	Correct Calls		Deviation		Correspondence	Deviation of Correspondence
		C <sub>1</sub>	C <sub>2</sub>	C <sub>1</sub>	C <sub>2</sub>		
2/11 .....	10	36	51	-14	1	64	14
2/26 .....	10	57	60	7	10	49	-1
3/5 .....	10	45	70	-5	20	41	-9
3/12 .....	20	90	92	-10	-8	95	-5
3/18 .....	10	42	54	-8	4	44	-6
Totals .....	60	270	327	-30	27	293	-7

INDIVIDUAL RECEPTION—SUBJECT C<sub>1</sub>—AGENT C<sub>2</sub>

2/11 .....	10	45	..	-5	..	
2/26 .....	10	51	..	1	..	
3/5 .....	10	45	..	-5	..	
3/18 .....	10	51	..	1	..	
Totals .....	40	192	..	-8	..	

INDIVIDUAL RECEPTION—SUBJECT C<sub>2</sub>—AGENT C<sub>1</sub>

2/11.....	10	..	55	..	5	
2/26.....	10	..	52	..	2	
3/5.....	10	..	60	..	10	
3/18.....	10	..	52	..	2	
Totals.....	40	..	219	..	19	

TABLE IV  
RESULTS WITH SET D.

## SIMULTANEOUS RECEPTION

Date	Runs of 25	Correct Calls		Deviation		Correspondence	Deviation of Correspondence
		D <sub>1</sub>	D <sub>2</sub>	D <sub>1</sub>	D <sub>2</sub>		
2/25.....	10	52	46	2	- 4	46	- 4
3/4.....	20	97	97	- 3	- 3	106	6
3/18.....	20	95	104	- 5	4	100	0
3/25.....	10	45	61	- 5	11	60	10
Totals.....	60	289	308	-11	8	312	12

INDIVIDUAL RECEPTION—SUBJECT D<sub>1</sub>—AGENT D<sub>2</sub>

3/25.....	5	31	..	6	..	
-----------	---	----	----	---	----	--

INDIVIDUAL RECEPTION—SUBJECT D<sub>2</sub>—AGENT D<sub>1</sub>

3/25.....	5	..	25	..	0	
-----------	---	----	----	----	---	--

TABLE V  
RESULTS WITH SET E.

## SIMULTANEOUS RECEPTION

Date	Runs of 25	Correct Calls		Deviation		Correspondence	Deviation of Correspondence
		E <sub>1</sub>	E <sub>2</sub>	E <sub>1</sub>	E <sub>2</sub>		
3/19.....	10	55	55	5	5	36	-14
3/23.....	20	94	100	- 6	0	107	7
3/27.....	30	167	185	17	35	144	- 6
Totals.....	60	316	340	16	40	287	-13

TABLE VI  
RESULTS WITH SET F.

SIMULTANEOUS RECEPTION							
Date	Runs of 25	Correct Calls		Deviation		Correspondence	Deviation of Correspondence
		F <sub>1</sub>	F <sub>2</sub>	F <sub>1</sub>	F <sub>2</sub>		
3/19.....	15	70	88	- 5	13	88	13
3/23.....	20	86	96	-14	- 4	105	5
4/2 .....	25	126	117	1	- 8	129	4
Totals.....	60	282	301	-18	1	322	22

For example, the largest positive deviation obtained by any subject over the entire series of fifteen hundred trials is 40 made by E<sub>2</sub>.<sup>2</sup> This gives an X value of less than 2.6 and is hardly significant. And the highest deviation of the correspondence score above chance is 22,<sup>3</sup> which would give a lower value of X than the above example. In some cases the deviations were negative, but, again, the largest, -30,<sup>4</sup> would not give a significant X-value. Thus the deviations both positive and negative, fall within the limits of chance expectancy.

As we said before there was no evidence of telepathic connection between the twins. Such connection would have been shown by a significant positive deviation in the correspondence column in the tables. There were, however, some indications of telepathy obtained in certain parts of the experiment; that is, some of the scores for the cards looked at by the agent. This score, be it remembered, was the only one for which the subjects were working.

Twin C<sub>2</sub> on March fifth acted as subject in twenty runs of twenty-five cards each.<sup>5</sup> The first ten were with C<sub>1</sub> as agent, the second ten with the author as agent. For the first five runs with her sister, C<sub>2</sub> made an average of 4.6 correct in twenty-five. The experiment was then interrupted while C<sub>2</sub> answered a telephone call. Upon resumption of the experiment C<sub>2</sub> raised her average to 7.4 correct calls in twenty-five for the ten runs with the author as agent. The X value as found by the formula given above for the entire twenty runs is 3.2. For the ten runs with the author it is 3.1. Both these values are significant.

The subject was asked if she had heard anything particularly pleasing when she was called to the telephone. She said she had not. It is interesting to note, however, that for three days preceding the ex-

<sup>2</sup> Cf. Table V.

<sup>4</sup> Cf. Table III.

<sup>3</sup> Cf. Table VI.

<sup>5</sup> Cf. Table III.

periment,  $C_2$  had been out of school because of a slight cold, and had spent most of her time sleeping and resting in bed.

The other particularly significant results were also obtained with a subject after a few days rest from study, although this time, it was during a vacation. The factor of general fatigue seems to be detrimental to extra-sensory perception.

Twin  $E_2$  on the afternoon of March twenty-seventh, went through thirty runs of twenty-five cards,<sup>6</sup> all with the author as agent. For the entire series of thirty runs he had a positive deviation of thirty-five above chance expectancy for an average of 6.7 correct calls in twenty-five. This gives an X value of 3.2. The runs were made in groups of ten. On the last of these groups the number of correct calls was 73, an average of 7.3 per twenty-five. The X value for this series is 3.6. And on the last two runs, twenty-seven correct calls out of fifty cards were recorded. This gives an X value for the two runs of 6.0 which surely can be accepted as significant.

An interesting feature of the results is the wide difference in both the simultaneous and the individual reception of subjects  $C_1$  and  $C_2$ . For 100 runs,  $C_1$  was 38 below chance and  $C_2$  was 46 above. The critical ratio of this is 3.0. There was a noticeable difference in stature, height, weight and general physique.  $C_2$  had the advantage in all these respects. These physical differences would warrant further investigation of the relations of temperament and intelligence to extra-sensory powers. In this investigation  $C_2$  was more poised, more phlegmatic and somewhat more introvert than  $C_1$ .

The following conclusions are drawn as a result of the experiment.

- (1) The existence of extra-sensory phenomena was exhibited in the instances reported above.
- (2) The factor of fatigue must be avoided, since it seems detrimental to telepathic phenomena.
- (3) There was no evidence of "spontaneous" telepathic communication between the twins. This is shown by a lack of significant positive deviation of the scores for correspondence.<sup>7</sup>

The lack of the evidence referred to in (3) may conceivably be due to unfavorable experimental conditions or to the fact that the twins did not know that the correspondence scores would be considered. More work is needed to learn if this knowledge would make a difference. It would also be desirable to include a larger number of identical twins in a further study of this kind.

<sup>6</sup>Cf. Table V.

<sup>7</sup>Cf. Tables I to VI.

## THE EFFECT OF DISTANCE IN *ESP* TESTS

J. B. RHINE

---

**ABSTRACT:** This is a review of the work done on the problem of the possible effect of distance on *ESP*. Although in some instances scores declined with distance, there were others in which there was no decline. In certain Duke distance tests, especially, was the latter true. New evidence is presented from experiments in which distance was an incidental condition. This supports the hypothesis that distance is not restrictive in its influence on *ESP*. In a series conducted at Tarkio College, subjects calling an unbroken pack at distances up to 1,400 miles showed no reliably different scoring rate at any distance. In a series conducted between Duke University and San Diego, California, the group at San Diego made higher scores than other subjects at points nearer Duke.

Dr. Rhine is Professor of Psychology at Duke University.

---

GIVEN a natural phenomenon or process of unknown character, one of the first questions asked is: What is its relation to space? What effect, if any, does distance have upon it? A satisfactory answer to this question goes far to clarify the relation which the unknown phenomenon bears to the already known one.

In the history of the investigation of extra-sensory perception, this question has frequently been asked. In the non-experimental studies, confined to the collection and classification of instances of reported personal experiences appearing to involve extra-sensory perception, the investigators observed the apparent independence of distance in the anecdotal material collected. That is, a dream (for example) coincident with an event that was presumably unknown to the dreamer was observed to be as likely to indicate correct knowledge of the event at a great distance as at a short one.

In the earlier experimental studies of the process of extra-sensory perception the reports touching on the question of the effects of distance have indicated that there *is* no appreciable effect. Some of this earlier work, now fifty years old, is difficult to evaluate precisely, particularly the experiments in hypnotization and hypnotic tests conducted with distance between agent and subject. But the standing of the experimenters is noteworthy, and the agreement of the results is general. Ac-

cordingly, the experiments cannot be justifiably disregarded on any ground of which I am aware.

One of the earliest of these experiments was that of Professor William Barrett, a physicist, who reported in 1876 that he had been able to produce reliable responses in a hypnotized subject at distances of hundreds of yards, presumably by an extra-sensory means of thought-transference.<sup>1</sup>

Among the French physicians and university professors who were actively experimenting with hypnosis during the last quarter of the century, there were several who reported success in reliably inducing mental states at distances (measurable in miles) that prohibited any possible sensory perception. Prominent among these names are those of Pierre Janet, Richet, Gibert, Hericourt, Dufay, and Dusart. And others too were involved as witnesses. Frederic Myers of England collaborated with Janet in a series of 20 trials at hypnotization at a distance, with 80% success. Myers concluded this was evidence of extra-sensory thought-transference, for which he proposed the name "telepathy." Janet has not committed himself to any explanation. The distances in these tests varied from a third of a mile to several miles.<sup>2</sup>

A number of other reports appearing in the *Proceedings* of the Society for Psychical Research of London might be listed as favoring the view that distance is not inhibitory to thought-transference, but in some of these cases it is not easy to judge the value. The Miles-Ramsden and Wales-Somers reports come in this class. A study, however, by Usher and Burt introduced the quantitatively measurable method of using playing cards. In 60 trials with distances of 120 to 960 miles, there were four complete successes (where one hit in 52 is expected on the average), and 14 were right in value (with one in 13 expected). They had better success with agent (or sender) and percipient in the same room; namely, nine complete successes in 36. Although the number of trials is small for the purposes of conclusion, such as it is, the results (14 right in value) are statistically reliable.

The trans-Atlantic series of tests in telepathy between Paris and New York, conducted by Warcollier and Gardner Murphy in the early 1920's, was regarded by the experimenters as promising but inconclusive. Two radio broadcast experiments in extra-sensory perception were conducted during the same decade, one by Murphy and Gault from Chicago, and another by Woolley and Lodge from London. No

<sup>1</sup> *Proceedings of the Society for Psychical Research*, 1882, I, 241-242.

<sup>2</sup> See review of this work in Frank Podmore, *Apparitions and Thought-Transference* (London and New York: 1915.)

conclusive results favorable to extra-sensory perception have been reported from these, to my knowledge. The London report<sup>3</sup> is negative. Likewise negative was the sequel series of tests at a distance conducted by Soal following the broadcast test.

It would not be correct to say that distance as a condition in an ESP experiment was under test in all the instances of the last paragraphs. Distance was merely one of a number of other conditions, any of which might have been responsible for the low results. For example, the following conditions might have interfered: (1) Lack of personal acquaintance between agents and percipients; (2) the mass approach, using many subjects at once and thus lowering the individual interest; (3) absence of specific acquaintance on the part of the percipients with the exact location of the agents; (4) no preliminary confidence-building, lack of which might have made the long-distance test seem overwhelmingly difficult or impossible to the subjects.

What is required for a study of distance effects is, of course, a comparative series, with the non-distant control series preferably first and as close to the distance series in time as possible. Therefore the failure of the distance series just reviewed may not be held as of weight against those like the Usher and Burt study and others to be described, where the tests were comparative.

The most conclusive series of tests of extra-sensory perception at a distance published up to the time of its appearance in print was that reported in *Mental Radio*,<sup>4</sup> by Upton Sinclair. The tests were made with the author's wife as percipient, and for the distance series a brother-in-law of the percipient acted as agent. The distance was twenty-five to thirty miles; the subject matter of the tests consisted of freely chosen objects drawn on paper by the agent and "concentrated upon" by him at the appointed time, at which the percipient was attempting to draw the object. The seven trials made were recognizably successful. This is a small number, and success is partly a matter of personal judgment; but while there may be something lacking for entire conclusiveness in the experiment, there can be little dispute as to the success shown in these trials. The distance results were among the most successful of the entire series.

In a monograph entitled *Extra-Sensory Perception*,<sup>5</sup> I reported in 1934 experiments made to distinguish the extra-sensory perception of objects, commonly called clairvoyance, from that of subjective states in

<sup>3</sup> *Proc. S. P. R.*, 1928, XXXVIII, 1-9. For Soal's Report see *Proc. S. P. R.*, 1932, XL, 165-362.

<sup>4</sup> New York: 1930.

<sup>5</sup> Boston: 1934; London: 1935.

another person, known as telepathy. In both experimental series (telepathic and clairvoyant) there were some tests attempted with distance as a condition, distance that varied from a few feet to a matter of miles. The condition of spatial separation of subject and object (or agent) served the additional purpose of further reducing the likelihood of sensory cues (in all distance tests, at least one wall separated the percipient from the agent or cards), and was often used with this feature in mind. In an article published later<sup>6</sup> I reported also the results of interposing short distances in tests for extra-sensory perception, both telepathic and clairvoyant, with a professional medium. A brief resumé of these experiments is needed for a judgment of the status of the problem of distance effects in ESP tests.

The first series of distance tests for ESP is the Pearce-Stuart series, in which Mr. C. E. Stuart was the experimenter. This was an attempt to measure the effect of distance on three different techniques (DT, BT, and PT<sup>7</sup>) used in the tests at that time. Without going into the comparison of the different techniques, which was not especially fruitful, I will mention only the gross comparisons at the different distances.

In 2,400 trials made with percipient and cards (or agent in PT) in the same room, the average was 6.7 hits per 25. At a distance of 8 to 12 feet, with one wall between, in 1,925 trials the average hits per 25 were only 5.65; while at 28 to 30 feet and two walls between, 1,075 trials averaged as low as 5.24 per 25. This looks like a marked decline with distance, even though not a proportional one. In spite of the fact that the 3,000 trials made with the two different distances averaged 5.5 hits per 25, which gives a critical ratio of 2.7 and is significant, there was so large a drop in scoring with the introduction of the distance condition that it would appear that this condition was in some way responsible.

A similar effect was obtained in the Cooper-Ownbey series, with Miss Sara Ownbey as observer, this time with PT technique. In 1,800 trials with both agent and percipient in the same room, the average was 9.2 hits per 25; whereas with the two in adjoining rooms and about ten feet apart, the score average for 300 trials dropped to 5.8.

It would seem here, as with the Pearce-Stuart series, that distance is an inhibiting condition for extra-sensory perception, but there is a subtle but logical alternative hypothesis to consider; namely, that perhaps the effect was an indirect one. The subject may have been affected

<sup>6</sup> *Character and Personality*, 1934, III, 91-111.

<sup>7</sup> See glossary, p. 000.

by his doubts, let us say, of his ability to work at a distance. Fear of distance, rather than distance itself, might have been responsible. Only further tests could determine this.

With the Bailey-Ownbey series the results were quite different. With agent and percipient in the same room, using the PT technique, the 275 trials averaged 11.4 hits per 25; about ten feet apart, in adjoining rooms, the 450 trials dropped to an average of 9.7 hits per 25. But 28 to 30 feet apart, and each in the second room away from the other, the 150 trials rose to an average of 12.0 hits per 25, or slightly above the normal.

Likewise, the Zirkle-Ownbey series yielded results showing no decline with distance. With both agent and percipient in the same room, the 950 trials yielded 14.0 hits per 25; at the next stage, with ten feet between, the 750 trials gave 14.6 per 25; and at the distance of 28 to 30 feet, the 250 trials averaged 16 hits per 25. Here again, as with the Bailey-Ownbey work, there was a slight increase of score average with distance.

The tests for the last two series mentioned followed the PT technique, which allowed for telepathy only, and excluded clairvoyant perception. But in the Garrett series both types of extra-sensory perception were tested, and in both types the distance trials gave a higher average than did the normal or close-up series. For the telepathic tests the normal average in 875 trials, consisting of only those on the same days as the distance runs, was 8.3; while in 1,075 trials at a distance between agent and percipient of 15 to 30 feet, the average was 9.0 hits per 25. In the clairvoyance tests there was an average of only 5.2 hits per 25 for the 3,750 trials with cards and percipient in the same room, but in the 3,725 distance tests the average was 5.8. The distance of 15 to 30 feet, then, was evidently not a hindrance in any case for this subject under the conditions. As with the two preceding cases, there was actually a slight favoring of the distance condition in score averages.

Passing on to an account of the greater distances introduced, we find again not a complete agreement in the results. For example, Zirkle, who did so well at close range, even improving up to 30 feet in distance from the agent, when separated from the same agent by 165 miles had no appreciable deviation from chance average.

On the other hand, Pearce, who got nearly chance results at short distances of from 8 to 30 feet, got fairly high ones at 100 to 250 yards in clairvoyant tests. Cooper, too, failed at distances of 8 to 30 feet, and later at a distance of two miles (in work which I am not including

because I am not personally able to vouch for it, though I know nothing wrong with the conditions) he too did splendidly when his mother was able to act as agent. It is well to keep in mind that the initial failures (or near failures) of Pearce and Cooper may have been due to their having been pioneers; confidence could not have been so high at the outset as it was later, after some success by others. This is a possible explanation.

The Pearce-Pratt series has been given in full detail elsewhere,<sup>8</sup> but deserves a summary here. This work was done entirely with the clairvoyant phase of extra-sensory perception and relied upon synchronized watches for keeping track of the order in the tests. Mr. J. G. Pratt handled the cards in one building, while the percipient in another, 100 to 250 yards away, made calls as to what the cards were as each was individually isolated and removed for later checking.

In the total 1,850 tests made with the subject 100 and 250 yards distant from the cards, an average of 7.4 was obtained; but this was less than the average of 8.2 which he got in the 900 trials made in the same room during the same period. A drop from 8.3 to 7.4, however, is not great. But, more to the point, an analysis of the distance series shows that the first 300 trials at 100 yards gave an average of 9.9 per 25; and all 750 trials made at that distance yield 8.7 hits per 25 trials, a figure above the control average.

It is true, the distance series at 250 yards fell to 6.7 hits per 25. But an examination of the scores shows that Pearce was quite as high on certain days at this greater distance as he was at 100 yards, or at one yard. For example, he scored 10 or better 13 times in 44 runs, usually having markedly high days and very low ones. The first day the scores were 12 and 10; the next, 6 and 4; the third, 10 and 10; the fourth, 2 and 6. For some reason, he was very unstable and swung from one to the other extreme. For instance, he made zero three times in this series; for him this was totally unprecedented, and it is very unlikely by chance. The distribution of scores is bimodal, with modes at 10 and 4 hits per 25. Hence it is difficult to regard the effect as due to distance. With the increased score average at 100 yards and this peculiar distribution, it would appear that distance does not, up to the limits tried, inhibit high scoring in the tests for extra-sensory perception. The unstable scoring at 250 yards rather suggests psychological disturbances.

Of the previously reported tests of distance effects in ESP, there

<sup>8</sup>J. B. Rhine, "Some selected experiments in extra-sensory perception," *Journal of Abnormal and Social Psychology*, 1936, XXXI, 216-228. (Also *Journal of Parapsychology*, 1937, I, 70-80.)

remains only the short Turner-Ownbey series of telepathy tests, with over 250 miles' separation between Miss Ownbey, the agent, and Miss Turner. The normal series averaged in 275 trials 7.7 hits per 25, while the long distance series gave an average of 10.1 hits per 25 in the 200 trials made. Only 25 trials per day were made, with three and five-minute intervals, with the use of synchronized watches to keep the order of the calls. After eight days the series was stopped, the scores having been 19, 16, 16, 7, 7, 8, 6, 2.<sup>9</sup> The first three days' scores were the highest ever obtained by the subject, and as a succession were very unusual by anyone. It could scarcely be concluded by those who accept these results that distance is a limiting condition in ESP.

These series of tests just reviewed represent the data upon which a conclusion was reached in the report of 1934 that ESP was not limited by space as is sensory perception. There were only a few other short series that gave approximately chance results. Four runs of PT at a still greater distance (over 300 miles) by Miss Turner and Miss Ownbey gave only chance average, 5.0 per 25. Five runs at card-calling at 300 miles averaged 6 hits per 25 until interrupted. Cooper with Miss Ownbey averaged 4.4 for 7 runs at 7 miles, and Zirkle 5.5 per 25 in 10 runs at 165 miles.

The failures do not, I think, affect the argument. If in any one series the distance tests hold up reliably to the same extra-chance level as the normal ones, this suffices to eliminate space considerations from ESP functioning. Such series were obtained in four short distance tests at 8 to 30 feet, and in two long distance groups at from 100 yards to 250 miles.

#### FURTHER EXPERIMENTS

There have been no further experiments designed purely and primarily to investigate the effect of distance on ESP. It was felt that the cumulative effect of the results of past investigations was adequate to show that distance was not a limiting condition in ESP tests. Other urgent exploratory problems lay ahead, and the question of distance was no longer a primary issue. Such tests involving distance as were made afterward were conducted with spatial separation as a secondary aspect. In none of these, therefore, are there comparative or control series made under close-up conditions (i.e., comparable except for distance).

<sup>9</sup>The safeguards and alternative hypotheses of these series are treated elsewhere: Rhine, *op. cit.* Those who question the safety of stopping a series when extra-chance scoring ceases would have to argue that because a gambling slot-machine gave a good yield one time it would be likely to do so on the next. Warner has dealt with this run-of-luck theory in "The Rôle of Luck in ESP Data," *J. Parapsychol.*, 1937, I, 84-92.

In the experiments to be reported, the actual primary objectives of the tests will not be described here, since to do this would involve prohibitive detail and digression. They will be reported later with the main bearing in view. Here will be reported only what is essential to the evaluation of the bearing of distance on success in ESP tests.

There have been conducted, since the last distance work was reported, a total of 102,075 trials that have had distance as a condition. As in the earlier work, there will be no selection of the data in publication. All tests made will be included.

Although a few of these trials were made at distances to be measured in feet, the great majority were made at distances ranging from 70 to 3,000 miles. It is these comparative distances that throw light on the problem of the effect of distance on ESP scores. Unfortunately, the different distances did not obtain for tests with the same subjects. The subjects were different for each distance. Only the relatively large number of cases, therefore, prevent the results at different distances from being ascribable to possible individual differences.

The total of 101,450 trials have an average of 5.105 hits per 25 and a positive deviation of 430 from the mean. This is 3.4 times the standard deviation of the mean. Moreover, the extra-chance character of the series is supported by other treatments of the results. If the frequency of occurrence of the various scores is evaluated by the chi-square method,<sup>10</sup> it yields (with sub-series combined by Fisher's chi-square technique) the probability of .00234 that the results are due to chance. By the sign test modification<sup>11</sup> of the chi-square method, the probability of chance causation is still lower. It is .00014; that is, the odds are about 7000 to one that the results are not due to chance.

The main portion of the 4,083 runs was made with the GESP method, with an agent looking at the ESP cards and thus providing two possible stimuli for the percipient; i.e., both the card and the mental image of it. Second in number were the runs made with the pack untouched throughout the run; i.e., with DT. The few exceptions will be accounted for below in the detailed presentation of the different sub-series.

Under the circumstances, the alternative hypothesis of sensory cues is not applicable. Barring chance, improper selection of results, sensory cues, and an inadequate method of evaluation, it is difficult to find an

<sup>10</sup> See R. A. Fisher, *Statistical Methods for Research Workers*, 3rd Ed. (London: Oliver & Bond, 1934).

<sup>11</sup> J. A. Greenwood and C. E. Stuart, "Mathematical Techniques used in ESP Research," *J. Parapsychol.*, 1937, I, p. 217 ff.

explanatory hypothesis for these results, except to say that, since some perceptual response is indicated under conditions in which the senses are not capable of operating, extra-sensorial perception is apparently occurring. This tentative, non-committal position is taken as the minimal conclusion that can describe the results obtained.

Entirely apart, then, from their bearing on the distance problem, the results of this series add still further support to the establishment of the occurrence of ESP.

#### DISTANCE COMPARISONS

##### *Tarkio Series.*

In 26,125 trials conducted during the summer of 1937 by Mr. James MacFarland of Tarkio College, the comparison of score averages in relation to various distances indicates again that distance is not a limiting condition.

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. 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.

The results are given by zones in Table I, with averages per 25 for easy comparison. Note that the fifth zone yielded an average higher than the fourth and the third higher than the second. This could hardly occur if that distance were a hindrance to ESP. Otherwise, these results could only be accounted for by large individual differences sufficient to overcome the supposed distance effect and happening to be distributed so as to do so. That this might have been the case is possible, but not very likely.

##### *Duke Series.*

This series was suggested by Dr. X, a well known psychiatrist of San Diego, California, who had in preliminary tests discovered a number of high-scoring subjects in ESP tests among her friends and acquaintances. The records of these first experiments were not kept by Dr. X, but a later series, made after the distance tests, averaged 6.9 hits per 25 for 5 subjects in a total of 3,500 trials, with a critical ratio of 11.2. These tests were made by the DT method, the records were

TABLE I  
SCORE AVERAGES AT VARIOUS DISTANCES, TARKIO SERIES.

Zone	Distance in miles	Number of Subjects	Total Trials	Total Deviation	Average Hits Per 25
1.....	0— 50	16	12,125	282±44	5.58
2.....	50— 100	2	2,750	47±21	5.43
3.....	100— 300	7	6,625	170±33	5.64
4.....	300— 600	2	2,125	17±19	5.19
5.....	600—1,400	3	2,500	30±20	5.30
Total.....	0—1,400	30	26,125	546±66(CR=8.3)	5.52

double-checked, and Dr. X was the experimenter as well as one of the successful subjects.

It was arranged as a result of the suggestion of Dr. X that a series of tests using the GESP method be conducted with the cards and agent at Duke University and the subjects at San Diego, and that after a time, in a second sub-series, the order be reversed. A third sub-series was carried out, with a return to the first arrangement; namely, the cards and agent at Duke. These three sub-series are thrown together to get larger numbers for comparison.

At the same time that the Duke-San Diego series was under way, there were others invited to participate from other localities and at other distances from Durham. These afford the material for comparison in relation to distance. The results of all the series up to a certain point will be thrown together without omissions, for purposes of comparison. At that point results which had been definitely positive took a turn toward negative deviation which was maintained until the end. Thus from the beginning on November 13, 1935, to March 13, 1936, the 23,825 trials were not only positive in the deviation of hits from the mean, but were significantly so, giving a critical ratio of 2.8. After March 12, during the five more weeks that the experiment was continued, a total of 34,450 trials was made; but the results averaged either chance expectation or slightly below it throughout this period.

There is, of course, no point in including these latter data in our efforts to compare scores made at various distances. (They were, however, included in the total given above, 102,075.) If average scores are not above chance, there is nothing to say. Hence the plan to base this comparison on only that section of the experiment in which positive scoring occurred, and so to take all scores from the start on November 13th up to March 12th. In this section we find the following results, presented in Table II.

TABLE II  
SCORE AVERAGES AT DIFFERENT DISTANCES, DUKE SERIES.

Zone	Distance in miles	Total Trials	Total Deviation	Average hits per 25
1.....	0— 500	2,750	+ 32	5.29
2.....	500—1,000	7,300	+ 58	5.20
3.....	1,000—2,000	700	- 2	4.93
4.....	2,000—3,000	13,175	+ 85	5.16
Total.....	0—3,000	23,925	+173	5.15

The strongest point that could with safety be made from this data bearing on the question of distance effects is that the difference between the results of Zone 4 and Zone 2 is too slight to represent a decline with distance. There is a difference of nearly 2,000 miles in distance between these two zones, and so the score difference of .04 is insignificant and quite out of proportion to the distance involved. The same is true of the difference between Zones 1 and 2.

The fact is, the best scores, on the whole, made during the Duke series were made by the San Diego group of subjects, although this group was located at the greatest distance from Duke. In this Duke-California work there were three special conditions under test which will be reported in a later paper. In these distance was only incidental. But in the best of these conditions the San Diego group averaged 5.54, which is better than the comparable average secured by any group in any closer zone.

While, then, this Duke "broadcast" series was relatively low in scoring level, it is entirely in line with the other evidence on the distance question, showing, that in so far as it has been tested distance is not in itself a limiting condition to ESP.

#### *Other Series.*

There were some series of tests included in the total figures given at the beginning of this section which have no very conclusive bearing on this problem of distance and so need not be discussed in detail. They gave, on the whole, either chance or sub-chance scores. They will be reported in detail in later articles covering the specific problems concerned.

#### DISCUSSION

The many difficulties of this field of research have made it impossible as yet to get an exact and elaborate study of the question of whether distance affects ESP scoring, but the evidence that it does not influence it is nevertheless considerable.

The case would be much more doubtful if there were any good evidence pointing in the opposite direction. If any suggestion of inverse correlation with distance had been found, the situation would be different. Such instances as we have of failure under distance conditions seem to be due to factors other than distance, such as loss of confidence. This is shown by the fact that, in most instances, the same subject later is able to succeed under conditions of distance sometimes greater than that with which he failed at first. In any case, failure by one subject and success by another under similar distance conditions indicates that the failure could not have been due to the spatial conditions. If distance does limit, even one real exception could not occur. But the evidence offers a number of such exceptions.

While, therefore, the nice demonstration that is to be desired on this problem is still wanting, the question itself is apparently answered effectively. The demonstrations best meeting the requirements are the Pearce-Pratt and the Turner-Ownbey series. But the evidence from these is considerably strengthened by the results of the shorter distance series and of the Tarkio and Duke series, in which distance was only an incidental condition.

If distance has no limiting influence upon ESP, the logical bearing of this spatial freedom is extremely far-reaching. It is so far-reaching that every alternative must be examined before it is decided that distance is not a possible limiting factor.

Is it possible that the distances tried in the tests herein reported were all too short? They would not be too short to test the energies we know. In all mechanical effects such as might conceivably be involved in transmitting symbols from cards, the distances of 100 and 250 yards in the Pearce-Pratt series is entirely adequate. 250 miles for the Turner-Ownbey series is enough more to add good measure.

Competent physicists appear to agree that nothing known at present in our spectrum of electromagnetic energies, nor outside of it, is applicable to the results of the ESP tests. Some of these maintain, however, that the phenomena should not hastily be relegated to the realm of the extra-physical, since physics may expand and since there must be some physical basis for a perception of an object. This apparently reasonable position is doubtless where the problem must be left until further results throw better light on the ESP process. Finding no mechanical explanation (as mechanics is today) need not necessarily exclude a physical explanation by a physics which may yet develop.

The problem of a physical theory of ESP is focussed more sharply

upon the next logical step in the experimentation itself. If distance in space is not a limiting condition in ESP tests, what about remoteness in time? In this space-time world, wherein any measurable objective event is necessarily both spatial *and* temporal at once, if anything could escape space it might be expected to be free from time as well. If an ESP subject can in some way transcend spatial barriers, might he not be able likewise to transcend temporal limits?

The summary of distance work in ESP presented in this paper is made to preface a series of reports on research which has been made into the problem of precognitive or previsionary ESP, which will follow.

#### SUMMARY

This report is a review of experiments made to ascertain the effect of distance on success in ESP tests. The work already reported seems to show that distance does not limit ESP as it does sensory perception. Particularly those Duke experiments known as the Pearce-Pratt and the Turner-Ownbey series demonstrate this by reason of the control series needed for comparison. In a total of 101,450 trials in which distance was a condition and which gave as a whole a significant deviation, there was shown in the two series in which there were score averages at different distances that there was no proportionate falling off of average with distance, such as a mechanical analogy would require. In fact, it would appear that no falling off whatever directly occurs with spatial separation.

## CHANCE AND EXTRA-CHANCE RESULTS IN CARD MATCHING

DOROTHY R. MARTIN

---

**ABSTRACT:** Thirty-nine university students made a total of 76,525 trials with standard ESP cards. Of these, nineteen subjects were carried through a minimum of 1,000 trials. Among these, five were found to make averages of high statistical significance, yielding critical ratios of 10 or over. The 'DT' or 'down through' technique was quite generally used.

Two control series were employed: The first involved 25,000 trials of 'pure chance' matching of two packs of ESP cards. The second control consisted of a 'reverse check,' in which the guesses of each subject were checked against the reverse order of the pack, thereby yielding for each subject a chance score on each run of 25. For most subjects, the guessed or 'direct' average was found to be above the average for the reverse check.

Miss Martin is an instructor in Psychology at the University of Colorado.

---

RECOGNIZING the necessity for widespread verification of the findings of Dr. J. B. Rhine and his associates, we have undertaken experiments closely resembling those performed in the Duke University laboratories, and have used statistical techniques outlined by Dr. Rhine, with the purpose of ascertaining whether or not results comparable to his could be obtained in our laboratory. The investigations here reported we regard as distinctly exploratory in character.

### GENERAL CONDITIONS

The materials used throughout the tests consisted of standard ESP cards. These cards were arranged in packs of twenty-five, there being in each pack five cards of each design. Thus, pure chance would yield an expectancy of five correct guesses out of each twenty-five.

The thirty-nine subjects participating were all University students, most of whom were Sophomores taking the required course in General Psychology. The group was quite unselected; the individuals volunteered. They were almost equally divided as to sex.

*General Procedure:* After certain preliminary tests with experimental procedures, the method usually referred to as 'DT' or 'down through'

was selected for general use. This method admits of rather perfect control, since the possibility of sensory cues seems here eliminated.

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. (See Table II.) It should be noted that neither the subject nor the experimenter had any knowledge of the actual order of the pack until after all twenty-five guesses were made.

#### CONTROL SERIES

(1) *'Pure chance' Series*: Two control series were used. The first is one which has been reported elsewhere, namely: Two well shuffled packs were turned up simultaneously, and the coincidences noted.<sup>1</sup> 25,000 such trials or 1,000 runs of 25 were made. The results quite closely coincide with those reported by Carpenter and Phalen for 10,000 trials. (*Journal of Parapsychology*, I, pp. 31-43). They also parallel fairly closely the theoretical curve of the binomial expansion.

(2) *'Reverse check' Series*: The second control series permits a striking comparison of a particular subject's 'direct' score with his own chance record for the same series. In it, the subject's guesses for each run of 25 cards 'down through' were checked against the reverse order of the pack, or 'up through,' and the hits noted. Through this device, scoring peculiarities due either to unusual card sequences or to possible idiosyncrasies of the subject in choice (e.g., design preference) and in ordering of guesses should become evident. Should such factors be negligible, these 'reverse' hits may on the average be expected to coincide with chance expectancy. (So they did. See Table I.)

#### RESULTS

Table I presents all results and the 'reverse check' for each of thirty-nine subjects. It will be seen that in the total series of 76,525 trials, there was an average of 5.8 hits per 25 trials. This gives a critical ratio of 22.2. In Table II are given the results of those trials recorded under double observation.

<sup>1</sup>We realize that our 'pure chance' procedure, namely matching two packs of cards, is not strictly comparable to the 'guessing' of a pack by a subject who does not attempt to get precisely five of each design. We are planning to remedy this defect in 'pure chance' matching in future experiments.

## DETAILED DISCUSSION OF THE RESULTS OF FIVE SUBJECTS AND OF THE EXPERIMENTAL TECHNIQUES USED

For Mr. C. B. and Mr. C. J. there was no departure from the general 'DT' methods previously outlined. Many of these trials were observed by a third person. A record of these doubly observed trials is presented in Table II. The record of Mr. C. J. showed marked improvement in scoring with practice. His average for the first 2,500 trials was 5.3, for the last 2,500 trials it was 7.66. Neither was there any change in the methods used with Miss B. C., though one idiosyncrasy of hers is worthy of note. She always guessed exactly five cards of each design, and usually, though not always, scattered all of the symbols for one design throughout the column before writing another symbol. The experimenter encouraged her in the use of this method as it seemed an interesting variation from the usual techniques. Such a method compares closely with the 'pure chance' matching of two packs. Accordingly her average of 6.6 and critical ratio of 11.6 is especially interesting.

With Miss E. W., and with her alone, there was a major departure from the 'DT' technique. She preferred our continuing a method tried in preliminary runs and usually referred to as GESP. The experimenter, holding the pack, looked at each card as she removed it and placed it face downward on the table with an audible click; the subject then recorded a guess. After 25 cards were guessed, the hits were checked against the order of the pack. This method, the author is aware, is subject to criticism from the point of view of controls. There exists the remote possibility that an occasional card might be seen by the subject, together with the possibility of occasional reflection from the experimenter's eyes. (Here, however, the subject rarely glanced in the direction of the experimenter, and frequently sat with back turned, getting signals only from the click of the cards.) Again, there is the further possibility of unintentional cues given by the experimenter. A screen, which should eliminate most of these factors, was once introduced. On this occasion her average for 250 trials dropped to 4.0. (The subject, however, had reported fatigue and disinclination to work.) This series was followed by a return to the old method, with subject's back turned, and a re-establishment of high scoring, (averaging 7.1 for 1,250 trials). Screened trials will be resumed with this subject in the fall. For the present, her total average of 7.44 and critical ratio of 15.4 should be considered with some skepticism.

With Miss D. W. distance was introduced as an experimental factor. Her results (see Table III) are divided into three series on the basis of the degrees of distance which separated her from the cards.

*Series 1.* This preliminary series of 101 runs or 2,525 trials involved no particular controls other than those imposed by the 'DT' method and the usual caution in experimentation. The average per 25 of 6.9 and the critical ratio of 9.9 gave ample warrant for further experimentation.

*Series 2.* This series of 60 runs or 1,500 trials (completed in just two hours!) was conducted with the subject and experimenter as far apart as a room 20' x 12' would permit. The pack was checked with both subject and experimenter observing the results. This series yielded an average of 7.6 and a critical ratio of 10.0.

*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 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. This series yielded an average of 7.8 with a critical ratio of 7.6.

In series 2 and 3 of Table III the high consistency of scores above chance seems worthy of note:

No. of hits below 5.....	5
No. of hits of 5.....	15
No. of hits 6-9 inclusive.....	50
No. of hits 10-14 inclusive.....	20
	—
Total No. of Runs.....	90

A summary of the data in these series is recorded in Table III.

#### DISCUSSION

Table I shows that five subjects out of thirty-nine<sup>4</sup> have averages of sufficient statistical significance (their critical ratios being 10 or more) to be regarded as quite beyond the operation of pure chance factors alone, granted that the mathematical logic is sound. Two others have critical ratios above the usual 2.5 criterion of significance. One subject, J. K., has an average low enough to suggest slight negative significance.

It may be emphasized that in this study, subjects making 'chance' scores were not dropped, as was the practice in many previous studies. Thus, since we have a large number of 'chance' scorers, and since the 'reverse' hits for most subjects fall below their 'direct' hits, (Table I)

<sup>4</sup>Other experimenters have found a similar relative frequency of 'good' subjects. Carpenter and Phalen found four out of sixty-five subjects who scored significantly; Woodruff and George found three out of seventeen. (*J. Parapsychol.*, 1937, I, 18-43.)

CHANCE AND EXTRA-CHANCE RESULTS IN CARD MATCHING 189

TABLE I\*

TEST SERIES							REVERSE SERIES			
Name	Trials	Score	Av./25	Dev. from np	S	X	Score	Av./25	Dev. from np	X
D. W.....	4,775	1,394	7.3	439	27.64	15.9	963	4.9	+ 8	+ .29
E. W.....	4,500	1,311	7.2	411	26.83	15.3	950	5.2	+50	+1.86
C. J.....	9,500	2,438	6.4	538	38.98	13.8	1,889	4.97	-11	-.28
B. C.....	4,750	1,270	6.68	320	27.56	11.6	995	5.2	+45	+1.63
C. B.....	8,500	2,071	6.09	371	36.87	10.0	1,668	4.9	-32	-.86
E. A.....	4,500	1,008	5.6	108	26.83	4.0	852	4.6	-48	-1.78
C. A.....	1,000	244	6.1	44	12.65	3.4	223	5.5	+23	+1.81
C. D.....	1,750	389	5.5	39	16.73	2.3	324	4.6	-26	-1.55
E. R.....	250	62	6.2	12	6.32	1.8	47	4.7	- 3	-.44
A. B.....	3,000	637	5.3	37	21.91	1.6	576	4.8	-24	-1.09
J. H.....	6,500	1,353	5.2	53	32.25	1.6	1,311	5.0	+11	+ .34
D. M.....	7,000	1,451	5.18	51	33.46	1.5	1,426	5.0	+26	+ .77
L. F.....	250	59	5.9	9	6.32	1.4	46	4.6	- 4	-.63
B. P.....	1,500	318	5.3	18	15.49	1.1	274	4.6	-26	-1.67
L.D.G.....	1,000	215	5.3	15	12.65	1.1	200	5.0	0	0
R. LaF.....	575	126	5.4	11	9.59	1.1	110	4.8	- 5	-.52
R. M.....	500	110	5.5	10	8.94	1.1	89	4.4	-11	-1.23
B. N.....	250	57	5.7	7	6.32	1.1	42	4.2	- 8	-1.26
M. W.....	1,000	213	5.3	13	12.65	1.0	192	4.8	- 8	-.63
V. M.....	1,000	212	5.3	12	12.65	.9	208	5.2	+ 8	+ .63
J. W.....	250	55	5.5	5	6.32	.7	48	4.8	- 2	-.31
C. McS.....	1,400	290	5.1	10	14.96	.6	255	4.5	-25	-.67
M. DeB.....	750	157	5.2	7	10.95	.6	143	4.8	- 7	-.63
G. B.....	750	157	5.2	7	10.95	.6	164	5.4	+14	+1.27
L. L.....	250	53	5.3	3	6.32	.4	45	4.5	- 5	-.79
E. J.....	2,750	557	5.0	7	20.97	.3	580	5.2	-30	+1.43
R. K.....	250	52	5.2	2	6.32	.3	56	5.6	+ 6	+ .94
L. B.....	750	152	5.0	2	10.95	.1	132	4.4	-18	-.64
A. L.....	750	151	5.0	1	10.95	.09	159	5.3	+9	+ .82
G. D.....	750	149	4.7	- 1	10.95	-.09	140	4.66	-10	-.91
H. P.....	625	122	4.8	- 3	10.00	-.3	120	5.0	- 5	-.50
F. G.....	500	95	4.7	- 5	8.94	-.5	100	5.0	0	0
K. K.....	500	94	4.7	- 6	8.94	-.6	98	4.9	- 2	-.22
A. Ber.....	250	45	4.5	- 5	6.32	-.7	57	5.7	+ 7	+1.12
W. G.....	250	43	4.3	- 7	6.32	-1.1	45	4.5	- 5	-.79
W. K.....	1,000	184	4.6	-16	12.65	-1.2	193	4.8	- 7	-.55
R. W.....	900	163	4.5	-17	12.00	-1.4	181	5.0	+ 1	+ .08
S. N.....	1,000	181	4.5	-19	12.65	-1.5	204	5.1	+ 4	+ .31
J. K.....	750	121	4.0	-29	10.95	-2.6	159	5.2	+ 9	+ .82
Totals...	76,525	17,759	5.80	+2,454	110.64	+22.18	15,265	4.99	-40	-.37
Pure.....										
Chance Series..	25,000	4,978	4.98	-22	63.24	-.34				

\*X' signifies the Critical Ratio, determined by the formula  $X = \frac{d}{s} = \frac{h - np}{\sqrt{npq}}$ , wherein 'd' denotes the deviation from chance expectancy and 's' denotes the standard deviation of chance.

the data suggest the possibility that some slight but widespread ability may be postulated. The need for experimental confirmation or refutation of such a postulate is apparent.

TABLE II  
SOME TRIALS MADE WITH TWO OBSERVERS PRESENT.

Subject	Trials	Score	Deviation from n.p.	S	X
E. W.....	2500	745	+245	20.00	+12.24
C. J.....	2750	772	+222	20.97	+10.58
C. B.....	1750	445	+ 95	16.73	+ 5.67
Total.....	7000	1962	+562	33.46	+16.79

TABLE III  
SUBJECT: MISS D. W.

	DIRECT						REVERSE			
	Trials	Score	Av./ run	D	S	X	Score	Av./ run	D	X
Series 1...	2525	704	6.9	199	20.10	9.9	540	5.3	+35	1.7
Series 2...	1500	456	7.7	156	15.49	10.0	277	4.6	-23	-1.4
Series 3...	750	234	7.8	84	10.95	7.7	146	4.9	-4	-4
Total.....	4775	1394	7.3	439	27.64	15.9	963	4.9	+8	+ .29

#### CONCLUSION

The extra-chance results obtained in this preliminary study suggest the following hypotheses: 1. There may exist some constant error in the experimenter's procedure which is at present unknown. This might include hypersensitivity of the subject to unknown cues which were uncontrolled. 2. The mathematical analysis may be at fault. Such fault might rest with the limited number of trials, or with the statistical interpretation of scores. 3. There may exist in certain subjects an 'extra-sensory' perceptual ability such as Rhine postulates.

.24  
.58  
.67  
.79

## SOME PSYCHOLOGICAL RELATIONS OF EXTRA-SENSORY PERCEPTION

MARGARET H. PEGRAM

---

**ABSTRACT:** The writer made 184,970 calls of ESP cards over a long period of time. The method used was that of calling down through a shuffled pack of 25 or more cards before checking. Of these calls 148,270 were made with the previously specified intention to score above chance average (high-aim), 36,700 were made with the previously specified intention to score below chance average (low-aim). The average scores were relatively close to chance average in both series, but in view of the large number of trials, both series ("high-aim" and "low-aim") are highly significant. The results indicate that extra-sensory perception is amenable to voluntary direction.

X  
1.7  
1.4  
.4  
.29

Analysis of the data showed that the subject was more successful in hitting cards at the ends of the deck than in the middle when trying to score high. The opposite held when low scoring was attempted.

Miss Pegram, now a member of the staff of the Parapsychology Laboratory at Duke University, was undergraduate assistant in Psychology at Guilford College at the time these experiments were conducted.

---

### INTRODUCTION

gest  
in  
ght  
un-  
ult  
ter-  
ra-

**EXTRA-SENSORY PERCEPTION**, like any other little understood phenomenon, needs explanation and interpretation, particularly in regard to its relationship to better known phenomena. Psychologists who recognize ESP as a genuine occurrence are particularly likely to be interested in relating it to known mental processes.

The general aim of this research was to inquire into the nature of ESP.<sup>1</sup> Of special interest at the outset of the work was the question of the rôle of volition in the direction or orientation of ESP. Considerable evidence that ESP may be voluntarily directed is already on record. In fact, some direction of ESP is inherent in the very demonstration of it. This research was undertaken originally with a view to

<sup>1</sup>The writer wishes to express her appreciation to Dr. J. B. Rhine both for his frequent suggestions and criticisms throughout the course of the work and for the materials he supplied for the experiments; to Dr. and Mrs. C. A. Milner for their encouragement and for permission to use the psychology laboratory at Guilford College for the experiments; and to Mr. C. E. Stuart for his assistance with the statistical treatment of the data.

throwing more light upon the respects in which the *direction* of ESP is subject to voluntary control.

The inability of subjects to give an adequate introspective account of ESP processes leaves an impression that ESP is essentially an unconscious function. The possibility of the conscious direction of what at present appears to be an unconscious function demands further consideration. The technique used in studying this aspect of ESP consisted in comparing the results obtained when the subject attempted to guess a pack of cards correctly at will, with the results obtained when the subject attempted to avoid guessing them correctly; specifying, of course, at the start which choice was made.

Two of Rhine's subjects<sup>2</sup> had shown in relatively short series that by trying to miss the card series they could score significantly *below* the theoretical mean. Pearce made 275 calls averaging 1.81 hits per twenty-five where five was chance expectation, and Stuart made 1,300 calls with an average of 3.50 hits per twenty-five. It seemed important to know whether this success in deliberately making either a correct or an incorrect call would hold through a long series for another subject, and what the relative success for these two methods would be.

A phenomenon that can be controlled by will (i.e., by conscious volition) is on a more secure and understandable basis than one which cannot be so controlled. Such a phenomenon serves, in addition to the mathematical evaluation, as a reliable exclusion of the chance hypothesis. Briefly, voluntary control is not a chance control.

Other problems arose in the course of the research and were investigated as far as possible when they were encountered. Still other points of interest reported below arose as the results were studied after the close of the experiment. Some of the results of this experiment were entirely unforeseen by the investigator.

The general bearing of the results, both of the main objective and of the incidental features, tends to indicate further relations between ESP and the more recognized and better understood processes of mind, particularly in regard to phenomena which are of a configurational or Gestalt character.

#### GENERAL SURVEY OF METHODS AND CONDITIONS

The experiments reported here were conducted at Guilford College, North Carolina, during an experimental period extending from December, 1934, to September, 1935.

<sup>2</sup> Rhine, J. B., *Extra-Sensory Perception*, Bruce Humphries: Boston, 1934, pp. 70, 80.

The work is divided into two parts. One part of the work consisted of trials in which the subject's aim was to guess correctly the symbols on the cards towards which the attention was directed; this intention of scoring positively will be denoted by the term "high-aim." The second division is composed of those runs which were made with the aim of avoiding naming the symbol on the card; that is, making scores averaging below chance. This intention of avoiding correct calls will be designated as "low-aim." The direction of the deviation desired was indicated before the calls were made; that is, the subject specified whether the calls were to be made with the desire to attain above-chance scores (high-aim) or below-chance scores (low-aim) before the guesses were made.

The usual ESP cards were used.<sup>3</sup> All computations, unless otherwise stated, are based on a standard pack of 25 cards composed of five each of the five different symbols.

The general experimental procedures are similar to those already described in earlier reports. The procedure most frequently used was that known as DT or "down through," in which a shuffled deck of cards is placed face down before the subject, who records his guesses for each card in order from the top card down through the bottom card of the pack before the pack is disturbed. The calls are then checked and the results recorded.

Less frequently used in this experiment were the BT or "before touching" and the OM, or "open matching" procedures. In BT the top card of a shuffled pack is called, the call recorded, and the card removed and laid aside, face down, still unexamined. The next card is called and the procedure is repeated until calls are made for all the cards. The results are then checked. In OM the subject places each card opposite the one of the five symbols which he feels it to match, placing all the cards in this manner before the results are checked.

In addition to those runs for which the standard pack of twenty-five cards was the basis, there were other runs for which packs containing fifteen, fifty, and seventy-five cards were used. An equal distribution of the five symbols was always maintained. The number of cards in the pack is indicated by a numerical subscript: For example, "DT<sub>75</sub> high-aim" designates that procedure in which the subject calls a pack of seventy-five cards with the aim to score above chance without removing any of the cards from the pack until the calls for the whole pack are completed.

<sup>3</sup> See frontispiece, *Journal of Parapsychology*, Vol. 1, No. 1, March, 1937.

The shuffling customarily used consisted of two or more dovetail shuffles and a cut before each run. New packs of cards were frequently substituted in order to diminish the likelihood of discrimination on the basis of sensory cues. As far as possible, the subject avoided looking at the pack of cards after it had been shuffled. DT procedures were used almost exclusively because they were thought to be more easily controlled to eliminate sensory cues.

The writer served both as subject and observer in all the runs reported here, working alone most of the time, without witnesses. Since the question of error, deliberate or accidental, arises frequently in research, particularly when the subject works alone, it is of interest to mention that since these experiments were completed, the writer has done significant witnessed work in the Duke Laboratory, with averages as high as or higher than those included in this report.

#### RESULTS

*General Results for High-Aim:* During the experimental period, a total of 184,970 single calls were made. Of this number, 148,270 were made with high-aim, or the previously specified intention of scoring above chance. For the high-aim there were 1,789 more correct calls than the mean expectation; the average per 25 was 5.30 where 5 is the mean expectation. This is an average of three hits above chance for every ten runs of 25 calls each.

This average is low compared to the average of 7.1 per 25 for 91,174 calls which Rhine reports for his subjects.<sup>4</sup> But even the comparatively small average of 5.30 obtained in this experiment gives very high odds against chance because of the large number of trials. The deviation of high-aim scores alone may be expected by chance on the average only once in  $10^{30}$  times.<sup>5</sup>

There is no need, in view of the large amount of evidence already presented, to offer this paper as independent proof of the existence of extra-sensory perception, but as additional evidence in favor of ESP it has some weight. Its more important function, however, is to throw new light on ESP as a process related to other psychological phenomena.

*Volitional Dirigibility of ESP.* The original object of these experiments was to investigate more thoroughly the question as to whether or not ESP can be directed by will so that the subject can hit or miss the

<sup>4</sup> Rhine, *op. cit.*, p. 162.

<sup>5</sup> For a brief description of the application of the standard deviation to ESP data see J. G. Pratt's article, "Clairvoyant Blind Matching," *The Journal of Parapsychology*, Vol. 1, No. 1, March 1937, p. 11.

cards as he chooses. The results demonstrate that ESP *is* subject to voluntary control and direction. Runs made with the conscious intention of avoiding correct calls have significantly below-chance averages, while those calls made with the aim of guessing the cards correctly show scores averaging significantly above chance.

36,700 calls of the total of 184,970 calls made were done for low-aim. For this group of runs there was an average of 4.35 per 25, with a critical ratio of 12.31 (see Table I), and anti-chance odds of  $10^{34}$  to 1.

TABLE I  
COMPARISON OF HIGH-AIM AND LOW-AIM RESULTS.

Condition	Number of calls	Number of hits	Average per 25	Deviation	C.R. = D/S.D.
High-aim.....	148,270	31,443	5.30	+1,789	11.61
Low-aim.....	36,700	6,396	4.35	- 944	12.31

The results demonstrate that ESP *can* be directed by will. Not only is the above chance deviation for high-aim and below-chance deviation for low-aim to be observed in the totals, but the same direction of deviation is to be found consistently with the various sized packs used.

The most obvious differences in the results were quantitative: (1) The average deviation from mean expectation for low-aim was over twice as great as the average deviation for high-aim. While the average for the total high-aim was 5.30, or an average of .30 hits above chance expectation for each 25 calls, the average for low-aim was 4.35, or an average deviation of .65 from chance expectation. (2) Because of this greater deviation for low-aim, the anti-chance value of the low-aim is even greater than that for high-aim, although there were only about one fourth as many calls.

This disparity in number of calls is due both to the fact that the low-aim condition was introduced later in the experimental period, and to the fact that it was used proportionately less than high-aim during the time both methods were employed. It is possible that there was a favorable selection of working periods for low-aim. Therefore the difference in average deviation between the low-aim and the high-aim ESP may be at least partly to the fact that there were fewer opportunities for the occurrence of the inhibiting effects of satiation and of unfavorable working conditions.

There are indications, however, that differences intrinsic to the subject's attitude toward the two conditions might have produced the difference in results.

That successful scoring with the two methods is not highly correlated is shown by the data below. For 34,350 consecutive calls for high-aim, the most successful long period of scoring, there was an average score of 5.56 per 25. During the same period there were 16,625 consecutive calls for low-aim with an average of 4.36 per 25. That is, the average deviation from chance with high-aim was .56, while for the same time, the average deviation for the runs with low-aim was .64. Following the same principle used in the above selection, a similar group representing the most successful long period of low-aim was selected. For 3,775 consecutive calls for low-aim there was an average of 4.18 per 25. Thus, for the most successful long period of runs made with the aim of avoiding correct calls, an average deviation of .82 below the mean expectation of 5 was obtained, while the runs aiming towards high scores made during the same period showed an average deviation of only .26 above chance expectation.

It is possible that some intrinsic difference either between the abilities to score above or below chance at will, or between the degrees to which the scores for the two methods are affected by the various aspects of the total situation, is indicated by these results.

*Relation of the Length of the Working Period to the Level of Scoring.* The number of runs done during a single working period varied widely. Forty-eight consecutive days' work were analyzed to discover any possible relationship between the number of runs performed on a single day and the success of the performance. The periods of work for high-aim may be classified as follows: (1) Twelve days for which the number of runs for high-aim ranged from one to 25; (2) Ten days for which there were from 25 to 50 runs; (3) thirteen days for which the number of runs varied between 50 and 100; (4) eight days when the number was from 100 to 200; and (5) five days for which more than 200 runs were done.

The results are presented both in the form of a table (Table II) and of a graph (Graph I).

A similar treatment of all the low-aim ESP for the same period gives the results presented in Table III and Graph II.

The data comparing the average scores of high-aim for working periods of different lengths seem to indicate a positive relation between the number of runs done per day and the averages obtained: The most successful days are those on which from 25 to 100 runs for high-aim were done on a single day. Those days on which that number was exceeded show a decided drop in scoring average. Either a fatigue effect

TABLE II

AVERAGE FOR HIGH-AIM ACCORDING TO THE NUMBER OF RUNS DONE PER DAY.

Number of working periods	Number of runs done per day	Total hits/runs	Average per 25
12.....	1 - 25	1008/ 185	5.44
10.....	25 - 50	2121/ 374	5.66
13.....	50 - 100	5420/ 973	5.67
8.....	100 - 200	5718/1081	5.18
5.....	200 - 300	6637/1282	5.18
Total 48.....	1 - 300	20905/3895	5.36

TABLE III

AVERAGES FOR LOW-AIM ACCORDING TO THE NUMBER OF RUNS DONE PER DAY.

Number of working periods	Number of runs done per day	Total hits/runs	Average per 25
17.....	1 - 25	1159/258	4.49
8.....	25 - 50	1224/282	4.34
4.....	50 - 100	1193/275	4.33
2.....	100 - 200	1035/237	4.36
Total 31.....	1 - 200	4611/1052	4.38

or a loss of interest is suggested as a possible explanation. However, this apparent optimum for the number of runs to be done on a single day does not withstand further analysis.

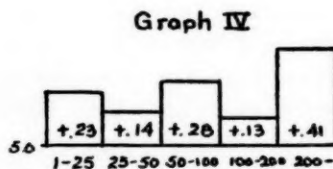
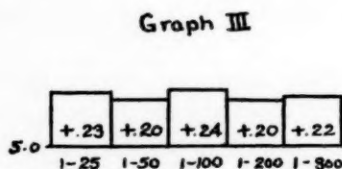
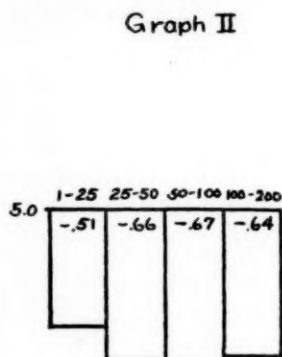
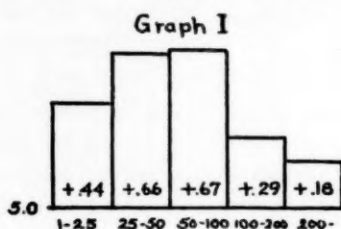
If this decline were due to fatigue effect, the average for the first 25 runs done on any given day should correspond to the average for those days on which the number of runs done ranged from one to 25; the group of runs between the first and the fiftieth should be similar to a comparable group of runs for those days on which the total number of runs does not exceed fifty, etc.

To see whether this relationship *did* hold, those runs for the thirteen days on which more than 100 runs for high-aim were made were analyzed. Averages were obtained for the first 25 runs, for the first 50 runs, etc., using the same groupings that were used for the first study. That is, the divisions are made and averages are stated in terms of what would have been the totals if the work had been discontinued at these various intervals. This analysis, as is shown in Table IV and Graph III, does not confirm the apparent relation between the number of runs done and the results obtained: The level of scoring is remarkably constant throughout. Further analysis of the same data was made by obtaining the averages for various groups of runs throughout the day;

TABLE IV  
RELATION OF POSITION OF THE RUNS IN THE SERIES TO THE AVERAGE SCORE FOR SERIES CONSISTING OF MORE THAN 100 RUNS.

Position in series	Hits/Runs	Av.
1 - 25.....	1,701/ 325	5.23
1 - 50.....	3,382/ 650	5.20
1 - 100.....	6,816/1,300	5.24
1 - 200.....	10,828/2,081	5.20
1 - 300.....	12,355/2,363	5.22

that is, averages for the first 25 runs, for the runs between the twenty-fifth and the fiftieth, etc. (not the cumulative average of the first 50 runs, as in the preceding study.) The results of this analysis (see Table V and Graph IV) show that the highest average for the whole



GRAPH I. Average deviation from chance for high-aim according to number of runs done per day.

GRAPH II. Average deviation from chance for low-aim according to the number of runs done per day.

GRAPH III. Average deviation from chance for series consisting of 100 runs or more when the average up to a given number of runs is considered separately.

GRAPH IV. Average deviation from chance for high-aim of runs according to position in a long series.

group of runs is for those runs which were done last in the day; that is, for the runs done after 200 runs had already been done during that day.

Furthermore, reference to observations made at the time of the experiment reveals that the number of runs done during these days was in part determined by the results. Notes were made of instances in which, because of some mood or emotional condition of the subject, the average previously obtained was lowered; and of other cases in which such inhibiting conditions prevented the establishing of a successful level of scoring. In either case, the number of runs was prolonged by the subject's reluctance to discontinue work with a below-chance average. The high average for the last part of the work is indicative of the positive effect of the subject's determination to remedy the below-chance average.

TABLE V

AVERAGE FOR HIGH-AIM ACCORDING TO POSITION IN SERIES OF WORKING PERIOD.

Number of working periods	Position of run in series	Total hits/runs	Average per 25
	1 - 25	1701/ 325	5.23
	25 - 50	1681/ 325	5.17
	50 - 100	3434/ 650	5.28
	100 - 200	4012/ 781	5.13
	200 - 300	1527/ 282	5.41
Total 13.....	1 - 300	12355/2363	5.22

A similar treatment of the low-aim ESP would be inadequate because there were only two days on which more than 100 runs were done on one day. Neither is it necessary, because of the comparatively small fluctuation in scoring rate for the periods of various lengths (see Table III and Graph II).

This discussion may be summarized very briefly: There is evidence that the number of calls made at a single experimental sitting is not in itself a determining factor for the level of scoring.

*The DT Curve.* In the earlier Duke work, Rhine<sup>6</sup> reported that the rate of scoring throughout the pack in work with the DT method was not constant. Some subjects made more correct calls at the beginning and end of the run, with fewer hits for the cards at the center of the pack, so that when the results were graphed, using groups of five calls as units, a U-shaped curve was obtained. One subject reversed this order, getting fewer hits for the first and last groups of five calls and

<sup>6</sup> Rhine, *op. cit.*, pp. 122 ff.

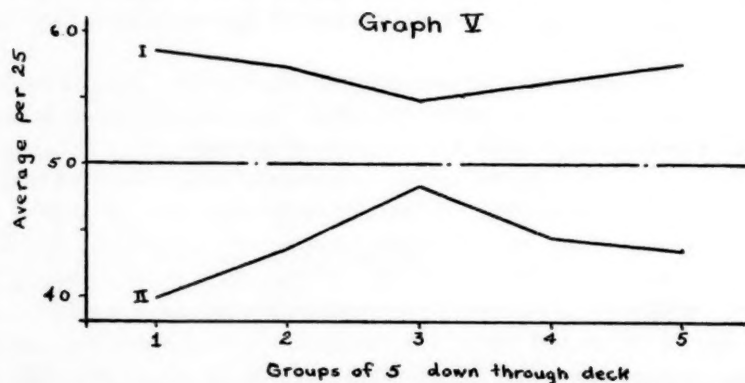
more for the center five, so that when the results were plotted, an inverted U-shaped curve resulted. Rhine tentatively calls these curves "attention curves"; he regards them as further evidence of the purposive extra-chance nature of ESP scoring.

The variety of methods used in the experiments reported here provides data for several studies of variations in scoring with the DT method: Comparisons of high-aim and low-aim are possible, both for  $DT_{25}$  and  $DT_{75}$ .

Four groups of 100 consecutive runs each of  $DT_{25}$  high-aim and  $DT_{25}$  low-aim were selected for treatment. The  $DT_{25}$  high-aim and  $DT_{25}$  low-aim were in each case done during corresponding periods of time. This selection was necessary because of the unwieldiness which would be encountered should an attempt be made to treat the total number of runs for each condition. One group of 100 consecutive runs each includes the first 100 runs of  $DT_{25}$  low-aim and a corresponding group of  $DT_{25}$  high-aim. Two similar groups were selected for periods occurring about the middle of the experimental period, and one group of 100 consecutive runs each of high-aim and low-aim from the last period of work. In this way a representative selection of data was obtained.

In the graph below (Graph V) the results of these four groups have been totalled and the averages plotted. Averages are stated on the basis of the average number of hits per 25. Groups of five calls are used as units; i.e., the first unit is the average number of hits per 25 made for the first five calls of each pack; the second, the average for the second five calls, etc.

For the  $DT_{25}$  high-aim, proportionately more correct calls occurred

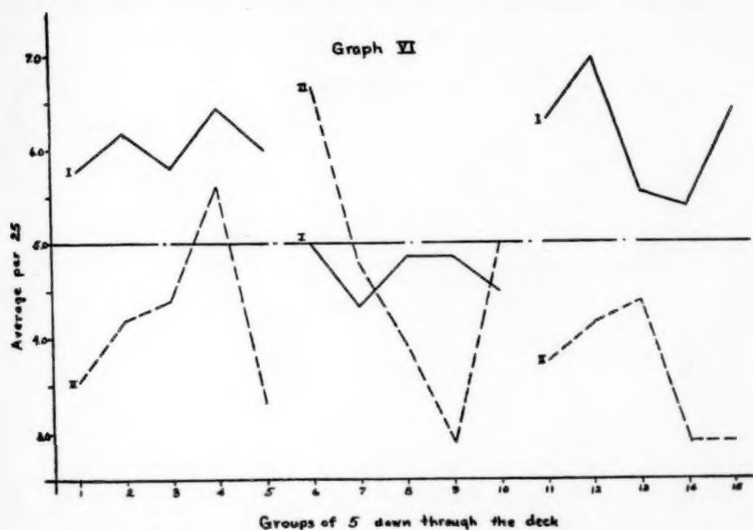


GRAPH V.  $DT_{25}$  curves showing averages of I, high-aim and II, low-aim.

in the first and last groups of five; that is, more of the calls for the top and bottom parts of the pack were correct. As might be expected, the curve for  $DT_{25}$  low-aim is the inverse of this curve, with fewer hits for the end groups and more for the middle of the pack. In terms of success in achieving the results aimed for, the two curves are similar. The subject was more successful either in calling or in avoiding calling correctly the cards near the bottom and the top of the pack and less successful for the cards in the center of the pack.

This suggests the possibility of receiving sensory cues from the first and last cards in the pack, since these cards are more likely to be exposed to the subject's sensory field. This possibility is successfully countered by the fact that the deviations for the second group of five calls for the high-aim  $DT_{25}$  and for both the second and fourth groups of five for the  $DT_{25}$  low-aim are independently significant.

That this variation in levels of scoring is a function of the entire task rather than of the number of cards called is indicated by the curve for  $DT_{75}$ . In Graph VI the first 114 runs of  $DT_{75}$  high-aim and the entire 72 runs for  $DT_{75}$  low-aim (on the basis of 25 calls per run; actually there were 38 runs for high-aim and 24 runs for low-aim, using a pack containing 75 cards) are given the same analysis given  $DT_{25}$ . Here the averages are plotted for the fifteen groups of five calls.

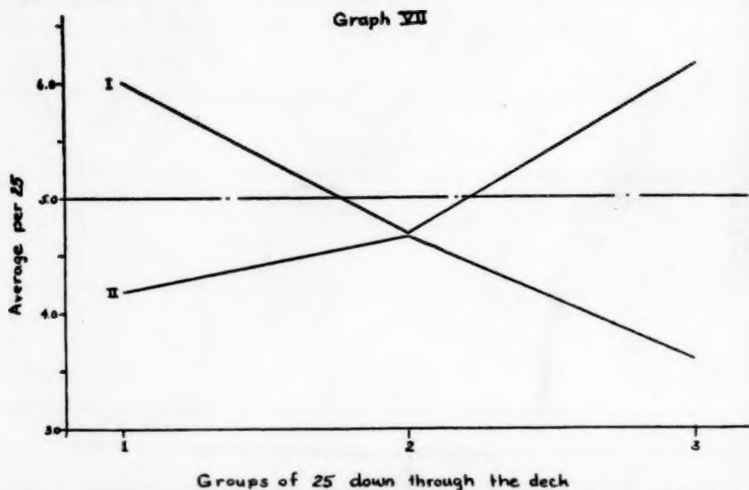


GRAPH VI.  $DT_{75}$  curves plotted to show the difference between I, high-aim, and II, low-aim in successive groups of 25 in the deck. (An analysis of Graph VII.)

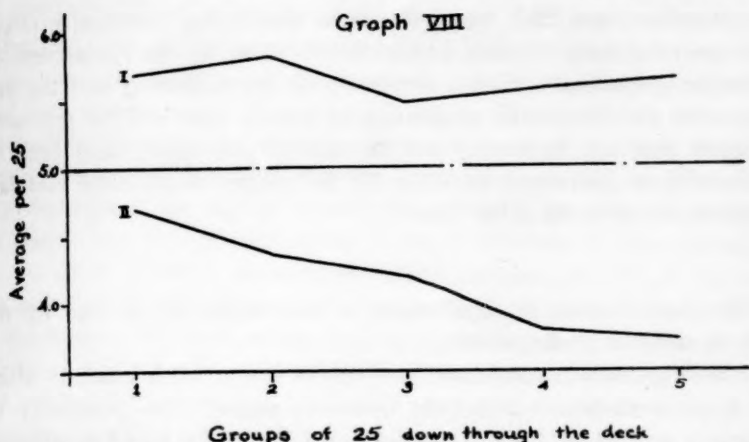
The  $DT_{75}$  curves are distinctive in several ways. First, three groups of 25 calls stand out as units, although in the deck of cards there was no separation of one group of 25 from another. Secondly, the composite curves are not a repetition of the  $DT_{25}$  curves; that is, when group one of the first 25 calls is added to group one of the second and third, etc., so that each 25 calls is treated as a regular run of  $DT_{25}$ , the curve is not that obtained for the  $DT_{25}$ , but is one which shows little variation of distribution, as is shown in Graph VII. The most striking feature is that when each group of 25 calls is treated as a unit, the resemblance to the  $DT_{25}$  curves is marked; the curve for  $DT_{75}$  high-aim is similar to the curve for  $DT_{25}$  high-aim, and the  $DT_{75}$  low-aim curve is a reversal of the  $DT_{75}$  high-aim curve (Graph VIII). That is, the first 25 cards are to the 75 card pack what the first five or ten cards are to the 25 card pack; the center 25 to the 75 pack as the center five to the 25 card pack (Graph VIII).

A point of possible significance is that while there is not perfect correspondence of the four  $DT_{25}$  high-aim curves representing the four groups of 100 consecutive runs, the three curves with the higher averages approximate the U-shaped curve most nearly while the one group with the low average of 5.14 approximates the inverted U-shaped curve typical for lows.

The same relationship is borne out by the  $DT_{75}$  curves. For the  $DT_{75}$  high-aim it is not possible to draw conclusions because none of



GRAPH VII: Averages in each group of 25 calls down through the  $DT_{75}$  deck for I, high-aim and II, low-aim.



GRAPH VIII.  $DT_{25}$  curves with groups of 25 calls as units for I, high-aim, and II, low-aim.

the curves for the three separate groups of 25 calls in the  $DT_{75}$  curve resemble the  $DT_{25}$  curve very closely. The most that can be said of them is that the first and third groups of 25, which are the high points in the  $DT_{75}$  curve, have relatively low averages for the center, as is true of the curve for successful  $DT_{25}$  high-aim; while the middle group of 25, which represents a below-chance average for the  $DT_{75}$  curve, has a relatively high average for the center calls, as does the typical  $DT_{25}$  low-aim curve.

For the  $DT_{75}$  low-aim, the relation is more striking; the two more successful groups of 25, the first and last, have exaggerated  $DT_{25}$  low-aim curves, while the middle, or less successful group of 25 has a curve like an exaggerated U-shaped curve for  $DT_{25}$  high-aim.

It seems likely that this variation in distribution of the hits throughout the pack is dependent on variations of attention given to various parts of the task. It has been noted that in sensory perception items near the beginning and end of the series stand out in contrast to the other items of the series. A striking resemblance to memory curves is shown by the DT curves. Bigham, Calkins, Binet, Pohlmann, Ebbinghaus,<sup>7</sup> and others have noticed that more errors have been made in trying to recall nonsense syllables or other items which occur in the middle of the series than for those which occur at the beginning or the end of the series. Probably the remarkable similarity between the memory curves and the DT curves is not indicative of a relation be-

<sup>7</sup>Frobes, Joseph, *Lehrbuch der Experimentellen Psychologie*, Herder and Company, Freiburg, 1923, p. 535.

tween memory and ESP, but is due to an underlying function of attention operating both in recall and in ESP. That is, the phenomena of a greater proportion of items recalled near the beginning and the end of a series and the greater proportion of correct calls in ESP for cards similarly near the beginning and the end of the series, may both be attributable to the closer attention of the subject to the first and last parts of the series in either case.

#### SUMMARY

The main points of significance of this work are derived by the various analyses of the results.

The first analysis considered was that of the problem towards which the experimentation was largely directed; namely, the possibility of voluntary control of ESP. It was shown that ESP could be directed at will so that the subject could obtain a significant positive deviation from chance when the task was to try to call the cards correctly and a significant negative deviation from the mean expectation when the task was to avoid calling them. The primary importance of this observation lies in its relating ESP to other voluntarily controllable and directable mental functions.

The rate of scoring was found to be independent of the number of runs performed within each period. A comparison of average scores for working periods of different lengths seemed to show that there was a correlation of the number of runs done for the day and the average score obtained, the lowest scores occurring for days on which the largest number of runs were done. Study of those days for which a larger number of runs were done, however, showed that there was not a high average for the first part of the day's work and a decline with later parts, but that factors other than the number of runs were responsible for the low average and that the very last runs of the day had the highest average scores.

Low-aim showed more consistent deviations from chance than did high-aim. High deviations for high-aim were consistently accompanied by comparable deviations for low-aim. But the converse was not true: Large deviations for low-aim might accompany very small deviations for high-aim. This difference might have been due to the fact that low-aim was used less frequently than high-aim and for shorter periods.

One phenomenon observed seemed to be of distinctly Gestalt nature, that of the DT curves. More correct calls were made for the cards at the beginning and the end of the pack and fewer for the center cards

for DT high-aim and a reverse of this order for the low-aim; that is, for either high-aim or low-aim more successful calls were made (according to the aim) for the first and the last part of the pack. This relationship held both for packs of 25 and of 75 cards. The similarity of results obtained in these experiments to those observed in so-called memory curves led to the conclusion that the variation in scoring level was due to a variation in the amount of attention given to the different parts of the task. The lack of conformity of the distribution in the successive groups of 25 in  $DT_{75}$  to the distribution obtained for  $DT_{25}$ , and the similarity of the distributions in relation to the total pack rather than to the number of cards called demonstrated that the variation of distribution was not physically determined, but was a psychological function related to the task as a whole and not to the size of the pack or the numerical position of the card in the pack.

## MATHEMATICAL TECHNIQUES USED IN ESP RESEARCH

J. A. GREENWOOD and C. E. STUART

---

**ABSTRACT:** The statistical methods applied to ESP research are means of finding whether a given score is different enough from that expected by chance alone to show that a relation exists between the calls made and the symbols on the cards indicated. The critical ratio method measures the deviation in terms of the standard deviation. The chi-square method measures the variation about mean chance expectation in terms of that mean. Both results may be interpreted in terms of a fraction of probability or odds to one against a chance hypothesis. Application of various methods to the results of a distance ESP test taken from the literature shows that, with reasonably high scores, any of the methods of analysis tried will demonstrate reliable difference between ESP scores and chance control scores.

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

---

THE MATHEMATICAL treatment of the data in tests for extra-sensory perception provides the most fundamental basis for the validity of the conclusions. While mathematical treatment may not substitute for careful initial observation, the fact that the results of a test may be expressed in objective logical and mathematical terms serves to increase their security as reliable observation. In studies of extra-sensory perception the phenomenon under observation is held by many to be antecedently highly improbable, by others an obvious fact of everyday experience. Under such a circumstance an objective method for evaluating evidence bearing upon the reliability of these judgments is very necessary.

Of all points in the report of his research to establish extra-sensory perception, Rhine has been most frequently criticized on the mathematical treatment (16, 6, 15). Many of these criticisms have been met in publication (14, 18). There has, however, been no generally available discussion of the mathematics applied to the results of ESP card tests in such a research. The writers propose in so far as it is possible to meet that need in the following discussion.

The subtleties of the mathematics of probability cannot be dealt with here even sketchily. One point, however, must be made clear at the outset. "Probability" as a concept is useful only when it is considered in its relative aspect. When one says that any bridge hand is just as probable as any other and therefore one is just as likely to get thirteen spades as any other particular hand, the statement is quite true, but only if we interpret it very strictly. What we mean is that if a player predicted the exact cards which he was going to hold in the next deal, the exact probability of the cards matching his prediction would be just as small whatever thirteen cards he chose to predict. But if one argues upon this fact that very improbable events continually occur, he is using "probability" in an absolute and meaningless sense. We may speak of probability only from the standpoint of a particular moment wherein an individual had a limited basis for rational knowledge about an event; the probability is an expression of the relation of that moment to the succeeding event.

While the statistical treatment of any quantitative research data should be fitted more or less exactly to the problem under consideration, there is no justification for the view that research in extra-sensory perception demands a very special mathematical technique that is strikingly different from the commonly used statistical methods, nor for the more radical view that no mathematical technique whatever is valid in application to experiments in such a field. The antecedent probability of the existence of the phenomenon under discussion is a question for individual judgment, but the mathematical techniques used in evaluation of the research data must stand or fall on the basis of their own internal consistency and their applicability to quantitative problems in general.

Early workers in the field of parapsychology used playing cards, numbers, and letters of the alphabet as measurable material. Rhine introduced the use of ESP symbols, five easily distinguished forms. These were printed on cards and used in shuffled decks of 25 cards of five each of the symbols.<sup>1</sup>

#### THE CRITICAL RATIO METHOD

When a card was presented to a subject to be called, the probability of his hitting it by a chance guess was  $1/5$ . And in general the average chance number of hits to be expected was  $1/5$  the total number of calls made.

In Rhine's experiments the actual number of correct calls exceeded

<sup>1</sup> See plate opposite p. 1, *Journal of Parapsychology*, Vol. I.

the average chance expectation. To find whether such a deviation from chance was greater than random fluctuation about the average to be expected, it was necessary to measure it in relation to a standard unit or probable deviation. Rhine (12, p. 31 ff.) used as this unit the probable deviation from the mean when a number of independent events have a probability of success of  $1/5$ . The probable error is the amount of deviation that is expected by chance in about 50 per cent of observations. It is easily computed from the formula  $.6745 \sqrt{npq}$ , in which  $n$  is the total number of calls,  $p$  is the probability of a successful hit, and  $q$  is the probability of failing.

The observed deviation is divided by the probable error to find an X-value or critical ratio. Tables of the probability integral, when adjusted to use with the probable error, give the probability of a deviation as great or greater than the observed one occurring by chance in a normal or nearly normal population.<sup>2</sup>

For example, Rhine reports (13, p. 220) a DT series of 1,625 trials, or 65 runs of 25 cards, each with HP as subject. The total number of hits was 482. The expected number ( $np = 1,625 \times 1/5$ ) is 325. The deviation above chance was 157. The probable error ( $.6745 \sqrt{npq} = .6745 \sqrt{1,625 \times 1/5 \times 4/5}$ ) is 10.09. When the deviation is divided by the probable error, the resulting critical ratio of 14 indicates that the probability of such a deviation occurring by only chance variations is about  $10^{-20}$ .

Pratt (11) and other investigators have modified this evaluative method slightly by using the *standard deviation* (or standard error) in place of the probable error. As the probable error is directly derived from the standard deviation by the formula: P.E. =  $.6745$  S.D., the only change is in the elimination of the fractional constant term. This usage, which follows a general trend in statistical method (Kelley, 5, p. 98) has since become standard in studies of extra-sensory perception.

When the standard deviation is used as the measuring unit for the observed deviation, the probability of the chance expectation of such a deviation may be read directly from tables of the Probability Integral, such as Pearson's (9).

*Application I. Test.* To allow comparison of the methods to be dis-

<sup>2</sup>The "curve of normal distribution" is discussed in any statistical textbook. In a group of measurements made of a nearly normal population most of them will cluster about the average and few will vary a great deal from the average. For example, if the measurements are of the heights of men, most of the measurements would be between 65" and 73", with the average at about 68"; nearly all would be between 59" and 76"; and very few would be below 54" or above seven feet.

cussed, a single experimental example will be used. We choose the Pearce-Pratt distance clairvoyance series (13, p. 221) for that example. It has been considered an unusually good case for the establishment of extra-sensory perception. With the experimenter and subject in different buildings, 1,850 trials, or 74 runs of 25 cards each, yielded 557 correct calls.

The chance expectation,  $1/5$  the number of trials, is 370. The observed deviation is + 187. The standard deviation ( $\sqrt{1,850 \times 1/5 \times 4/5}$ ) is 17.20. The critical ratio (187/17.20) is 10.87. Pearson's tables indicate that the probability that such a positive deviation occurred by chance is about  $10^{-27}$ .

*Control.* The method when applied may have varied from the theoretical assumptions noted above. To test this hypothesis another experiment, identical with the example except for the possibility of extra-sensory perception determining the call-card correspondences, is necessary. To get this data the calls of each run were checked against the cards of the third run following. That is, the calls made for the first run were checked against the card sequence of the third run, the calls of the second against the cards of the fourth, and so on. The two final call series were checked against the first two card series. In 74 runs the hits totalled 385, a deviation of 15 above the chance expectation of 370.

The observed deviation (+ 15) divided by the standard deviation (17.20) yields a critical ratio of .87. Pearson's tables show that this deviation could occur by chance about once in five times.

The results of the control are valuable other than as a comparison with the test case. They show that the experimental conditions *approximate* the theoretical assumptions of the method. These are (1) that the probability for a chance success on each card is  $1/5$ , (2) that each trial is independent of the others, and (3) that the expected hit distribution is nearly normal.

The assumptions themselves must bear analysis, however. The first, that the probability for a chance hit on each call is  $1/5$ , has been challenged by Willoughby (16), Kellogg (6), and Weisberg (15). The following demonstration seems the easiest to follow of several available.

Given a shuffled deck of ESP cards, we draw the top card and lay it face down upon the table without looking at it. The probability that it is a *star* is  $1/5$  (as the five symbols are equally distributed in the deck.) But it either *is* a star or *is not* a star. If it *is*, the probability that the next card drawn (the top card of the pile remaining) will be

a star is  $4/24$ . If the first card drawn is *not* a star, the probability that the next card drawn will be a star is  $5/24$ . If the first card is a star, the probability that the next card is *not* a star is  $20/24$ ; if the first card is *not* a star, the probability that the next card is *not* a star is  $19/24$ . Tabulated (\* = star, o = not star), the possibilities after two drawings are:

- a. \*\*  $5/25 \times 4/24 = 20/600$
- b. \*o  $5/25 \times 20/24 = 100/600$
- c. o\*  $20/25 \times 5/24 = 100/600$
- d. oo  $20/25 \times 19/24 = 380/600$

Following the same argument, we may draw a third card and double the size of the table.

- a. \*\*\*  $5/25 \times 4/24 \times 3/23 = 60/13,800$
- b. \*\*o  $5/25 \times 4/24 \times 20/23 = 400/13,800$
- c. \*o\*  $5/25 \times 20/24 \times 4/23 = 400/13,800$
- d. \*oo  $5/25 \times 20/24 \times 19/23 = 1,900/13,800$
- e. o\*\*  $20/25 \times 5/24 \times 4/23 = 400/13,800$
- f. o\*o  $20/25 \times 5/24 \times 19/23 = 1,900/13,800$
- g. oo\*  $20/25 \times 19/24 \times 5/23 = 1,900/13,800$
- h. ooo  $20/25 \times 19/24 \times 18/23 = 6,840/13,800$

The table (which may be extended to any number of drawings desired) gives the probability of all sequences of the first three cards. A star occurs on the first card in cases a, b, c, and d. The sum of the probabilities is  $(60 + 400 + 400 + 1,900)/13,800 = 2,760/13,800 = 1/5$ . A star occurs on the second card in cases a, b, e, and f, whose probabilities summed gives  $(60 + 400 + 400 + 1,900)/13,800 = 2,760/13,800 = 1/5$ , and similarly on the third drawing  $a + c + e + g$   $(60 + 400 + 400 + 1,900)/13,800 = 2,760/13,800 = 1/5$ .

The selection of a *star* as the called symbol was wholly arbitrary. The same argument would hold for any other symbol.

Let us consider any rank order in the deck, say the  $i$ th card. What is the probability that the  $i$ th card is a star? This probability is the sum of the probabilities of all possible arrangements of all distinct sets of  $i$  cards from the ESP deck with the  $i$ th card a star. Any such arrangement has a determinable probability of occurrence, say  $p_1$ . The sum of these,  $p_1 + p_2 + p_3 + \dots + p_n = P_i$ , is the probability that a star will occupy the  $i$ th place. The same demonstration holds for a circle in the  $i$ th place, except that each of the arrangements will differ from the above by having stars and circles interchanged. Due to the symmetry of constitution of the deck, the probabilities of each of these arrange-

ments will be the same as for the star, and their sum will equal  $P_i$ . As there are five symbols, for which this arrangement holds, and as one of them *must* occupy the  $i$ th place, it follows that  $5P_i = 1$ , and  $P_i = 1/5$ .

This theoretical probability has been uniformly substantiated in empirical checks. When the subject's intended relation between call and card is ruled out, the number of hits is negligibly variant from  $1/5$  of the total trials. The control series above is an example. The number of hits, 385, divided by the number of trials, 1,850, gives the fraction .208, which is very close to the expected .200.

That a chance success on each trial is independent of all other trials is the second assumption of the critical ratio method. Does this hold in the calling of a fixed deck of ESP cards? From the point of view that the subject is free to call any symbol for any given card, the correspondences might be considered wholly independent. But the number of symbols in the ESP deck is fixed at five each. The subject is aware of this limitation. He may, if he wishes, call just five of each symbol every run. While the subjects are instructed to pay no attention to previous calls, the fact that they may do so suggests that the trials might not be independent.

Since it is possible for the chance hits to be interdependent, it is necessary to find the magnitude of this dependence and what effect it will have upon the hit score.

It has been shown by Greenwood (4) that when one deck of ESP cards is matched card for card against another, the intercorrelation of hits due to the fixed number of cards is  $1/24^2$  or  $+ .001736$ .

Such a correlation does not affect the  $1/5$  probability of hitting an individual card, nor the average expectation of five hits per 25. But it does change the *variance* of these statistics, the amount any chance observation might differ from the average expectation.

The variance of a distribution is the square of the standard deviation. When the events are correlated, Yule (19, p. 286) gives the formula for the variance as:

$$\sigma^2 = npq[1 + (n-1)r].$$

Substituting:  $n = 25$ ,  $p = 1/5$ ,  $q = 4/5$ , and  $r = 1/24^2$ ,

$$\begin{aligned}\sigma^2 &= 4.167 \\ \sigma &= 2.041\end{aligned}$$

As the standard deviation upon the assumption of complete independence is 2.000, the subject by calling each symbol exactly five times may

increase the standard deviation about two per cent. Greenwood (4) has shown that 2.041 is the maximum standard deviation for a run of ESP cards. That is, every variation in calling from the strict five of each symbol *decreases* the variance, the limiting case being a zero variance if the subject chooses to call only one symbol all the time.

In principle, therefore, the second assumption is violated when ESP cards are used in fixed packs of 25 each. But in practice the error introduced is so small as to be almost negligible in its effect upon an experimental conclusion. For the number of runs usually considered in experiments with ESP cards, the increase in the value of the standard deviation is well within the variance of the standard deviation itself.

It is suggested, however, that for complete safety in drawing conclusions from a deviation, the maximum standard deviation should be used. The true standard deviation will always be less than this figure. Large critical ratios will be affected only slightly; and borderline cases, those just meeting the criterion, will have the added margin of safety. In place of the customary formula for the standard deviation:

$$\sigma = \sqrt{npq},$$

the formula for the maximum standard deviation with ESP cards is suggested:

$$\begin{aligned} \sigma &= \sqrt{25/24} \sqrt{npq} \\ &\text{or} \\ \sigma &= 1.02 \sqrt{npq} \end{aligned}$$

A third assumption, that the expected distribution of chance hit scores is approximately *normal*, is implicit in the critical ratio method. Whether chance distribution of ESP cards approximate a normal distribution has been the subject of widely variant opinions. Willoughby (16, 17) holds that because the 1/5 probability of a success against 4/5 for failure is different from the 1/2 probability of both success and failure in the pure case of the normal curve, and because of *skewness* introduced by this difference, the critical ratio method is wholly inadequate. Lemmon (7) finds that distribution of hits on ESP cards is "at least as close to normality as those commonly treated as normal in other branches of psychological research." Check series have in general borne out Lemmon's view. Willoughby's contentions have been met elsewhere by Stuart (14), who found that skewness was negligible in the chance hit distribution of ESP cards.

The exact theoretical hit-frequency distribution per matching of one deck of ESP cards against another is at present unknown. It is

nearly normal. It is closely approximated by the binomial expansion  $(4/5 + 1/5)^{25}$ . Empirical curves, when tested for Pearson's types, have indicated that the theoretical curve may be a Type I. It is to be hoped that the exact distribution will soon be available. It will allow more exact statements of the probability of results. Practically, however, the curve will approximate the normal so closely that it is doubtful whether any conclusion from experiments evaluated by the critical ratio method will be affected.

#### CRITERIA OF SIGNIFICANCE

When a score varies enough from chance expectation to exceed a number previously agreed upon, it is said to be significant. In the first work with ESP cards, Rhine used the generally accepted value of four times the probable error as a significant deviation. The probability that a positive deviation might equal or exceed such a value by chance is about  $1/143$ . When the probable error was replaced by the standard deviation, the roughly equivalent 2.5 times the standard deviation was agreed upon as the criterion of significance. The probability that a deviation might equal or exceed this value by chance is about  $1/162$ .

These values are necessarily arbitrary. Suggestions have been offered that the criterion be raised from 2.5 to 3.0 times the standard deviation, to conform to the general usage among American psychologists. This would decrease the chance probability of a significant deviation from  $1/162$  to  $1/740$ . Statisticians who consider the problem at all, however, tend to find this standard too high, especially in the evaluation of differences between means. For example, Davenport and Ekas use 2.00 (2, p. 38), Peters and Van Voorhis suggest 1.78 (10, p. 346), and R. A. Fisher recommends 2.00 (3, p. 45).

The criterion of significance for use in the critical ratio method remains, therefore, a matter for the judgment of the experimenter. Carefully observed check series have been consistently well within the 2.5 limit, and with a reasonably large number of runs such a criterion should be wholly adequate.

*Significance of a Negative Deviation:* When an experiment has been designed to produce a negative deviation, as in the case when a subject is instructed to score low or to miss the target symbols, the same criterion of significance may be used as in the case of a positive deviation. Occasionally, however, when a deviation is opposite in sign from that expected in the experiment, or when on principle one may expect either a positive or negative deviation, the criterion of sig-

nificance must be raised. If 2.5 is the criterion for a positive deviation, the criterion for an equally small probability of chance deviation in *either* direction is 2.82. If 3.0 is the criterion for a positive deviation, 3.2 is an equal criterion for a deviation either positive or negative.

#### THE CHI-SQUARE METHOD OF EVALUATION

The critical ratio method measures the deviation of an observed score from an expected score by comparing the algebraic sum of all the deviations with a unit of standard deviation. The chi-square method, developed by Karl Pearson, is similar in principle but disregards the sign of the deviation. A complete discussion of the method will be found in R. A. Fisher's *Statistical Methods for Research Workers* (3).

The critical ratio method assumes that chance results from the conditions of an experiment will produce hit frequencies that are approximately normal. The chi-square test may be applied to distributions that are not normal.

Willoughby (16, 17) suggested and first applied a chi-square test to data from ESP cards. Because he contended that the expected distribution was not normal and was more accurately described by the frequency distribution of the expansion of the binomial  $(4/5 + 1/5)^{25}$ , he grouped his data to find whether they "fitted" the expected distribution. (The fact that they fitted very well did not support his contention that the chi-square test was the only valid statistical method for evaluation of such data. They would have fitted any nearly normal curve quite as well. That is, for the number of runs considered, 200, the expected *chance* variations from the expected frequencies were very much greater than the actual difference in the theoretically expected frequencies between the normal and binomial curves.)

In the application of the chi-square test, the data under consideration is broken up into groups, chosen arbitrarily or according to the principle to be tested. The expected number of hits for each group is calculated. The difference between the observed number of hits and the expected number is *squared*. This result is divided by the expected number. The quotient is the chi-square value for the group. The chi-square values for all the groups are then added. The probability of this total chi-square value being due to merely chance fluctuations in the data is given in tables, either Elderton's Tables of Goodness of Fit presented in Pearson's *Tables for Statisticians and Biometricians* (9, p. 26 ff.) or in the more generally available tables by Fisher (3). The probability depends upon the "degrees of freedom" of the group

chi-squares that make up the total. The "degrees of freedom" are usually one less than the number of groups.

The general formula is

$$\chi^2 = \sum \left[ \frac{(f - f_0)^2}{f} \right]$$

in which  $f$  = the expected frequency of hits in the group,  $f_0$  = the observed number of hits in the group, and  $\sum$  denotes that the values for the groups are added together.

*Application II. Test.* The data of the Pearce-Pratt distance series are to be examined to find whether the hits per run were merely random variations from the frequencies that might be expected upon the theory that chance hits will follow the distribution of the binomial  $(4/5 + 1/5)^{25}$ .

The 74 runs are separated into runs in which one hit occurred, two hits, three hits, and so on. The expected frequencies may be found from the expansion of  $74(4/5 + 1/5)^{25}$ . (The small frequencies on the two ends of the expected distribution are grouped together.)

TABLE I  
CHI-SQUARE FIT OF TEST SERIES TO BINOMIAL

Hits per Run	Expected Frequency	Observed Frequency	Difference	Difference Squared	$\frac{(f - f_0)^2}{f}$
0 - 2.....	7.25	6	1.25	1.56	.22
3.....	10.04	5	5.04	25.40	2.53
4.....	13.81	8	5.81	33.76	2.44
5.....	14.50	4	10.50	110.25	7.60
6.....	12.08	8	4.08	16.65	1.38
7.....	8.20	4	4.20	17.64	2.15
8 -15.....	8.03	39	30.97	959.14	119.44

$$\chi^2 = 135.76$$

Degrees of freedom = 6  
 $p = 7.86 \times 10^{-27}$

As Elderton's tables do not go above a chi-square of 70, the probability of this chi-square must be found by formula. It is  $7.8 \times 10^{-27}$ , a figure so infinitesimal that there is no question that the hits made were not chance variants from the frequency distribution of the binomial. (Note how closely the probability corresponds to the approximate  $10^{-27}$  found by the critical ratio method.)

*Control.* The same test may be applied to the cross-check series.

With a chi-square of 7.60 and six degrees of freedom, we find from Elderton's tables that  $p = .26$ . That is, about one fourth of chance distributions theoretically based upon the binomial series would vary

TABLE II  
CHI-SQUARE FIT OF CONTROL SERIES TO BINOMIAL.

Hits per Run	Expected Frequency	Observed Frequency	Difference	Difference Squared	$\frac{(f - f_0)^2}{f}$
0 - 2.....	7.25	9	1.75	3.06	.42
3.....	10.04	7	3.04	9.24	.92
4.....	13.81	13	.81	.66	.05
5.....	14.50	16	1.50	2.25	.16
6.....	12.08	10	2.08	4.33	.36
7.....	8.20	5	3.20	10.24	1.25
8 - 15.....	8.03	14	5.97	35.64	4.44

$$\chi^2 = 7.60$$

$$\text{Degrees of freedom} = 6$$

$$p = .26$$

from it as much or more than this series. The data of the control series is a "good fit" to the binomial expectation.

In the above examples the chi-square method does not assume that the distribution of hits will be normal about a mean of 5 per run. It assumes only that the chance deviations about the expected frequency in each group will be normal. The method may be applied to other arbitrary groupings of the material upon the same assumption.

*Application III.* Test. The 74 runs of the Pearce-Pratt series may be broken up into 14 consecutive groups of five runs each, and one group of four runs. The expected number of hits in each group will be 1/5 the number of trials; 25 hits for the five-run groups, and 20 hits for the last group.

TABLE III  
CHI-SQUARE GROUPING OF TEST SERIES.

Groups	Expected Hits	Observed Hits	Difference	Difference Squared	$\frac{(f - f_0)^2}{f}$
1.....	25	35	+10	100	4.00
2.....	25	59	+34	1,156	46.24
3.....	25	34	+ 9	81	3.24
4.....	25	22	- 3	9	.36
5.....	25	29	+ 4	16	.64
6.....	25	33	+ 8	64	2.56
7.....	25	41	+16	256	10.24
8.....	25	48	+23	529	21.16
9.....	25	35	+10	100	4.00
10.....	25	44	+19	361	14.44
11.....	25	24	- 1	1	.04
12.....	25	40	+15	225	9.00
13.....	25	38	+13	169	6.76
14.....	25	34	+ 9	81	3.24
15.....	20	41	+21	441	22.05

$$\chi^2 = 148.0 \text{ with } 14 \text{ degrees of freedom}$$

$$p = 1.82 \times 10^{-24}$$

The high chi-square value again falls outside the available tables. The very small probability indicates practical certainty that the variations from expectation within the groups are not chance fluctuations.

*Control.* The same method applied to the cross-check series gives the deviations per group noted in Table IV.

TABLE IV  
CHI-SQUARE GROUPING OF CONTROL SERIES.

Groups	Expected Hits	Observed Hits	Difference	Difference Squared	$\frac{(f - f_0)^2}{f}$
1.....	25	24	- 1	4	.16
2.....	25	23	- 2	8	.32
3.....	25	35	+10	100	4.00
4.....	25	27	+ 2	4	.16
5.....	25	23	- 2	4	.16
6.....	25	25	0	0	.00
7.....	25	32	+ 8	64	2.56
8.....	25	31	+ 7	49	1.96
9.....	25	19	- 6	36	1.44
10.....	25	26	+ 1	1	.04
11.....	25	19	- 6	36	1.44
12.....	25	25	0	0	.00
13.....	25	30	+ 5	25	1.00
14.....	25	28	+ 3	9	.36
15.....	20	16	- 4	16	.80

$\chi^2 = 14.40$  with 14 degrees of freedom  
 $p = .42$

As there are 15 groups, we find in Elderton's table that a chi-square of 14.4 has a probability of .42; that is, we may expect a chi-square value as great or greater than this in about 42 per cent of chance series. The fluctuations within the groups are merely normal chance deviations.

*Sign test.* As the chi-square test considers only the magnitude of the deviation and not its sign, the above method may be further analysed with respect to the direction of deviation. We may ask: What is the probability that such a chi-square value will occur with the ratio of positive and negative signs observed? In Table IV, where the signs of the deviations are about equally positive and negative, this would not be a matter of interest. But in Table III the signs are predominantly positive and the consideration of them might give a more accurate value of the grouping.

Since on the chance hypothesis the deviations should occur equally positive and negative,<sup>3</sup> the expected frequencies will be half the total

<sup>3</sup>When the signs are predominantly negative, it is suggested that the ratio 52 to 48 for negative to positive expectation be used.

number of deviations, and the chi-square computed will have a single degree of freedom.

For example, consider the differences of Table X. There are 13 positive deviations and 2 negative deviations. The expected number of each is 7.5. The chi-square evaluation will be as follows:

	Expected	Observed	Difference	Diff. Squared	$\frac{(f-f_o)^2}{f}$
Positive .....	7.5	13	5.5	30.25	4.03
Negative .....	7.5	2	5.5	30.25	4.03
				$\chi^2 = 8.06$	
			degrees of freedom	1	
				$p = .004$	

When the probability found for the grouping test is multiplied by this probability of sign distribution (this multiplication is valid because of the independence of the chi-squares from which the two probabilities are derived), the resulting  $p = 7.3 \times 10^{-27}$  is again close to the results of the critical ratio and the fit to the binomial tests.

#### THE CHI-SQUARE CONTINGENCY TEST

The most complete analysis of ESP results may be made with chi-square studies of a *contingency table*.<sup>4</sup> With this table, both the hits and the misses are recorded and evaluated. A general discussion of the analysis of contingency tables is given by Kelley (5) and Fisher (3).

A five-fold table (25 squares, five on a side) of all the call-card correspondences is made. In Table V, for example, the number of circles called "circle" is listed in the first square of the top row, the number of rectangles called "circle" in the second square of the row, the number of circles called "rectangle" in the first square of the second row, and so on until all the correspondences are listed. There are 25 *cells* to the table. The diagonal cells include all the hits. The 25 cells are groups to which the chi-square test may be applied.

If the occurrence of the cards is listed in the vertical columns, and the calls in the horizontal rows, then the sum of the items in each column will be the same, as the ESP deck is made up of five suits equally distributed. Since there are five columns, the probability that a particular trial (call-card correspondence) will fall in a given column is 1/5. The calls are not so equally distributed. The subject does not try to call

<sup>4</sup>Kellogg (6) suggests that this method is the only valid means for establishing conclusions in tests of extra-sensory perception. This contention, like Wiloughby's similar statement in favor of the "goodness of fit" test, remains a matter for individual opinion.

exactly five of each symbol per run, and naturally varies from an exactly equal distribution. The sums of the trials in the rows will likely be different. But the probability that a given trial occurs in that row is the number of trials in the row divided by the total number of trials. As the probability for the column is known, and the probability for the row is known, then the product of the two gives the probability that an item will occur in the intersecting cell of that row and column. This fraction, times the number of trials in the experiment, gives the expected number of correspondences for that call.

The chi-square evaluation follows exactly as in the above cases. The observed frequency in each cell is compared with the expected frequency. The square of the deviation is divided by the expected frequency and the total chi-square value evaluated by reference to Elderton's table.<sup>5</sup>

*Application IV. Test.* Table V lists the call-card correspondences

TABLE V  
CONTINGENCY TABLE FOR TEST SERIES.

Calls	CARDS					* Total of Rows
	Circle	Rectangle	Waves	Plus	Star	
Circle.....	118 75.6	67 75.6	63 75.6	70 75.6	60 75.6	378
Rectangle.....	62 69.8	101 69.8	75 69.8	54 69.8	57 69.8	349
Waves.....	57 73.2	74 73.2	109 73.2	61 73.2	65 73.2	366
Plus.....	73 81.4	62 81.4	60 81.4	127 81.4	85 81.4	407
Star.....	60 70.0	66 70.0	63 70.0	58 70.0	103 70.0	350
Total of Columns.....	370	370	370	370	370	1850

$\chi^2 = 132.31$ , with 16 degrees of freedom.  
 $p = 2.27 \times 10^{-20}$ .

<sup>5</sup> The number of degrees of freedom is, however, not 24, but 16. The degrees of freedom are the number of groups to which arbitrary scores can be given without changing the total score. In the contingency table the totals of both the rows and columns are fixed. In the first row we might put any number in the first four

of the Pearce-Pratt test data. The figure in the upper right hand corner of each cell is the observed frequency of such a correspondence, and the figure in the lower left hand corner the expected frequency of such a correspondence. It will be noted at once that the diagonal cells (the correspondences which are hits) have larger observed frequencies than any of the other cells.

The probability of  $2.27 \times 10^{-20}$  again indicates that there is greater variation among the cells than could occur by chance.

*Control.* Table VI lists the control series in the same way. The chi-square of 18.41, with its resultant probability of .30 shows that 30 per cent of chance arrangements of cells in this fashion would give differ-

TABLE VI  
CONTINGENCY TABLE FOR CONTROL SERIES.

Calls	CARDS					Total of Rows
	Circle	Rectangle	Waves	Plus	Star	
Circle.....	76 75.6	60 75.6	96 75.6	70 75.6	76 75.6	378
Rectangle.....	68 69.8	89 69.8	59 69.8	67 69.8	66 69.8	349
Waves.....	77 73.2	71 73.2	67 73.2	76 73.2	75 73.2	366
Plus.....	83 81.4	78 81.4	82 81.4	83 81.4	81 81.4	407
Star.....	66 70.0	72 70.0	66 70.0	74 70.0	72 70.0	350
Total of Columns.....	370	370	370	370	370	1850

$$\chi^2 = 18.41, \text{ with 16 degrees of freedom.}$$

$$p = .30.$$

cells, but to make the row add up to the proper total the number to be placed in the fifth cell would be determined by the sum of the preceding four numbers. This holds true for all five rows; only four cells of each row are free. It also holds true for each column. The number in the fifth cell of the column is determined by the sum of the first four cells, and the total number of cards. So if 16 cells of the table were filled arbitrarily, the number in the rest of the cells would be fixed. Therefore in entering the chi-square table, the total chi-square has 16 degrees of freedom.

ences as great as, or greater than, those observed. The variations are chance fluctuations.

The criterion of significance for chi-square tests is uniformly the standard set by Fisher,  $p = .01$ . A chi-square yielding a  $p$  value as small as or smaller than this indicates that the variation of the data from chance expectation has been too great to be explained as a sum of chance variations.

#### EVALUATION OF A DIFFERENCE

Most experimental problems using ESP cards consist of two or more series, the results of which are to be compared. If the average hit score per run of one series differs from the average of the second series, it is necessary to know whether this difference is significant, or whether it is within the range of random variations customary when two samples are taken from the same population.

The average difference may be found by subtracting the average hits per run of one series from the average hits per run of the other series. The standard deviation of the difference is computed by the formula

$$\sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2} \quad (1)$$

in which  $\sigma_1$  is the standard deviation of the mean of the first series and  $\sigma_2$  is the standard deviation of the mean of the second group. As  $\sigma_1 = 2.04/\sqrt{R_1}$  and  $\sigma_2 = 2.04/\sqrt{R_2}$  ( $R_1$  and  $R_2$  being the number of runs in the two series respectively), the above formula in terms of the number of runs reduces to

$$\sigma_d = 2.04 \sqrt{\frac{1}{R_1} + \frac{1}{R_2}} \quad (2)$$

The average difference is divided by the standard deviation of the difference to find the critical ratio.

*Application V.* The difference between the averages of the two series we have used for previous examples may be tested for significance. The test series yielded an average of 7.53 for 74 runs, the control series an average of 5.20 for 74 runs. The average difference is 7.53—5.20, or 2.33.

$$\sigma_d = 2.04 \sqrt{1/74 + 1/74}$$

$$\sigma_d = 2.04 \sqrt{.027}$$

$$\sigma_d = .327$$

The critical ratio is  $2.33/.327 = 7.12$ , which indicates that the probability that the two series were random samples of the same population is about  $1/1,000,000,000,000$ .

Instead of the theoretical 2.04, the empirical standard deviation of each distribution may be substituted in the formulas above. (The empirical standard deviation may be found by summing the squares of the amount each hit score of a run deviates from the average for all the runs, dividing by the number of runs, and taking the square root of the result. Any statistical text gives directions for and examples of this computation.) Except for extremely unusual distributions, the standard deviation of the difference so found will be very close to that based upon the theoretical standard deviation.

When the tests of the two series to be compared are presented in pairs, or may be paired off according to some principle, some correlation may exist between the series. For example, if two different kinds of ESP tests are presented alternately to the same subject, he might hit high or low scores on both of them at a given time but favor one or the other such as to produce a marked difference in total scores. The scores would be correlated in that some ESP ability entered into the deviations of both series.

The runs should be paired off and a product-moment correlation coefficient  $r$  computed. (See any statistical textbook for the method; for example, Davenport and Ekas, page 79 ff., or Askin and Colton, page 85 ff.) The standard deviation of the difference is then found by substituting in the formula

$$\sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2}.$$

#### METHODS OF COMBINATION

The data for an experiment using ESP cards is usually collected from a number of experimental periods ranging over many days and including the work of several subjects. When a critical ratio test is applied, the assumption is implicit that any extra-chance factor is reasonably uniform throughout the data. In practice, however, different subjects may have different average scoring levels, and experiments may vary with uncontrolled factors. It is sometimes necessary to study the data as coming from its natural sub-groups rather than pooling it together.

Rhine (12, p. 34) suggested a formula for combining critical ratios which, while difficult to defend mathematically, was very practical for use with the range of values of his data. He summed the squares of

the critical ratios from groups with a positive deviation, subtracted the sum of the squares of the critical ratios from groups with a negative deviation, and found the square root of the result. The formula can be written

$$X_T = \sqrt{|X_{1+}^2 + X_{2+}^2 + X_{3+}^2 + \dots - X_{1-}^2 - X_{2-}^2 - \dots|}$$

in which the + subscripts indicate X values from a positive deviation and - subscripts indicate X values from a negative deviation.

In certain theoretical cases this is equivalent to the critical ratio of the pooled data. Applied to several check series, it gives a slightly lower probability than the pooling process. This would be expected because in a series in which a critical ratio is sensibly zero, the formula ignores the data from which it arose. The formula gives the larger X values much greater weight in determining the X value of the combination.

The Rhine formula is logically defensible in terms of hypotheses of the variability of the factor being tested. We suggest, however, that until the range of its usefulness can be determined, it be used only in comparison with other combination methods.

A chi-square method for evaluating a set of sub-series has been proposed by R. A. Fisher (3, p. 97). The method has the authority of generally accepted statistical usage. In application to some cases it tends to be too severe (for example, in a distribution studied by Norton, 8, p. 350), but this tendency is on the safe side; that is, the computed probability is apt to be greater than the actual probability.

The probability for each of the sub-series is ascertained. Then the sum of the natural logarithms of these probabilities is multiplied by  $-2$  to find the chi-square. The degrees of freedom are twice the number of sub-series. If  $p_1, p_2, p_3, \dots, p_n$  are the probabilities of the sub-series, then

$$\chi^2 = -2 (\log_e p_1 + \log_e p_2 + \log_e p_3 + \dots + \log_e p_n)$$

and the degrees of freedom are  $2n$ .

If the probabilities of the sub-series are obtained by a chi-square evaluation of sub-groups within the sub-series then the additional test for the signs of the sub-groups (see p. 217) may be applied after combination by the Fisher method.

A thoroughgoing interpretation of the methods of evaluation discussed here would require more consideration of the theory of probabil-

ity than the limitations of a single article permit. There are, however, some practical questions which may be dealt with briefly.

An anti-chance probability in favor of a particular hypothesis must be considered in respect to *all* the data of the experiment. For example, if 200 subjects were given a test of 100 trials each, and one of the scores yielded a significant critical ratio of 2.5, the conclusion that the given subject had demonstrated an extra-chance factor would be without basis. One such critical ratio at least would be expected by chance in the population of 200.

When data is selected for the study of some hypothesis, the selection must follow a principle independent of the results. Any such selection is admissible. The data may be grouped by subjects, by experimental periods, by days of the week, or by any other principle so long as the selection is, on a chance hypothesis, independent of the scores of the runs selected.

The mathematical techniques discussed here are essentially operations. They test the assumption that a group of data varies only in a random chance manner. They can answer the question only in the words, "probably yes" or "probably no." Any other conclusions based on them are matters for the logic of science. It is afiel from the purpose of this paper to defend the hypothesis of extra-sensory perception. But it is hoped that the foregoing applications of methods will effectively eliminate two alternative hypotheses: That ESP results are chance results, and that their significance is an artifact of the mathematical treatment.

#### REFERENCES

1. ARKIN, H. and COLTON, R. R. *An Outline of Statistical Methods*. New York: Barnes & Noble, 1934. Pp. 177.
2. DAVENPORT, C. B. and EKAS, M. P. *Statistical Methods*, 4th Ed. New York: John Wiley & Sons, 1936. Pp. xii + 216.
3. FISHER, R. A. *Statistical Methods for Research Workers*. 4th Ed. London: Oliver and Boyd, 1932. Pp. xiii + 307.
4. GREENWOOD, J. A. The variance of ESP call series. (In manuscript for early publication in the *Journal of Parapsychology*.)
5. KELLEY, T. L. *Statistical Method*. New York: Macmillan Co. 1923. Pp. xi + 390.
6. KELLOGG, C. E. Dr. J. B. Rhine and extra-sensory perception. *Journal of Abnormal and Social Psychology*, 1936, XXXI, 190-193.
7. LEMMON, V. W. Extra-sensory perception. *Journal of Psychology*, 1937, IV, 227-238.

8. NORTON, H. W. Use of scores in an exact test of significance in a discontinuous distribution. *Annals of Eugenics*, 1937, VII, 349-354.
9. PEARSON, K. *Tables for Statisticians and Biometricians*. 3rd Ed. London: University College, 1930. Pp. lxxxiv + 143.
10. PETERS, C. C., and VAN VOORHIS, W. R. *Statistical Procedures and their Mathematical Bases*. State College, Pa.: 1935. Pp. vii + 363.
11. PRATT, J. G. Clairvoyant blind matching. *Journal of Parapsychology*, 1937, I, 10-17.
12. RHINE, J. B. *Extra-Sensory Perception*. Boston: Bruce Humphries, 1934. Pp. xiv + 169.
13. RHINE, J. B. Some selected experiments in extra-sensory perception. *Journal of Abnormal and Social Psychology*, 1936, XXXI, 216-228.
14. STUART, C. E. In reply to the Willoughby "Critique." *Journal of Abnormal and Social Psychology*, 1935, XXX, 384-388.
15. WEISBERG, L. Note in "Personal and Otherwise" section of *Harpers*, February, 1937.
16. WILLOUGHBY, R. R. A Critique of Rhine's "Extra-Sensory Perception." *Journal of Abnormal and Social Psychology*, 1935, XXX, 199-207.
17. WILLOUGHBY, R. R. Prerequisites for a clairvoyance hypothesis. *Journal of Applied Psychology*, 1935, XIX, 543-550.
18. WRIGHT, E. H. Note in "Personal and Otherwise" section of *Harpers*, February, 1937.
19. YULE, G. U. *An Introduction to the Theory of Statistics*. London: Charles Griffin & Co., 1912. Pp. xv + 381.

## LETTERS AND NOTES

To the Editors of the *Journal of Parapsychology*.

Sirs:

I was glad to receive a copy of your admirable *Journal* for March, and to find in it an interesting account of my work in Extra-Sensory Perception, written by Mr. B. M. Smith. By your kind invitation, I am sending the following comments.

Mr. Smith has skilfully outlined my method and main results in a clear and sympathetic account, but there is one curious property belonging to the humanly, or, as he calls them, "subjectively" selected events, as contrasted with mechanically selected events, which I do not think he has entirely grasped. Of this property, discovered by Mr. G. W. Fisk, Mr. Smith says: "He found that if he knew the success or failure of each trial, he could score by system. He need only avoid a box for a time after a hit occurred. This was due, he found, to Mr. Tyrrell's tendency to avoid duplication of boxes in his selection." (p. 67).

Actually, Mr. Fisk's method was to begin by opening any box, and to go on opening it until a success occurred. Then he passed to another box and continued to open that until a success occurred there, and so on in strict routine. It was found that, when the lights in the boxes were humanly selected, the average percentage success given by the method was nearly 30: but when the lights were mechanically selected, the average percentage success was only 20. But it turned out that the reason for this did not lie in any tendency on my part, as operator, to avoid duplicating the lights. On the contrary, duplications and triplications were about as frequent with the human as with the mechanical selector. (Other operators found the same.) The difference lay in the fact that from time to time the mechanical selector would avoid a particular box for quite a large number of trials together, whereas the human selector tended to distribute his choices *more* evenly among the boxes and never neglected any one box for a long period of time.

It is the *maximum interval before repetition* which is *less* in the humanly selected events than in the mechanically selected events; and it is in this particular that humanly selected events differ from a purely random order.

The same thing applies to a pack of 25 "Zener" cards. Such a pack, however it is shuffled, does not give a truly random sequence, because the condition is imposed that five cards of every kind must enter into 25 trials; and this would only rarely happen with cards in a truly random order. However, the Fisk method of utilising this fact to score above 20%, depends on a knowledge of each result as it occurs, and this has never been the case in card-experiments. There is no reason to suppose that humanly selected events are not as good as mechanically selected events, as far as this particular kind of departure from randomness is concerned, so long as the subject does not know the result of each trial.

Mr. Smith then goes on to say: "The crucial question would seem to be, 'Did Miss Johnson share this peculiarity with him [Mr. Fisk.]'? It would be only natural to do so, since it requires training to avoid the tendency." (p. 67).

But Mr. Smith has misapprehended this point. We were not dealing with Mr. Fisk's natural way of scoring but with a deliberate experiment which he was trying; and there is no evidence to show that anyone else ever used this method. In fact, there is a good deal of evidence that no one did, as I have tried to show in my Report. When it was most unfortunately explained to Miss Johnson that this method existed, she tried to use it in order to test the result; and the result was that she mingled it with her own natural way of scoring and merely reduced her scores to chance value. She never did succeed in isolating it and using it even as a test. It only introduced a confusion which has spoilt a great deal of valuable work.

There is, however, no possibility of mistaking results which are attained by the Fisk method with results attained by the method of E.S.P. The Fisk method on the machine gives a steady rate of scoring averaging about 10% above the ordinary chance, with anyone who tries it. Miss Johnson's results, on the contrary, have always fluctuated in an incalculable manner and have been in the habit of going down to chance sometimes for days together. No scoring by any kind of system of number-habit would do this.

It is unfortunate that this Fisk experiment not only had the effect of seriously upsetting the subject; it also seems to have directed attention from the main results which had been obtained. The principal figure was, in fact, delayed by the former cause, so that it had to be inserted in a foot-note after the report was in type. Mr. Smith says: "Tyrrell reported a long series of 3,261 trials at an average of 21.7 per cent., and odds of 125 to one, carried out using the mechanical selector to determine the choice and using the keys, with choice unknown to the operator. These odds (125 to one) are taken as significant by Tyrrell, though they are somewhat below the standard used in the Duke experiments and other similar work." (p. 66).

If Mr. Smith had looked at the foot-note on p. 151 of my Report, he would have seen that the figures for this class were 7809 trials at an average of 23.5% and odds of about a hundred million million to one; and it was these odds which I was quoting as significant.

Subsequently to this report, this class of mechanically selected numbers, the lamps lit being unknown by the operator, has risen to 10,959 trials at an average rate of scoring of 23.8%, and probability that it is due to chance is about 10<sup>-22</sup>.

The Fisk method could not, of course, possibly apply to these. Nor has any adequate explanation except an extra-sensory faculty been suggested.

With regard to the possible lack of impartiality of the Selector, it was subsequently tested more than once and the results of 2,000 trials have now been subjected to mathematical analysis without any significant impartiality being discovered.

One fact which, I think, was not made clear in my Report is that the wires from the Selector pass through the Commutator on their way to the lamps. As the commutator is continually being altered, any impartiality in the Selector, if it existed, would be distributed amongst the five lamps.

But as a matter of fact, number preferences, even if they exist on both sides of the experiment, cause surprisingly little disturbance of the probability of success. In diagrams, I. and II. (See my Report in *Proc. S. P. R.*, Vol. xlv, p. 167) will be seen a distribution characteristic of myself on the left and of Miss Johnson on the right. In both cases there is a fairly marked tendency to favour the middle box; yet when worked out it is found that it only alters the probability of success from 0.2 to 0.2008. (See pp. 130, 131.) In order to give a score about the same as that of a first class subject, both operator and subject would have to favour the middle box 50 times in 100 trials. So that slight irregularities in impartiality could not possibly account for the results actually obtained.

One final point I should like to mention is that it seems to have been rather assumed that I have advocated the motor response as superior *in general* to the verbal response. It is true that I designed my apparatus with this as one of the points in view; but I also designed it with one particular subject in view, viz. Miss G. M. Johnson. And in her case, everything goes to show that the motor response is easier. Probably the same is true with some other subjects, but I should not like to be taken as holding that it is necessarily true generally.

Yours faithfully,

7 June, 1937.

G. N. M. TYRRELL,  
6, Carlton Hill,  
London, N. W. 8.

## 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*,  $S. D. = 2 \sqrt{\text{no. of runs}}$ .
- 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*.