

THE PROCEEDINGS OF THE AMERICAN SOCIETY FOR PSYCHICAL RESEARCH

Tests of Extra-Sensory Perception Among A.S.P.R. Members

BY DR. GARDNER MURPHY AND ERNEST TAVES

Columbia University

During the academic year 1937-38 the present writers were carrying on a study of extra-sensory perception among normal persons, whose task it was to try to "perceive" various concealed materials. Sometimes the material was a few feet away; sometimes at distances ranging up to 3,000 miles. Much of this work has already been published in the *Journal of Parapsychology*.¹ The Executive Secretary of the A.S.P.R. participated in a number of these experiments, visiting our laboratory at Columbia University, and also making a large number of long-distance calls on material in our laboratory. Owing largely to her interest, several other members of the A.S.P.R. began to take part in the work during the spring and summer of 1938. Their work, with one type of material, up to November, 1938 is included in the publication mentioned in Footnote 1. The main bulk of the data to be reported here, however, is new, consisting of the results of work done at the New York offices of the A.S.P.R. between November 1938 and March 1940.

The method of organizing a research group at the A.S.P.R. offices was as follows: The Executive Secretary reported to the President the willingness of the present writers to conduct experiments at the Society offices under carefully controlled experimental conditions. The President arranged a meeting at which the senior author made a brief address on some current ESP problems and the junior

¹ Murphy, G. and Taves, E., Covariance methods in the comparison of extra-sensory tasks, *J. Parapsychol.*, 1938, vol. 3, no. 1, 38-78.

author assisted in the demonstration of a number of current methods. Fifteen volunteers expressed a readiness to come once or twice a week in the late afternoon to take part in telepathy and clairvoyance tests under the present writers' direction. The meetings, immediately began, the usual times being Monday at 4:30 p.m. and Friday at 4:30 p.m. There were usually five or six subjects present at the Friday meetings, and eight or nine at the Monday meetings, some persons at times attending both. Subjects were free to participate at any time, and at no time was any subject discouraged from participating in the experiments. Neither was any person subjected to undue pressure to participate. The subjects were, that is, entirely free to participate or not at any time.

The general background of our problem was this: Prior to 1938 large-scale investigations of the capacity of ordinary normal persons to read or guess material not accessible through the ordinary channels of sense had, as far as we can ascertain, never given a clearly positive result. The results of J. E. Coover² at Stanford University, making use of one hundred subjects, each of whom called playing-cards a thousand times, are admittedly ambiguous; the results of S. G. Soal³ in the analysis of material broadcast in England appeared completely negative; the 1924 radio test made by the present senior author did not lend itself to statistical treatment, while the statistical treatment of the Zenith Radio data obtained during the winter of 1937-8 is notoriously difficult to interpret.⁴ As far as we have ascertained, clear-cut, positive results in clairvoyance, under good experimental conditions, have regularly been reported not from an ordinary, normal, unselected population, but from *specially gifted* persons. This is true, for example,

² Coover, J. E., *Experiments in psychical research at Stanford University*, Stanford University Press, 1917, xxiv + 641.

³ Soal, S. G., Experiments in supernormal perception at a distance, *Proceedings of the S.P.R.*, 1931-32, vol. 40, 165.

⁴ Goodfellow, L. D., A psychological interpretation of the results of the Zenith radio experiment in telepathy, *J. Exper. Psychol.*, 1938, vol. 23, 601-632. See also, Stuart, C. E., A review of recent criticisms of ESP research, *II. J. Parapsychol.*, 1939, vol. 3, no. 2, 194-205.

of the classical British, French, Dutch, and German work, and, as Dr. Rhine has pointed out over and over again, is true even in the exceptionally successful Duke University experiments.⁵ To the present writers, then, there seemed to be plenty of room for one more effort to ascertain whether extra-chance results are obtainable from a group of subjects distinguishable from the general public *only in terms of their interest*, not in terms of previous demonstrations of high scoring ability.

It is, of course, to be granted immediately that members of a society for psychical research may, in a sense, be a select group, selected because of some actual gift in themselves or in their relatives. The point is, however, a hazardous one. There is nothing in our contact with the group, or in accounts of their experiences, to suggest that they differ clearly from an ordinary sample of the population, except that their educational level and range of intellectual interest are above average, while at the same time their membership in such an organization doubtless means also a less skeptical attitude towards the existence of such phenomena than is ordinarily found in comparable groups of intellectuals.⁶

Procedure

One of the great needs in psychical research for a half century has been the development of some simple method of experimentation which any competent investigator could follow, so that the results from many investigators working with many subjects could be compared. If one man with one method gets positive results in the study of telepathy, clairvoyance, precognition, and other allied phenomena, and another man is unable to get the same results, the data are

⁵ Rhine, J. B., *Extra-sensory perception*, Boston S.P.R., 1934.

⁶ Dr. W. F. Prince, in a report on *Human Experiences*, Bulletin of the E.S.P.R., 1931, found a large number of eminent persons, most of them college graduates, who reported psychic experiences, but these did not constitute by any means a majority of the group to whom the questionnaire was sent. In an unpublished questionnaire among Columbia graduate and undergraduate groups, the senior author found about 60% willing to respond, and of these about two-thirds reported positively on having had one or more "psychic experiences" (without any commitment as to the explanation of the experience).

likely to seem ambiguous to other investigators and to most readers. In any border-line field a *repeatable method* is important in order that basic incontestable facts can be independently ascertained.

One feature of such a repeatable method would be the selection of a type of material which could be used with many different groups of subjects. Another important criterion would be that the method be usable with the ordinary run of normal persons, not merely with special sensitives. A positive finding in one laboratory based on the use of a specially selected group of sensitives might be quite valid in itself, but since the very question at issue is likely to be the genuineness of the result, the method should be one which can give clear results with any ordinary normal population, so that other normal populations can be compared with it.

When the opportunity presented itself to carry out experimental studies of extra-sensory perception at the New York offices of the A.S.P.R., the present writers undertook to start from the very beginning in formulating a method which would meet the stated requirements. This meant, in the first place, working with a group of subjects who did not claim to be and who, in fact, were not exceptional persons or "gifted sensitives." It meant, secondly, that a very large variety of simple tests of telepathy and clairvoyance should be devised, in the conviction that capacities of the normal person might well appear in relation to some test methods which would be altogether absent with other test methods.

Our present report is a description of our quest for such a method, rather than a final report on any spectacular results. It is felt that the data obtained are of value and importance only in so far as they clarify, both for us and for other investigators, the general problem of methods of testing for extra-sensory capacities in ordinary normal persons. If we move one short step in the direction of devising a universally valid and suitable method which will

in subsequent years lead to more general agreement among investigators as to what kind of results can be obtained with a uniform method, our present report will have served its purpose.

Now as regards our general procedure. It had been G. M.'s practice in earlier years to carry out with each new subject a series of crude *preliminary experiments* in order to awaken the subject's interest and to gauge roughly the likelihood of finding anything worthy of sustained research. Such preliminary experiments, analogous to those reported by Upton Sinclair⁷ and by René Warcollier⁸, are seldom capable of giving results which can be clearly appraised in terms of chance coincidence. Sinclair's experiments, for example, though brilliant in results, can only be evaluated strictly if the total number of impressions correctly transmitted can be compared with a definite number indicating the successes which might conceivably be attributed to the operation of chance alone. As pointed out by E. T. in the March number of the *Journal of the A.S.P.R.*,⁹ a statistical method for the treatment of chaotic and casual material of this sort is not only unavailable but inconceivable. In practice, crude preliminary experiments may actually get the subjects into bad working habits and make strict work more difficult to achieve at a later time. We determined, therefore, from the beginning to use material which could be evaluated strictly in terms of the theory of probability. This is not at all a difficult thing to do, nor, which is important, does it necessitate that the material be entirely without interest to the subjects.

A second requirement was that the material must be categorically out of the range of the known senses. This was achieved by the following procedure: The subjects in most cases sat in the south room (room A in Figure 1) of the A.S.P.R. offices on the 9th floor at 40 East 34th Street,

⁷ Sinclair, U., *Mental radio*, Pasadena, 1930.

⁸ Warcollier, R., *Experiments in telepathy*, Boston S.P.R., 1938.

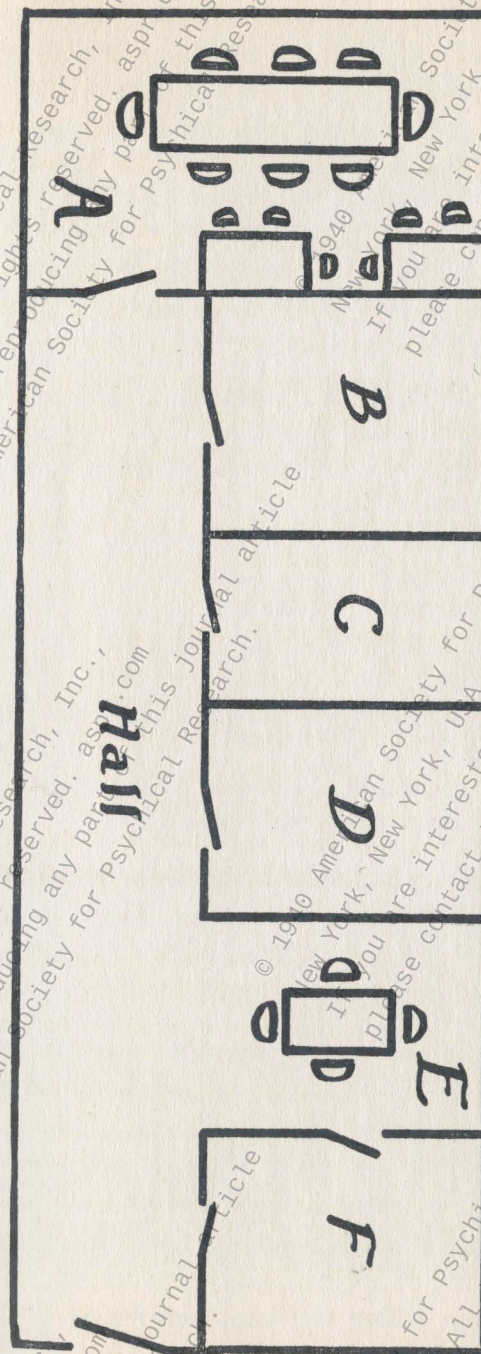
⁹ Taves, E., Coincidence in psychical research, *Journal of the A.S.P.R.*, 1940, vol. 34, 65.

New York City, either gathered around a central table or in other chairs nearby, as shown in the floor plan. On some days, when the number of subjects was small, it was found more convenient to use the smaller room E. There is no indication in the results that this variable was of any importance.

In all tests of pure clairvoyance in which no human mind was or had been occupied with the nature and order of the stimulus-material, the material to be called or guessed was concealed in a cardboard box or other opaque container. This material consisted in most cases of decks of cards of various sorts which had been elaborately shuffled and cut, face down, by a person other than the experimenter. Thus, for example, E. T. would shuffle a deck of cards and would place the deck in a cardboard box, and the box would be given to G. M. as he started down-town to the A.S.P.R. office, or vice versa. In some experiments, in which the same deck of cards was used several times at one session, the experimenter retired to room B, C, or D, shutting the door, where he shuffled the cards thoroughly, cut the deck while not looking at it, placed the deck in the box, and took the box into the experimental room. As the empirical controls of shuffling in a later section of this report indicate, all persons who shuffled the cards achieved a random order of cards in the deck, i.e., the shuffling was adequate no matter who performed the operation.¹⁰

¹⁰ In addition to the tests of clairvoyance in which the actual arrangement of the material to be guessed was known to no one, there were two other procedures: (1) The "pure telepathy" (PT) procedure in which the experimenter acted as agent or asked another member of the group to act as agent, choosing the material and concentrating upon it, but having no physical reference point or object on which to concentrate. (2) There were also some instances of the procedure which Rhine calls GESP (general extra-sensory perception), in which the agent gazes at the material to be transmitted. In this case correct perception of the material might conceivably be due either to the telepathic factor (i.e., contact with the agent's mind), or to the clairvoyant factor (i.e., contact with the stimulus material itself). In general, most telepathic and GESP material, depending as it does upon voluntary selection or some very inadequate method of selecting what is to be sent, lends itself only very poorly to statistical appraisal. A small amount of work has been done with GESP in which the material to be transmitted was chosen by a "pure chance" method (involving the use of dice and similar material). These data will be reported in due time; they are not yet sufficient to warrant any conclusions.

FIG. 1. FLOOR PLAN OF THE A.S.P.R. OFFICES



Sometimes both experimenters attended the meeting, but as a rule only one was present. As the tables show, E. T. gathered well over half of the data. On two occasions Dr. Gustave Gilbert, who was familiar with our methods, took charge of the meeting when neither of us was free at the appointed hour. A comparison between results obtained by the different experimenters will be made later. Also, Dr. Gilbert at times assisted G. M. and shuffled the cards. As just noted, however, the identity of the *shuffler* does not appear to be an important variable.

The subjects, then, seated themselves informally at the tables in the experimental room, and followed the instructions given them by the experimenter, namely, to "guess" card orders in various kinds of decks. Figure 2 indicates the way in which the subjects proceeded. The group is shown seated around a table, recording their guesses, or "calls." It is important here, we believe, to stress the informality of the laboratory situation. Without relaxing actual experimental control, the experimenters attempted to make the subjects feel completely at ease; all subjects were extremely cooperative, and the resulting atmosphere was conducive to the collection of considerable quantities of data under rigid, yet informal, conditions.

It is now necessary to define in detail the special materials used in the clairvoyance tests, bearing always in mind that these were varied somewhat from week to week in accordance with the judgment of the experimenter.

A large amount of "50-50" material was made up; that is to say, material in which each guess should have, on a purely statistical basis, an even chance of being correct. Here we used red and black cards, red and yellow cards, yellow and black, black and white, red and white, and, less frequently, numerous other combinations. We used photographs of men and of women, and we used several pairs of symbols from the Rhine set, e.g., permitting a choice be-

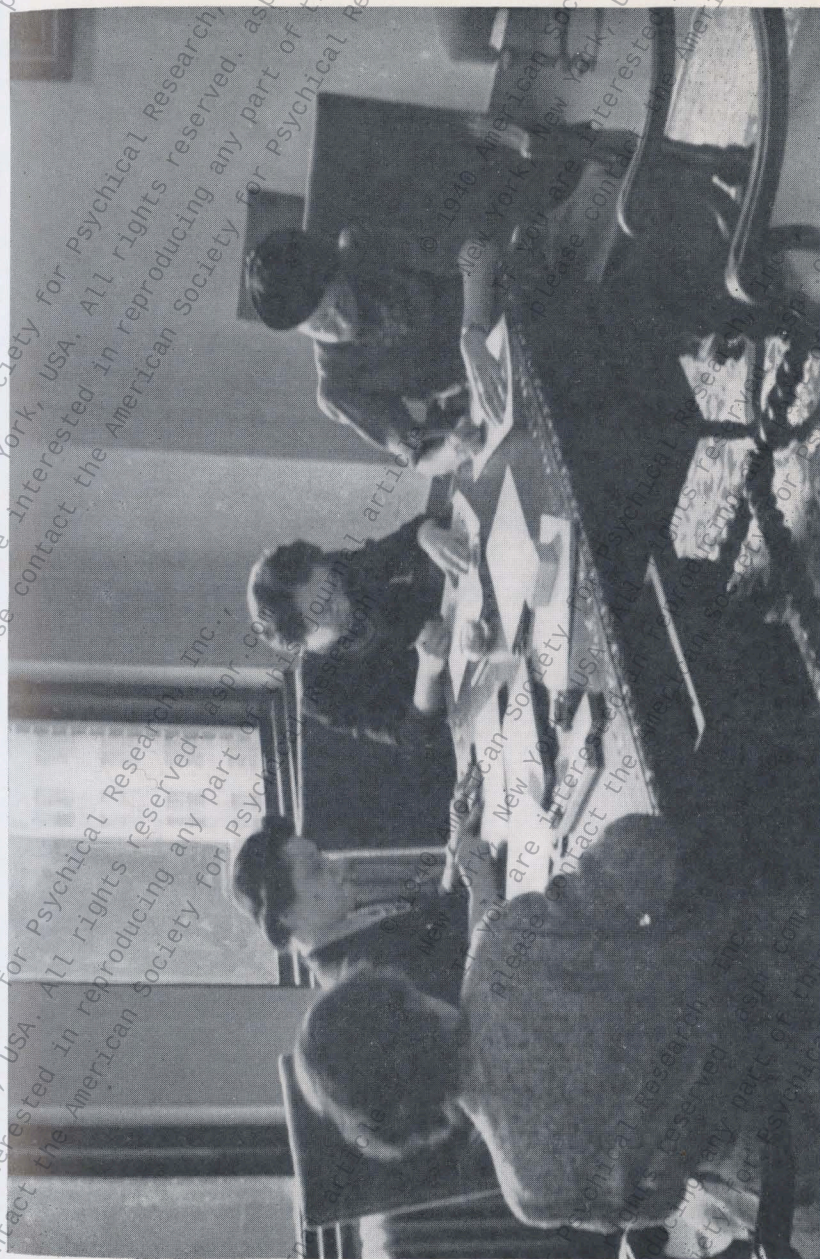


FIGURE 2. A.S.P.R. SUBJECTS SEATED AT TABLE. FROM LEFT TO RIGHT ARE MRS. HINES, MISS WELLMAN, MISS MUNT,

Tests of Extra-Sensory Perception

[9

tween circles and crosses, or squares and wavy lines. In addition to this 50-50 material, we used playing-cards, where the odds are 1 in 4 of getting the suit correct, 1 in 13 of getting the numerical value correct, and 1 in 52 of guessing the card exactly.

We also made a great deal of use of "Rook" cards, the deck consisting of fifty-six cards with 4 suits (printed in different colors), and the cards being numbered simply from 1 to 14 in each suit. The subjects were usually required to call the top 14 and the bottom 14 cards of this deck, although occasionally the entire deck or the top 28 cards were called. Most of the Rook data are from calling either the top and bottom 14 cards, or from calling just the top 14 cards, and in this paper we are reporting only on such calls; we are obtaining more data on Rook calls when the subjects are required to call either 28 or 56 cards in a row, but these experiments are not yet complete.

We also made considerable use of what we call the "5-dot" task, an outgrowth of a method used with interesting results at Columbia. Upon 5 completely blank cards, plain white stock with no mark on either side, large black dots were marked in ink. These 5 were then shuffled into a deck of 45 identical blank cards, bearing no markings of any sort. After a thorough shuffling, and cutting, face down, the 50 cards were placed in a cardboard box and the subjects were required to indicate on which of the 50 cards there was a black dot, being allowed only 5 guesses in which to locate the marked cards—guessing, for example, that cards 2, 18, 35, 40 and 49, counting down from the top, were the marked ones. The chance expectation here is one correct guess in every two complete executions of the task. (The chance that an individual call is right is one in ten, and this is multiplied by 5, since there are 5 guesses. In the course of two executions of the task, 10 guesses are made as to the location of a card bearing a dot, and, by chance alone, one of these ten guesses may be expected to be correct.)

The subjects were provided with record sheets, with a carbon paper and duplicate attached, so that all their "calls" were recorded in duplicate. Before the experimenter opened the box of cards, each subject handed in one copy of his calls to him. This item seems worthy of some emphasis here in view of current criticisms of published ESP research. It is important to guard against errors in recording, either by subjects or by the experimenter. The routine was, therefore, for the experimenter to record the target order as he opened the box, and *before* he looked at anyone's calls. The method of scoring for *official* purposes, when the data were assembled at Columbia, was to juxtapose the target order and the call order as written down by the subjects on record sheets. For *unofficial* purposes of informing the subjects as to their progress, it was common to read off to them the actual order, allowing them to check their own scores after each task, but the carbon of the call order had, of course, previously been handed in, permitting a completely independent checking against the official target order.

Results

We are now ready for the results. We shall here give and discuss the significant findings. Complete data will be found in the Appendices.

Regarding the general problem of method which we raised in the introduction, it should be made clear first of all that, under the conditions described, all of the 50-50 tasks *except the red-white* must be regarded as giving results so completely comparable with chance expectation as to justify no discussion whatever. In most of the 50-50 tasks this group of normal subjects working with this group of experimenters obtained nothing of parapsychological significance. As an example, Table I shows the number of guesses regarding faces of men and of women in photographs, the number of successes, and the number of successes expected if chance alone were operating. With-

Tests of Extra-Sensory Perception

TABLE I. Sets of photographs of men and of women, shuffled and placed in boxes, and called in "down-through" order.

Number of calls	1600
Number of correct calls	807
Expected by chance	800
Deviation from chance	+7

out statistical discussion, it is sufficient to report here that these scores are well within the limits of chance expectation. The same is true for the other 50-50 tasks; there is no evidence for the operation of other than chance factors. The complete data for all 50-50 tasks are given in Appendix I. Of the tasks with odds other than 50-50, none gave significant results except the Rook cards.

The data from the *Rook* and the *red-white* cards, however, are in a different category. These results are very surprising, and appear to warrant the view that the general problem of method raised at the beginning has been somewhat advanced towards the point of ultimate solution.

On February 20, 1940, when all of the data in hand at that time were brought together, it became evident that the correct calls on the top 14 and bottom 14 Rook cards in the deck of 56 were considerably in excess of chance expectation. Table II presents the relevant Rook data up to this date; the letter P (probability) indicates the likelihood of getting such a result by chance alone ("random sampling"), i.e. there are only four chances in 1000 of getting such high Rook scores through the operation of nothing more than chance.

The Rook scores, then, suggested that we were making progress towards a method. But the data in hand by February 20th pointed to another result, just as striking, namely, a very consistent tendency for our subjects to go *below* chance on the red-white decks. The complete batch

TABLE II. Data on Rook calls to February 20, 1940.

Number of calls	14,826
Correct calls	1,143
Chance expectation	1,059
Deviation from chance	+84
P	.004

TABLE III. Data on red-white calls to February 20, 1940.

Number of calls	2,650
Correct calls	1,230
Chance expectation	1,325
Deviation from chance	-95
P	.0002

of data collected up to this time, including all tasks, would be *extremely* unlikely to give any task as high as the Rook, or any task as low as the red-white through the operation of chance factors. As seen in Table III, the red-white calls on February 20, 1940, numbered 2650, and the successes numbered 1230 as against the 1325 expected by chance. The P value here is .0002, which means that there are only 2 chances in ten thousand that a chance (random) sample would give such low scores on the red-white deck.

(For evaluating the obtained deviations from chance expectation the binomial statistics are used. The P value is determined by computing in the case of the Rook data, for example, the moments of the distribution $(1/14 + 13/14)^{14}$ and substituting in the Charlier expression of the generalized normal curve. In this connection see Huntingtons' paper, A rating table for card-matching experiments, in the *Journal of Parapsychology*, 1937, vol. 1, no. 4, 198.)

With these data in hand, it was immediately determined to gather two types of additional information: (1) Both E. T. and G. M. must gather considerably more material with the Rook and red-white decks with the same group of subjects to ascertain whether the *positive* trend for Rook and the *negative* trend for red-white would continue. (2)

Completely independent investigators, preferably those having no convictions on the question of extra-sensory perception, should be asked to collaborate, working entirely independently and with fresh groups of subjects.

Accordingly, a considerable amount of additional Rook and red-white data were gathered at the A.S.P.R. meetings, and the research carried forward through March 31, 1940, at which time the method was finally terminated so far as the present report is concerned. Concurrently, G. M. wrote to professional psychologists of his acquaintance teaching at universities and colleges in various parts of the United States, instructing them in the preparation of the Rook and red-white cards, and asking them to get their classes to call them by the down-through (DT) method. The results from these classroom experiments will be reported when available.

The additional data from the A.S.P.R., however, have clearly continued to support the results already indicated, and in the following section of this report the complete data, up through March 31, 1940, will be presented.

Analysis of data collected through March 31, 1940

I. *Rook*. Approximately 3,000 additional Rook calls (first and last 14 cards in the deck) were collected between February 20 and March 31. This additional sample of data also revealed the positive trend which was previously in evidence, and the entire number of calls and successes is indicated in Table IV. This table represents the entire Rook data for the method described, and the P value lies beyond the ordinary limits of chance variation.

TABLE IV. Complete Rook data, top and bottom decks of 14.

Number of calls	17,612
Correct calls	1,361
Chance expectation	1,258
Deviation from chance	+103
P	.0015

Table IV presents most of the relevant Rook data, but it still leaves a number of questions unanswered. We know that we have a positive deviation from chance expectation, a deviation of sufficient magnitude to be designated significant, but we do not know to what kind of performance on the part of the subjects the deviation is to be attributed. Does the deviation result from many scores just a little over chance, or from relatively few extremely high scores? Do the subjects score consistently a trifle above chance, do they score mostly at the chance level, with occasional extreme scores, or are both phenomena present? These questions are of considerable interest, and in order to answer them it is necessary to tabulate the obtained frequencies of the various scores.

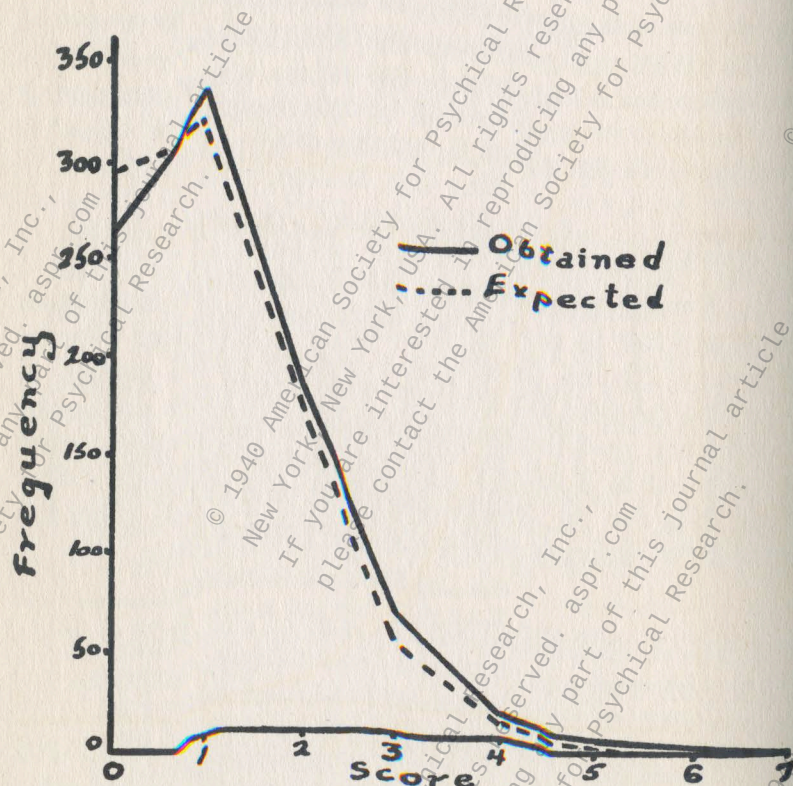


FIGURE 3. DISTRIBUTION OF ROOK SCORES — TOP 14

Chance expectation for the deck of 14 Rook cards is 1, but this does not mean that all scores above 1 are definitely extra-chance. Rather, chance will produce a distribution of all the scores from 0 to 14 (although no scores as high as 14 are expected by chance unless the number of trials is astronomically large), the average of which will be 1. It is possible, then, to compute the expected chance distribution, which may then be compared with the obtained distribution of scores by means of standard statistical techniques. This process yields the additional information we desire. In Figures 3, 4, and 5 are shown the expected and obtained distributions of the various Rook scores. From these curves and from the data presented in Table V it is apparent that

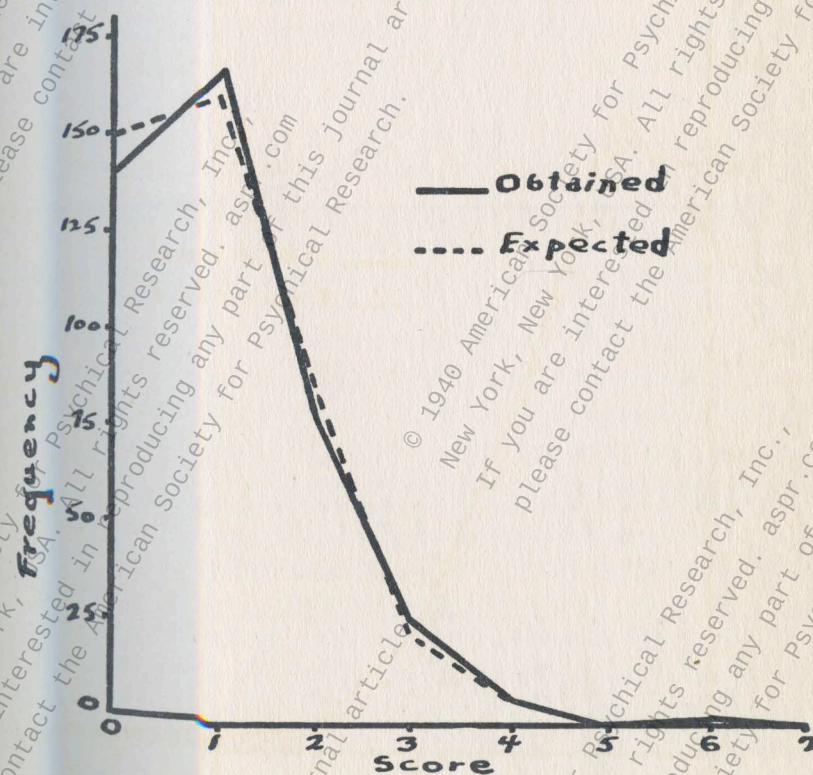


FIGURE 4. DISTRIBUTION OF ROOK SCORES — BOTTOM 14

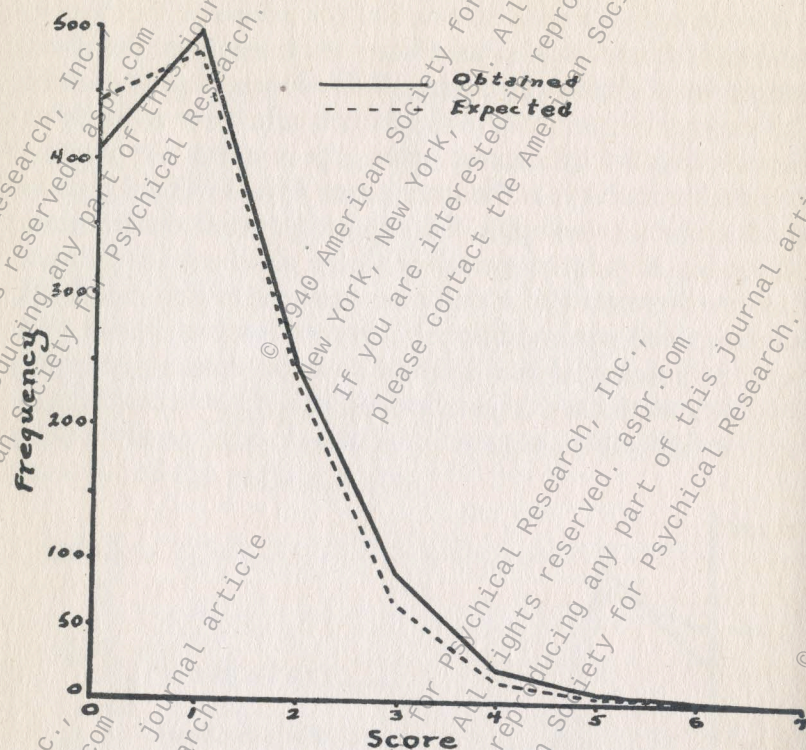


FIGURE 5. DISTRIBUTION OF ROOK SCORES — TOP 14 AND BOTTOM 14 COMBINED

the deviation from chance expectation shown in Table IV is attributable both to a deficiency of scores of 0 and a surplus of scores of 4, 5, and 6 (6 being the highest score we have ever obtained on a deck of 14 Rook cards).

TABLE V. Comparison of expected and obtained distributions of Rook scores — top 14 and bottom 14.

Score	Obtained Frequency	Expected Frequency	Difference
0	404	446	-42
1	493	480	13
2	247	240	7
3	89	74	15
4	20	16	4
5	3	2	1
6-14	2	1	1

Scores of 1 and 2 are near the expected frequencies, scores of 0 are less frequent than expected by chance, and scores of 4, 5, and 6 are in excess of chance expectation. In interpreting Table V it is important to note that the actual size of the deviation from chance must be considered with respect to the magnitude of the score. The deviation of 4, for example, for the score of 4 is considerably more significant than the deviation of 13 for the score of 1. Deviations of the obtained frequencies from expectation are considerably more significant where the score is high than where the score is only slightly deviant from chance.

Next in the analysis of the Rook data comes the question whether the differences between the obtained frequencies and the frequencies expected by chance represent a significant deviation, or whether these differences are of such small magnitude that they might well be obtained in a procedure involving nothing more than random sampling. The Chi-square procedure provides such a test of significance, and obtained values are presented in Table VI.

TABLE VI. Chi-square and P for the various Rook distributions.

Distribution	Chi-square	P
Top 14	9.105	.06
Bottom 14	2.255	.70
Both	9.447	.05

The P values of .06 and .05 in Table VI are not small enough to be termed significant, but they are suggestive. What we have so far, then, is the knowledge that the total deviation from chance of the Rook scores is significant, and that the departure from expectation occurs not in one place in the distribution of scores, but is more or less spread out over the entire distribution, so that the obtained distribution still does represent a pretty fair "fit" to the expected distribution. Stated differently, this means that our significant deviation from chance was not obtained in spectacular spurts of high scoring, but was obtained more or less con-

sistently during the research. This is, perhaps, an indication that normal people have generally available a small quantity of some psychic factor. This is speculative, but it indicates the next step in the analysis, which is to break down the scores in terms of individual subjects to determine whether the positive deviation is from all subjects more or less equally, or principally from a few "gifted" subjects. This analysis was performed, and the results are shown in Table VII. Forty-two subjects participated in the Rook experiments; some subjects called the deck but once, whereas one subject called the deck 222 times. Of the 42 subjects, 14 are below chance expectation, 23 are above, and 5 subjects show no deviation whatever. The table shows that over a third of the total deviation came from one subject, J. G., whereas considerably less than a tenth of the total number of calls came from this subject.

The average score of the entire group is 1.081 correct calls per deck of 14 cards, chance expectation being 1.000. It is of interest to know whether the entire group, excluding the exceptional case of J. G., scored significantly above chance. The average score for the remainder of the group, then, is 1.056, and if J. G.'s scores are reduced to this level, statistical analysis reveals that there are 92 chances in 100 that the "true" difference between expected and obtained scores is greater than zero, and is positive. The result is suggestive, but since it still leaves 8 chances in 100 that the difference is not greater than zero, no conclusion that the group as a whole scored significantly above chance is indicated. Rather, the writers are of the opinion that there is a certain amount of evidence that the entire group did manifest some small ability to score above chance; without the results of J. G.'s calls the data are merely suggestive, but the entire mass of data is quite striking.

The analysis of the Rook data was almost complete at this point, but there was one other question in which we were interested — the question at which points in the

TABLE VII. Breakdown of Rook data by individual subjects.

Subject	N*	D**	P	Subject	N	D	P
LA	1	0		RL	10	7	
AB	155	0		CM	31	-7	
EB	2	1		EM	2	0	
MB	4	1		GM	18	3	
OB	2	0		HM	100	-3	
VB	4	2		MM	118	15	.09
GD	6	2		SM	6	1	
KD	10	-1		AN	5	-1	
LD	6	-2		JO	1	-1	
BF	12	2		AP	20	-2	
HF	4	10		JP	47	2	
RF	30	8		MP	46	-4	
SF	80	4		ER	9	-2	
CG	6	1		HR	1	-1	
GG	10	8		IR	54	8	
JG	90	37	.00007	MR	2	-1	
AH	8	-2		BT	4	1	
DH	34	-7		ET	6	4	
GH	68	5		CV	2	-1	
MH	19	1		AW	222	13	
EK	2	0		FW	1	2	
Total deviation			103				
Total N			1258				

* N is the number of decks of 14 called.

** D is the deviation from chance expectation.

deck was a correct guess most likely. In other words, when guessing the order of cards in a deck, is spatial position in the deck a factor of importance? Rhine¹¹ had previously reported interesting results with respect to this question, and we wondered whether our data contained any such result. For this particular treatment of the data we used only the calls from J. G., since here, apparently, a genuine extra-chance factor was producing high scores, and here would be the most promising field in which to search for any such effect. J. G.'s correct guesses were therefore tabulated according to position in the deck; the results are shown in Table VIII and in Figure 6.

¹¹ *Ibid.*, p. 124 ff.

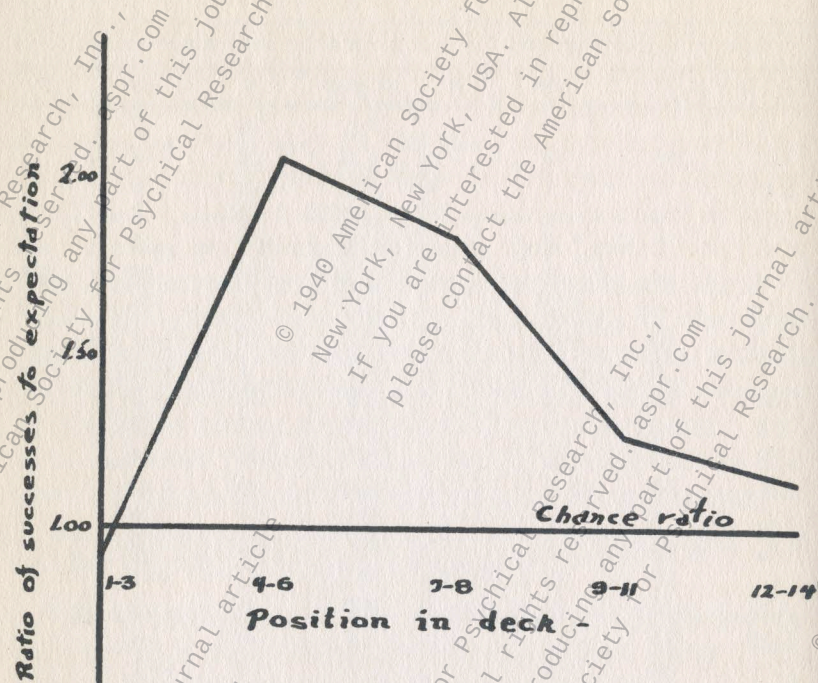


FIGURE 6. GRAPH SHOWING DISTRIBUTION OF J. G.'s ROOK SUCCESSES

TABLE VIII. Tabulation of J. G.'s correct calls at each position in the deck of 14 Rook cards.

Position in deck	Total correct calls
1	10
2	3
3	5
4	15
5	10
6	15
7	6
8	17
9	6
10	6
11	9
12	12
13	7
14	3

The table reveals a general tendency for successes to cluster toward the first third of the deck, and when the values are averages and plotted as in Figure 6 the relationship becomes apparent. This "inverted U" curve is one of the types of curves reported previously by Rhine, although most of the Duke University subjects characteristically produced a regular U curve, opposite to that shown in Figure 6.

2. *Red-White.* As shown in Table III, the red-white data up to February 20, 1940 were distinctly extra-chance, correct calls being so far below expectation that there was no question as to the reliability of the trend. This, of course, was an interesting result, but to us it had little value unless it could be repeated. Consequently we decided to collect more red-white data under similar conditions, to see whether the P value would remain significant or would increase until it became of little value. Accordingly, over 4,000 additional red-white calls were collected, so that the data were almost trebled. These additional calls also revealed the negative trend present in the original red-white data. In collecting these additional data, decks of both 25 and 50 cards were used, the trend being equally negative in each case; to expedite statistical analysis, the decks of 50 are considered as two decks of 25. The complete data are shown in Table IX.

TABLE IX. Complete red-white data as of March 31, 1940.

Number of calls	6,975
Correct calls	3,343
Chance expectation	3,487.5
Deviation from chance	-144.5
P	.000,276

It is apparent that the entire red-white data are, by this analysis, beyond the limits of ordinary chance variation—the negative deviation is statistically significant, indicating that in the present research when subjects were asked

to guess red-white cards, they scored below expectation to such an extent that chance is excluded as an explanation.

The next step in the red-white analysis was to plot the frequency distributions of obtained and expected scores. These curves are shown in Figure 7. From visual inspection it is plain that the obtained curve is not simply skewed to the left, as might be expected, but is moved bodily to the lower end of the score scale. For almost every score value below the chance expectation of 12.5 the obtained scores exceed the expected, whereas for every score value above the expected mean the expected scores exceed the obtained,

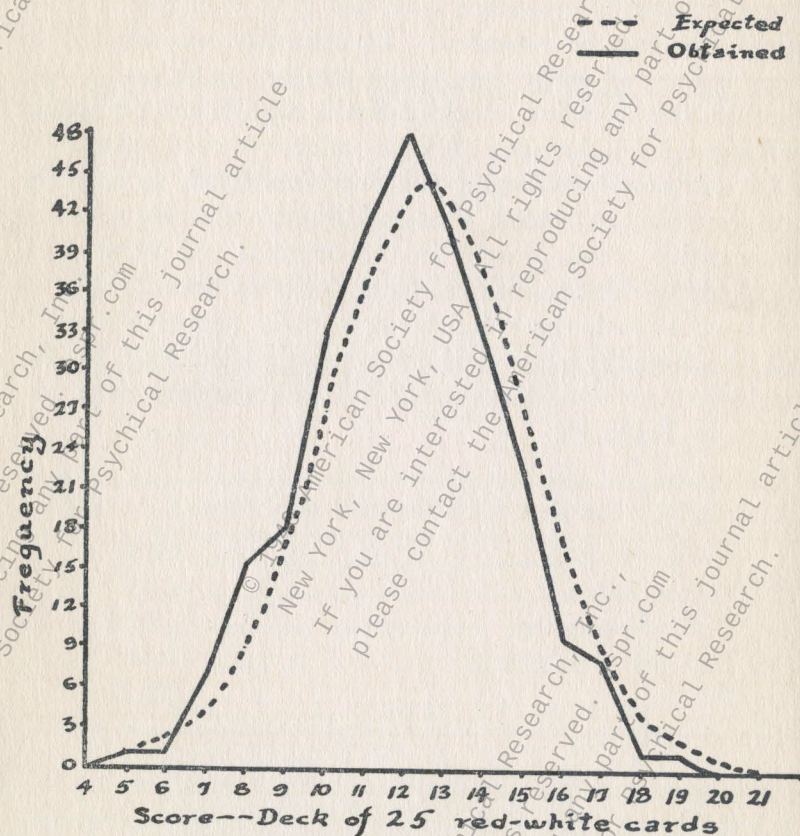


FIGURE 7. OBTAINED AND EXPECTED FREQUENCY DISTRIBUTIONS OF SCORES ON DECK OF RED-WHITE CARDS

as shown clearly in Table X. This is an interesting result, indicating a linear departure from chance expectation. The obtained negative deviation is not attributable to a large number of extremely low scores, nor to a spectacular deficiency of extremely high scores—it is rather that some constant factor is apparently at work on *all* scores, high or low, making each score just a little less than it should be, resulting in the bodily shift of the entire distribution in a negative direction. Since the departure from the expected curve is still a good fit for the obtained curve as the Chi-square and P values in Table XI indicate.

TABLE X. Table showing comparison of expected and obtained distributions of red-white scores—deck of 25.

Score	Obtained Frequency	Expected Frequency	Difference
19-25	1	2	-1
18	1	4	-3
17	8	9	-1
16	10	17	-7
15	22	27	-5
14	32	37	-5
13	41	43	-2
12	48	43	5
11	41	37	4
10	33	27	6
9	18	17	1
8	15	9	6
7	7	4	3
6	1	2	-1
5-0	1	1	0

TABLE XI. Values obtained in Chi-square analysis of red-white scores.

Chi-square	16.34
n	13
P	.22

In the case of the red-white data, as in the case of the Rook data, the total number of hits, or correct calls, was tabulated by individual subjects, with the results shown in Table XII. Here the group effect is considerably more marked than in the case of the Rook data. Twenty-four subjects took part in the red-white experiments, some calling the deck but twice, one subject calling the deck 59 times. Of the 24 subjects, 17 are below expectation, only 6 are above expectation, and 1 subject shows no deviation. As in the case of the Rook data, here also is one exceptional subject, this time A. W., whose red-white scores up to April 1, 1940 were so low that the chances are 250 to 1 against an explanation in terms of chance alone. Excluding for the moment the data from A. W., the remainder of the group scored so low that the chances are about 170 to 1 against a chance explanation. When the entire data are considered together, however, the negative deviation is so large as to be expected by chance only once in 4,000 such periods of experimentation.

A negative result of this kind, from a purely statistical standpoint, is precisely as significant as an equally positive deviation would be. One must, in a sense, know what the cards are in order to miss them enough to produce a significantly negative deviation. Such a negative result is rather striking, and demands an attempt at an explanation. The present writers have tentatively adopted a hypothesis, but further research is necessary before much can be said about it. It would appear, at least, that there may be something about the red-white situation which arouses negativism. This may be because of obscure affective factors—the symbolic values of red and of white are deeply ingrained in our culture, at least, and this general affective tone of the material used may have something to do with the type of result obtained. Beyond that, however, the authors are not at present willing to proceed.

TABLE XII. Breakdown of red-white data by individual subjects.

Subject	N	Correct calls	Deviation	P
MB	2	22	—3	.239
AB	12	133	—17	.028
BB	3	30	—7.5	.053
LD	4	46	—4	.242
KD	6	75	0	.500
RF	12	165	15	.042
SF	36	430	—15	.166
GG	3	40	2.5	.363
JG	8	92	—8	.145
GH	30	368	—7	.318
KH	6	74	—1	.468
DH	9	109	—3.5	.345
MH	9	106	—6.5	.212
EK	2	28	3	.239
HM	9	102	—10.5	.092
MM	15	197	9.5	.176
CM	9	99	—13.5	.036
JP	14	179	4	.359
AP	12	136	—14	.053
MP	4	45	—5	.184
IR	11	139	1.5	.452
ET	3	35	—2.5	.322
AX	3	31	—6.5	.082
AW	59	662	—50.5	.004

Empirical Controls

The results just reported, however interesting, would nevertheless mean little unless proper empirical controls were made. It is necessary, that is, to determine the degree to which the shuffling of the various decks of cards was adequate—to determine whether the materials used were truly *random*, or were in some systematic way deviant from random sampling. If in the red-white decks, for example, there was a consistent tendency for the red cards to be found at the top of the deck, and a subject similarly tended to guess more red cards at first and more white cards later, a spurious positive result would be obtained.

While it is true that in the subject's calls such deviations from random sampling are present, results will yet be genuine provided that the target, or deck of cards to be called, represents a random sample. To make this determination, then, either of two methods may be used: (1) Each target deck may be checked against the succeeding target deck, the correspondences being noted. A consistent extra-chance tendency in the decks will then produce extra-chance frequency distributions. (2) The frequencies of each kind of material used at each point through the deck may be tabulated. Such an analysis would reveal any tendency for the fifth card in the red-white deck to be red more than it should be by chance, or any similar tendency. Either of these two controls is probably adequate for all practical purposes; in any case, negative results with both types of control should satisfy most readers as to the adequacy of shuffling. Both such controls were made, the results of which are shown in Appendices III, IV, V, and VI.

The obtained P values indicate that the shuffling, for all experimenters, in both the case of the Rook and red-white data, was adequate.

Summary

We should like now to recapitulate the main trends of the work done in the two winters' investigation at the A.S.P.R. offices. Most of the work appears to confirm the conclusion voiced by Coover that ordinary subjects, working in an ordinary mental state, with statistically measurable material, do not average significantly above chance expectation.

Some of the results, however, do indicate a process which the statistician would not permit us to treat as chance effects (nor, in view of the procedure, as effects due to sensory cues, imperfect shuffling, etc.). First, the scores on the simple choice of red or white cards are paradoxically low, not merely low enough to arouse interest, but so extremely low that such a result would be expected to appear by

chance only once in the course of 4,000 such experiments. Second, when Rook cards are guessed 14 at a time, the scores are high to a degree to be expected by chance but once in the course of 670 such experiments. It is possible that the results are partly due to the personalities of the experimenters, or to fatigue, excitement, health, diet of the subjects, and other such multitudinous factors, but to the writers it seems more likely that the attitude of the subject toward the task is the most important single variable.

It cannot be too strongly emphasized that we have striven chiefly to develop a method, and that until our results have been independently confirmed they cannot be incorporated into the general body of science. They may have been due in part to tangible or intangible aspects of the procedure, such as the type of material used, the manner of preparing it, or the attitudes consciously or unconsciously entertained by the experimenters or by the subjects. We believe that the method can be successfully repeated, but nothing is less characteristic of science than premature declarations of success.

A few words about plans for subsequent experimentation seem in order. Our experience has been that relatively few subjects are willing to give the time to carry out rigidly controlled experiments permitting statistical evaluation over a long period of time. A given type of material has to be used a long time before one has a clear-cut result. Nevertheless, two forward steps are possible. Perhaps a few members of the Society, reading this report, will be sufficiently interested to join and work with the group. Second, there is now enough material in hand from some of these procedures to justify starting with other types of objects to be guessed, involving other colors and pictures and various materials with more direct and unambiguous feeling tone or emotional value. It is true that such material quickly loses its special appeal, but the recent study of Pratt and Woodruff¹² shows that a change of method every

¹² Pratt, J. G. and Woodruff, J. L., Size of stimulus symbols in extra-sensory perception, *J. Parapsychol.*, 1939, vol. 3, no. 2, 121-158.

few trials, provided only that the statistical treatment remains uniform, is even advantageous. It is of the utmost importance that the current work of Pratt and Woodruff just mentioned be repeated somewhere outside of the Duke University laboratory. Here high scores are apparently obtained from normal persons by a rigid method of control.

Third, the ingenious method of Whately Carington,¹³ informally reported from Cambridge, England, could well be adapted to our group performances and urgently calls for repetition. These are among the things to be done during the coming season.

¹³ Carington, W., Experiments on the parapsychical cognition of drawings, to appear in *J. Parapsychol.*, vol. 4, no. 1.

Appendices

A few words as to the most convenient method of reading the tables may be in order. The column at the left in Appendix I designates the kind of material used in a given set of experiments. The capital N in the second column indicates the number of individual tests given by a given method. There were, for example, 2794 playing-cards upon which guesses were made. The next column, "hits expected," tells the number of right guesses to be expected on the basis of chance alone. For example, the likelihood of getting the color right when one is limited to ordinary guessing is 50-50, so that half of 2794, or 1397 should be correct. The next column, "hits obtained," tells the actual number successfully guessed, which, in this case, deviates from chance expectation by only 3, as the next column shows. The next column, " $\sqrt{npq}$ " is a standard measure to tell within what range the scores may be expected to vary; n is the number of calls,

APPENDIX I. Summary of all 50-50 card data.

Material	N	Hits Exp.	Hits Obt.	Dev.	$\sqrt{npq}$	$D/\sqrt{npq}$	P*
Playing Cards..	2794	1397	1400	+3	26.43	0.11	.46
Circle-Blank ...	3275	1637.5	1652	+14.5	28.61	0.51	.31
Square-Waves..	775	387.5	381	-6.5	13.91	0.47	.32
Yellow-Blue ...	270	135	121	-14	8.2	1.71	.04
Yellow-Black ..	1085	542.5	527	-15.5	16.47	0.94	.17
Circle-Cross ...	8010	4005	4084	+79	44.75	1.77	.04
Masc.-Fem.	1600	800	807	+7	20.00	0.35	.36
Black-White ..	2275	1137.5	1124	-13.5	23.85	0.57	.28
Red-Yellow	625	312.5	300	-12.5	12.5	1.00	.16
Red-White	6975	3487.5	3343	-144.5	41.7	3.46	.0003
TOTAL.....	27684	13842	13739	-103	83.2	1.24	.10*

* The values in this column are not completely accurate, being based simply on $D/\sqrt{npq}$. Actually, as indicated on page 12, the statistical procedure for a precise evaluation is considerably more complex. These P values, however, are sufficiently good approximations to indicate whether significant results are being obtained. When a critical case is found, the more precise analysis is needed.

APPENDIX II. Results obtained checking each red-white deck against the succeeding deck.

N	1350
Hits expected	675
Hits obtained	667
Deviation	-8
P	.34

p is the probability that a given call is correct, and q is the probability that it is incorrect. This measure, the *standard deviation*, will include about 68% of all the scores which would appear in a large chance series. The procedure here is exactly as if we said, "If these are Japanese, I should expect them to be about 5 feet 5 inches" but if we know that about 68% of all adult Japanese men are between 5 feet 3 inches and 5 feet 7 inches in height, the 2 inches, measured in either direction from 5 feet 5 inches is the σ or standard deviation, the measure that we are here using. The reader, without puzzling about statistical detail, can see that the further a deviation is from the expected number, the more standard deviations or units it is from the expected number, and the more surprising it is. A value of more than 3σ above chance arises by chance alone only once in about 700 cases. If the reader glances down this next-to-the-last column he will see that with all of our materials except the red-white the data are within the category of ordinary chance expectation. The results with the circle-cross are suggestive, and with the yellow-blue almost the same, but no statistician would accept such values as conclusive, and certainly no psychological researcher should do so either.

The column which means the most, and which fortunately can be explained more easily, is the last column, the one captioned *P*. The *P* value is simply the likelihood that a random chance collection of material would give such a deviation from the expected mean as is actually obtained.

APPENDIX III. Red-white target control, showing frequencies of red cards at each position in the deck for each experimenter.

Position in deck	Experimenter		
	GM	ET	GG
1	15	3	1
2	18	4	2
3	24	6	0
4	17	6	3
5	21	6	3
6	19	6	1
7	17	9	2
8	24	4	3
9	21	5	4
10	21	4	2
11	28	8	2
12	23	7	2
13	22	5	2
14	22	7	0
15	22	5	1
16	16	6	1
17	17	3	2
18	19	6	4
19	21	9	1
20	21	8	1
21	19	5	3
22	14	4	1
23	16	5	3
24	21	6	1
25	15	6	2

The *P* value, of course, reveals nothing whatever of the cause of a phenomenon, but it permits a concise statement as to whether a result is reasonably attributable to chance or not, provided one has a large amount of material permitting uniform treatment, and provided one makes full use of the necessary statistical safeguards.

APPENDIX IV. Distribution of 5 Dot Scores.

Score	f_g	f_t	D	D^2	D^2/f_t	D	fD	fD^2
0	148	166	-18	324	1.952	0	0	0
1	106	92	+14	196	1.849	1	106	106
2	24	20	+5	25	1.136	2	48	96
3	3	2				3	9	27
TOTAL	281	280			4.937		163	229

$$\chi^2 = 4.937$$

$$n = 3$$

$$P = .08$$

$$\sigma_0 = \sqrt{\frac{229}{281} - \left(\frac{163}{281}\right)^2} = .692$$

To be within chance expectation, the values in the GM column should range between 12 and 30, in the ET column, between 1 and 9. Inspection shows that all the obtained values in these two columns lie within the expected ranges. In the GG column the amount of data is so small that no single position in the deck could reveal a significant deviation, i.e. values of 0 and 4, which are the maximum and minimum values possible of achievement, are within a 3σ range. The table indicates that the shuffling of red-white cards was adequate for each experimenter.

APPENDIX V. Rook control, GM's shuffling.

Denomination of Card	Position in Rook Deck														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
14	6	3	3	7	3	2	2	2	5	3	7	6	7	6	62
13	7	6	4	2	7	1	4	5	4	5	3	4	3	4	59
12	2	7	8	6	4	3	10	2	4	4	7	1	4	4	66
11	7	2	3	6	4	2	8	4	7	5	3	2	4	4	62
10	6	2	3	2	4	6	4	4	2	1	5	3	6	7	55
9	4	3	6	5	4	6	3	4	6	8	3	6	1	3	62
8	3	4	3	6	4	5	3	5	4	3	3	5	2	3	53
7	6	4	6	2	4	1	3	3	6	7	7	3	3	4	59
6	2	10	5	5	6	3	3	3	2	2	5	3	7	5	56
5	3	3	3	5	8	5	4	1	6	4	9	7	3	7	70
4	1	1	5	5	1	2	8	8	4	3	2	6	5	6	61
3	3	5	3	6	4	3	6	6	6	2	2	4	3	5	58
2	5	4	7	3	5	8	6	4	5	5	3	2	7	3	66
1	6	7	2	6	5	5	2	3	8	3	8	3	4	3	65
Total	61	61	61	61	61	61	61	61	61	61	61	61	61	61	854

$\sqrt{2\chi^2} - \sqrt{2N-1} = 0.38$. This value would have to exceed 2.00 materially before any extra-chance factor would be indicated. The obtained value of 0.38 indicates that GM's shuffling was adequate.

APPENDIX VI. Rook control, ET's shuffling.

		Position in Rook Deck														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Denomination of Card	14	12	9	16	9	10	9	12	12	12	19	10	12	10	12	164
	13	17	14	10	16	7	8	12	11	15	9	12	6	12	10	159
	12	12	3	9	5	6	12	9	9	13	11	15	11	4	134	
	11	11	14	9	15	11	14	12	8	10	9	11	9	14	10	157
	10	5	7	9	11	10	14	9	12	10	12	16	13	8	9	145
	9	11	12	15	7	14	14	10	18	7	17	13	15	11	14	178
	8	15	10	6	14	12	13	10	9	13	8	3	10	12	15	150
	7	13	13	10	10	10	6	17	9	8	7	6	8	10	12	137
	6	10	12	19	12	16	12	12	14	12	11	14	4	9	8	165
	5	7	12	9	12	14	10	9	15	18	9	15	12	10	12	164
	4	9	17	12	10	9	8	11	12	11	9	10	18	12	13	161
	3	13	12	12	9	13	11	9	7	9	15	9	12	15	13	159
	2	12	10	6	13	12	15	9	7	11	11	12	11	14	7	150
	1	9	9	12	11	10	8	13	11	5	7	8	9	6	15	133
Total		154	154	154	154	154	154	154	154	154	154	154	154	154	154	2156

$\sqrt{2\chi^2} - \sqrt{2N-1} = 1.77$. As in Appendix V, the obtained value indicates that the shuffling of the Rook cards was adequate.