The Journal *of*
Parapsychology

Volume 1 JUNE, 1937 Number 2

EDITORIAL COMMENT

IT IS PREMATURE yet to judge the general reaction of readers to the first number of the JOURNAL. It was to be expected that on the whole the more friendly commentators would have kindly hastened to offer their favorable words. They have done so and have done it in a way and degree that is most gratifying. The adversely critical, on the other hand, may not have had time yet, since they may be expected to be more hesitant.

As a matter of fact, it would appear that the time is ripe for an exhaustive critical examination of the present status of the entire array of evidence for extra-sensory perception, by someone (or more than one) reasonably well prepared for the task. Such a thesis, by someone thoroughly acquainted with the extensive literature of the last sixty years and familiar with the special problems involved, would be an important contribution to the field and editorially very welcome to these pages.[1]

Likewise there is need for an historical outline of the field of experimental parapsychology, in order that currently reported research may be better comprehended and the general state of the subject more fully appreciated. The preparation of a series of such historical articles is planned. Among the projected articles will be one outlining the main problems awaiting study, another describing methods in use or available for use, and a third briefly summarizing the evidential status of ESP to date.

Already the suggestion that more variety and scope be given the

[1] The critical review entitled "Extra-Sensory Perception," by Vernon W. Lemmon (*Journal of Psychology*, 1937, 4, 227-238), was brought to our attention as this manuscript went to press. The author evaluates the criticisms hitherto published and at the same time offers an interesting physical hypothesis of extra-sensory perception.

contents of the JOURNAL has begun to be heard. It was anticipated. It comes from those who feel that the case for ESP had already been proved and overproved and who are looking for further advances and wider bearings.[2] But this monotony, as it may seem to those mentioned, will not continue indefinitely. The first number necessarily, we think, centered attention almost exclusively on the question of the occurrence of ESP, in order to afford some grounds for judgment to those who have been waiting to see whether earlier reports of ESP were to be confirmed by other experimenters.

The reports in the present number throw more light upon the functioning of extra-sensory perception. The actual magnitude of these steps cannot be evaluated here. Among psychological relationships may be noted Clark and Sharp's observation of the importance of social factors in the experimental situation, especially the rôle of the experimenter. These latter observers, as well as Bond, conducted tests with subjects in groups, in the classroom or otherwise. Change in the size and number of the stimuli (card symbols) seems to have no effect, as noted by L. E. Rhine; no more does distortion of the symbols, as noted by MacFarland and George, except for an effect explainable perhaps wholly on "psychological" grounds.

The age range in which extra-sensory perception occurs has been widened greatly by studies reported in this number. L. E. Rhine discusses the youngest children thus far systematically studied as ESP subjects.

Following and overlapping with this in age range is a study of fourth and fifth grade children by Bond and one of adolescent boys by Price and Pegram. Further child studies are on record for later publication, and it becomes evident that a new and promising field of "subject material" is being opened in these experiments at the lower age levels.

What will be for many an important feature of the Bond study is the fact that the large majority of her subjects were retarded children. This at least introduces a problem in which there is considerable psychological interest, the possibility of a relation between ESP and intelligence. It does not attempt here to solve it.

The Price and Pegram report will have a special interest in that it is the first systematic research in print dealing with ESP in blind subjects. Capacity for ESP is obviously of special significance with reference to blindness. In this first report, however, the relative ability of

[2] They, as well as others, will doubtless find of interest the recent article in *Harpers* magazine, June, 1937, by Mr. J. W. N. Sullivan, entitled "What and Where Are We?"

the blind for ESP is not a primary point of attack. The case is made for the simple fact of the occurrence of ESP in blind subjects.

On the ever-present question of the adequacy of the conditions to eliminate all possibilities alternative to ESP, there is a notable piling up of safeguards in some of the studies here published. In the MacFarland and George study there was added to the STM method the DT condition of leaving the pack untouched till the end of the run. Price and Pegram added to the blind-subject tests the use of sealed, opaque envelopes enclosing the cards. Some may feel that this is ridiculous, but blindness is relative and is sometimes doubtful. The opaque envelopes were more reliable and in any case were a good supplement.

And for those who are interested in the important rôle of the mathematics and in the question of "luck" as related to the data of ESP tests, Warner's treatment giving an empirical statistical control equivalent to over 100,000 trials in ESP tests on 2,000 "subjects" will be of special interest.

It will appear, then, that these studies constitute an advance in the variety of problems dealt with, while at the same time the primary question of adequacy of conditions for establishing the occurrence of ESP is not neglected nor lost sight of. These primary considerations must continue to demand vigilant attention, for no matter how great the evidence from the past or from the work of others may be or become, no individual contribution is any stronger than the adequacy of the conditions under which it was obtained to exclude all alternative hypotheses.

The present number of THE JOURNAL OF PARAPSYCHOLOGY includes a glossary of terms that are used in a specialized or technical fashion in reports of experimentation. The meanings tabulated are tentative working definitions only. Organized research in the field of parapsychology is too young to dictate a fixed terminology.

THE ROLE OF LUCK IN *ESP* DATA

LUCIEN WARNER

ABSTRACT: Can it be that there are "runs of luck" which are not measurable by statistical method? Is it possible that investigators of extra-sensory perception have just selected subjects who were having a "period of luck"? The experimenter attempted to answer these questions by adopting a technique designed to rule out extra-sensory perception and with which he was able to study "runs of luck" with himself as the subject. Every score of eight or more in a preliminary series was immediately followed by other runs to see if it were the beginning of a "run of luck." Very high scores were followed by a series of fifty runs. The technique was comparable to that used in actual experimentation for ESP subjects. The scores came out very close to what might be expected by chance variation.

Dr. Warner is co-author of an article entitled, "Telepathy in the Psychophysical Laboratory" which appeared in the preceding number of this Journal.

A CONSIDERABLE portion of my time during the past year and a half has been devoted to the study of what has been termed extra-sensory perception. Before studying your bird you must catch it, so my first task was to locate subjects who appeared able to demonstrate clairvoyance and telepathy. My procedure in this search was, chiefly, a repetition of Rhine's basic experiments with cards. In brief, I discovered two things.

First, I found that it is a simple matter to obtain negative results, i.e., to find people who could not demonstrate the ability. Rhine already knew this, for most of his subjects gave no evidence of ability and were therefore dropped. Only a minority were retained for long investigation. Several of Rhine's critics[1] have reported investigations yielding negative results. It is evident, then, that there are lots of people who can sort cards and get chance scores only, so my first finding was not news.

In the second place I found that by persevering I could unearth a

[1] For example, Willoughby, R. R., "Prerequisites for a Clairvoyance Hypothesis," *Journal of Applied Psychology,* 1935, XIX, 543-550; and Cox, W. S., "An Experiment on Extra-Sensory Perception," *Journal of Experimental Psychology,* 1936, XIX, 429-437.

few subjects who could, and did, over long periods score somewhat better than chance. One or two scored far better. Before assuming that such cases demonstrate the operation of an unknown sensory capacity one must, naturally, eliminate all other possible explanations, however far-fetched. One which occurred to me might be called the "run of luck" hypothesis. Can it be that investigators of telepathy and clairvoyance have discarded all subjects except those who, at the time, were having a long streak of luck? A fact which lent weight to this suspicion was that sooner or later all subjects lost their ability to guess cards correctly more often than they should by chance.

Of course, conventional statistical methods are presumed to measure this factor of luck. No matter how many consecutive cards were to be guessed correctly there would be a remote possibility that this occurred through chance and the odds could be computed. A number of investigators have submitted data wherein the odds are so overwhelmingly against the chance explanation that it really must be discarded. But can it be that there are runs of luck which are not measurable by conventional statistical methods? This possibility troubled me from the beginning of my work. It was independently suggested to me by two prominent psychologists, and so I feel I must take it seriously.

Perhaps the reason some find it possible to believe in a luck factor unrecognized by orthodox statistics is that they have lost money at bridge, roulette, or the like. All gamblers are familiar with runs of luck. They are superstitious and devise various far fetched explanations as to why at times the luck is with them, and at times against them. Usually when one gambles the play is so rapid that it is impossible to compute statistical probabilities. When a succession of long shots wins, for example, one is struck by the deviation from expectation, but there is never time to calculate whether these deviations are, in a scientific sense, significant. Were the deviations significant they might prove the existence of some factor beyond chance: cheating on the part of the dealer, for example. But the gambler because of his ignorance of the statistics involved in his game may often have a hunch that his cards are running significantly untrue to form when they are really *not* deviating significantly. If he trusts the dealer, he feels called upon to postulate some other explanation and any explanation based upon statistical ignorance is likely to be pure superstition. However, since there exists the possibility of superstatistical runs of luck, I have made a small start at investigating the matter empirically.

First, we must reduce runs of luck to concrete terms. Such runs

consist of the usual data, but unusually distributed. If I were to toss a coin 1,000 times I would get approximately 500 heads and 500 tails. If it so happened that my first 500 throws were heads I would be said to have had a run of luck even though the next 500 averaged out the series by being tails. Statistically the chances of my throwing the 500 heads were $1:2^{500}$ but the run-of-luck hypothesis would increase my chance immeasurably *by postulating interaction between successive throws.* This is the essence of the run of luck hypothesis:—that having been lucky and thrown ten heads the eleventh is more likely to be a head than a tail. This sounds preposterous. Yet critics say of a subject whose ESP score varies significantly from expectation, "Well, you probably caught that subject when he was having a lucky streak. After all, you give tests to scores of subjects and you continue work only with the ones who do well." Does a subject who has guessed ten cards correctly improve his chances of guessing the eleventh? To summarize then, the run of luck hypothesis is based upon the assumption that scores made immediately following an exceptionally large deviation from expectation are likely to deviate in the same direction.

The experimental procedure which was used duplicated (though on an abbreviated scale because of the time involved in the work) the method commonly used to find promising ESP subjects and to obtain from them data deviating significantly from expectation. My procedure differed from this method notably in two respects: (1) I was my own subject, but I played the rôle of 2,000 subjects in turn: (2) the procedure was designed so as to eliminate (so far as we know) the operation of any ESP influence. It is obvious, then, that (presuming I succeeded in ruling out any ESP influence) if I were able to obtain any data deviating from expectation by anything like the extent that the obtained ESP data does it is very likely that there is a fallacy in the ESP work, or at any rate in such of it as I have done.

Before describing in detail the procedure in the present experiment I shall briefly note the method I have used in seeking for good ESP subjects. First, I selected subjects from among my friends and by advertisement in newspapers. The factors governing the selection included their good nature and amount of leisure time and, also, their interest in psychic matters. Second, I gave each some form of ESP test, using the usual ESP cards. The extent of this preliminary testing was *not* uniform but varied widely, this variation being determined principally by the subject's willingness to work. Third, I eliminated subjects who appeared unable to score significantly and continued only with

the promising subjects. This elimination was not based upon a fixed criterion. In general, however, few subjects worked very long who could not produce deviations of over one Standard Error. Fourth, I continued with these selected subjects only so long as they continued to score well. As soon as a subject's score dropped to chance, work with him was usually terminated.

To summarize then, my ESP work (and, I believe, that of most workers) consists of four steps:

1. Find people willing to be tested.
2. Give these a preliminary test.
3. Test further only the few who score high.
4. Drop work on each of the high-scorers when their scores drop to the chance level.

By stating the case thus baldly I think I make clear one reason critics are inclined to doubt the validity of the work. It looks to them as though we catch the few in the population who are having a lucky streak (and a certain minority can always be expected to be lucky, although the constitution of the minority shifts constantly). Having caught those who are at the moment lucky, we study them until their luck changes and then drop them like hot potatoes. Obviously, some critics will argue, you are not playing fair. You are overlooking the long flat stretches between people's runs of luck. Thus your ESP evidence is spurious.

In the experiment here described I have done my best to produce just such spurious data. I sat down at a table calling myself subject No. 1 and gave myself a preliminary test. If, as was usually the case, this test resulted in an insignificant deviation from expectation I said to myself, "Subject No. 1, you're no good. I'm through with you." Next time I took up a deck I called myself Subject No. 2. Let us presume that in his preliminary test Subject No. 2 meets our specific criterion by deviating adequately from expectation (more than one Standard Error). Subject No. 2 is not excused but is forced to continue work, taking the test proper. This testing is not continued until a given number of decks have been dealt but, rather, until the subject's deviation from expectation (the computations including preliminary test and test proper) falls below a certain standard (one Standard Error, or still above chance). At that point, his luck may be said to have changed and the test proper terminates. Nevertheless, if the subject has *at any time* scored three or more Standard Errors above chance the test proper is immediately followed by a supplementary test of 50 decks with the thought that if

capable of such good luck (even though followed by a break in luck that brought him below one Standard Error), the subject may run into another good streak. This procedure corresponds to the fact that in my ESP work I usually retest subjects who have at any time made phenomenal scores. Returning to the present experiment, I must note that the records for the Preliminary test, the Test proper, and the Supplementary test were kept separate, even though the tests were given in succession (and in that order) to a given "subject."

DETAILS OF THE TESTS

1. The Preliminary Test

The preliminary test was given as follows: seated at a card table I shuffled a deck of 25 ESP cards six times and cut it once. Then I laid the cards out successively from left to right in five piles, of five cards each. Thus the leftmost pile consisted of the 1st, 6th, 11th, 16th, and 21st card; the next one of the 2d, 7th, etc. I then shuffled and cut a deck which I kept on the shelf behind me and dealt off the top cards face up, opposite the five piles in order, dealing off extra cards to replace any duplicates. Each of the five piles was then checked against the corresponding card taken from the second deck and the number of hits was recorded.

If there was any process, ESP or otherwise, which permitted me to control the number of hits so obtained I am unaware of it. I would have preferred a machine to obtain the scores but, lacking one, I feel that the present method reduced my part in the procedure very nearly to that of a machine. Analysis of all the decks checked shows the scores to conform rather closely to chance expectation, thus supporting my belief that no ESP factor affected my work.

If the Key score obtained in the Preliminary test was not greater than one Standard Error (if, in other words, no more than 7 hits were recorded), work ceased. If, on the contrary 8, or more hits occurred, the Test proper was immediately given.

2. The Test Proper

The Test proper was conducted in exactly the same way except that it was continued for deck after deck (these being shuffled in advance and cut before using) until the luck turned and the deviation above expectancy fell below one Standard Error for the combined Preliminary test and Test proper. Computations necessary to determine when this point was reached were simplified through the use of prepared tables.

3. The Supplementary Test

The Supplementary test was given in exactly the same way except that it consisted of 50 additional decks regardless of the scores made. The Supplementary test was given only to "subjects" who had at any time scored as much as three times the Standard Error above chance (considering the scores cumulatively).

RESULTS

I originally set out to give one thousand "subjects" the Preliminary test which formed the basis of selection, but eventually increased the number to two thousand. The scores of these tests I have called key-scores. Their distribution is given in Table I. It follows closely expectation. Scores for 1,780 of the 2,000 key decks were 7 hits or less. Thus these fell below the criterion of luck which had been established (a positive deviation of 1.0 SE or more) and work with these 1,780 "subjects" was discontinued. The remaining 220 keyscores surpassed the criterion so work on them was continued to determine whether scoring following such lucky scores was better than expectation; to determine, in other words whether there was any tendency toward "run of luck."

TABLE I

DISTRIBUTION OF 2,000 PRELIMINARY TEST SCORES, COMPARED TO EXPECTED DISTRIBUTION.*

Hits	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Observed key scores	9	58	170	269	341	391	326	216	121	60	26	8	4	1
Expected distribution (Binomial)	8	47	142	272	373	392	327	222	125	59	24	8	2	1

*A Chi-square tests of goodness of fit of the observed data to the expected distribution indicates that 36 per cent of all series described by the binomial might vary from it by chance as much or more than this observed one.

I have grouped the data for the Test proper according to the key-score made on the Preliminary test. A summary of these data so grouped is given in Table II. Now *if* the run of luck hypothesis was

TABLE II

ANALYSIS OF DECKS FOLLOWING KEYSCORES OF 8 OR MORE FOR 2000 RUNS.

Key deck group	*Dks.*	*Hts.*	*Exp.*	*Dev.*	*SE*	*X*
8	529	2574	2645	— 71	46.00	— 1.54
9	445	2202	2225	— 23	42.19	— 0.54
10	268	1324	1340	— 16	32.74	— 0.49
11	70	339	350	— 11	16.73	— 0.66
12	114	590	570	+ 20	21.35	+ 0.94
13	7	32	35	— 3	5.29	— 0.57
Total	1433	7061	7165	— 104	75.71	— 1.38

operative we would expect to find a positive deviation for all six groups (following keyscores of 8, 9, 10, 11, 12 and 13). Yet we find small negative deviation in five and a small positive deviation in the sixth. In considering this fact, one should remember that work was continued just so long as scoring was good and was discontinued the moment it fell off so that the *total* deviation for the "subject" (i.e. deviation of cumulative score including Preliminary test) fell below a *positive* deviation of 1.0 SE. Thus the Test proper followed immediately a score so high as to suggest the beginning of a run of luck and the test was terminated as soon as such a run failed. Despite this fact, *none* of the groups show significant positive (or negative) deviation, nor does the combined score for all Tests proper. A total of 1,433 decks were dealt in the Tests proper and 7,061 hits were recorded. This is 104 hits (or 1.38 SE) below expectation. A deviation of this magnitude is not uncommon. It or a larger deviation would be expected about 16 times in 100. Thus the Tests proper offer no evidence for superstatistical runs of luck.

Yet 19 of the 220 "subjects" given the Tests proper did show, at one time or another, a positive deviation of 3.0 or more SE and each of these 19 was given the Supplementary test of 50 decks immediately following the Test proper. (Of these 13 were "subjects" who made 11 or more on the key decks, and were therefore automatically eligible; and the remaining six were "subjects" who, after a good start on the key deck, raised their deviations to 3.0 SE. None of the 220 "subjects" reached 3.0 SE in the Test scores alone.) These data were combined according to keyscore grouping and are summarized in Table III. In

TABLE III

SUMMARY OF SUPPLEMENTARY SERIES FOR THE TWO 1000's COMBINED.

Key deck group	*Dks.*	*Hts.*	*Exp.*	*Dev.*	*SE*	*X*
9	150	714	750	— 36	24.49	— 1.47
10	150	730	750	— 20	24.49	— 0.816
11	400	2025	2000	+ 25	40.00	+ 0.625
12	200	979	1000	— 21	28.29	— 0.74
13	50	281	250	+ 31	14.14	+ 2.19
Total	950	4729	4750	— 21	61.64	— .034

one series of 50 there is a fairly large positive deviation, being 31 hits, or 2.19 SE. This would be expected about 29 times in 1,000. The other deviations are all far smaller and the results combined for all the Supplementary tests show an insignificant deviation of 21 hits, or 0.34 SE. So again I failed to find demonstration of run of luck.

In Table IV will be found a summary of scores from Tests proper and Supplementary tests combined. The 2,383 decks yielded 11,790 hits, or 125 (1.28 SE) below expectation. Again no evidence of run of luck is found.

TABLE IV

DECKS FOLLOWING KEY DECKS (TABLE II) AND SUPPLEMENTARY SERIES (TABLE III) MAY JUSTLY BE COMBINED SINCE BOTH REPRESENT A RANDOM SELECTION.

Key deck group	*Dks.*	*Hts.*	*Exp.*	*Dev.*	*SE*	*X*
8	529	2574	2645	— 71	46.00	— 1.54
9	595	2916	2975	— 59	48.78	— 1.21
10	418	2054	2090	— 36	40.89	— 0.88
11	470	2364	2350	+ 14	43.35	+ 0.32
12	314	1569	1570	— 1	35.44	— 0.03
13	57	313	285	+ 28	15.09	+ 1.13
Total	2383	11790	11915	— 125	97.63	— 1.28

Results for the entire 4,383 decks (the equivalent of 109,575 trials in ESP tests) which were shuffled, dealt, and scored in the present study are combined in Table V. A total of 21,699 hits were recorded. This is 216 short of expectation. Since the Standard Error is 132.41 this represents a negative deviation of 1.63 SE.

TABLE V

COMBINATION OF DECKS FOLLOWING KEY DECKS, PLUS ALL KEY DECKS, PLUS SUPPLEMENTARY SERIES IS ALSO JUSTIFIABLE SINCE ALL THREE SERIES ARE INDEPENDENT AND RANDOM.

	Dks.	*Hts.*	*Exp.*	*Dev.*	*SE*	*X*
All decks following key decks	1433	7061	7165	— 104	75.71	— 1.38
All key decks	2000	9909	10000	— 91	89.44	— 1.02
All supplementary series	950	4729	4750	— 21	61.64	— 0.34
GRAND TOTAL	4383	21699	21915	— 216	132.41	— 1.63

CONCLUSION

I fail to find evidence for the run of luck hypothesis in these data but submit them so that others may study them, if they wish.* The chance element does not appear to favor a "subject" who has just scored high or disfavor one who has just scored low. So far as chance is concerned, every deck is independent of every other and the conventional statistical treatment is justified and adequate.

* [Ed. note: The raw numerical data, the bulk of which prohibits its inclusion here, is on file in the Duke Parapsychology Laboratory in Durham and with Dr. Warner in New York City.]

ADDENDUM

A Statistical Pitfall. It is customary to test "subjects" for ESP ability, to discontinue tests with those whose records do not deviate appreciably from chance expectation, and to continue work on the few who do score high. In defense of this practise it has been said that ESP ability is present to a demonstrable degree in only a minority, that we are justified in selecting the gifted few and studying them only. I believe this defense to be adequate, but only with this important proviso. The Preliminary tests which result in the selection of the gifted few must be carefully distinguished from the actual tests made of them thereafter or else the Preliminary tests of the rejected subjects must be also included in the treatment. One is *not* justified in combining the Preliminary test data obtained on an individual with his records made after he has been selected as a promising "subject" if his selection has depended upon high Preliminary test scores. Such selection amounts simply to starting the experiment with already favorable scores.

On the other hand if *all* the Preliminary test scores both of selected and rejected subjects are included with *all* records made with those who have been selected, there can be no fault. There is no unfair selection because the selection is that of *subjects before* they take the tests. It is the selection of *data after* obtaining them that is hazardous.

EXTRA-SENSORY PERCEPTION OF NORMAL AND DISTORTED SYMBOLS[1]

J. D. MACFARLAND and R. W. GEORGE

ABSTRACT: Thirteen student subjects were given 20,450 trials in matching tests for extra-sensory perception and averaged 7.1 hits per 25. The cards were screened and were left untouched until the entire 25 calls were made. An additional observer was present during most of the tests.

The trials were divided between two conditions, one half made with the use of standard ESP cards, the other with the ESP symbols grossly distorted. The aim was to test for a difference in results obtained from normal and distorted symbols. Scores of only one subject showed a reliable difference in favor of the normal cards. He was one of the investigators, whose "preference" was presumably reinforced by interest in the experiments.

An empirical check was made to test the applicability of the hypothesis that the results were due to chance. All the calls were cross-checked in a way to allow for "chance" but not for extra-sensory perception. The average was 4.95, compared to 7.1 for extra-sensory perception.

The research was conducted at Tarkio College, where Mr. MacFarland is a senior student in psychology and Dr. George is Professor of Psychology.

IN MOST OF THE experiments in extra-sensory perception conducted since the appearance of the Duke report in 1934, the standard ESP cards have been used. It is important to know what effect variations from these would have upon the results, since experimental variation often leads to the discovery of new relationships. Variation of the stimulus might occur in a number of ways, such as color, size, composition, physical location, and the like.

The present study concerns the effect of changing the shape of the forms used as stimuli, i.e., the comparison of results when "normal" and distorted stimuli are used.

If such distortion of stimuli should have any effect upon the results, the direction of the change might advance our general understanding of extra-sensory perception. Theoretically a difference in favor of either normal or distorted symbols presents two possibilities. Such a difference

[1] This research was undertaken as a project in a course in experimental psychology at Tarkio College, Tarkio, Missouri.

might arise if one set of symbols is actually easier to perceive; that is, if the extra-sensory stimuli are more readily organized into the proper cognitive pattern in one or the other of the situations.

The more likely possibility is that simple *preference* for one set of symbols might account for the difference; that is, that the factors making for the difference might be mainly conative.

It was our plan to begin with a modest departure from the normal ESP symbols, barely sufficient to introduce a distinct choice. This step consisted in so distorting the symbols in printing them on the cards that they would present an impression of disfigurement.

GENERAL CONDITIONS

The subjects used in the research consisted chiefly of volunteers from students who had already developed considerable interest in extra-sensory perception. Most of them were familiar in part with the results which have been obtained from other studies in the same field. They cooperated in a most satisfactory manner and showed an active interest throughout the period of the investigation.

The materials used in this work included several decks of the regular ESP cards. These are called the "normal" cards. In addition to them, several decks of "distorted" cards were used. The cards in these decks were similar to the ESP cards except that the five designs were drawn in a noticeably irregular manner. See illustration, opposite p. 95.

Special care was taken to arrange and to control the conditions under which the work was done. In the experimental situation, the subject and the experimenter sat on opposite sides of a table with an upright screen between them. The screen which was used was rectangular in shape, 18 inches high and 24 inches long. It had an opening at the bottom which was 2¼ inches high and 12 inches long. As an additional precaution, a small shield (2½" x 16") was built in front of this opening in such a way that the subject could not possibly see beyond the screen even with eyes brought to the level of the table, and yet both subject and observer were able to see key cards placed in the opening in the screen. Two sets of key cards were used, one normal and the other distorted. Each set consisted of one card of each of the five designs mentioned above. The procedure which was followed consisted of a combination of the STM (screened touch matching) and the DT (down through) techniques. In this procedure the experimenter placed the deck of cards which was to be used face-down on his own side of

PLATE I: Combination of STM and DT Techniques. The subject indicates with a pointer his call for each card down through the unbroken deck on the experimenter's side of the screen.

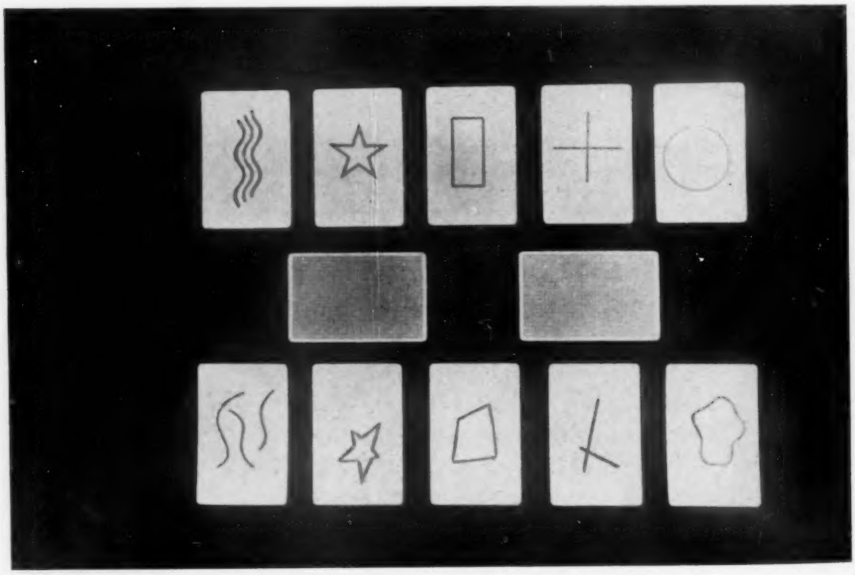


PLATE II: The normal (above) and distorted (below) symbols used by MacFarland and George.

UNIV. OF MICH.

the table and out of the subject's sight. The key cards were then placed in the opening in the screen. The subject pointed to the key card which he believed matched the first card in the deck, then to the key card which he believed matched the second card in the deck, and so on "down through" the entire deck of 25 cards. The cards in the deck *remained in position* until all 25 responses were completed. This constituted what we refer to hereafter as a "run." As the subject pointed to each key card, the experimenter recorded the response on a data sheet prepared for the purpose. The responses were checked and the score recorded at the end of each run. Two decks of normal cards and two of distorted were used throughout the experiment. Normal decks were alternated with distorted in the order of presentation. When the normal decks were used, the key cards were also normal; when the distorted decks were used, the key cards were distorted. Each deck was carefully shuffled after the run in which it was used. At the beginning of the day's work with each subject the decks were again shuffled, except in those instances referred to later in this report. Although the screen would seem to minimize the importance of the caution, old decks were always exchanged for new ones whenever the old ones became soiled.

It will be understood, of course, that under no conditions were the subjects made aware of their success or failure in matching until after the run of 25 had been completed. Each subject did, however, observe the checking of his responses at the end of the run. In all but a very few cases, at least one other observer was present. There was therefore double witnessing of each check-up, as well as of the behavior of the subject throughout the test. The experimenter for the entire series was MacFarland, except when he was acting as subject in the tests. As second observers, George, and several students[2] took part at different times. All data sheets were double-checked.

The conditions under which the experiments were conducted were kept as constant as possible. Subjects worked at approximately the same time each day; distracting influences were reduced to a minimum. The aim was to keep the motivation constant. At the same time a generally relaxed and unrestrained atmosphere was maintained. The tests were conducted informally, with little or no technical discussion and with a generally favorable and optimistic attitude on the part of the experimenter.

[2] Messrs. Fred Omer, Ralph Walker, and William McKinley.

RESULTS

The total scores for both conditions in the 20,450 trials given to the 13 subjects during this experimental series were 5,794, or 1,704 above the number most likely from chance alone. This represents an average of 7.1 per 25, exactly that of the 85,000 trials first reported by Rhine. The critical ratio (deviation divided by standard deviation) for a deviation of 1,704 is 29.77, which represents odds of 10^{197} to one against a chance explanation.

When the results of the normal and distorted symbols are compared, it appears there is not a great difference shown. As may be seen in Table I, the average for the normal figures was 7.28, while for the distorted it was 6.88, a difference in average in favor of the normal of .40. The standard deviation of this difference is .25 and the ratio only 1.6 (where 2.5 is required for a reliable difference). The odds against this being a chance occurrence are in the region of 20 to one. This would appear to indicate that the use of the normal as against that of the distorted figures did not in these tests show a significant difference in score totals.

But an examination of the individual scores shows that two subjects contributed most of the difference in favor of the normal. The fact is that the subject who contributed the most was the experimenter himself (MacFarland), presumably the one most interested in the outcome of

TABLE I

A SUMMARY OF ALL RESULTS OF THE STUDY, SHOWING SEPARATELY THE RESULTS IN THE NORMAL AND DISTORTED SYMBOLS.

Subject	Total Trials	Hits		Av. per 25		Av. normal minus distorted	Dev. of totals	CR of totals
	(Evenly divided)	Normal	Distorted	Normal	Distorted			
W. S.	6,750	1,072	1,008	7.94	7.48	.46	730	22.21
E. S.	3,000	261	229	8.7	7.63	1.07	190	12.26
M. W. (Miss)	3,250	502	512	7.72	7.87	— .15	364	15.96
W. M.	1,500	164	174	5.46	5.80	— .34	38	2.45
L. B.	750	119	108	7.93	7.2	.73	77	7.03
W. S. (Miss)	250	25	36	5.00	7.2	—2.2	11	1.74
D. H.	1,150	132	117	5.73	5.08	.65	19	1.40
J. Mc.*	700	105	39	7.5	2.78	4.72	4	.38
C. M.	750	89	86	5.93	5.73	.20	25	2.28
W. A.	350	32	36	4.57	5.14	— .57	— 2	.26
D. C.	750	78	86	5.2	5.73	— .53	14	1.27
M. M. (Miss)	2,500	377	363	7.74	7.26	.48	240	8.98
F. O.	250	23	21	4.6	4.2	— .4	— 6	.94
Totals	20,450	2,979	2,815	7.28	6.88	.40	1,704	29.77

*One of the experimenters.

the experiment. His average for the normal is 7.5 and for the distorted 2.78. The former is significantly high and the latter significantly low. In view of the fact that the subject always knew which kind of pack he was calling, and since subjects have been reported (by Rhine) as able to score high or low at will, it would appear that MacFarland's "preference" was exercised even to the point of suppressing extrasensory impressions of the distorted symbols. It is probable that experimental zeal augmented the normal preferential discrimination. Aside from this outstanding case, the remaining twelve are equally divided into those favoring and those disfavoring the normal symbols. Normality and distortion of symbols with them are at most minor matters as far as ESP under the conditions of this experiment is concerned.

DISCUSSION

The special safeguards taken to secure the results against experimental defects, together with the averages obtained, warrant some discussion. The results, in terms of number of trials, averages, and percentage of successful subjects, are comparable to those of the original Duke work. It is of greatest importance to know what conditions prevailed. The presence of a second observer and the continuous double witnessing of the check-up reduced observational error to a minimum. The use of the screen to prevent visual cues, and the policy of leaving the deck untouched (after shuffling) until the end of the run, make sensory alternatives unlikely. A cross-check control for the entire series of 20,450 trials checked all factors empirically from the point of view of chance.

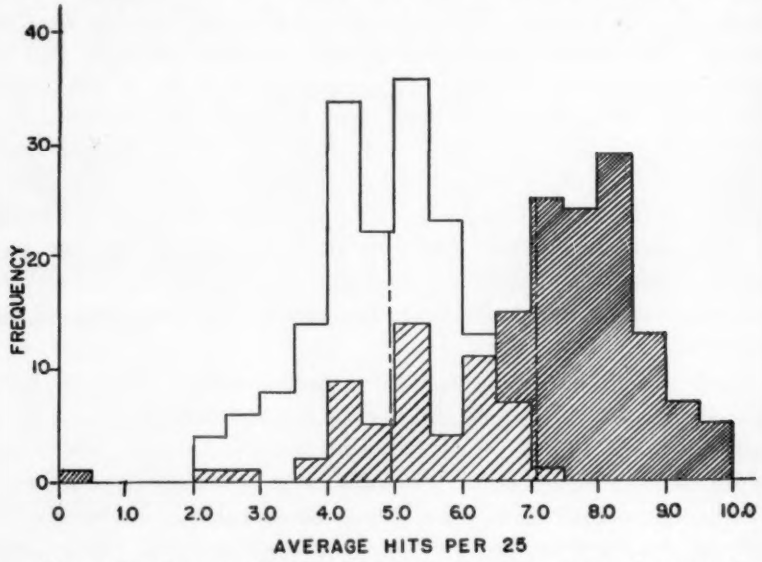
Each of the data sheets provided space for recording the results of ten runs, five for the normal and five for the distorted material. As an empirical check on the results reported above, the data sheets were divided into sets of two each. The calls made by the subjects in runs of 25 recorded on each sheet were then checked against the order of the cards for the corresponding run on the companion sheet. This control series differs from the test series only in that in the one case the calls of the subjects are unrelated to the cards with which they are compared, whereas in the other case they are matched against the cards the subjects intended to match. This empirical check gives an average score for the normal material of 4.93 as compared with an average of 7.28 for the normal ESP test series (409 runs). This is a difference of 2.35 between the two averages. The standard deviation of this difference is .28 thus giving us the very significant critical ratio of 8.4.

For the distorted material the average score for the empirical check was 4.96 as compared with an average score of 6.88 for the distorted ESP test series (409 runs). This is a difference between averages of 1.92. The standard deviation of the difference is .33 giving a critical ratio of 5.8.

When the scores for the normal and distorted material are combined, the empirical check gives an average score of 4.95 as compared with an average of 7.08 for the combined ESP test series (818 runs). (See Graph I.)[3] The difference between averages of empirical and ESP scores is 2.13. The standard deviation of this difference is .135 giving the very significant critical ratio of 15.8.

SHUFFLED AND UNSHUFFLED DECKS

During the course of the experiment there was noted some evidence of higher scores on the first two runs than on succeeding runs the same



GRAPH I.—A comparison of frequencies of averages per five runs for combined normal and distorted symbols, with results when the call series were *checked against another than the intended* card series; i.e., given a cross-check control. The shaded histogram represents the frequency of averages for the ESP calls, the unshaded that of the cross-check. The half-shaded portion is that area belonging to both distributions.

[3] This graph represents frequency distributions based upon the average scores for each five ESP runs and the average scores for each five cross-check runs, except that in the case of the four data sheets not completely filled, the averages were based upon the number of runs recorded on these sheets.

day by the same subject. It was the custom to shuffle decks after they were put away after an experimental series. Since subjects had no access to the cards between testing periods, they were frequently presented without shuffling when the new test series began. *The apparent tendency* to score high on the first two decks suggested, among other possibilities, that such a phenomenon might occur because the decks had remained unshuffled and had, in some manner, made a greater impression on the mind of the subject on that account. Although two decks of both normal and distorted cards were always available for use, one deck of each was used for the first ten runs and the other two decks were then used if the experiment for that day continued beyond the first ten runs. Thus under the conditions just described, the first two runs (one normal and one distorted) made use of unshuffled cards and the remaining eight runs made use of shuffled cards. Care was taken to be sure that the decks were not molested during the interval between the testing periods, and two checks were employed in an effort to discover if our supposition were correct. The first check consisted of a comparison of the scores on the first two runs (one on normal and one on distorted symbols) of each test series (unshuffled) with the scores on the third and fourth runs of the same test series (shuffled); the second involved a comparison of the first two runs of a number of series when the cards were not shuffled, with the first two runs of a number of other series in which the cards were shuffled immediately preceding the test period. It was only after a considerable part of our data was in that the possible significance of these comparisons was noticed, so that we do not have an equal number of runs representing each condition. All comparison involving runs of *unshuffled* decks No. 1 (normal) and No. 2 (distorted) and shuffled decks No. 3 (normal) and No. 4 (distorted) are based upon 32 runs for each deck. Those involving decks No. 1 (normal) and No. 2 (distorted) when *shuffled* are based on 17 runs each.

Table II summarizes results obtained from comparison of scores on decks which have been shuffled immediately before use with scores on decks which have remained unshuffled for some time before they were used. All of the runs which relate to this comparison, beginning with the time when the point in question arose, are included so that the selection is an entirely proper one. The results of our comparison, although based on a relatively small number of trials, are sufficiently significant to warrant giving this problem further attention. Other than this, no conclusion is ventured.

TABLE II

COMPARISON OF SCORES ON SHUFFLED AND UNSHUFFLED DECKS.

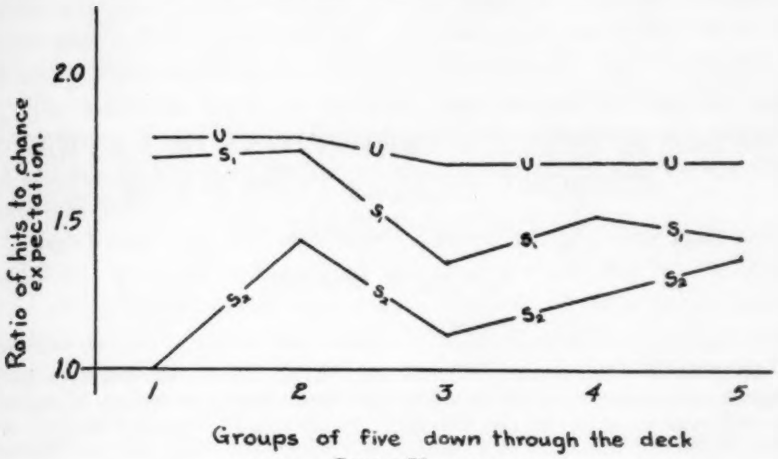
	Average Number Hits				Dif.	Crit. Ratio (D/S.D. dif.)
	Unshuffled		Shuffled			
	Trials	Hits	Trials	Hits		
Normal Symbols						
A. Run 1 (unshuffled) with run 3 (shuffled) on same day....	800	8.90	800	8.06	.84	1.87
B. Run 1 (unshuffled) with run 1 (shuffled) on different days.	800	8.90	425	6.29	2.61	5.33
Distorted Symbols						
A. Run 2 (unshuffled) with run 4 (shuffled) on same day.....	800	8.43	800	7.28	1.15	2.98
B. Run 2 (unshuffled) with run 2 (shuffled) on different days.	800	8.43	425	6.05	2.38	3.91
Normal and Distorted Combined						
A. Runs 1 and 2 (unshuffled) with runs 3 and 4 (shuffled).	1,600	8.67	1,600	7.76	.91	2.92
B. Runs 1 and 2 (unshuffled) with runs 1 and 2 (shuffled) on different days.....	1,600	8.67	850	6.17	2.50	6.02

Graph II illustrates the frequency of hits in sections of five cards each down through the deck. (The only similarity of these results of the combined STM and DT techniques to the U-shaped curves reported by Rhine for unmodified DT is the drop in the middle section of each group.) The curve designated U represents hits on 64 runs with unshuffled decks. The S1 curve represents the hits on 64 runs with these same decks after they have been shuffled and used again. The S2 curve represents the hits on 34 runs with decks shuffled just before the experiment.

SUMMARY AND CONCLUSIONS

(1) The data presented in this report, covering in all something over 20,000 trials, definitely confirm the results of other investigations which have indicated an ESP factor which enables subjects to obtain information, under strictly controlled conditions, in some extra-sensory fashion. The reliability of our data is affirmed by an empirical check in which all conditions are identical with the test situation except that the ESP factor is excluded. With the ESP factor thus excluded scores obtained are only those expected by chance. When the ESP factor is present scores are decidedly above chance.

(2) A comparison of scores obtained when normal and distorted material is used, indicates a small difference in favor of normal material



GRAPH II

largely contributed by two subjects. Even though inconclusive, this would indicate the importance of further attention to situations in which "preference" might be expected to play a more important part.

(3) It was found that when cards remain unshuffled for a considerable period of time before being used in a test series the scores of subjects will be higher than when they are shuffled immediately preceding the series. This is offered, however, only as a further problem and in no sense as an established generalization.

SOME STIMULUS VARIATIONS IN EXTRA-SENSORY PERCEPTION WITH CHILD SUBJECTS

LOUISA E. RHINE

ABSTRACT: An experiment involving 14,200 trials with seventeen child subjects was conducted to see the effect upon extra-sensory perception of varying the card symbols as to size and also as to the number of them presented at one time. In the first case large card symbols were mixed with small ones and matched without the subject's knowledge as to which was which. In the other case cards bearing five symbols were mixed with others bearing only one and were used similarly.

No significant discrimination was found in either case, although the total scores were high enough to be indicative of an ESP factor in operation. Some suggestions are given toward the technique of using child subjects.

Dr. Louisa E. Rhine is a member of the staff of the Parapsychology Laboratory at Duke University.

THE OCCURRENCE of extra-sensory perception under certain conditions has been reported for many adults, but at the time of the beginning of the research here reported no extensive work had been published on subjects of an age much below the adult level. The Creery sisters,[1] whose ESP was studied by the Society for Psychical Research in the 80's,[2] ranged in age from 13 to 21. Guthrie, also in the 80's, was puzzled by his ten-year old son's apparent ESP capacity, but the child was not formally tested. Richet reported[3] an anonymous schoolmaster who found that some of his pupils could guess letters of the alphabet at a rate slightly but consistently above chance.

Some results[4] Miss Esther Bond obtained with fourth grade school children in 1935 yielded evidence of ESP. The case of Ilga K. in Latvia, aged 13 and reported to be capable of telepathic perception, came

[1] *Proc. Soc. Psy. Res.*, Vol. I, 13-34; 35-42; 43-46; 70-98.

[2] *Proc. Soc. Psy. Res.*, Vol. II, 24-42.

[3] Richet, Charles, *Thirty Years of Psychical Research* (London: Macmillan and Co., 1922), p. 91.

[4] Bond, E., "General Extra-Sensory Perception With a Group of Fourth and Fifth Grade Retarded Children," *Journal of Parapsychology*, I, 114-122.

to the attention of the world through the report of Neureiter during the same year.[5] Further studies with children were clearly warranted by these instances, and more results were needed on the "age" question.

The need also for easily available experimental subjects for work in ESP tests, coupled with the interest in the question of the relation of age and ESP, led to the use of young children for the experiments here reported.

Initial tests with my two older children, Robbie, then aged seven, and Sally, aged five and one-half, had given encouraging results. Children of the neighborhood were readily available as subjects for experimentation and a project of testing for ESP among them was planned. It was hoped, however, to extend the goal of the study beyond the mere testing for the ability of ESP. Some inquiry into the conditions affecting ESP scoring was undertaken in order to make whatever data obtained doubly meaningful.

Considerable interest has arisen as to whether any relationship between sensory and extra-sensory perception exists, and as to the effect on the process of ESP of varying the stimuli in various ways; for example, of varying the kind of object, and the conditions of its presentation.

A fairly large range of stimuli has been reported in the literature as having been used in tests for ESP with some measure of success. In the early work reported by the Society for Psychical Research, playing cards, lotto blocks, numbers, and letters of the alphabet printed on paper were tried, as were also names of people and of objects.[6]

Some experimenters went still farther and tried the telepathic transference of various sensations such as those of taste, smell, pain, and touch.[7] Some degree of success has been reported for each of these. In all of them the possibility of a sensory analogy had held, since in every case there was an object as well as the subject who sensed it.

In the early work reported from Duke University[8] this possible sensory parallel was, for the first time, ruled out when a purely subjective stimulus was used. This was done in the so-called PT, or pure telepathy tests. In these tests there was no objective record of the symbol chosen for sending until after the call was made and recorded by the

[5] Von Neureiter, Ferdinand, *Wissen um fremdes Wissen* (Gotha: Klotz, 1935), pp. 1-56.

[6] *Proc. Soc. Psy. Res.*, Vol. I, 13-34; 35-42; 43-46; 70-98.

[7] *Proc. Soc. Psy. Res.*, Vol. III, 424-452.

[8] Rhine, J. B., *Extra-Sensory Perception* (Boston: Bruce Humphries, 1934).

percipient. The agent merely thought of the symbol and the percipient succeeded to a significant degree in apprehending his thought.

Another block of work reported at the same time consisted of a number of tests in which the percipient was separated from the stimulus by distances as great as several hundred miles without interfering with ESP. This distance work, like the pure telepathy experiments, seemed to mark a fundamental distinction between sensory perception and ESP and made the reality of a perfect analogy between them impossible. However, there might still be a chance that the physical condition of the stimulus might affect the process in some way not yet known. Size of stimulus, for example, might be important and if not, experimental data to prove it would, of course, afford surer ground for decision than that deriving from the logical consideration above.

Partly from the same line of reasoning it seemed desirable to test also the effect of varying the number of stimuli presented at one time. Would a group of stimuli be more effective than a single one? The question concerning this point will seem more understandable to those who have studied some of the earlier experiments in the extra-sensory perception of free drawings. In a number of instances only a fragment of the stimulus drawing was reproduced by the percipient. In some cases an associated thought from the agent's mind was caught instead of the intended one. In other words, it appeared there might be difficulty in *directing or aiming* ESP and hitting the target. A card with many symbols might possibly be easier to hit.

In many instances, too, in recent work at Duke and elsewhere, in using ESP cards, observers have remarked the unusual frequency of the transposition of symbols from their correct sequence. With some subjects this may involve getting two symbols in reverse order, for instance, and continuing to do this far beyond the number of times which might be expected to occur by chance. Others will get symbols in correct sequence, but one place ahead of or behind the order of the target pack.

The bearing of all such types of vaguely defined impressions from the experimental data is to suggest that perhaps in ESP the "aim" is bad; that, on the analogy of visual perception, the focus is poor. Perhaps, if this should be the explanation, then offering a large number of symbols of the same size at any given time would give a greater advantage to the percipient. It seemed a point worthy of being tested.

In making comparisons of two or more conditions in tests in ESP it is difficult, though of course necessary, to control other factors that

may be important. Preferences of the subject and suggestion from the experimenter are among these.

It has become increasingly apparent to many ESP workers that suggestion can play an immense part in the outcome of tests of ESP. Unless guarded against with extreme care it can do much to confuse the issue and obscure the true situation. If a subject be prejudiced beforehand against one type of stimulus, it is entirely possible for him to fail in ESP tests with that particular stimulus; another subject, not prejudiced, might have no difficulty with the same stimulus.

Adult subjects are very prone to this type of distortion by suggestion. They quickly get apriori notions about symbols. Of the ESP symbols, some prefer stars, some do not like waves, etc. Many say at once upon seeing the small symbols used in the tests to be reported below, "I know I can't get those little ones."

It seemed likely that children, being more naïve and less analytical than adults, might be particularly suitable as subjects for an investigation of the effect of variations in the condition of the stimulus. This proved to be the case. The children accepted any type of card given them almost without question. Since they were not in any case permitted to see the faces of the cards when working with them, they seemed to assume that it would not matter to them what kind of symbol they bore. It seems likely, therefore, that results obtained from them are as free from bias as could be hoped for from any type of subject.

I. METHODS AND PROCEDURE

The children who served as subjects in these experiments were not selected, beyond the fact that they consisted of all those of suitable age in the immediate neighborhood. They were asked to come in to try a "new kind of card game." Seventeen different children ranging in age from three to fifteen years took part to some degree or other, the amount of work from each depending upon his interest and availability on the particular days I was able to "play."

The work on size of symbols is divided into two main groups: Series A done the first year without a screen to conceal the cards from view, and Series B done a year later with a screen. Series A represents 246 runs of 25 cards each, covering various experimental periods from October 1935 to February 1936. Series B was done in two experimental periods, one April 16, 1937, the other April 23, 1937. In the Series B group, 190 runs of 25 cards each were made.

In Series A the technique used was regular OM (open matching) in which two children, working usually in competition, sat opposite each other at one end of a square cornered table. I sat next to and between them at the end of the table. Five key ESP cards, one of each symbol, lay exposed between and in front of the children. Their task was to distribute a shuffled pack of cards face down into five piles corresponding to the key symbols. I did all the shuffling and all the recording, and watched the procedure throughout. When cards showed signs of soil, new ones were used. None of the children attempted to peep. I purposely gave them the impression that I was watching their every move so that no temptation to do so would arise. None of them gave any evidence whatever of paying particular attention to the backs of the cards. They all seemed as a matter of course to fix their attention mainly on the key cards before them rather than on the card in hand. During most of the tests I gave them perfect freedom to talk and laugh as they pleased while working, and they usually availed themselves of this freedom.

By the spring of 1937 a convenient type of screen had been devised for the Parapsychology Laboratory and Series B was run with the screen being used as an added precaution against possible sensory cues. The children sat or stood as their height demanded, in positions at the table as before, but each had the screen before him which hid from his sight his hands and the pack of cards he held. The key cards were still in sight before him between the screens. This technique was the regular SOM, or screened open matching.

The screens used in this work were made of artificial board, and were designed to allow the subject to hold a pack of cards concealed from his view while looking at the key cards. The shape was that of a small inclined table, at one side rising to the level of the chin of the subject seated at the table on which the screen rested, at the other falling away to the level of the table. The subject at work was required to be so close to the screen that vision of the cards was effectively cut off by it. The observer, seated at the side, was ideally located to detect at once any infraction of the rules. See photograph opposite p. 107.

The block of tests using the multiple symbols was mostly made in January, 1936, and mostly in one experimental period. The OM technique was used as in Series A above.

The cards used in Series A were made up especially for this experiment. They were made in decks of 25 from smooth white cardboard 4" x 5½", and bore the regular ESP symbols, five of each. Each set of

PLATE III: The Screened Open Matching Technique allowing competition. Key cards are in the center boxes, and the children place the cards to match them in the corresponding sections under the screens.

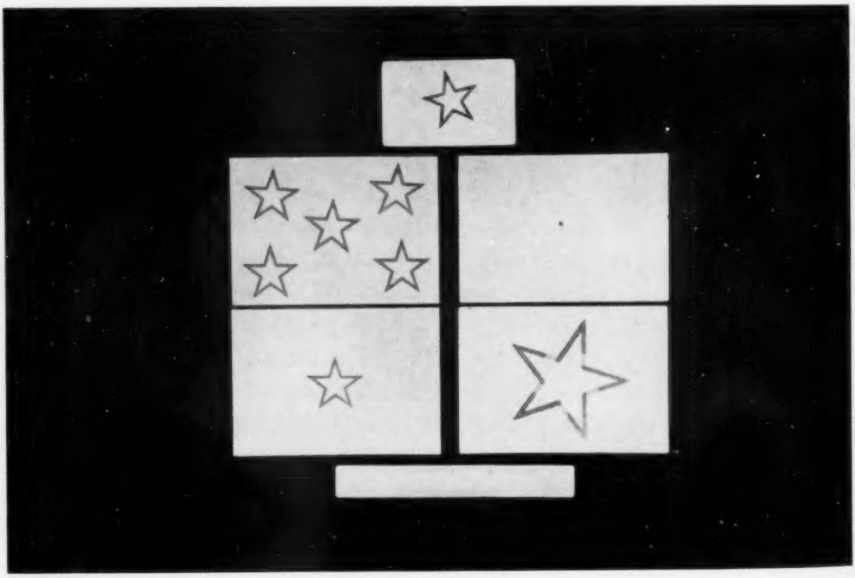


PLATE IV: Various cards used in the L. E. Rhine study of extra-sensory perception in children. Top: ESP card; above left: Multiple symbol; above right: Small symbol; below left: Normal size symbol; below right: Large symbol; (Below: 6 inch rule.)

UNIV. OF MICH.

two decks, however, consisted of one deck of large symbols drawn in India ink as large as the card would conveniently hold, and another bearing the same symbols drawn quite small. The contrast in size may be gauged from the diameter of the circles, which were 3½" and ⅛" respectively. In area the proportions were 784 to one.

These two decks were divided into two packs, one containing 12 large symbols and 13 small ones, the other just the reverse. The children used them in packs of 25 for convenience. (See Plate IV.)

The results were checked at the end of each run and the packs exchanged between the two subjects so that the symbols played would be equally divided between the large and small ones. In a few cases children stopped working with an uneven number of runs, and in these cases, of course, there is a slight inequality in number of cards of each size played. This, however, is slight enough to be negligible.

The cards used in Series B were made up from the regular ESP stock. The symbols on the large pack, instead of being extra size as in Series A, were normal in size, circle 1⅝" in diameter; the symbols of the other deck, drawn with a fine pointed hard pencil, were very small, the circle being under 1/32" in diameter. This gives areas in the ratio of 2,704 to one, approximately. These small symbols were only distinguishable in good light when held quite close to the eye. I found from constant checking them, however, that I learned to distinguish them much more easily and quickly than at first. These two decks were mixed and run in sets of 50 each, each set being recorded as two runs. Two such sets were in constant use and always exchanged between players after being checked and shuffled at the end of each run.

In the multiple series, large white cards like those of Series A were used. The multiple deck bore five normal sized symbols which quite completely filled the card. The single ones bore only one normal sized symbol in the center.

II. RESULTS

There were thirteen children participating in each series of the large vs. small symbol tests. Of these two groups of subjects, nine were the same for both. This made a total of seventeen children in all. In Series A two children obtained scores above an X value of 2.5, which is the standard set for significance. In Series B no individual child achieved scores which were independently significant, but in the total scores given by both series, another child's scores gave an X value of 2.46, which is only slightly below the standard. (See Table I.)

In the A series, however, the total scores for all thirteen subjects reached a deviation sufficient to give an X value of 4, and in the B series a value of 3; the total value of both is $X = 5.1$. The odds against a chance theory represented by this figure are 5,000,000 to one.

The difference between the number of large and small symbols hit in Series A was 52 in favor of the large, in 246 runs of 25 cards each. In Series B it was 8 in favor of the small, in 194 runs of 25 cards each. Since both these numbers yielded by so many runs are quite insignificant and easily within the range of chance, it was thought that possibly any preference might be obscured by the comparatively large number of chance or sub-chance runs that the total results include. Accordingly, the results were separated by runs into two groups. All runs that yielded a score of five or under were put in a "chance or sub-chance" group, all those of six or higher into an "over chance" group. The point was that any preference, especially since it was apparently slight, might be more easily detectable in the high runs, where a possible ESP factor would more certainly be operative than in the lower score runs.

No preference, however, was detectable even by this analysis. In Series A the slight total advantage that the large symbols had is well divided between these two groups of runs, 35 falling in the "over chance" runs and 17 in the "chance" runs. In the B series, as will be seen in Table I, the total of 8 in favor of the small symbol is divided between 20 in favor of the small in the "over chance" group and 12 in favor of the large in the "chance" group. That makes a total in both A and B series of 15 in favor of the large in the "over chance" group and 29 in favor of the large in the "chance" group. These slight variations are within the range of chance variation. Just as the total results are indicative of no preference, so too do the individual scores show only chance variations between the two sizes of symbols, with no preference for either.

In the experiment on multiple *vs.* single stimuli, only five subjects took part, but each of them obtained significant scores. The X value for the entire group of 128 runs is 6.5. (See Table II.) In this case, as in that of the experiment on size, no preference was indicated by the total number of multiple cards hit as against the total number of singles. The slight majority lay with the multiples, but it was entirely within a range reasonably to be expected from chance factors. The actual figure, as shown in Table II, is 20 for the excess of multiples. This is about the same proportion as the difference obtained with the large and small symbols.

Accordingly, this data was also divided into "over chance" and

"chance and under" runs, as in the previous experiment, and the results totalled. Here again no "over chance" preference was found. The total difference of 20 in favor of the multiples was divided into an "over chance" surplus of 31 in favor of the multiples as against a surplus of 11 in favor of the singles in the chance runs. Again no significant preference is discernible.

In this case, too, most of the subjects show no marked individual preference for either type of stimulus. There is one exception however. Tommy registered a score in the "over chance" column of 26 in favor of the multiple symbols in 23 runs. His score in the chance runs is three in favor of the multiple symbols, which divides his total of 29 only slightly. The X value of this difference being due to chance is 2.1. Certainly further work with this subject would be desirable. The figures for this series are given in Table II.

There were in the combined series on size and that on multiplicity of symbols a total of 568 runs of 25, or 14,200 trials. The average per run was 5.6 and the total deviation was 362 above mean chance expectation. That gives an X value or critical ratio of 7.6 for the entire work covered in this report. The odds against this being a case of chance only are 10^{13} to one.

TABLE I

LARGE *vs.* SMALL SYMBOLS.

SUBJECT	SERIES A					SERIES B				BOTH SERIES COMBINED			
	Age at start	Number of runs	Hits large	Hits small	Diff. in terms of large	Number of runs	Hits large	Hits small	Diff. in terms of large	Total hits	Dev.	Av. per run	X Value
1. Sally	5½	26	81	80	1	16	50	43	7	254	44	6.69	3.4
2. Robbie	7	21	68	72	— 4	.....	.....	.....	.....	140	35	6.7	3.8
3. Nancy	8	19	50	50	0	16	44	53	9	197	22	5.6	1.8
4. Ruth	6	26	73	65	8	18	46	40	6	.....	.....	.....	.....
5. Lela May	7	11	37	20	17	26	71	65	6	193	8	5.2	0.6
6. Tommy	9	48	128	117	11	26	62	70	— 8	377	7	5.1	.0
7. Howard	4	7	21	24	— 3	14	40	41	— 1	126	21	6.0	2.2
8. Jack	7	44	128	115	13	32	81	99	— 8	423	43	5.9	2.46
9. Roderick	14	22	60	53	7	.....	.....	.....	.....	113	2	5.1	.0
10. Bruce	15	10	28	28	0	.....	.....	.....	.....	56	6	5.6	0.9
11. Pat	14	6	12	13	— 1	.....	.....	.....	.....	25	— 5	4.2	.0
12. Betsy	3	3	12	8	4	4	12	15	— 3	47	12	6.7	2.2
13. Betsy Ellen	10	3	6	7	— 1	12	41	32	9	86	11	5.7	1.4
14. Patsy	6	.....	.....	.....	.....	6	17	10	7	27	— 3	4.5	.0
15. Mariella	6	.....	.....	.....	.....	12	32	40	— 8	72	12	6.0	1.8
16. Jimmy	8	.....	.....	.....	.....	8	21	18	3	39	— 1	4.9	.0
17. Yarborough	10	.....	.....	.....	.....	4	8	7	— 1	15	— 5	3.8	.0
Totals	.....	246	704	652	52	194	525	533	— 8	2414	214	5.5	5.1

TABLE II

MULTIPLE *vs.* SINGLE STIMULI.

Subjects	Runs	Hits Multiple	Hits Single	Multiple Above Singles	Dev.	X Value
Sally	47	132	139	— 7	36	2.7
Robbie	32	87	102	—15	29	2.49
Ruth	12	42	38	4	20	2.8
Jack	14	51	42	9	23	3.0
Tommy	23	92	63	29	40	4.1
Total	128	404	384	20	148	6.5

TABLE III

TOTALS FROM TABLE I AND TABLE II.

	Runs	Large or Multiple	Small Single	Diff.	Dev.	X Value	Av. per Run
I. Large *vs.* Small							
Series A	246	704	652	52	126	4.0	5.5
Series B	194	525	533	— 8	88	3.	5.4
II. Multiple *vs.* Single	128	404	384	20	148	6.5	6.1
Totals	568	....	....	..	362	7.6	5.6

III. DISCUSSION

Since each of the three series into which the experimental data are divided is independently significant mathematically and since in Series B the use of a screen in addition to other precautions against the operation of sensory cues did not prevent significant scoring, there would appear to be in these tests something occurring besides "chance factors" and beyond sensory perception. By definition, then, the phenomenon must have been that of extra-sensory perception.

Out of 17 different children who took part in the tests, five obtained independently significant scores. Since the children were not selected but included any who were available, the data indicates not only that some children are capable of ESP, but that a somewhat larger proportion of them show this capacity than has been reported for adults. In this case nearly one third of those tried demonstrated ESP.

It is difficult to believe that all of the remaining twelve children who did not score significantly would with further testing be found incapable of passing the mark. Of these twelve there are six who have average rates within the range of those who made significant deviations, but they did not have enough trials at this rate for significant deviation. The

significant scorers averaged 46 runs on the large and small comparison while the six mentioned averaged only 16 runs. Besides, the evident importance for success of psychological setting and "atmosphere" leads one to suspect that a failure may be relative, and may simply be a case of difficulty in finding the most favorable set of conditions. This point, however, is intended as a suggestion and not as an argument.

Neither the large and small stimulus variation nor the multiple *vs.* single comparison yielded results reliably favoring either side of the comparison. In both cases the small preferences recorded are well within the mathematical range of chance fluctuation and are inconclusive, unless the case of Tommy indicates a trend which further experimentation would uncover.

As far as this experiment goes, then, it appears that size of symbol within the range tried does not affect ESP, nor does a multiple stimulus offer any advantage over single. In as far as any sensory analogy between sensory and extra-sensory perception may have been supposed to exist, this data shows it to have been quite groundless. Small symbols are as easy to get by ESP as large ones, single ones as easy to hit as multiple ones. The variations tried out in these tests line up with the phenomena of pure telepathy and with distance effects in ESP in further distinguishing ESP from the sensory type of perception. Distance and size and kindred conditions of the stimulus within the ranges studied are not for ESP what they are for sensory perception. Whatever may be found to limit ESP, it will hardly be any of these things, unless beyond these ranges. Any relationship kindred to that of the Weber-Fechner law in sensory perception must be sought elsewhere, for thus far there is no sign of it in variations of the type herein studied.

The fact that the subjects were children, and that the technique of handling them had to be worked out as the work progressed, accounts for some peculiarities of the results. For instance, the scores were much higher on the multiple and single cards than on the large and small ones. But that fact may not mean, as it may superficially seem, that the multiple cards were intrinsically easier. Although this is possible, the fact seemed to be, instead, as will be explained later, that better psychological conditions prevailed when that work was done than existed when many of the large and small tests were made.

There had not yet been developed a technique suitable for handling a group of children in work of this kind. At first it was taken for granted that certain of the conditions found favorable for adult subjects would be good for children as well. Thus it seemed logical to suppose

that they would work best in quiet, if not solitude, in order to favor a certain amount of concentration. It was, however, recognized that their interest would be more evanescent and less dependable than that of adults.

Accordingly, the first tests, and they were made on the large and small cards, were conducted by calling one child into the house while the rest played outside. He was seated alone with the experimenter in a quiet room and the task explained to him; he was encouraged to complete it in silence. But it was soon evident that such a method was too formal for children. It worked better to use two of them at once and let them compete with each other. It was evident too that the highest scores were not obtained by concentration; as one child expressed it when surprised by a particularly (for him) high score, "I just hurried, and played while I listened to you-all talk." And so very soon the method was developed of working two children in competition while the rest of the crowd came and went and talked and played at will in the room round about. I no longer tried to keep the children quiet or to encourage concentration, but endeavored in every way to maintain the atmosphere of ease and lack of self-consciousness appropriate to a game, except that I did not relax my watch over them as they handled the cards, or my care in doing all shuffling and score keeping. Under this greater freedom of attitude, higher scoring occurred. From the first, a small money reward was found effective.

For a time tests were made any time two or three children were available. After the novelty was worn off somewhat, however, the interest in earning pennies seemed to become the main motive, rather than spontaneous interest in the game for its own sake, and then scores dropped off. Later at a birthday party the best results up to that time were obtained when the children were keyed up to a greater than usual degree of excitement and interest. Some who had before done little work above chance went considerably higher. The major part of the work on multiples was done at the birthday party. I allowed the group to use the new multiple cards for the first time. Robbie and Sally whose scores on the large and small symbols had sometime before reached the point of significance had already been working with these cards and the rest of the children were all the more eager to try them. Thus it would seem that the novelty of the cards and the excitement of the party probably had more to do with the rate at which they scored than did any intrinsic ease of perceiving the particular type of symbol.

One other point of technique was developed to help maintain interest. At first each child had been permitted to keep on playing until

he wished to stop, regardless of the kind of scores he made. In this way some of the large chance scores like those of Tommy, Ruth, etc., in the Series A work, were piled up. But with the introduction of the "party" method the rule was introduced that the subjects take definite turns playing, always two at a time and each one holding his place only as long as his scores were better than 5 per 25. One score of 5 or less eliminated a child until his turn came again. This helped to increase the interest considerably, and of course automatically allocated the available time more largely to those children who were at the time capable of scoring.

The age of the children did not matter, apparently. Betsy, the youngest, was not quite three years old when her first tests were made. Howard was only four. Both made very few runs as low as 5. *Their interest was short-lived, however,* and so only a few runs were obtained from them. The oldest—Roderick, Pat, and Bruce, only worked a time or two. Apparently so much association with smaller children was beneath their dignity. None of them obtained above chance scores, although Roderick began high and gradually fell off. Possibly a party of boys his age would provide a setting in which he too could score.

It is apparent, then, that in handling children the easy, spontaneous type of interest like that called forth in their play must somehow be maintained if they are to succeed in demonstrating ESP. Apparently this interest is more determinative than age. Concentration and quiet are in general not necessary, and desirable only if the children fall into them naturally. Any imposed restraint seems to reflect itself in lowered scoring. And because of the short-lived wave of interest often occurring in children, a rule for stopping after low scores are obtained is advisable.

CONCLUSIONS

The occurrence of ESP is again demonstrated, in this case in 14,200 trials with 17 children aged from 3 to 15. An average of 5.6 hits per 25 were given, and a critical ratio of 7.6.

Within the scope of the experimentation herein reported size variations ranging to proportions of 2,704 to 1 in stimuli did not result in a significant preference in scoring.

Variations in number of stimuli presented 5 to 1 at a given time did not result in a significant preference.

Children of a group of 17 made scores in ESP tests giving significant deviations under conditions believed to exclude sensory cues.

In general the children's highest scoring occurred when the conditions approximated most nearly those which obtained in their spontaneous play.

GENERAL EXTRA-SENSORY PERCEPTION WITH A GROUP OF FOURTH AND FIFTH GRADE RETARDED CHILDREN

ESTHER MAY BOND

ABSTRACT: A class of 22 retarded children from the fourth and fifth grammar school grades guessed numbers upon which the teacher concentrated. The number of correct responses from the group was significantly above the chance expectation. Four of the 22 children made individually significant scores.

It was noted that extra-chance scoring did not occur on days following an absence from school, nor when the subjects used rational schemes to guide their choices. No reliable correlation between scoring ability and intelligence rating was found.

Miss Bond is a student in psychology in the Graduate School of Duke University. She conducted the experiments while teaching in an elementary school in Florida.

THE PURPOSE of this experiment was to determine whether or not extra-sensory perception with teacher as agent might occur in the classroom situation.

The subjects were 22 retarded children from the fourth and fifth grades of a grammar school whose total enrollment was 450. The age range was 9 to 14 years and the educational achievement range was 3.5 to 6.5. According to the Metropolitan Achievement Test, Intermediate A, all except two of the subjects were retarded. The teacher was the agent throughout. She had had no appreciable practice, training, nor acquaintance with extra-sensory perception at the time of the experiment. She had been the teacher of this group about four months.

The material used as stimuli was the integers from 1 to 10, inclusive. Each stimulus was chosen by shuffling a deck of ordinary playing cards from which the face cards had been removed. This was done the night before for convenience and for diplomatic reasons, since playing cards in the schoolroom might lead to misunderstanding. The selections were listed and carefully concealed until after the experimental period. A run of 10 trials was made each day.

The experiment was conducted in the morning after a 15-minute arithmetic drill. This drill was according to the usual schedule; but from the point of view of the experiment it allowed a possible weakness. Since the stimuli were selected the night before each day's experimenting, it was possible for the teacher to remember what they were and to stress in the drill those numbers coming up most frequently in the ESP tests and thus influence scoring. The results, however, do not show that this was done. Had it been, the most frequently appearing numbers should have been most successful in the scoring. On twelve days, totalling 2,160 responses, the same number was the stimulus three or four times. But on these 2,160 trials only 217 responses were the same as these most frequently appearing stimuli; 216 is the number expected by chance. In the total responses, then, the numbers that *might* have been unconsciously stressed in the drill gave only the 1:10 proportion which should be obtained by chance.

The subjects were told that the purpose of the experiment was to find out whether or not guessing numbers was a fair way to choose pupils for the much desired special duties such as passing out materials or running errands. The subjects seemed to enjoy the experiment and showed great interest in their scores. Before concentrating on each successive number, the agent would say, "Ready for the—one," using ordinal numbers to indicate the trial. The concentration period was about four or five seconds for each trial; but it was lengthened if any subject had not made his response at the end of that time. Responses were usually made in a second or two. On May 23 through 31 the agent stood at the front of the room, arms at her sides, lips closed, eyes shut. From June 3 to June 13, she stood at the back of the room and did not keep her eyes shut all the time. The subjects were asked for introspective reports at the end of the experiment.

RESULTS BEARING ON EXTRA-SENSORY PERCEPTION

Since there were 10 possible responses, chance alone would make one response out of 10 correct. This is equal to an average daily score of 1.00. The average obtained was 1.29. In the 2,870 trials made for the period, the total number of correct responses was 371, 84 more than the mean chance score. The deviation divided by the standard deviation is 5.2, which gives a probability of explanation in terms of chance of 0.000,000,1. There were more individual daily scores between 2 and 6 than chance would give. The number of times each score would appear in a chance distribution and the number of times it appears in the exper-

iment are given in the accompanying table. Whereas chance gives 77.5 scores of 2 and over, 106 such scores were obtained. A chi-square evaluation of these results gives a probability of explanation in terms of chance of less than 0.000,001.

TABLE I

COMPARISON OF FREQUENCY OF HITS EXPECTED FROM CHANCE WITH OBTAINED FREQUENCY.

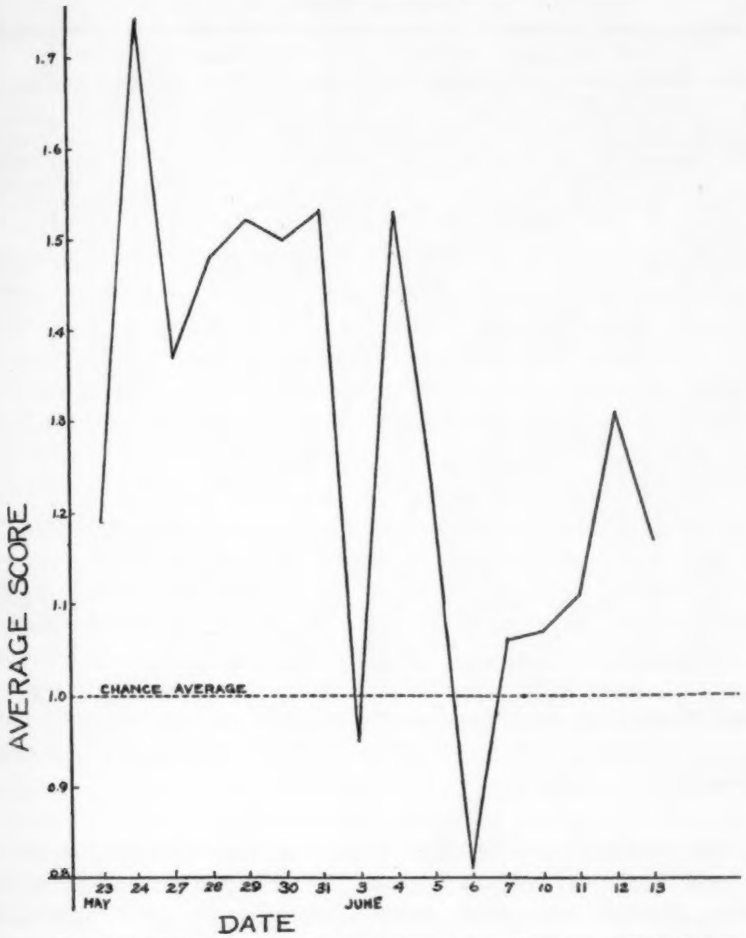
Score	F Given by Chance	Obtained F
0	100.1	80
1	111.2	101
2	55.59	70
3	16.47	21
4	3.20	9
5	0.43	5
6	0.04	1

Of the 21 individual average daily scores that were different from 1.00, chance would give 10.5 greater than 1.00. There were actually 14 such averages. Four of these were greater than 1.00 by more than 2.5 times the standard deviation, and two others were from 2.3 to 2.5 times the standard deviation. It may be concluded, then, that four of the 22 subjects perceived extra-sensorially and that two others may have done so; that is, about 1/6 of the subjects demonstrated extra-sensory perception in this situation.

The total number of hits by all subjects for each of the 10 daily trials except the sixth were greater than chance. The first trial was highest, with a total of over twice chance expectation; and all the other totals were much lower than this maximum.

Daily group averages are shown in Graph I. The average dropped to less than chance when the change in procedure of having the agent stand in back of the percipients was introduced; but there was a partial recovery.

In Table II are shown the individual daily scores of all the subjects. Of special interest are the abrupt changes noticeable when the experimenter changed from the front to the rear of the room during the experimental period, as she did on June 3. In the first series (seven days) in 1,320 trials, 194 hits were made, averaging 1.47 per day per subject, and giving a critical ratio of 5.7, better than that for the total results. There was a marked drop, as already mentioned, on June 3, following the change of position. During this second series (nine days) the average dropped to 1.14, and the 1,550 trials yielded a total score of



GRAPH I.—The average score of the total group on successive days.

177 and a critical ratio of 1.9. Four subjects (with a borderline fifth) obtained significant deviations during the first seven days; two (with a borderline third) did so during the second nine-day period.

It is necessary to consider whether secondary cues were responsible for the results. Of the known means of communicating number cues conceivable under the conditions, unconscious whispering on the part of the experimenter would be the most likely to occur. If such was the case we should expect, first, that the subjects nearer the agent would have the higher scores. However, good and bad subjects seemed to be scattered randomly about the room. Second, if the subjects used sen-

TABLE II

INDIVIDUAL SCORES.

	MAY							JUNE									Average**
	23	24	27	28	29	30	31	3	4	5	6	7	10	11	12	13	
CH	0	2	2	1	1	3	0	1	1	0	2	0	1	2	2	0	1.13
HH	0	1	2	1	0	0	1	1	....	....	....	....	....	....	....	....	0.75
EH	2	1	2	0	1	1	1	0	3	1	0	1	1	2	0	0	1.00
AH	2	0	2	4	5	6	3	4	0	1	0	0	1	0	2	0	*1.88*
EL	2	0	....	0	1	0	2	1	2	1	0	2	1	0	0	1	0.87
CP	2	2	0	0	4	1	2	0	3	3	3	1	1	0	0	1	1.44
JT	0	0	1	0	1	1	0	0	3	1	0	2	1	0	2	0	0.75
BW	1	2	1	4	2	....	2	2	1	1	1	0	1	2	0	2	1.47
BB	1	2	1	2	1	1	4	0	1	2	0	0	1	0	1	1	1.13
ED	....	1	1	* 2	0	....	0	....	1	....	1	0	....	....	....	....	0.80
JD	1	4	0	* 1	2	0	....	2	3	4	1	5	....	3	0	3	*2.32*
CD	0	2	2	1	1	1	0	1	1	1	1	1	2	1	0	2	0.95
AF	2	3	0	2	0	1	5	....	0	....	....	....	....	....	....	....	1.63
EP	2	1	1	2	1	....	....	0	0	0	0	2	0	1	2	0	0.86
FP	1	2	1	....	2	3	1	0	....	1	0	1	0	1	1	1	1.07
BR	2	3	2	5	2	....	....	2	3	2	....	....	1	2	2	3	*2.42*
JR	0	2	3	0	2	....	....	1	5	1	....	....	2	1	1	2	1.67
AS	1	3	1	0	0	0	1	....	....	....	....	....	....	....	....	....	0.86
TS	1	2	2	2	1	4	....	0	0	0	....	0	....	0	....	1	1.08
BD	2	....	....	2	4	....	....	2	1	2	0	1	....	3	3	2	*2.00*
EM	1	....	0	1	....	1	1	0	0	0	2	1	0	1	2	2	0.86
DS	2	....	2	1	1	1	....	1	1	1	2	....	3	1	3	2	1.62
Total hits	25	33	26	31	32	24	23	18	29	22	13	17	16	20	21	21	
	132 runs, 194 hits, av. = 1.47 dev. = + 62 ± 10.9; CR = 5.7							155 runs, 177 hits, av. = 1.14 dev. = + 22 ± 11.8; CR = 1.9									

Total: 287 runs, 371 hits, av. = 1.29; dev. = + 84 ± 16.1; CR = 5.2.

*Five trials only.
**Those italicised are significant; i. e., $X > 2.5$ where $X = d/\sigma$.

sory cues which were as far from perfect as to give below two hits per ten trials on the average for the first day, it seems very probable that learning through the sixteen days of testing would have taken place. On the contrary, neither the curve of the first seven days' scores nor that of the second nine (see graph) shows evidence of learning. Third, it seems likely that such cues would be stronger for some numbers than others. The results show that the ratios of the frequencies with which each number was correctly given to the frequencies with which it was used as stimulus are not the same for all numbers; but no ratio is preponderantly higher than the others (see Table III). Such additional checks as may be had, therefore, do not favor the view that sensory cues were operative.

TABLE III

RATIO OF HITS ON GIVEN NUMBERS CALLED TO THEIR FREQUENCY AS STIMULI, SHOWING THAT NONE UNDULY STOOD OUT ABOVE THE REST.

Integer....	1	2	3	4	5	6	7	8	9	10
Ratio.....	2.20	1.67	2.29	2.41	2.00	2.94	2.00	2.85	2.14	2.58

SOME SPECIAL RELATIONSHIPS SUGGESTED

Several conditions prevailing in these tests appeared to limit scoring; at least, under them scoring above chance did not in general occur. Since this is important to the general understanding of extra-sensory perception, these conditions are reported for what they may be worth.

When the children had been absent from school prior to a test day, their averages were close to chance. Thus the average score for Monday was 1.07, and the average individual score after a day's absence through the week was 0.95. (The general average for all periods was 1.29.)

Records of some subjects seem to indicate a following of some habitual or rational scheme in making their calls. For these calls the average is even less than mean chance expectation. Two types of schemes were found. In one of these, which we might call the ordered type, the subject was found to follow the scheme of counting forward, counting backward, or using the same number on all 10 trials. These were, of course, not accidental. Twenty-two such schemes occurred, and their total score was only 13, which represents a negative deviation that is twice its standard deviation. The average score of these "schemers," where they "schemed," is 0.59, where 1.00 is chance. Although this negative trend is not significant, it suggests the problem of whether the selection of a rational scheme might be influenced extra-sensorially. The point of interest would be whether, when a subject chose a scheme, he chose that one of the 12 from which he might select whose score would be less than chance expectation. One subject, A. H., whose average was 3.0 before he used a scheme made an average of 0.6 for the remaining period, during which he used schemes four times out of seven. This decrease in score has an anti-chance value of 4.6, which corresponds to a probability of 0.000,002,2. In giving his introspective report, he said, "I just started at first and put all 9's, then all 10's, 1's, 2's, 3's, 4's, like that. Tomorrow I will make 10's, then 1's, 2's, 3's, and on like that." When asked if he did this when we first started, he said, "I would just think of them, put them down, and some were right." It was noted that a subject employed scheming only after a low scoring

period. It would, however, be hazardous to speculate much upon so little evidence.

The second type of scheming was that of using *each* of the ten integers, but in an unordered pattern, in the run of ten responses. We might call these non-repetitive schemes. These differ from those mentioned above in which all 10 integers were used, in that they are not in any recognizable order. The chance expectation frequency in 287 runs for the occurrence of a non-repetitive scheme (involving use of all the ten integers) is 0.1. The obtained frequency was 77, or almost one-third of the entire number of runs. The tendency toward following this type of scheming is thus seen to be very strong indeed. It seems likely that this tendency was due to failure to appreciate the instructions and understand the experimental situation adequately. After each day's run the children were permitted to see what the stimuli had been, and most of them were interested in doing so. But seeing the random order of stimuli in which the same integer might recur by chance did not in some cases result in giving a random order of guesses with recurring integers. No set of stimuli failed to have at least one recurrence. There were more non-repetitive schemes on the first day than on any other day; but the total proportion of such schemes for the second eight days was only 4% lower than that for the first eight days.

The 77 runs in which non-repetitive schemes were followed contained only 76 hits, an average of 0.99. It seems unlikely that anything other than chance was operative in these runs. Deducting these 77 daily runs from the total of 287 for the sixteen days, there are left 210 other runs. These 210 runs contained 295 hits, an average of 1.40. The difference between the average daily scores of the two groups of runs is 0.41, which is 3.23 times its standard deviation. The probability of obtaining this critical ratio by chance is 0.000,62.

There was a total of 99 scheming runs of both types with 89 hits. This averages 0.90. Deducting them, we have left 188 runs with 282 hits. This averages 1.50 hits per run of 10. The difference between the scheming and non-scheming averages was 0.60, which is 5.08 times its standard deviation. The probability of obtaining this ratio by chance is 0.000,000,2. We may therefore conclude that something was operative in non-scheming runs which was not operative in scheming runs.

A third point arising in connection with possibly rational scheming was the tendency to avoid calling the stimulus which was the same as the number of the trial. That is, on the fourth trial, 4 would be avoided as a response or call; on the third trial, 3, etc. The stimulus and the

ordinal number of the trial were the same in 13 of the 160 times. This is close to the expected number, which is 16. The call number and the number of the trial were coincident 146 of the 2,870 times as compared with 287 coincidences given by mean chance expectation. This indicates an unmistakable tendency toward the avoidance of calling the number which gives the order of the trial. Further, when the ordinal number of the trial was called for the trial, the scoring was close to chance. The 146 calls which coincided with the number of the trial would give a chance score of 146 x 13/160, or 12. The obtained score was 13. In the present experiment, therefore, only chance results obtained when the announced ordinal number of the trial was followed by the corresponding cardinal number as a response.

It appears from the analysis of these results that extra-chance scoring generally obtained but not under the following conditions: (1) when the subject had been absent from school the preceding day; (2) when the number to be called by the subject was the same as the number of the trial; and (3) when the subject followed rational or habitual schemes in selecting his calls or responses, whether he followed a fixed order or merely avoided any repetition.

An interesting point arises in connection with the subject B. R., whose average score of 2.42 is the highest, and his older brother, J. R., whose average of 1.67 ranks fifth. These two children made the same response 23 times in a possible 120, whereas chance would give them the same response only 12 times. This difference is 3.3 times the standard deviation. If these coincidences were due to chance, the percentage of successes would be no greater than that of B. R.'s other responses. On 23 trials, B. R.'s number of successes would be 2.42 x 23/10, or 5.6. Actually he had 10 successes. That is, he got more hits among those which he called in common with his brother than on the others; namely, an average of 4.3. While the number of cases is not enough upon which to base conclusions, it seems probable that there is some extra-sensory rapport between the two brothers.

As pointed out above, 20 of the 22 children used in the tests were rated as retarded. This situation affords opportunity for some comparison of scoring ability in the tests for extra-sensory perception and rating in the tests for intelligence. The comparison, however, for this group shows no reliable evidence of relationship between these two sets of measurements. From this small amount of evidence, then, intelligence differences cannot be said to correlate with differences in extra-sensory perception.

A rank difference correlation was computed for extra-sensory perception averages and intelligence ratings. Achievement test averages were no longer available for six of the subjects at the time it was decided to make this correlation. The intelligence ranking is only roughly approximate and is arrived at on the basis of achievement test averages and teacher judgments. The correlation was $+0.14 \pm .14$. For this intelligence range, then, we can draw no conclusion as to whether or not extra-sensory perceptive ability is correlated with intelligence. The highest four individual averages for the extra-sensory perceptual tests, however, came from the upper half of the intelligence distribution. This point in itself would suggest a connection between the two functions, were it not for the fact that the two highest of these four subjects held only tenth and eleventh positions in the intelligence ranking.

SUMMARY

It is most likely that extra-sensory perception occurred in the situation of these experiments, with the subjects here described, and, if so, that at least one subject in six gave evidence of it. Rationalized schemes appeared to interfere with scoring, as did also certain other conditions noted. There was no reliable relation found between intelligence and the scoring ability of the children tested.

GROUP TESTS FOR EXTRA-SENSORY PERCEPTION

VERNON SHARP and C. C CLARK

ABSTRACT: ESP card tests were given (1) to college classes, (2) to groups voluntarily coming into the laboratory, and (3) to individuals from those groups. The aim was to find whether individuals capable of extra-sensory perception could be found in such groups, and what effect isolating them from the group would have on scoring ability.

Of 69 subjects tested, 10 ultimately indicated individual ability in extra-sensory perception. The individual tests gave higher scores than the group tests. It was observed that marked variation in group scoring coincided with changes in the experimenter's attitude. Individual scoring showed notable changes with different experimenters.

The study was conducted in the Science Department of the School of Commerce, Accounts, and Finance of New York University where Dr. Clark is Associate Professor of General Science and Mr. Sharp is his assistant.

EXTRA-SENSORY perception (or ESP) has been defined as the ability to perceive without the use of the known senses. It includes both telepathy and clairvoyance. Recent publications have presented strong evidence that the ability occurs in certain subjects. It therefore seemed desirable to determine if a repetition of these experiments under different conditions would confirm the previous findings, and if so, to attempt to learn more as to the distribution of ESP throughout the population. Might it be expected that some persons in almost any group could demonstrate some extra-sensory ability under the proper conditions? Would groups and individuals chosen at random have tendencies similar to those previously discovered? To answer these questions, a series of experiments involving group and individual tests was conducted.

GENERAL CONDITIONS

The experiments should properly be divided into two series. The first series began March, 1936, and continued for eight weeks. The second series, likewise continuing for eight weeks, began in November, 1936, and was concluded January, 1937. There was an important difference in experimental conditions between the two series.

With few exceptions, volunteer students from New York University were used as subjects. In the main they were students who were taking the course in General Science. The greater part of these students were freshmen and sophomores in the School of Commerce. All who volunteered were used without any other criteria for selection.

In the first series, composed largely of group tests, the size of the groups ranged from two to twenty-four. The average size was six. In the fifth week eight psychology students joined the groups. In the eighth week of this series there were some brief tests of nine large groups, consisting of about fifty people each. They composed the entire classes taking the science courses, and were given the tests during the week when the regular course assignments related to alleged parapsychical phenomena. These large groups included some of the subjects who were working in the regular series. The statistical method of testing for ESP ability was carefully explained to all these groups, and a description of the present tests was given. All the students were given a chance to call the order of cards in two or three decks at a sitting.

In the second series both group tests and individual tests were made. However, all the groups in this series were small. The sizes ranged from 2 to 15. No class tests or large groups are included. Most of the testing done in this series was with individuals working alone. These scores may be compared with the scores secured in earlier series, since they included all individuals volunteering. This study was primarily made to discover if ESP ability could be detected in a substantially large number of people, whether they worked in groups or by themselves. The group method was a convenience in securing larger numbers of data, and offered an interesting comparison with individual tests.

All tests of the first series took place in the General Science classroom. Students were seated so that there was considerable space between them. During most of the tests the room was free from disturbances. At times, however, there were some interruptions due to people coming into the room for other purposes. The tests of the second series were conducted in a special room set aside for this purpose. Here the tests were free from the disturbances noted above. If a person was present only once, as a beginning subject, records have been withheld pending further work. In tabulating results, only those of individuals who have been present in more than one experimental period have been included. The only exception to this is the comparison of class results. The observer in most of the tests was Mr. Sharp.

The cards and the general methods used are those developed at Duke

University.[1] They were used because of their simplicity, and because otherwise it would be impossible to compare the results obtained here with those of Dr. Rhine. The technique which was followed throughout the first experiments, with the exception of a few preliminary trials, is known as the DT method. This is one in which the packs are shuffled and placed face down on the table, and all calls made before a single card is removed. If the pack is properly handled so that the bottom card is not made visible, the only part that the subject could see is the back of the top card and the edges.

Although the subjects in general preferred the method in which each card was held up for the members of the group to concentrate upon, the DT method was used for several reasons. The first was that it minimizes sensory cues for the subject. Second, this method permitted more tests to be given in the same space of time. Third, it avoided the annoyance which the other method caused the subject by moving at a speed that was not in time with his own scoring tempo. It prevented, too, the discomfort which the other method produced in some subjects as they looked in the direction of each isolated card and then down to record their calls. An hour of this would be likely to produce annoyance and muscular fatigue.

Five decks of cards were used at one time. The five decks were each shuffled by the observer, and laid out in a row on the desk or table where everyone in the group could see them. The students were given record sheets which had columns for the calls made, and for the correct calls. The students recorded their guesses as to the order of the cards in the deck on these record sheets. As soon as all had finished one deck, the correct order of the cards was called by the observer. These calls were written in the correct column and then checked against the guesses made. The deck was then shuffled by the observer while the students recorded their calls on the next deck. The shuffled deck was then placed at the back of the table. At the end of five decks these previously shuffled cards were moved forward without lifting and placed at the front again. They were rearranged so that the first in the old order might be the third or fourth, etc. in the new order.

At all times students were told not to attempt to keep any logical order in their guessing, but to put down the first impression that came into consciousness. It was explained that the fact that there were five of each symbol in the deck did not mean that they should count up their

[1] J. B. Rhine, *Extra-Sensory Perception* (Boston: Boston Society for Psychic Research, 1934.) Pp. 36-41.

cards to see that they came out to the exact number; that if they had any ESP ability, there would be no need to keep track of the symbols called. Every attempt was made to keep the gatherings informal. The only stipulation was that if there was any comment, to restrict it as far as possible to the intervals between the runs. In addition, they were told that they did not have to continue the work merely because they had started it, and that they were free to leave the group at any time. However, most of them worked regularly. There were frequently new subjects coming into the groups and usually they continued the work.

There are some obvious weaknesses to the procedure in this first series of tests. One is that students might intentionally or unintentionally give themselves credit for more hits than they had actually made. On the group work this has been eliminated by comparing the correct calls with a master record made up of the calls on which the group as a whole agreed. Any errors of this sort have, therefore, been eliminated. Another possible objection is that students might alter their records while checking them or have blank call spaces to be filled in as the correct symbols are called. In the second series we tried to eliminate these possibilities.

The group work in the first two weeks of the second series was done in the same way as that in the first series, but thereafter all members of the groups were required to exchange papers before the correct symbols were called back.[2] In the individual DT work of the second series the experimenter marked down the correct calls and had the cards always in full view as the subject (for convenience) turned them over in the check-up.

Some other techniques were used in the individual work of the second series, mainly TM and STM. TM consists of having the experimenter hold the cards while the subject touches or points to one of five key cards to indicate where he believes the top card on the deck should be placed. Cards are piled up behind these key cards and checked at the end of the deck. This procedure was followed with 13 subjects, through 166 runs. STM consists of essentially the same process with the additional feature that the cards are held behind a screen so that the subject does not see them. For one week (the 4th) a vertical screen with a 3-inch aperture between the lower edge and the table was used.[3] As no secondary screen was used covering the aperture, the subject

[2] These first two weeks contributed nothing to the evidence. The scores were close to chance.

[3] For a picture of this set-up, see *Journal of Parapsychology*, Volume I, March 1937, facing p. 18.

could see the experimenter if he brought his head to the level of the table. After that time an inclined screen blocking vision to experimenter's shoulders and extending toward the subject down to table was used. This permitted the experimenter to keep the subject in view at all times while holding the cards themselves in a position such that subjects could not possibly see them. This was used with 20 subjects calling 384 decks.

A few other techniques were used: OM with 5 subjects for 45 runs, BM with 2 of these for 15 runs and BT with 1 subject for 15 runs. GSTM was used with 5 subjects for 97 runs.

In OM (Open Matching), five cards bearing the five ESP symbols are placed face up in a row before the subject. He is given a pack of 25 ESP cards and instructed to "match" each card, face down, to one of those before him. After the 25 are matched the piles are turned over and checked for the number of hits.

In BM (Blind Matching) the key cards (five cards each bearing a different symbol) are placed face down before the subject such that he does not know the symbols on them. He matches a pack of 25 cards to them as in OM.

In BT ("Before Touching") the pack of ESP cards is placed face down before the subject. He calls the top card; a record is made of the call, and the card removed and laid aside by the experimenter (or in some cases by the subject). This procedure is continued through the 25 cards. The removed cards are kept in order, and at the close of the run are checked against the calls.

In GSTM the procedure was to have one experimenter look at the cards (held out of sight under the inclined screen). The experimenter holding the cards did not know the location of the key cards, which were placed by a second experimenter and the subject. The key cards were frequently rearranged, so that the first experimenter might not deduce the location of the key cards from the number of hits made on each deck. Miscellaneous procedures were used on 34 runs.

The following problems were principally in mind in this research. First, as stated above, the occurrence of ESP, and its distribution; second, the relative influence of the presence of the group upon the individual during the tests; third, the influence of the experimenter himself; and fourth, the occurrence of any periodic variations in correct calls in the use of the BT and DT methods such as have been previously observed.[4]

[4] RHINE, *op. cit.*, pp. 122-137.

RESULTS BEARING ON THE QUESTION OF THE OCCURRENCE OF ESP

Tables I and II summarize the data secured in the first and second series. The students taking the tests have been classified as follows: *Beginners* are students who were working with the cards for the first time. All their calls during their entire first experimental period are included. Experienced groups include those who had done individual work on the cards before beginning group tests. Experienced individuals are persons who had previously worked with the cards and who later did individual work. A group was a minimum of two people working on a deck of cards at the same time. Critical Ratio = Deviation $/\sqrt{npq}$. In order to be accepted as statistically significant the results must produce a critical ratio equal to 2.5 or higher. Where the CR is designated by (*) it means that the ratio is less than .5.

Table I summarizes all of the calls observed by Sharp in the two series; also, a few calls made under miscellaneous observers. This data affords some evidence for the existence of ESP ability. In Series I the

TABLE I

SHOWING SUMMARY OF ALL CALLS IN SERIES I AND II.

Week	Decks	Dev.	Av/25	CR	Classification
1.....	119	+3	5.00	*	
2.....	131	+50	5.38	2.2	
3.....	125	+10	5.13	*	
4.....	286	+104	5.40	3.1	Series I
5.....	549	−12	4.98	*	
6.....	904	−16	4.98	*	
7.....	639	−3	5.00	*	
8.....	674	+64	5.10	1.2	
Total.....	3427	+200	5.06	1.7	
1.....	40	+8	5.2	.6	
2.....	88	−8	4.9	*	
3.....	47	+3	5.06	*	
4.....	245	+42	5.17	1.3	Series II
5.....	342	+132	5.39	3.6	
6.....	89	+34	5.38	1.8	
7.....	129	+17	5.13	.8	
8.....	229	+218	5.95	7.2	
Total.....	1209	+446	5.37	6.4	
Total I-II.....	4636	+646	5.14	4.7	
Total...	35	+33	5.94	2.8	Preliminary Series, Misc. Obs.
Total.....	372	−39	4.82	1.0	Misc. Obs., Sec. Series
Grand Total....	5043	+640	5.13	4.5	Total

total deviation on 3,427 decks was 200 above chance expectation. The CR of this deviation is 1.7, and taken by itself might be a chance variation. In Series II 1,209 decks were called with a positive deviation of 446. The CR in this case is 6.4. The odds against this deviation being due to chance are greater than 10,000,000,000 to one. When the results of both series are combined, the deviation of 646 on the 4,636 runs reported is 4.7 times the SD, giving odds of 500,000 to one that the result did not arise by chance.

In addition, there was a short exploratory series where each subject was alone with the experimenter. In this case there was a positive deviation of 33 calls above chance on 35 decks run by 5 students. The high average of 5.94 for these five students is due largely to one subject who averaged 7.5 on the 8 runs he made.

Also, a number of decks were run in the second series under miscellaneous observers. On the 372 decks called there was a deviation of —39. The average score is 4.82, and the CR of this is 1.0.

If all of these results are totaled, as shown in Table I, we find that on 5,043 decks of cards consisting of 126,075 calls the subjects made 640 hits above chance. This gives 4.5 for the CR and represents odds of 300,000 to one against chance.

It would thus seem that enough ability of some sort existed in this sample to provide a significant positive deviation for the total number tested. The results as a whole indicate that there is some extra-chance factor present in these tests.

* * *

Table II is a record of all subjects observed by Sharp in both the first and second series. These are arranged in order from the lowest average to the highest average.

How well distributed is ESP ability? A recent survey by W. F. Prince[5] of a large group of persons listed in *Who's Who in America* indicated that one out of four admitted having experiences which they took to be of a parapsychical nature. What percentage of persons will succeed in ESP tests? Is ESP a common possession, varying only in degree, or is it as rare as the photo-electric effect in the field of physics? Preliminary to asking whether a large number of people have such an ability, let us propose instead a question that can be answered with a greater degree of certitude from the data we have available. What proportion of people can make scores, under the conditions described, that are in excess of those expected by chance? Can any of these individual

[5] *Human Experiences,* Boston Society for Psychic Research, 1931.

TABLE II

SHOWING TOTAL SCORES OF INDIVIDUALS COMPOSING GROUPS.

SERIES 1					SERIES 2				
Subjects	Decks	Dev.	Av./25	C.R.	Subjects	Decks	Dev.	Av./25	C.R.
L. Zuckerman	20	— 9	4.50	1.12	T. Seabrook	24	—24	4.00	2.46
H. Epstein	35	—17	4.51	1.44	La. Dyson	14	— 7	4.50	.94
M. Denecke	99	—44	4.55	2.21	A. Ignatow	10	— 4	4.60	.63
A. Thompson	78	—27	4.65	1.53	J. G. Pratt	6	— 2	4.67	*
L. Goldstein	30	— 9	4.70	.82	M. Adler	30	— 8	4.73	.73
L. Sternfeld	44	—11	4.75	.83	V. DeSmitt	11	— 3	4.73	*
J. Marsich	19	— 4	4.79	*	Li. Dyson	24	— 6	4.75	.61
M. Slutzkin	45	— 9	4.80	.67	J. Stern	9	— 2	4.77	*
L. Blum	96	—16	4.83	.82	M. Iceland	6	— 1	4.83	*
D. Dalton	69	—11	4.84	.67	L. Zucker	10	0	5.00	*
Li. Dyson	95	—15	4.84	.77	E. Scher	97	+ 2	5.02	*
T. Seabrook	111	—17	4.85	.81	P. Greenberg	20	+ 1	5.05	*
L. Rosenberg	39	— 5	4.87	*	D. Holtzman	70	+ 5	5.07	*
S. Dworetz	149	—12	4.92	*	H. Nelson	58	+ 5	5.09	*
G. Paganelli	92	— 6	4.93	*	H. Fertig	35	+ 3	5.09	*
H. Steiglitz	65	— 3	4.95	*	A. Fertig	36	+ 4	5.11	*
A. Slobodin	69	— 3	4.96	*	R. Davidson	59	+ 7	5.12	*
J. Pearle	83	— 1	4.99	*	T. Martin	35	+ 5	5.14	*
J. Weiss	115	— 1	4.99	*	H. Aronson	62	+12	5.19	.76
M. Eisgraw	11	0	5.00	*	S. Dworetz	53	+14	5.26	.96
L. Ratnofsky	34	0	5.00	*	J. Mottershead	3	+ 1	5.33	*
H. Hawley	5	0	5.00	*	D. Pond	15	+15	5.33	.65
R. Davidson	234	+ 1	5.00	*	J. R. Welch	84	+32	5.38	1.75
W. Seigle	80	+ 1	5.01	*	F. Bridal	15	+ 7	5.47	.89
F. Davis	44	+ 3	5.07	*	M. Denecke	17	+ 8	5.46	.97
H. Gardlin	57	+ 5	5.09	*	F. Egner	30	+15	5.50	1.19
B. Coplon	41	+ 4	5.10	*	C. Worden	30	+17	5.57	1.55
T. Sloan	35	+ 5	5.14	*	A. Thompson	24	+14	5.58	1.43
C. Worden	122	+20	5.16	.95	E. Orhlich	17	+10	5.59	1.21
T. Martin	55	+ 9	5.16	.60	H. Banks	20	+12	5.60	1.34
R. Fuld	110	+18	5.16	.89	F. Spector	25	+16	5.64	1.6
L. Einsidler	22	+ 4	5.18	*	H. Hawley	72	+44	5.61	2.60
J. Mottershead	105	+19	5.18	.93	M. Slutzkin	45	+32	5.71	2.38
R. Richelson	125	+22	5.18	1.00	M. Reese	11	+ 9	5.82	1.35
E. Swid	175	+34	5.19	1.28	C. C. Clark	56	+65	6.16	4.34
E. Gaston	70	+13	5.19	.78	E. Sharp	10	+13	6.3	2.06
F. Spector	117	+23	5.20	1.07	E. Golish	5	+ 8	6.6	1.79
E. Scher	83	+17	5.20	.94	E. Swid	61	+137	7.2	8.8
L. Sajor	30	+ 8	5.27	.73					
P. Solomon	63	+16	5.25	1.02		1209	+446	+.37	6.4
R. Burgess	37	+10	5.27	.82					
L. Wackwitz	46	+12	5.27	.89					
E. Rivkin	35	+10	5.29	.85					
L. Wolff	68	+21	5.31	1.27					
M. Iceland	17	+ 6	5.35	.73					
H. Nelson	30	+11	5.37	1.1					
La. Dyson	97	+38	5.39	1.94					
H. Banks	44	+18	5.41	1.36					
V. DeSmitt	11	+ 5	5.46	.75					
J. G. Pratt	21	+11	5.52	1.20					
M. Reese	41	+26	5.63	2.03					
H. Jacobsen	5	+ 4	5.80	.90					
E. Sharp	4	+27	11.75	6.74					
	3427	+200	+.06	1.71					

subjects make calls that are *significantly* above chance, in the commonly accepted sense?

Of the 69 subjects reported in Table II, 23 scored at chance or below chance. Forty-six were above the chance level. There is a 2 to 1 ratio of above-chance scorers to the chance and sub-chance group. This of itself, of course, does not mean that all 46 have shown ESP ability since the variation in 42 of the 46 cases taken individually is within the limits of fluctuation from the theoretical mean to be expected by chance about once in 150 times, the CR of their deviations being less than 2.5. However, the preponderance of these variations on the positive side may, since the work as a whole is very significant, be taken to indicate that many of the subjects have some ability.

In these tests there were 4 subjects whose individual positive deviations each have a CR of 2.5 or above and are, when treated alone, statistically significant. They were: H. Hawley, who made 44 calls above chance on 77 runs. The CR of 2.5 is just on the border, representing odds of about 150 to 1 against chance causation. C. C Clark, one of the authors, comes next in significance, with a deviation of 65 for 56 runs with a CR of 4.3. E. Swid, with a total of 171 calls over chance in 236 runs, has a CR of 5.6; and E. Sharp, with a total deviation of 40 on 14 decks, has a CR of 5.5.

In this study, therefore, it is seen that a group of 69 people selected at random has produced a few who have this ability. The number of such persons is probably variable. However, further study is necessary before conclusions regarding the number may be made.[6]

[6] Further tests in a third series made with the same subjects in progress since this report was prepared, have added six more to the group whose scores have positive deviations with a CR above 2.5. These subjects with their scores are shown in the table below. In addition, 8 others have made scores with a CR above 2.0, which is used by some statisticians as a criterion of significance.

Subjects	*Decks + Dev.*	*CR*	*Av./25*
H. Aronson	499 + 232	8.2	5.47
M. Adler	151 + 94	3.8	5.62
F. Bridal	69 + 50	3.0	5.73
H. Banks	77 + 40	2.3	5.52
La. Dyson	224 + 59	2.0	5.26
R. Davidson	834 + 258	4.5	5.31
L. Einseidler	224 + 79	2.6	5.35
J. Marsich	55 + 32	2.2	5.58
J. Pearle	285 + 141	4.2	5.49
V. De Smitt	64 + 36	2.3	5.56
F. Spector	248 + 75	2.4	5.30
C. Worden	167 + 54	2.1	5.32
L. Wackwitz	74 + 37	2.2	5.50
J. R. Welch	150 + 56	2.3	5.37

EFFECT OF EXPERIENCE ON SCORES

Another of the questions was: Do subjects improve with experience? Some clue to this might be gained by comparing the work of beginning and experienced subjects. There are two avenues of approach to this question. One is a comparison between these beginners and experienced subjects in the two series. Another is the comparison of the beginning and experienced subjects in the classes. A subject is considered a beginner for the duration of his entire first test period.

TABLE III

COMPARISON OF BEGINNING AND EXPERIENCED SUBJECTS.

	BEGINNERS		PRACTICED	
	Decks	Dev.	Decks	Dev.
First Series	407 Av. 4.95	−20	2,999 Av. 5.09	+209
Second Series	116 Av. 5.26	+40	1,087 Av. 5.38	+408
	523 Av. 5.04	+20	4,086 Av. 5.15	+617

In Table III it will be seen that the experienced subjects gave better results over both series. The same thing seems to be the case in the large classes as shown in Table IV. The number of decks done in the regular series by beginners is 523. The positive deviation on this is 20. This is an average of 5.04. Later calls by the same subjects as experienced are + 617 on 4,086 runs, with an average of 5.15. It would therefore seem that there is a difference between the beginning and experienced periods of work with these subjects. Something is apparently acquired with experience.

If we take the class groups from the same series (I) and compare the scoring of beginners and experienced subjects who were simultaneously making calls under the same conditions and on the same decks, we find the same favorable result of experience. Experienced subjects here made 52 hits on 86 decks called, with an average score of 5.6 per

H. Hawley moved from his borderline position by maintaining a rate of scoring for 344 decks which yielded a positive deviation of 170. The CR of this is 4.6. No subject dropped from significant level in the first two series to chance level by his additional work. Thus in the first two series, 4 out of 69 scored significantly in the series. In the third series the number rose to 10, which means that number reported at this time. Thus 1 out of 7 has been statistically accepted as evidence of this extra-chance factor in perception.

deck. The CR is 2.8 and may be accepted as probably being significant of something other than chance. Beginners, in the same classes, with conditions the same, except for the fact that they had not previously made calls on the cards and were therefore unfamiliar with the procedure and the symbols, made calls that were 36 above chance on the 535 decks called. Their average score was 5.07 per deck, and the CR of .82 might be a chance variation. Beginners here gave insignificant deviations, while experienced subjects yielded deviations that were significantly above chance. See Table IV for further details.

TABLE IV

COMPARISON OF BEGINNERS AND EXPERIENCED SUBJECTS IN CLASSES.

Class Period	Beginners	Experienced
Monday and Wednesday 11 to 12	92 + 9 Av. 5.10	14 + 4 Av. 5.29
Tuesday and Thursday 10 to 11	109 — 8 Av. 4.93	16 + 4 Av. 5.25
Tuesday and Thursday 11 to 12	99 + 15 Av. 5.15	30 + 35 Av. 6.12
Tuesday and Thursday 12 to 1	141 — 1 Av. 5.00	20 + 6 Av. 6.12
Tuesday and Thursday 2 to 3	94 + 21 Av. 5.22	6 + 3 Av. 5.50
Total	535 + 36 CR = .8	86 + 52 CR = 2.8

It appears from the results shown in the last two tables that in this experiment experience proved to be an advantage. It is realized that a number of possibilities lie within the general term "experience." Greater familiarity makes for better attention, and may remove inhibitions accompanying the beginning stage. Interest might grow as the feeling of competition or desire to achieve develops. The discussion of the group and of friends might lend a more favorable attitude and doubts possibly decline as good results of others are related. Finally, some development of ESP capacity with practice may occur, but it is no more established by these results than are the other possibilities. It is, however, an important question.

SOME OBSERVATIONS ON EFFECT OF OBSERVER ON SUBJECTS' SCORES

Is it possible that the observer may, consciously or unconsciously, influence the scoring for better or worse by his own attitude? During a long experimental period, if the mental attitude of the observer were altered, would it be reflected in the scores? If different observers were used would results given by same subjects vary with the observers,

corresponding with variations in their attitude towards the work in question?

During the fifth, sixth and seventh weeks of the first experimental period, Sharp's wife was in the hospital. In the fifth week the physician in charge informed him that an immediate operation was necessary. Internal hemorrhages made it questionable whether she would live. During these weeks the experimenter was working, then, with divided attention. At the end of the seventh week, Mrs. Sharp had recovered enough to be taken home. We have, therefore, classified the material as being done under "normal" or "abnormal" conditions (that is, for the observer).

The total scores of the groups for the first, second, third, fourth and last weeks give a deviation of + 231 on 1,235 decks. For the three weeks the experimenter was so disturbed, the deviation is —31 on 2,092 decks. It is interesting to note that each week under "normal" conditions gave a positive deviation; each "abnormal" week, a negative one. The CR for all five normal weeks is + 3.28, well above the standard accepted for significance. For the other three weeks the CR is —.34. A comparison of the difference in the two means shows that the CR for the difference is + 2.87. This contrast is not, of course, an experimental one. But the results are nevertheless highly suggestive and are submitted in that light.

Table V gives a further analysis of the work done during the two periods mentioned above. It is divided into three parts. The first consists of all subjects who gave negative deviations for the entire first series. The second consists of all subjects who gave a positive deviation for the entire series. These two groups consist of all calls made by practiced subjects who did work in both periods. All of these students were practiced before the fourth week. The third group is made up of all students who did not begin until the fifth week or later and the

TABLE V

COMPARING OBSERVER'S "NORMAL" AND "ABNORMAL" PERIODS.

Group Number	Number Subjects	"Normal"	"Abnormal"
I	10	349 + 0 Av. 5.0	475 — 135 Av. 4.72
II	13	504 + 195 Av. 5.38	603 — 19 Av. 4.97
III	8	198 + 34 Av. 5.17	404 + 56 Av. 5.14
Totals	31	1,051 + 229 Av. 5.22	1,482 — 98 Av. 4.93

observer was in the "abnormal" state. These classifications were chosen as probably containing a more homogeneous set of data than random comparison between the two periods.

In the group whose total scores were below chance, the total score for the normal period is exactly chance on 349 decks. During the abnormal period, the scores are 135 hits below chance on 475 decks. This is 3.1 times the sigma, and negatively significant. The subjects with total positive scores were 195 hits above chance on 504 decks during the normal period. The CR is 4.35. This is positively significant. In the abnormal period the calls were 19 hits below chance on 603 decks. The CR is —.39. With those students who started after the fifth week and would be unable to detect a difference in the attitude of the observer, due to the fact they had no previous standard for comparison, we find that in the normal period the deviation on 198 decks was + 34 averaging 5.17. In the abnormal period, the deviation was + 56 on 404 decks averaging 5.14. Here, too, there is a difference in the average score but it is not pronounced. A greater increase than .03 per deck might have been expected with a return to "normal" state. This is but a trivial difference.

The above observations are not reported as conclusive evidence within themselves. It is possible that other unobserved factors were operating to produce or influence these results. It seems likely that some factor or group of factors operated to produce a significant negative shift in chance scores, and to reduce significant positive scoring to chance level. The positive verification of the causes of such variability in ESP scoring would be a valuable contribution. The possible cause here reported is the only one noted in this study.

Another avenue of approach to the question of the influence of the observer is obtained from the second series. In this case, the observing was done by several of the assistants in the laboratory, as well as by Sharp. An analysis shows that with work done under the observation of Myers there was a negative deviation of 55 on 79 decks. This is significantly negative. In one case, a subject, working with Davidson, had a positive deviation of 15 on 17 decks for an average of 5.88 per deck. Myers took over the subject at this point, and on the following 10 decks the subject had a negative deviation of 9, for an average of 4.1 per deck. This is not significant, however. In another instance, Myers worked with another subject first, and the subject had a negative deviation of 11 on 15 decks, for an average score of 4.23 per deck. Sharp then took over the subject and the scoring improved gradually

until she averaged 5.53 per deck for the last 15 decks called. Again suggestive, though not significant. Another assistant, Berger, obtained a negative deviation of 26 on 238 decks with a beginning subject. All subjects observed by Berger were .84 sigma below chance in the calls made for him.

In the work observed by Myers the total scores are so definitely negative that they have a CR of 3.1. One of our subjects, who has done a great deal of work, stated privately that Myers distracted her to such a point that she was unable to think of what she was doing. While conducting the tests Myers kept swinging his watch chain and talked about extracurricular activities. He manifested little interest in the work at hand. At that time, he had little or no interest in the experiments.

TABLE VI

SHOWING SCORING OF SUBJECTS UNDER DIFFICULT OBSERVERS.

	MYERS		DAVIDSON		BERGER		SHARP		
Subject	Decks	Av.	Decks	Av.	Decks	Av.	Decks	Av.	Technique
S. Dworetz.....	10	4.10	17	5.88	..	....	20	5.90	STM
La. Dyson.....	16	4.75	..	....	10	4.70	4	4.00	DT
Li Dyson.....	16	4.69	..	....	..	....	4	5.25	DT
T. Martin.....	37	4.00	..	....	..	....	33	5.24	STM
H. Aronson.....	..	....	..	....	15	5.33	8	5.75	STM
F. Bridal.....	..	....	..	....	10	4.60	15	5.47	DT
F. Egner.....	..	....	..	....	40	4.85	30	5.50	DT
P. Greenberg.....	..	....	..	....	20	5.05	20	5.05	DT
H. Hawley.....	..	....	..	....	63	4.70	30	5.20	STM
E. Orlick.....	..	....	..	....	20	5.35	7	6.12	DT
L. Zucker.....	..	....	..	....	20	4.55	10	5.00	DT
Total.....	79	4.30	17	5.88	198	4.86	181	5.36	

A comparison of the scoring ability of 13 subjects using the same techniques with different observers is shown in Table VI. It indicates that the response of a subject may vary with different experimenters. Myers was definitely skeptical; Davidson was favorable; Berger not committed; and Sharp is, of course, very favorable to the research. It would be necessary to have many times the available data to come to a definite conclusion, but on the basis of that reported, there seems to be some evidence that the observer in experiments of this sort is more than a mere passive recorder. His attitude and actions would seem to be to some extent mirrored in the scores of his subjects.

INFLUENCE OF GROUP WORK ON INDIVIDUAL'S SCORES

Is there a difference between the scores of subjects working in groups, and the same subjects working as individuals? Tables VII and VIII present material bearing on this question. Two different attacks have been made on the same problem. Table VII gives a contrast between group and individual work done in the two series. It will be seen that the total number of hits in group work in the series is 194 above chance on 3,277 decks. The CR on this is 1.6. Individual work, on the other hand, of these same subjects gave a deviation of 388 on 1,156 decks, with a CR of 5.7.

TABLE VII

SHOWING A COMPARISON OF GROUP AND INDIVIDUAL WORK.

	GROUP				INDIVIDUAL			
	Decks	Dev.	Aver.	CR	Decks	Dev.	Aver.	CR
First Series.....	2950	+178	5.06	1.6	246	+39	5.16	1.2
Second Series.....	327	+16	5.04	.4	910	+348	5.38	5.7
	3277	+194	5.06	1.6	1156	+388	5.34	5.7

A few subjects did work in groups and as individuals under uniform conditions. The scoring of Mr. Denecke showed a marked contrast. Working with the groups his scores were consistently below chance, averaging only 4.42 per 25. There was a negative deviation of 55 on 94 decks which is 2.82 times the sigma. One day he made some calls on the cards after the group had stopped work. In this case he had a positive deviation of 11 on 5 decks. The CR for this is 2.46. Where he was making the calls alone under our direct observation he averaged 7.2 for 5 decks. (This subject seems not to score well working in groups.) His second series, consisting entirely of individual work, adds weight to this evidence. In this he made eight calls above chance on 17 decks. The CR on the 22 decks of individual work was 2.02. The comparison of the two means shows the difference to be 3.1 times the sigma.

Table VIII presents a comparison of practiced subjects who did individual and group work under uniform conditions. The techniques are the same (DT) and the observer was the same in all cases. Mr. Denecke's beginning efforts have been eliminated and his practiced work only included. This conforms with the other three subjects listed.

TABLE VIII

SHOWING COMPARISON OF GROUP AND INDIVIDUAL WORK UNDER UNIFORM CONDITIONS.

Subject	Group	Individual
Denecke	84 — 41	22 + 19
Scher	55 + 31	23 + 19
Spector	37 + 41	10 + 8
Swid	5 — 1	164 + 64
Total	181 + 30	219 + 110
	Av. 5.17	Av. 5.5
	CR. 1.1	CR. 3.7

PERIODIC VARIATION IN CORRECT CALLS

One of the curious things discovered by Dr. Rhine[7] was the fact that when working on a deck of 25 cards in the DT procedure there was a tendency to make the greatest number of hits at the beginning and ending of the deck. The smallest number was made in the center. This was true of each of four individuals tested in this way and of the composite curve made by adding their results together. Another subject, Miss Ownbey, gave a curve which was an inverted U as contrasted to the conventional U in the majority of cases.

We have drawn graphs of the scoring done by all of the students in the work described in this paper where such scoring has been labeled in all detail. In the first series our DT work was done in blocks of five decks, three decks, and one deck. In most of the cases they were simply labeled as DT work. This makes it impossible to use the first series as a whole. There are some cases where we have the material definitely labeled and this can be used. The DT work in our entire second series was in units of one deck, or 25 cards. We have the records of the groups in the class tests as an additional source of information on DT 25's.

An examination of the curves for the DT of all the classes, beginning subjects in both cases, shows a U curve similar to that of Rhine except that three points are below the mean. The curve is shown in Fig. 1. When all our DT 25 results from the classes and from the first and second series are added we also get a U curve, shown in Fig. 2. The data are reported in Table IX. Since the beginners' results in class calls were very near the chance level, it would seem that, in spite of the fact that scores are approximately chance, whatever factor or fac-

[7] RHINE, *op. cit.*, pp. 122-125.

tors operate to make individuals get the highest number of hits at the extreme ends of the deck, also seem to be operating in their case.

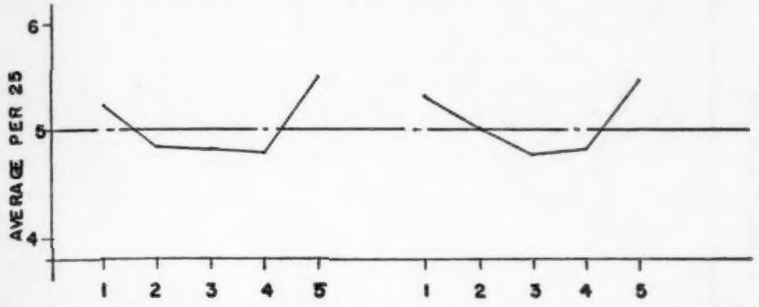


FIG. 1.—DT curve of average hits per 25 in groups of five cards in order down through the deck. Beginning subjects.

FIG. 2—DT 25 curve of average hits per 25 in groups of five cards in order down through the deck. All subjects.

Table IX summarizes this material. This is the record on nearly 500 people. There were 433 students in the classes who called the cards. Some of those in the classes were also in the first series.

TABLE IX

SHOWING DISTRIBUTION OF HITS BY FIVES IN DT TWENTY-FIVE.

Groups of Fives	1	2	3	4	5	Mean or np	Classification
Number of hits:	550	510	503	491	575	516	Classes
	70	66	56	63	74	66	Series 1
	525	508	467	486	521	494	Series 2
	1,145	1,084	1,026	1,040	1,174	1,076	Total
	5.31	5.04	4.76	4.83	5.46	....	Av. per 25

We have some runs or score sheets labeled as DT 75. This means that three decks were recorded before the cards were checked back. Taking each deck cross-sectionally by 5's, the same as in the DT 25's, we find that instead of a U curve we have an inverted U, with the minimum number of hits at the two ends of the deck. This is shown in Fig. 3. The DT 75 was done by six subjects, and it is compared with their own DT 25 records.

Comparing runs of 25 at a time with runs of 125 at a time we find again the inverted U curve. This is shown in Fig. 4. The material upon which this is based is composed of 5 runs by each subject under both conditions. This was all part of a single day's work and was done

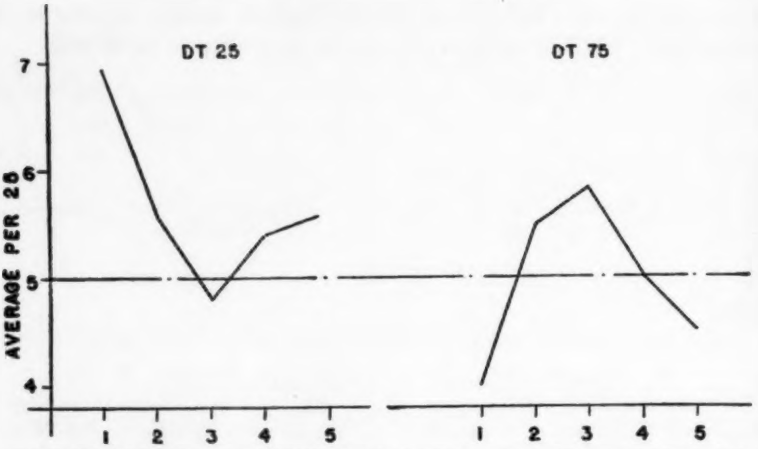


FIG. 3.—Comparison of DT 25 and DT 75 in average hits per 25 made in five sections down through the deck.

within the period of an hour's time. Here, again, we notice the curious inversion observed for the DT 75.

In our first series we have an instance where 24 students were working on the same deck of cards in BT procedure. The observer paused slightly at the end of each group of 5. Curiously enough, the hits on each card plotted against cross sections of 5 shows much the

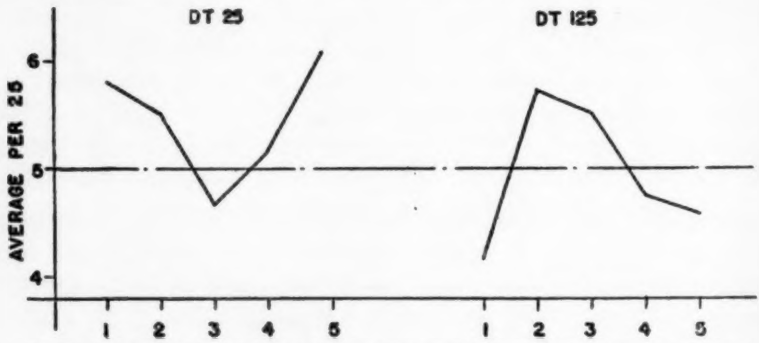


FIG. 4.—Comparison of DT 25 and DT 125 in average hits per 25 made in five sections down through the deck.

same rhythmic tendency that is apparent in Pearce's curve. The composite curve again is remarkably similar to Pearce's. This curve is given in Fig. 5. Hits were as follows in order through the deck: 13, *18*, 17, 12, 14; 12, *18*, 15, 13, 14; 11, 13, *19*, 13, 18; *22*, *19*, 14, 13, 10; 16, 14, 12, 10, 10.

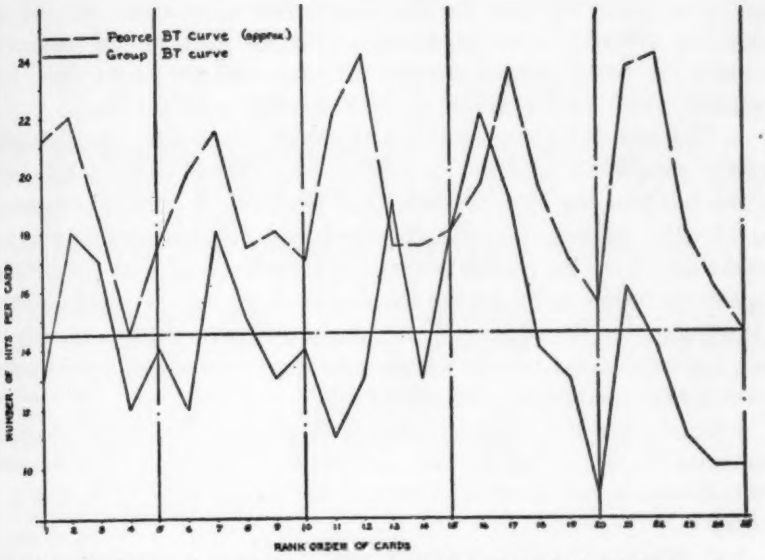


FIG. 5.—Comparison of hits per rank order of cards in BT procedure with Pearce curve.

SUMMARY

The experimental data compiled and analyzed in this study covers a total of 126,075 trials including all work done under the conditions here reported. The study leads to the following conclusions:

1. ESP ability as measured by the techniques employed in this investigation seems to be a real phenomenon. Sixty-six per cent of the 69 subjects tested over many periods have been able to maintain a scoring rate above the chance mean. Of the total there were 46 who had positive deviations from chance and 23 who had negative deviations. This yields a critical ratio of 4.5 for the total of 5,043 runs, or 126,075 calls made.

Four of these subjects had positive deviations which were individually significant. Later scoring has increased this number to 10 of the subjects included in the original 69.

2. Experience in taking the tests appeared to increase the subject's ability to call the cards correctly. The CR for all calls made by subjects after their first period of work is 4.8. How long additional experience would cause an increase in scoring ability was not determined.

Why this improvement occurs was not a part of this study. It may be the result of more familiarity with the technique and materials, more

familiarity with the way the perception manifests itself, loss of any inhibiting attitude, or an improvement on the part of the student in securing the proper mental attitude for extra-sensory perception. It is not clear whether ESP ability as such develops with practice.

3. The attitude and personality of the observer when present seem to have some effect upon the exercise of ESP ability. This is suggested by the fact that the observer who gave most of the tests experienced a period when he was mentally disturbed and distracted while carrying out the study. In the normal period, he secured results with a significant positive deviation, while during the distracted period the results showed a small negative deviation. While no conclusion is drawn, this difference is probably not accidental but may represent an actual difference in conditions surrounding the experiment.

Furthermore, the results given above suggest that some observers are definitely distracting to the subjects and may be unable to secure positive results, while other observers are better able to secure good results.

4. Subjects working individually with the observer are able to score higher than they do when working in groups. The CR of the entire group work of the Series II is .4 while that of all scores made by the subjects working individually is 5.7.

EXTRA-SENSORY PERCEPTION AMONG THE BLIND

MARGARET M. PRICE and MARGARET H. PEGRAM

ABSTRACT: In a study of extra-sensory perception in the blind, 66 subjects ranging in age from 8 to 35 years were given various card-calling tests. The total average above chance for all subjects significantly indicated that extra-sensory perception occurred among the subjects. Forty-four per cent of the subjects gave individually significant scores. No reliable correlation was noted between ESP scoring ability and age, extent of blindness, nor with the time in which the subject became blind. Some of the significant subjects were totally blind. When the subjects matched sealed cards their scores were higher than in similar matching of open cards.

Miss Price is a senior student in psychology and Miss Pegram is a member of the staff of the Parapsychology Laboratory at Duke University.

INTRODUCTION

THE WIDE-SPREAD popular belief that blind people are peculiarly endowed with a high degree of "intuition," or that they possess a compensatory "sixth sense" suggests a scientific study of extra-sensory perception in the blind.

Various questions arise in this connection: (1) Are blind subjects capable of extra-sensory perception? (2) If they do have such a capacity, how does it compare with the abilities of seeing subjects? (3) Is there any correlation between the extent of visual impairment and the capacity for extra-sensory perception? (4) Is there any relationship between ESP ability and the length of time which the subject has been blind?

It is of particular interest to see whether subjects who have been totally blind from birth, i.e., who have never had visual experience, are able to respond to stimuli which are commonly experienced only through the visual sense in sensory perception.

The general problems considered in this paper are: (1) Whether extra-sensory perception can be demonstrated by blind subjects; (2) whether the occurrence of this capacity correlates with the degree of

blindness; (3) whether there is any difference in ESP scoring ability between subjects who have been blind since birth and those who have lost their sight later in life.

In addition, the following studies were made: (1) A comparison of scores obtained by methods which involved verbal indication of guesses with those which involved non-verbal responses; (2) a comparison of results obtained when the cards to be guessed were sealed in envelopes (sealed card packs) with results obtained when the cards were not enclosed in envelopes; and (3) a comparison of the average scores for each experimental period.

Further analyses of the data will appear in a later report, and a comparative study of seeing and blind subjects made under closely comparable circumstances is contemplated.

SURVEY OF SUBJECTS, METHODS, AND CONDITIONS

Subjects: Two groups of subjects were used in this investigation. One group was composed of 61 boys who were enrolled in the State School for the Blind, in Raleigh, North Carolina (the criterion for entrance to which is inability to read pica type at a distance of one foot from the eyes). The age range for this group was from 8 to 23 years. Approximately sixty per cent of the boys enrolled were included in this group. They were selected as a representative group by the superintendent of the school. The second group was made up of five employees of the Durham Workshop for the Blind who volunteered to act as subjects. Ages of this group ranged from 19 to 35 years.

Altogether, there were 66 subjects ranging in age from 8 to 35 years and varying educationally from first grade pupils to high school graduates.

There was, therefore, a large number of partially blind subjects with a wide difference in the amount of sight. Nine of the total number of subjects have been totally blind from birth, and five more became totally blind at a later date. For purposes of this experiment, the group listed as partially blind has been subdivided into two groups: (1) Those who were able under some conditions to see and recognize printed figures; and (2) those who were not able under any conditions to read printed matter. In the former group there were 43 subjects; in the latter, 9 subjects, 6 of whom have been too blind to see printed matter since birth. These facts are tabulated in Table I.

Materials: Two types of card packs were used in these experiments. (1) One consisted of the regular ESP cards[1] (2) the second set of cards was similar to the first in that the same kind of cards and the same symbols were used; they differed in that each card had a blank card glued to the face of it, so that the symbol was inside, and this was sealed in an opaque brown envelope. Each envelope was taped around the outside with gummed paper to make the sealing secure. These cards were identified by a numeral printed on the back of the envelope in a code unknown to the subjects.

In order to acquaint the subjects with the symbols, five sample cards with raised symbols, made by pasting figures cut from cardboard on the cards, were presented to each subject.

Methods: For GESP (General Extra-Sensory Perception), the cards are shuffled and the pack placed face down screened from the subject. The experimenter removes the top card and looks at it while the subject attempts to call it. The card is then placed aside and the next card is treated the same way. This is continued for the 25 calls. This was found to be the most satisfactory method for introducing subjects to the experimental procedures because these subjects seemed to think it more "reasonable" to try to name the cards when the experimenter was looking at them. For GESP, as well as for all the methods used, the subject is not allowed to know his score or what cards he has called correctly until the end of a run of 25 calls.

A shift to BT ("Before Touching") was usually made near the end of the first experimental period with each subject. BT is similar to GESP except that the experimenter does not look at the cards until after the 25 calls have been made. That is, in GESP both clairvoyance and telepathy are possible; in BT telepathy is presumably ruled out.

Three matching techniques were used: (1) OM (Open Matching), (2) BM (Blind Matching), and (3) MP (Match Piling). All the matching techniques have one feature in common: The subject makes his guesses by distributing the twenty-five cards in five different packs. For OM and BM there are five key cards consisting of one each of the five symbols. These cards are arranged in a row, and the subject indicates the guess he makes for each card by placing it in a pile opposite one of the five key cards.

In OM the key cards are placed face up and the subject is told the order of the cards. The cards with raised symbols were used as key cards for some of the subjects, particularly the younger subjects who

[1] Illustrated in plate opposite page 1, *Journal of Parapsychology*, Vol. I, No. 1.

TABLE I

DATA OF INDIVIDUAL SUBJECTS.

Subject	Age	Degree of Blindness*	With Experimenter M. M. P.				With Experimenter M. H. P.			
			Runs	Dev.	Av.	C. R.	Runs	Dev.	Av.	C. R.
E. G.	10	p. b.	42	81	6.81	6.42	35	20	5.57	1.69
Ri. M.	13	p. b.	40	61	6.53	4.82	15	10	5.67	1.29
J. Ma.	15	p. b.	40	56	6.40	4.43	11	— 7	4.36	.60
E. R.	19	p. b.	58	66	6.13	4.33	22	10	5.45	1.07
B. J. M.	8	t. b.	13	29	7.23	4.02	15	20	6.33	2.24
A. H.	16	p. 3	41	51	6.24	3.98				
P. M.	13	p. b.	33	43	6.30	3.75	24	20	5.83	2.04
B. R.	15	p.	26	36	6.42	3.63				
C. C.	13	p.	15	28	6.86	3.61				
M. H.	21	p. t.	26	36	6.38	3.53				
G. C.	13	p.	14	26	6.86	3.47	12	18	6.50	2.60
D. Cr.	16	p. t. b.	25	35	6.04	2.85				
E. Goo.	11	t. b.	24	32	6.33	3.26	12	2	5.16	.29
Ray Mo.	25	p.	22	30	6.36	3.20				
Pat. N.	12	p. b.	48	44	5.91	3.18	21	4	5.19	.44
W. S.	15	p.	21	29	6.38	3.17				
B. R. M.	13	p. b.	21	29	6.38	3.17				
T. A.	15	p. b.	57	47	5.82	3.11	27	4	5.14	.38
J. H.	13	p.	47	38	5.81	3.04				
H. C. M.	11	p. t. b.	29	33	6.13	3.06	6	— 3	4.50	.61
J. K.	13	p. b.	22	28	6.27	2.99	23	7	5.30	.73
Da. Ne.	35	t. b.	35	34	5.97	2.87				
J. T. Mi.	18	p.	70	46	5.66	2.81	6	3	5.50	.61
H. R.	17	t.	40	35	5.88	2.77				
D. F.	15	t. 10	48	37	5.77	2.67				
I. H.	19	t. b.	22	25	6.14	2.67				
P. O.	19	p. b.	38	35	5.92	3.11				
J. V. H.	18	p. t. 8	27	26	5.96	2.50				
H. E.	10	p. b.	22	21	5.95	2.23				
Jo. H.	13	p.	11	14	6.27	2.11				
D. Con.	12	p. b.	60	27	5.45	1.74	26	0	5.	.00
A. Hod.	16	p. 3	3	5	6.67	1.45				
Cl. C.	14	p. b.	18	12	5.66	1.43				
D. Cobb	11	p. b.	6	7	6.16	1.42				
N. C.	15	p.	15	11	5.73	1.42	18	— 3	4.83	.35
J. D. H.	13	p.	15	11	5.73	1.42	12	— 3	4.75	.43
S. E.	15	p. b.	16	10	5.62	1.25				
H. Mi.	17	p.	32	14	5.43	1.24				
J. B. H.	10	p. t. b.	24	9	5.37	.92	6	3	5.50	.61
V. E.	15	p. b.	19	8	5.42	.92				
L. G. T.	13	p. b.	10	5	5.50	.79				
J. D.	13	p. b.	11	5	5.55	.75				
G. P.	23	p. t. 15	18	6	5.33	.71				
F. R.	12	t.	10	4	5.40	.63				
Wo. St.	19	p. t. b.	32	7	5.21	.62				
C. Y.	15	p. b.	11	4	5.36	.60				
W. S.	15	p.	26	4	5.15	.39				
R. C.	24	p.	26	3	5.11	.29				
W. L.	12	p. b.	27	3	5.11	.29				
D. P.	9	p. 2	35	3	5.09	.25				
R. B.		p. b.	15	2	5.13	.25				
J. E.	12	p. b.	8	1	5.12	.18	6	1	5.17	.20
G. B.	8	t. b.	15	1	5.08	.12				
E. P.	16	p. 5	27	11	5.40	1.16				

TABLE I—CONTINUED

Subject	Age	Degree of Blindness*	With Experimenter M. M. P.				With Experimenter M. H. P.			
			Runs	Dev.	Av.	C. R.	Runs	Dev.	Av.	C. R.
H. M.....	24	t. b.	8	0	5.00	.00	.....	.....	.....	.....
B. C.....	19	t.	29	— 4	4.86	.14	.....	.....	.....	.....
B. B.....	6	p. b.	8	— 1	4.87	.18	.....	.....	.....	.....
D. Ca.....	21	p. b.	51	— 7	4.86	.48	3	4	6.33	1.15
P. T.....	17	p. b.	9	— 3	4.67	.50	.....	.....	.....	.....
B. Geo....	16	p. 11	21	— 5	4.76	.54	.....	.....	.....	.....
C. B.....	25	t. b.	20	— 5	4.75	.56	12	— 3	4.75	.43
B. M.....	18	p. l.	17	— 5	4.70	.61	.....	.....	.....	.....
R. B.....	18	p. t. b.	8	— 4	4.50	.70	.....	.....	.....	.....
J. Hall....	18	p. t. b.	11	— 6	4.45	.90	.....	.....	.....	.....
T. E.....	19	p. b.	12	— 7	4.41	1.01	.....	.....	.....	.....
L. E. Hall.	20	t. b.	39	—15	4.62	1.20	.....	.....	.....	.....

* t = totally blind, pt = practically totally blind (i. e., too blind to see the symbols), b = this condition of blindness existed at birth, numerals = age at which blindness developed. (When no second designation appears data is absent from the records.)

were totally blind, so that they might touch the raised figures while the experimenter placed the cards opposite the key card indicated. This procedure was discontinued when the subject became thoroughly familiar with the order of the key cards.

BM differed from OM in that the key cards were placed face down, so that the subject attempted to match symbols whose order he had not been told and could not know himself through the senses, against key cards that were likewise unknown.

For MP no key cards were used. The subject distributed the pack of cards in five packs and then, after all the cards had been placed he "named" the piles; that is, he touched each pile and said "This is the circle pile, that is the star pile," etc.

Sealed card packs were used predominantly for all the methods except for GESP, which was done entirely with open card packs.

Procedures and Precautions: The atmosphere during the whole experiment was kept as informal as possible. The subjects viewed the experiment as a game in which each wished to do as well as he could. Candy rewards were occasionally given, but they were used infrequently so that they did not come to be goals in themselves. That is, the subjects did not as a rule expect or work for the reward. Competition was used when the subjects seemed to respond well to it. Boys were allowed to stand around while they were awaiting their turns and to talk among themselves as long as they did not interfere with the subject or seem to be having a disturbing effect on him.

With the conditions of freedom that prevailed, it was necessary to

have certain strictly enforced rules. (1) No person was allowed to stand beside or behind the experimenter. (2) For all GESP and BT work a stiff-backed notebook (9" x 12") was used as a screen behind and close to which the experimenter held the cards so that the cards were invisible to the subjects and spectators while the calls were being made. (3) The subject was kept at least four feet from the experimenter, usually seated at the opposite side of the table.

Additional precautions were taken. (1) New cards were frequently substituted for old decks for all work in which the subject handled open cards. The cards, which are stamped by hand with rubber stamps, do not have perceptible ridges caused by the ink or stamping. (2) For unscreened work (i.e., for conditions other than BT and GESP), sealed packs were used predominantly. This is particularly significant as a precaution because (a) the subjects were never told what symbol card was in any envelope; (b) the code was never revealed to any subject; (c) the code numbers were so small as to be indistinguishable by any subject; (d) three different packs of sealed cards with three different codes were used.

Two further checks on the results were attempted. (1) Two experimenters worked independently with some of the same selected subjects.[2] (2) A considerable part of the work was witnessed by teachers at the school or was observed and checked by a second experimenter. (See Table II.)

RESULTS

Occurrence of ESP Among Blind Subjects: The results of this experiment give considerable evidence that extra-sensory perception does occur among blind subjects. For a total of 2,024 runs there were 11,518 correct calls, or a positive deviation of 1,398. This gives an average of 5.69 hits per 25 calls from a mean expectation of 5. This deviation is 16.99 times the standard deviation. The odds against so large a deviation occurring by chance in this number of trials are 10^{54} to one. 28 of the entire 66 subjects had individually significant total scores; that is, had positive deviations which were at least 2.5 times the standard deviation. This significant scoring from 44 per cent of the subjects tried is a higher percentage than any previously reported.

Relation of History of Blindness to Success in ESP Scoring: Subjects were classified in three groups: (1) Those totally blind; (2) those

[2] It would be unwise to conclude that the failure of one experimenter to obtain successful results with a subject who had succeeded with another experimenter in any way invalidated the good results; but if two experimenters do succeed with the same subject, the case is strengthened.

too blind to see the symbols stamped on the ESP cards; and (3) those partially blind, but with sight enough to distinguish the symbols on the cards. These groups were further subdivided on the basis of the time of onset of blindness, i.e., whether the condition existed at birth or developed later.

Table II shows the number of independently significant subjects for each group, i.e., those whose results show a critical ratio of 2.5 or more.

TABLE II

DISTRIBUTION OF SIGNIFICANT SUBJECTS.

Degree of Blindness	Total number of subjects	Number of "significant" subjects	Per cent "significant"
(1) Totally blind	14	6	42%
(a) From birth	9	4	44%
(b) Later	5	2	40%
(2) Too blind to see cards	9	4	44%
(a) From birth	6	2	33%
(b) Later	3	2	66%
(3) Able to discriminate symbols	43	19	44%
Total	66	29	44%

The results show very little variation for the different extents of deficiency of sight or for the time of onset of blindness. The number of subjects in each group is too small, however, for final conclusions to be drawn.

Relation of ESP to age: In order to see if there were any optimum age level for extra-sensory perception, the subjects were tabulated according to age. The significant subjects were tabulated in the same manner (see Table III). The results show the distribution of significant subjects throughout the age group from 6 to 35. The small number of subjects used in the younger and older ranges makes the drop in percentage unreliable for any conclusion as to a real difference.

TABLE III

THE AGE RANGES OF INDIVIDUALLY SIGNIFICANT SUBJECTS.

Age	Number Subjects Tested	Number Individually "Significant Subjects"	Per Cent "Significant"
6 — 10	7	2	29%
11 — 15	28	15	54%
16 — 20	22	9	41%
21 — 35	9	3	33%

Sealed cards: Sealed card work was introduced as a further precaution against the possible effect of hyperacuity of touch which the blind

may develop. In order to prevent the subject's considering the precaution an indication of suspected dishonesty (which would have created an unfortunate experimental atmosphere), the sealed cards were presented in a particularly favorable light as their own special problem at which they should succeed exceptionally well. No suggestion of any difficulty of the method was made. Thereafter, the changes between open and sealed card packs were made without comment.

Compared to the average for methods using open cards, the averages for sealed cards were high. For runs with open cards there were 911 runs averaging 5.57 for sealed cards, 778 runs with an average of 5.96.

This superiority of sealed card scores *may be attributed* to three causes: (1) The favorable introduction given the procedure; (2) that the sealed cards were more easily manipulated by the blind subjects for matching techniques and hence produced less distraction than the open cards; and (3) that the envelope renders the cards psychologically less available; that is, psychologically more distant. It is speculative but interesting to point to the parallel between this and the effect of increased distance upon scoring reported by Rhine (1, p. 106).

Comparison of Results for Various Methods.

The comparison of methods which involved verbal response to the cards (BT and GESP) and those which involved non-verbal indications (the matching techniques) shows very little difference between the two. For the methods which required the verbal calling of the cards there were 1,093 runs with an average of 5.75. In the non-verbal group there were 596 runs with an average of 5.77.[3] Those subjects who had significant total scores, i.e., all those subjects whose deviations had an X-value of 2.5, showed a very slight difference favoring the matching techniques, with an average of 6.26 for matching techniques and 6.21 for calling techniques.

There were interesting individual differences in the relative success attained by subjects for the various matching techniques. These methods seemed to present different tasks psychologically to the various subjects. There seemed to be two fairly distinct groups: (1) Those whose scores were highest on MP and lowest on OM; and (2) those who had a reverse of this order with highest scores on OM, lower scores on BM and lowest scores on MP. Most subjects could be placed

[3] The discrepancy between the total average of 5.69 noted on page 148 and the average of 5.76 of these and following figures is due to the former being the average for all the work, and the later averages being based upon the totals for the comparison study only.

in one of these two groups. More subjects fell in group one than in group two as indicated by the totals: OM, 251 runs averaging 5.84; BM, 133 runs averaging 5.95; and MP, 65 runs averaging 6.21. In this data, all OM using open card packs have been excluded in order to make the conditions more strictly comparable. These averages are for methods using sealed card packs exclusively.

The significant subjects were evenly distributed in the two groups; i.e., about half were highest on OM and lowest on MP, and about half were lowest on OM and highest on MP. The fact of wide individual variations is, of course, obscured in the total score, so that OM and MP are both brought to approximately the same averages and BM, which for both groups had an intermediate average, is left with the highest total. The higher BM scores are particularly interesting in the light of the fact that exactly the opposite was noted by Woodruff and George (3, p. 24). Averages for this group of significant subjects were: OM, 149 runs averaging 6.10; MP, 43 runs averaging 6.23; BM, 85 runs averaging 6.45. A possible explanation for these differences is that the methods seemed to present psychologically to the subjects graduating steps of restraint—open matching being most restricted and match piling being the least restricted. Since open matching presents five cards face up which are to be matched by the sealed cards in the subject's hands the subjects often seemed to feel limited and distracted because they paid too much attention to the key cards. With BM, when the key cards are face down, this restriction seems partially to disappear. The cards are to be matched to cards which the subject cannot be sure he knows and he usually abandons the attempt to "figure them out." Matching piling is the least limiting method of all, since the subject himself piles them at will and designates later as he chooses.

In general these relations of the scores for the three methods were consistent for all of the subjects throughout the experiment.

Comparison of Results of the Subjects for Their Successive Experimental Periods. The subject decline curves noted by Rhine (1, p. 159), and the bearing any general rise or fall in scoring level would have upon method comparisons introduced at various times suggest a study of the trend of scoring for the successive experimental periods.

For this study, totals were obtained for each testing period; i.e., the scores made by each subject were totalled for each day that he was tested, and averages were obtained. The results show that the averages increase fairly steadily for the first five periods of work after which there was a sharp decline. A summary of this data is shown in Table IV.

It is possible that the rise may be due simply to increased experience with the task. The fact that, because of the number of subjects, less individual attention was given the experienced subjects toward the end of the experimental period might well account for a decline of interest and the resulting drop in scoring noted.

TABLE IV

COMPARISON OF AVERAGES FOR ALL SUBJECTS FOR THEIR SUCCESSIVE TESTING PERIODS.

Testing Period	Runs	Deviation	Average	X-Value
1st	402	+224	5.56	6.12
2d	373	+246	5.65	6.97
3d	305	+290	5.95	9.09
4th	242	+181	5.75	6.37
5th	154	+158	6.03	6.97
6th	86	+ 77	5.89	4.54
7th	70	+ 74	6.05	4.84
8th	40	+ 9	5.22	.78
9th	17	+ 25	6.47	3.32

When treated similarly, the results for the significant subjects show the same trend: Averages increase steadily, reaching a peak during the seventh period and then declining (See Table V.)

TABLE V

INDEPENDENTLY SIGNIFICANT SUBJECTS' AVERAGES BY EXPERIMENTAL PERIODS.

Period	Runs	Deviation	Average	Critical Ratio
1	176	+183	6.04	7.55
2	198	+242	6.22	9.42
3	166	+214	6.29	9.09
4	148	+151	6.02	6.79
5	104	+113	6.09	6.06
6	62	+ 83	6.34	4.39
7	54	+ 76	6.41	5.67
8	18	+ 1	5.06	....
9*	17	+ 25	6.47	3.32

*One subject.

DISCUSSION

Alternative Hypotheses:

(a) Sensory Cues: The fact that all these subjects had impaired vision, some to the extent of total blindness, in itself is a safeguard against visual cues. Additional precautions against any possible visual cues were taken, however. All the work done with the calling methods was done with the subject at a distance of four or more feet and with the cards screened from the subject and from any other person (with

the exception of another experimenter observing the testing) who might be present.

Auditory cues were made improbable both by the distance of the subject from the experimenter and by the presence of numerous other noises. For all the methods except GESP, which was used only during the first working period with each subject, the experimenter did not look at the card until after the subject had made his call.

The use of the sealed card packs served as a further guard against the subject's learning to identify the cards by cues obtained from the cards, because the subject had no opportunity to learn when a card was called correctly, since he was not allowed to observe the checking and since he could not discover through the senses the identity of the cards. The use of this pack was a precaution also against visual and tactual cues.

(b) Error, Deliberate or Unconscious: The precautions discussed were taken to guard against the possibility of the subject's being able to obtain knowledge of the cards through the senses, either by means of unconscious utilization of sensory cues or by deliberate attempt at deception.

The possibility of explaining the results obtained by unintentional inaccuracies or by deliberate falsification of the scores on the part of the experimenters was rendered unlikely by further methods of validation: (1) Significant results were obtained when both the experimenters were observing the subject's work; (2) significant results were obtained by the two experimenters working independently with the same subjects; and, (3) significant results were obtained when various teachers at the school observed the experiments. Table VI gives a summary of the results obtained under these conditions.

Blindness: Although the observation of 44 per cent of successful subjects among the population tested is markedly higher than any previously reported it is not possible yet, because of the absence of strictly comparable data, to draw conclusions as to the relative frequency of occurrence of ESP subjects among the blind and the normal population. The high proportion noted cannot be considered significantly unusual until further research has been done with non-blind groups of comparable age and grade and under similar social conditions. The great interest shown in the tests by the institutionalized children who found it a break in routine would be a factor towards higher scoring. The specific effect of the experimenter in the test situation, such as was noted by Sharp and Clark (2. p. 135), seems a matter worthy of further study.

TABLE VI
RESULTS OF TESTS WITH ONE AND WITH TWO OBSERVERS.

Second Observer	Runs	Deviation	Average	Critical Ratio
DOUBLY OBSERVED				
Miss P.	24	+ 6	5.25	0.61
Miss K.	43	+39	5.79	2.96
Miss S.	21	+50	7.38	5.44
Miss J.	12	+15	6.25	2.17
MHP*	23	+30	6.30	3.13
SINGLY OBSERVED				
Observer				
MMP	1,676	+1,414	5.84	17.27
MHP	335	+ 114	5.34	3.14

*The senior author.

Another suggestion toward the unreliability of the conclusion of exceptional ability among the blind is the fact that no relation was shown to degree of blindness. Partially blind subjects performed as well as those totally blind.

SUMMARY

This study is a preliminary survey of experiments in extra-sensory perception with blind subjects. The following general problems were considered: (1) Whether blind subjects were capable of extra-sensory perception; (2) whether this ability correlates with the degree of blindness; (3) whether there is any difference in ESP ability between subjects who have been blind from birth and those who have lost their sight later in life.

Twenty-eight of the sixty-six subjects tested had deviations with critical ratios of 2.5 or more. There seemed to be no correlation between ESP scoring ability and age, extent of blindness, nor with the time in which the subject became blind. Subjects who were able to score significantly were found at all ages from 8 to 35 years; some of the significant subjects were totally blind, some were only partially blind.

Averages for sealed card work were considerably higher than those for open card work; this difference may have been partly due to a more favorable manner of presentation for the sealed card work and to the greater ease with which the cards may be manipulated. Other possible explanations may arise after further investigation.

The scores show that the average scoring rate improved for the first five periods of testing (six periods for the group of significant subjects); during the fifth period the highest averages for the entire

experimental period was obtained after this. The averages for the following testing periods declined.

A later report will present further analysis of this data, together with the results of the continuation of this investigation along related lines. Some of the questions raised in this report will receive consideration at that time.

REFERENCES

1. RHINE, J. B., *Extra-Sensory Perception,* Boston: Bruce Humphries, Inc., 1934, pp. xiv + 169.
2. SHARP, V. AND CLARK, C. C, "Group Tests for Extra-Sensory Perception." *Journal of Parapsychology,* 1937, I, 123-142.
3. WOODRUFF, J. L. AND GEORGE, R. W., "Experiments in Extra-Sensory Perception." *Journal of Parapsychology,* 1937, I, 18-30.

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 total (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 ⅔ 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.*

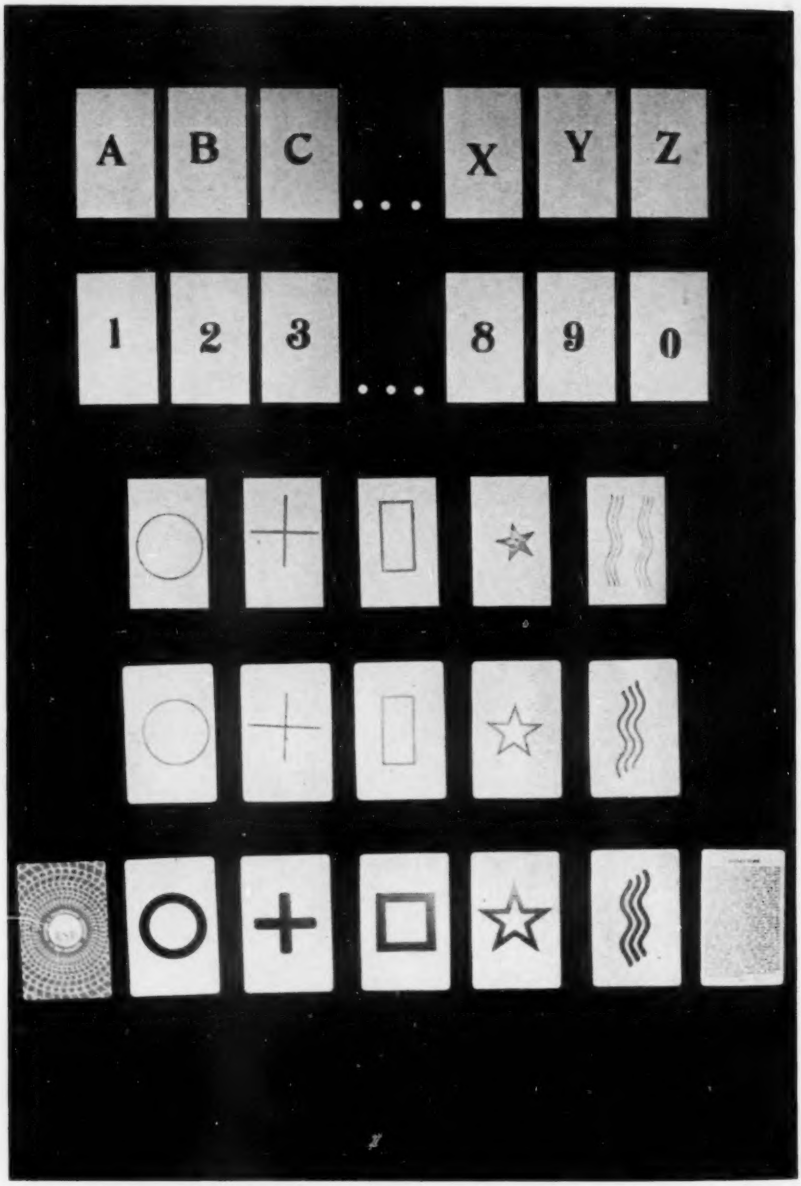


PLATE I.—Various cards used in ESP research at Duke University. *First and Second row:* Alphabet and number cards used in earliest experiments conducted in 1930; *Third row:* The Zener cards; *Fourth row:* ESP cards; *Fifth row:* New ESP cards, now published and generally available.

UNIV. OF MICH.