Volume 3 DECEMBER, 1939 Number 2

The Journal *of* Parapsychology

GARDNER MURPHY and BERNARD F. RIESS, *Editors*
ERNEST TAVES, *Managing Editor*

Contents

The Journal *of* Parapsychology

A SEMI-ANNUAL DEALING WITH TELEPATHY, CLAIRVOYANCE AND OTHER PARAPSYCHOLOGICAL PROBLEMS

The *Journal* is published in June and December. Contributions submitted for publication and all editorial communications should be addressed to Gardner Murphy, Columbia University, New York City, or to Bernard F. Riess, Hunter College, New York City. Correspondence with the editors is advised before submitting articles other than reports of experimentation. Since the *Journal* does not forward manuscripts by registered mail, it cannot guarantee that they will not be lost in transit, and contributors are urged to keep copies of their papers. All contributions should be typewritten, double spaced. References should be given in the form adopted by the *Journal.*

Reprints should be ordered when the proof is returned. Correspondence concerning subscriptions, change of address, back numbers, and other business communications should be addressed to Bernard F. Riess, Hunter College, New York City.

The subscription price is $3.00 a year; single current numbers $1.50. The rate for back volumes is $4.00; for single numbers of volumes 1 and 2, $1.00, for those subsequent to volume 2, $2.00. Missing numbers will be supplied free when lost in the mails if written notice is given within one month of the date of issue. All remittances should be made payable to Bernard F. Riess, Hunter College, New York City.

The Journal *of* Parapsychology

Volume 3 DECEMBER, 1939 Number 2

EDITORIAL

Among the various experimental studies in progress in relation to extra-sensory perception, it seems appropriate to select one extensive investigation from the Duke University Parapsychology Laboratory. Dr. Rhine and his associates feel that the investigation of Pratt and Woodruff reported here constitutes methodologically an important forward step, the rigor of controls being especially evident; and the reader will note that quite aside from offering evidence of ESP, these authors add significantly to the psychological analysis of the processes involved.

It is appropriate in the same issue to publish the current work of J. L. Kennedy who, as Fellow in Psychical Research at Stanford University, made a number of experimental and analytical attacks on the problem of ESP. The three papers from his hand published here are representative of the hesitancy of most psychological laboratories to accept the ESP hypothesis.

An important contribution to the methodological problem of setting up a *random* series of stimuli is contributed by J. H. Manley. The suggestion is timely, and the intellectual collaboration of physicists in our problem is warmly welcomed.

It was, in the past, the policy of this journal not to carry theoretical articles. Whether wisely or not, we do not know, the present editors are altering this policy and are publishing a theoretical discussion from a philosopher who is interested in the logical, physical, and biological questions which appear to be raised by the ESP problem.

As in the past, the Journal is delighted to carry a critical review by C. E. Stuart of the criticisms of ESP research.

The subsidy which previously supported this Journal is no longer available. We shall have to effect drastic economies; it is, for example, likely that in the future we shall have to resort to an offset method of publication. This would, however, probably permit increasing the number of pages while keeping within the income from subscriptions.

Partly because of the distance separating Durham and New York, and partly because of the necessary change in printing methods, we have severed the connection of the Journal with the Duke University Press. We want most gratefully to acknowledge the assistance given by the Press to parapsychological research and their kindness in assisting us with our immediate problems.

The most important contribution, after the research itself, made to the Journal is, at present, the indefatigable labor of its Board of Review, a group of professional psychologists who, under the chairmanship of Dr. S. B. Sells, carefully read and publicly criticize each piece of experimental research reported in the Journal. This has greatly improved the articles submitted to the Board. It has permitted rejoinder by authors, and the presentation to the reader of at least two views regarding major controversial issues. At the same time it has been enormously expensive of time. The Committee began its work on this issue in mid-November. Dr. Sells transmitted copies of papers to the various members of the Board, scattered from coast to coast, collated and integrated the various comments, and sent copies of the criticisms to the authors, who made use, in so far as they wished, of each suggestion. While new copy was being sent to the Committee, we were reading revised manuscripts from authors. We believe that quality has been improved.

But we are profoundly apologetic for the resulting delay in the appearance of the Journal, and will make strenuous effort to bring out the next issue in June, as scheduled.

One general word of editorial comment seems appropriate at a period when so much ESP work is being done, and opinions are still far from crystallized. The doors are open here to all work that seems genuine and serious, regardless of the degree of resemblance to the editorial opinions, or its "right" or "left" deviation from any half-articulate formulation of what "current science" believes. To quote the classical phrase used by the Society for Psychical Research, "The responsibility for both the facts and the reasonings in papers published in the Proceedings rests entirely with the authors."

G. M.

B. F. R.

SIZE OF STIMULUS SYMBOLS IN EXTRA-SENSORY PERCEPTION

J. G. PRATT and J. L. WOODRUFF
Duke University

Abstract: An investigation of four problems is reported: (1) Does ESP occur? (2) If so, what is the relation between level of scoring and size of symbols? (3) What is the effect of experience in formal ESP tests on rate of scoring? (4) What is the relation of "newness" of stimulus material to rate of scoring?

The entire research involved the participation of 66 subjects who were tested to the extent of 3,868 runs of 25 trials each with ESP cards. The total number of hits scored was 970 beyond mean chance expectation, an average of 5.25 hits per run, which gives a critical ratio (C.R.) of 7.80. These are total results from two series which are distinguished on the basis of differences in experimental conditions.

In one of these, Series B, two experimenters were present at every test and certain special safeguards against error were used here for the first time. In this series, 32 subjects made 2,400 runs with a positive deviation of 489 hits. The C.R. is 4.99.

No significant differences in scoring rates are found in relation to symbol sizes in the experiment as a whole. No direct relation is found between the experience of subjects and the rate of scoring.

The use of "new" material is found to give scores which are significantly higher than those obtained with "old" material. "Experienced" subjects scored as well with "new" material as "inexperienced" subjects. A decline in the effectiveness of "new" material with successive sessions of its use is noted. The advantage in favor of higher scores with "new" material was greater when the subjects knew what symbol size was being used.

INTRODUCTION

Background of the Research

Any research based on the hypothesis of ESP involves, in a sense, the problem of the re-testing of that hypothesis. Without adequate evidence that the phenomenon itself is present, any problem concerned with the nature of ESP has little chance of solution. In such a sense, this research was again a test of the primary hypothesis. At the same time, however, the major goal of the research at the point of its inception was to ascertain whether there is any relation between the size of symbols used as stimuli and the level of scoring in ESP tests.

The quantitative investigations reported and referred to in the pages of this *Journal* have been mainly those in which the well-known ESP symbols have been used. Because of the fact that these have varied little in shape and size, relatively little direct insight has been achieved concerning the role of the stimulus in ESP. However, some investigators have attempted to get at this question directly by making systematic changes in the testing materials. Carpenter and Phalen (1) found that their subjects could score as well with colors as with the ESP symbols. MacFarland and George (7) found no difference in success between the use of regular and of distorted symbols, with the notable exception of the results of one of the investigators who acted also as one of the subjects. He scored above chance on the regular symbols and below on the distorted—the effect which he had anticipated would be found. Murphy and Taves (8) used playing cards and special decks, some made up entirely of circles and blank cards, and others of circles and crosses, in addition to the usual decks of twenty-five ESP symbols, and found a tendency for the scores in various materials to vary together.

L. E. Rhine (12) varied the ESP symbols used, both as to size and as to the number of copies of each presented at each trial. In one series, she used symbols of 3½, 1⅝, ⅛, and 1/32 inches in diameter. (The measurements were all made upon the circle and the other symbols were of a proportionate size.) In another series, she compared the results from large cards stamped with a single symbol (1⅝ inches) and from cards of the same size upon which several copies of the same symbol, each 1⅝ inches in diameter, were stamped. She concludes that "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. Variations in number of stimuli presented 5 to 1 at a given time did not result in a significant preference."

The present research was concerned with making further systematic tests to determine the relation of symbol size to ESP scoring. On the basis of the above-mentioned studies, there would appear to be no reason to expect differences in scoring with variations in size or shape of ESP symbols unless such differences arise from the personal preferences of the individual subject. The importance of the hypothesis suggested—that ESP is not affected by the physical characteristics of the stimulus—would require prolonged research before such a statement could be advanced as a definite conclusion. Accordingly, a more extensive investigation of ESP in relation to symbol size appeared to be fully warranted.

Two further problems arose during the course of the investigation and were considered as fully as the general plan and scope of the research permitted. One was concerned with the possible relation of the amount of experience of subjects to scoring rate. The second involved the question of a relation between the amount of experience with a particular stimulus material and the rate of scoring. These problems seem especially apropos in view of the widespread opinion among experimenters that successful subjects decline in score averages after a period of some success. This has been noted by J. B. Rhine, (11), Pratt (9), L. E. Rhine (12), Price and Pegram (10), Gibson (3), and Riess (13), and is apparent in still earlier reports dealing with this field of research.

Restatement of the Problems

Four important problems are therefore considered in this report. They are: (1) Judging from the results of this investigation alone, does ESP occur? (2) Assuming the function of ESP, is there any relation between symbol size and the rate of scoring? (3) What is the relation between the amount of "experience" of subjects and the rate of scoring? (4) What is the relation of "newness" of material to rate of scoring?

EXPERIMENTAL CONDITIONS AND PROCEDURES

On the basis of differences in procedure, the work may be divided into two main series, both of which dealt entirely with the ESP of objects (clairvoyance). Series A was done in the period from March, 1938, to August, 1938. During this period the experimental set-up required the direct participation in the test of only one experimenter. This series was conducted by one of the writers (Woodruff) with only occasional introduction of other investigators to witness the procedure. Series B was done in the period from October, 1938, to March, 1939, and required the simultaneous participation throughout of both of the writers as experimenters. Important differences in the experimental conditions and procedures of the two series make it necessary to describe the two separately and to consider how the results of each bear upon the primary problem of the occurrence of ESP.

Series A

Subjects. Forty-two persons were tested in Series A. This number includes 14 members of Oxford (N. C.) Orphanage of high school age, 21 undergraduate students of Duke University, and 7 others ranging (in age) from adolescence to middle age.

Size of Stimulus Symbols. Throughout both series all the tests were conducted with the five ESP symbols in the usual balanced pack of 25 cards. In Series A, three sizes of symbols were used—the regular 1½ inch printed ESP symbols, ¼ inch symbols drawn in ink, and symbols not over 1/16 of an inch, drawn in ink. The smallest size required moderately close scrutiny to decipher. The measurements in each case are approximations and are given for the diameter of the circles. The size of the cards was in all instances that of the standard playing card, with a single symbol appearing on each card face. During all tests, each run of 25 cards consisted of symbols of one size. The 1½ inch symbol was uniform in design and was printed by a commercial process in black ink. The two smaller sizes were freehand drawings made with a fountain pen in dark blue ink.

Experimental Set-up. The card matching procedure known as the STM (screened touch matching) technique was used throughout Series A. The screen, which shielded the experimenter and the deck of stimulus cards from the subject, consisted of a piece of plywood 18 inches high by 24 inches wide, held in a vertical position by means of wooden supports. It rested on the table between the subject and the experimenter, who sat opposite each other. When the experimenter and the subject were seated normally with the screen in position, each could see the top of the head of the other. Between the bottom edge of the screen and the table top was an aperture two inches high and eighteen inches long. In this opening, five ESP symbols were located in a row in such a way that they were visible to both subject and experimenter. These five key cards were chosen from the regular brown-back ESP pack which has each symbol in a different color, in order that the key cards might not be identical with any of the symbols with which the subject was to be tested. On the experimenter's side of the screen, 3½ inches back of the aperture was a low vertical screen 3 inches high and 23 inches long. Its position in relation to the aperture was such that the subject could not possibly see the cards held by the experimenter.

Procedure. With the subject and experimenter both seated and with the screen and the key cards in position, the experimenter shuffled and cut a pack of 25 cards behind the screen, out of sight of the subject. With the pack face down in his hand in readiness for dealing, the experimenter then signalled to the subject to begin by saying "all right." The subject designated his choices (as to which symbol he thought was on the top card of the deck held by the experimenter) by touching, usually with the eraser end of a pencil, one of the five key cards lying in the aperture. This response of the subject was visible through the one-way aperture to the experimenter, who immediately placed the top card of the deck opposite the designated position but behind the second screen. Without waiting for a further signal, the subject then proceeded to touch the key card which he felt corresponded with the second card in the pack (by then the top one). The experimenter laid this card opposite the key card touched. This procedure continued until the 25 cards in the deck had been guessed.

As soon as the experimenter announced the end of the run, the subject removed the larger screen from the table. The key cards remained in position on the table. The experimenter picked up the pile of cards opposite the first key card, turned it over so that the symbols were facing upward and, while the subject watched the cards, laid those symbols which were "hits" nearer to the key cards and discarded the "misses"—at the same time counting aloud the number of hits. This procedure was followed for all five piles with a cumulative audible count of the score for the entire run. Following this, the cards segregated as hits were again examined and counted and the score was recorded by the experimenter in full view of the subject. The subject then replaced the screen and the experimenter shuffled and cut the cards preparatory to the next run. The number of runs done during a session with each subject varied somewhat with the rapidity of the subject's matching and the time at the disposal of the subject and the experimenter.

Methods of Selecting the Symbol Sizes to Be Used for Each Run. During the first part of Series A, only two symbol sizes were used—the regular 1½ inch printed symbol and the ¼ inch symbol. These sizes were alternated regularly from run to run so that the subject knew which symbol size was being used at any particular time. Likewise when the 1/16 inch symbols were introduced about midway through Series A and all three sizes were used, regular rotation among the three was followed from run to run.

However, for a short time in Series A, a variation in the alternating method of symbol selection was used. The experimenter, attempting to follow a random order in his selection, chose subjectively the cards (symbol sizes) for each run without letting the subject know until the check-up what size was being used. The experimenter restricted his choices in such a way that an equal number of runs with each size was made in each session. However, as the subject did not know exactly when the experimental period was to end, he had no dependable way of knowing which particular set of cards would constitute the last run for the day and he was therefore limited in his ability to infer the size of symbols to be used.

During a part of Series A, the tests were conducted with a third person present to witness the entire procedure. A second member of the Parapsychology Laboratory staff was present for 169 runs, either J. G. Pratt, Miss Margaret Price, or B. M. Smith. In addition, 63 runs were casually witnessed by other subjects.

Series B

Subjects. Twenty-four undergraduate college women of Duke University and 8 other adults participated as subjects in Series B. Of these subjects, 8 had participated in Series A.

Stimulus Symbols. During Series B, four different symbol sizes were used, the three already described for Series A and a still larger size with the circle 2¼ inches in diameter. The 2¼ inch symbols were drawn by hand with a broad-pointed pen, using india ink. The characteristics of the other sizes were the same as for Series A. The size of cards in all cases was again that of a standard playing card.

Experimental Set-up and Procedure. The STM condition as used in Series B was modified in several respects intended to safeguard the procedure against possible weaknesses present in Series A.

a. The Screen. The large screen used in Series B had the same dimensions as the former one except that the aperture was 20 inches long. However, it differed from that of Series A in two respects. A small shield 5 inches in width on the experimenter's side of the screen slanted up from the table at a forty-five degree angle from a point two inches back of the opening (Fig. 1). This sloping shield permitted the experimenter who handled the cards to see the subject's choices with greater ease and at the same time it was effective to prevent cues of a visual kind reaching the subject through the aperture without requiring the use of the small secondary screen (*cf.* p. 124). The shield was attached permanently to the screen and had the additional function of serving as a rest when the screen was turned on its side on top of the table for the check-up at the end of each run. The second new feature of the screen was a horizontal row of five wooden pegs which were placed about 4 inches above the top of the aperture and on the subject's side at intervals of about 2½ inches. The key cards, which again had the colored symbols, were each punched with a small hole near each end, by means of which they were hung on the pegs. The row of pegs permitted the use of the key cards in an order unknown to the experimenter who

Fig. 1. Side view of the experimental table.

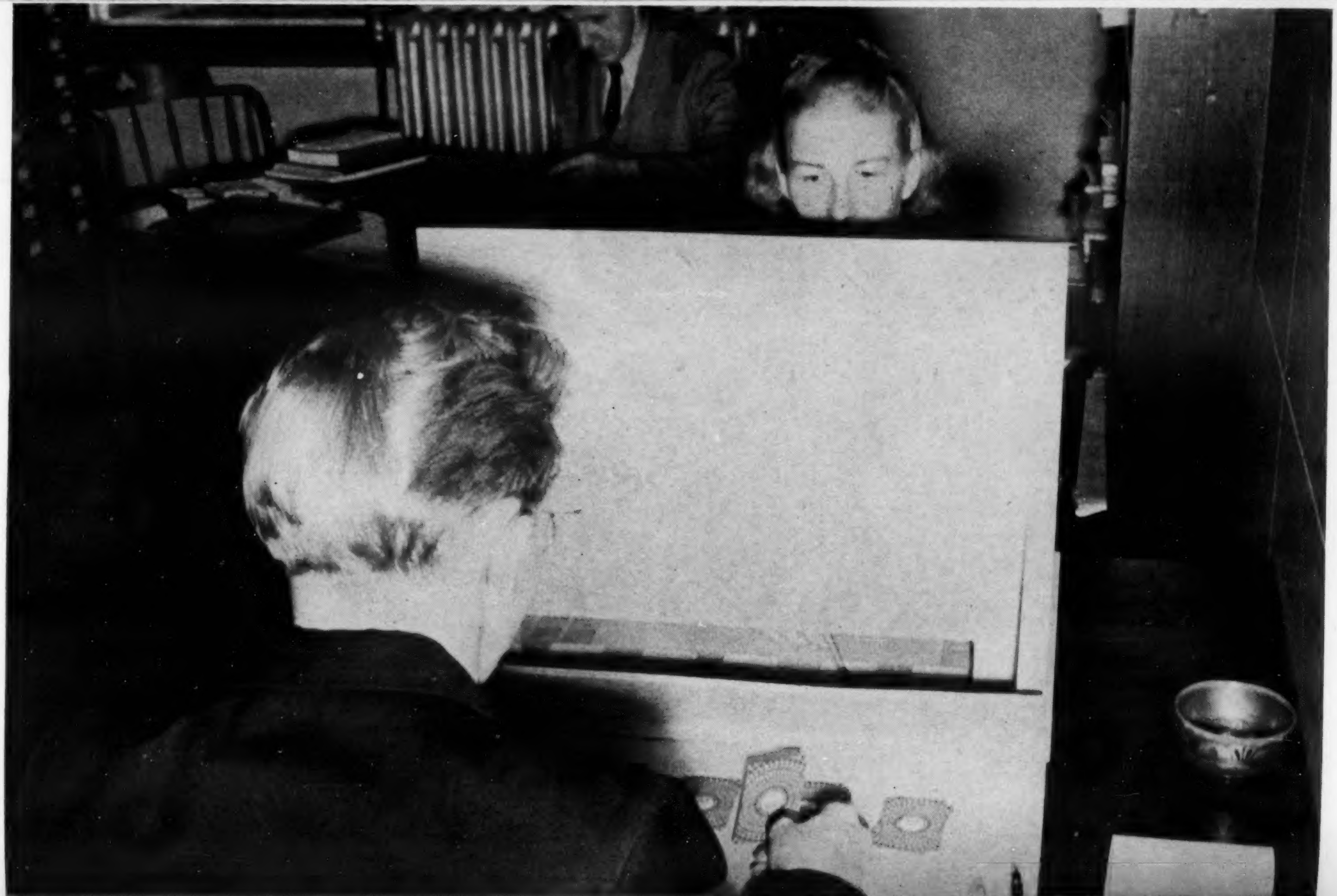


Figure 2. View of the experimental set-up from above and back of Woodruff.

Figure 2. View of the experimental set-up from above and back of Woodruff.

Figure 3. View of the experimental set-up from above and back of Pratt.

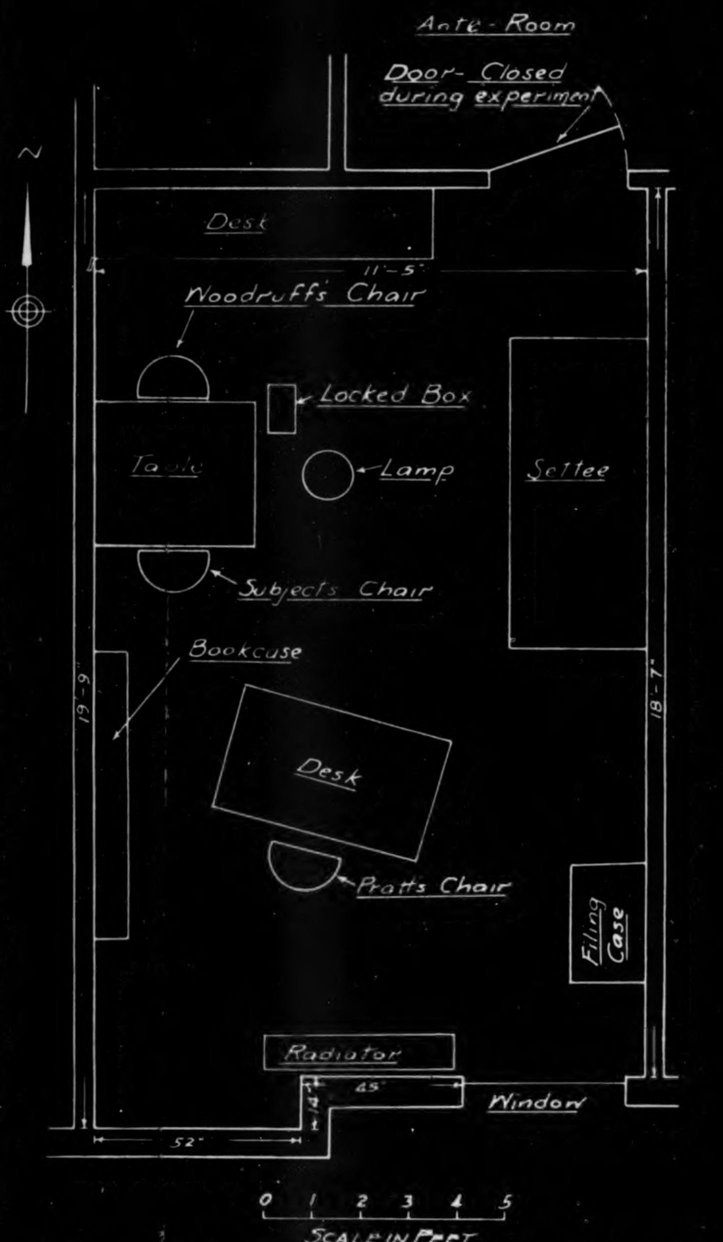


FIGURE 4. FLOOR PLAN OF THE EXPERIMENTAL ROOM

E.P.G.

handled the cards, as the key cards were arranged in a new order before each run and after the screen had been put in position. Five blank cards were placed in the aperture in position directly under the five key cards to facilitate the experimenter's dealing of the symbols in keeping with the subject's pointing.

b. Serially-numbered record sheets. Series B was broken into six sub-series involving the use of different combinations of symbol sizes, to be described shortly. The length of each sub-series was determined in advance. As a preparation for each sub-series, the experimenters wrote out a description of the length and general purpose of said series, keeping one copy for themselves and depositing another with one of the Laboratory secretaries.[1] Each experimenter was thereupon provided with exactly as many data sheets as the number of the projected runs. Each sheet had a serial number and a seal for identification purposes which was made with a special stamp available only to the secretary. The serial numbers on the record sheets of one experimenter were duplicated exactly by those of the other. The specific numbers assigned for the runs of this particular experiment were not used on any other record sheets issued in the Parapsychology Laboratory up to the time of making this report. Each experimenter was careful to use his record sheets in correct serial order. The purposes of recording required that each experimenter use one sheet for each run. Each run thus received a distinctive number at the time it was made.

c. Two experimenters: Roles during the run. The actual testing procedure for each run may be described as follows: One experimenter, Woodruff, and the subject sat facing each other across a table as in Series A. The second experimenter, Pratt, sat about six feet from and almost directly behind the subject (see Fig. 4). The screen was placed in position. While Woodruff shuffled and cut the pack of cards to be used, Pratt took the key cards from the pegs and handed them to the subject who changed their order and replaced them without giving Woodruff any indication of the new arrangement. In the last sub-series, Pratt re-arranged the key cards and put them on the pegs himself; during that period the experimenters were careful that the shuffling and cutting by Woodruff were not completed until Pratt had returned to his usual position, so that there would be no possibility of his seeing any of the cards held by

[1] The writers wish to express their thanks to Mr. E. P. Gibson for his assistance in preparing for the experiment and for his independent re-checking of all the results.

Woodruff after they were shuffled. Woodruff then gave the signal to the subject to start, and the subject proceeded to indicate his "guesses" by pointing to the blank cards in the opening under the screen. Woodruff distributed the cards, following the subject's pointer, but he was in complete ignorance throughout the run of the symbol designation intended by the subject.

d. Recording. At the end of the run, the screen was left in position on the table while Woodruff recorded the actual distribution of the 25 cards on the appropriate record sheets, and while Pratt recorded the order of the key cards on his record sheet bearing the same number. The order of key cards was recorded by Pratt in reverse order so as to make them correspond with Woodruff's record when the two sheets were juxtaposed later for checking. Pratt in addition recorded the name of the subject, the type of test, the date, and the initials of the experimenters. This recording was done without any communication between the experimenters or from the subject.

e. Locked record box. When Pratt finished his record, he carried it to the experimental table. Woodruff had usually finished his recording by this time. In case he had not, Pratt was careful to keep his record out of Woodruff's visual field until the other record was completed. Woodruff then clipped together the two independent records with the common serial number and deposited them without further marks or observation of the sheets themselves in a special locked box provided by the secretary for the purpose.

f. Counting. The screen with the key cards still on the pegs was then laid on its side, either by the subject or by Pratt, so that both the key cards and the 25 cards as distributed were visible to all three persons. Pratt then proceeded to sort out the hits from each pile, laying them nearer the key cards and counting aloud the number of hits for the run. This process was observed by Woodruff and the subject. The hits as segregated were then re-examined and re-counted. The score for the run as thus determined at the time from the cards themselves was recorded immediately by each experimenter in his personal record book.

To continue the test, the screen was again raised to its vertical position, the key cards were re-arranged upon the pegs, Pratt returned to his seat behind the desk, and Woodruff, having shuffled and cut the pack of cards, gave the subject the signal to begin.

g. Experimental periods and rates of performance. As in Series A, the length of each experimental session was not fixed but was adjusted to suit the convenience of the subjects. The speed of work varied somewhat from subject to subject, but for the average the number of runs performed within an hour was about thirty. Usually the subjects worked by appointment for from thirty to forty-five minutes at a session.

h. Obtaining subjects: Degree of selection. The first subjects to be used in Series B were those of Series A with whom the experimenter was able to make contact and who were interested in continuing the investigation. From time to time these subjects suggested the names of interested friends and in this way a considerable number of new subjects were brought in. No particular effort was made to select subjects on the basis of their performance or excellence in the tests. In general, however, those who did better were more interested to continue and were encouraged to do so. As far as possible, each subject worked one period each week.

i. Checking the scores from the record sheets. The record sheets were checked entirely independently by the secretarial assistant. After he had obtained the scores by juxtaposing the key cards (as recorded by Pratt) with Woodruff's record of the symbol distribution, a run by run comparison of the scores as recorded by the two experimenters at the end of the run and as found from this re-check of the record sheets was made. In case of a discrepancy, the written records were consulted again immediately to see whether the difference could be accounted for. If the difference were evidently an error from checking the record sheets, the secretary's score was adjusted to correspond to that of the experimenters. If, on the other hand, the re-check of the record sheets did not account for the difference, the lower score was accepted as the official one and was entered thereafter in all computations from the data.

Selection of Symbol Size to Be Used. During the first sub-series of 300 runs, the 1/16 inch, the ¼ inch, and the 1½ inch symbol sizes were selected subjectively by Woodruff in an attempt to approximate a random order, as for part of Series A. In the next 600 runs, the same symbol sizes were used, but the choice of the size for each run was determined by the cast of a die. Until this point in Series B, the subjects did not know until the end of the run what size of symbol was being used at any time.

In the next 300 runs only the regular 1½ inch symbols were used. During the last 1,200 runs, the 1½ inch and the 2¼ inch symbols were used alternately. During this period Woodruff was careful to inform the subjects which size was being used before they began their responses for each run.

Special Points of Procedure: a. Knife-cutting. During the last 830 runs of the experiment, the cards were shuffled by Pratt and were cut by Woodruff by means of a paper knife. The object of this variation in procedure was to determine whether extra-chance scoring might depend upon inadequate shuffling or upon peculiarities in the cards which make them cut by hand in a non-chance manner.

b. Shuffling methods. Woodruff used the method of shuffling in which the pack is held in one hand while cards are slipped out of it and re-inserted into the pack with the other. Pratt, on the other hand, divided the pack somewhere near the center and riffled the two halves together in a manner which, superficially, would appear to give a more adequate mixing of the cards. This was repeated five times for each deck of cards. Cutting with a paper knife theoretically permits of no effect of warping in favoring a division of the deck at particular points more than others.

c. Inverted keys: Blind STM. In the last sub-series of 400 runs, the key cards were placed facing inward upon the pegs so that the backs only were visible to the subject. During this series, the rearrangement of the key cards was always done by Pratt. The subject did not know the order of key cards, unless some of the symbols were recognized by cues from the backs of the cards. The purposes of the experimenters in making this innovation were, first, to introduce a novelty into the situation which might add to the interest of some of the subjects, and second, to provide an easy step toward a new experiment beyond the ones here reported which, it was feared, the subjects might consider to be too difficult without some transition.

Summary of Procedure: Series B. To help the reader fix in mind the experimental procedure for Series B, the essential steps of the plan of investigation may be reviewed: (1) Both experimenters were present during all the tests, each with a definite, pre-assigned role to facilitate the procedure and to safeguard against experimental error. (2) The subjects were tested for their ability to guess cards completely screened from sight and handled entirely by Woodruff. (3) The subject indicated his guesses by pointing in relation to five key symbols which were out of Woodruff's sight and unknown to him until after he had recorded the 25 cards as distributed at the end of the run. (4) Meanwhile, Pratt, without seeing Woodruff's cards, recorded the key cards and other essential data about the run.

(5) The two separate records were deposited at once in a locked box to be scored later by a third person. The record sheets were serially numbered in pairs and designated for the purpose of the investigation so that every run had to be clearly accounted for. (6) The two experimenters jointly checked each score from the cards and each entered the number of hits observed in his personal record of the run scores as obtained at the end of each run. (7) The laboratory secretary independently checked the scores from the record sheets. His scores were compared with those of the experimenter and in case of a discrepancy not immediately accounted for, the lower score was adopted.

EVALUATIVE PROCEDURES

The statistical methods used in the evaluation of the results are, in the main, standard procedures. The critical ratio method, the chi-square method, and the method for the evaluation of a difference are conveniently described with specific reference to the data of ESP research by Greenwood and Stuart (5). On the strength of Greenwood's empirical findings (4), the results were evaluated on the binomial hypothesis.

Methods of correcting a P-value derived by the critical ratio method for the possible factor of optional stopping—i.e., taking advantage of the trend of scoring throughout the experimental series to stop the tests when the total results "favor" a particular interpretation—have been devised by both T. N. E. Greville and J. A. Greenwood. A description of the latter's method, which has been applied in the present study, is awaiting publication, and a full explanation cannot be undertaken here. In general terms, the applicability of the method is based upon the assumptions that the total experimental series to be evaluated consists of sub-series having two characteristics: (1) a stated maximum number of such sub-series beyond which the experiment would not go; and (2) a fixed length for each sub-series. It is important that neither of these characteristics be influenced by the results, a requirement which is most clearly met if they are determined before the experiment is started. This is not to say, however, that otherwise the essential conditions for applying the optional stopping method are necessarily lacking. If it can be established that the maximum number of sub-series and the length of each one were not affected by the preceding scores in the experiment, that is all that is required to make the optional stopping method applicable. When the probability of chance occurrence for the results from the beginning of the experiment to the end of a particular sub-series is obtained, the optional stopping correction converts this value into the probability of a chance occurrence of the same deviation ratio at the end of any one of the possible stopping points (end of each sub-series). The application of this method to the results of Series B is discussed later.

DEFINITIONS

For purposes of the presentation of the results and discussion in later sections of the paper, the following definitions are given.

Those subjects will be designated as "experienced" subjects for a given series (A or B) who had participated previously in formal ESP tests, irrespective of the size of symbols used, either with the writers or with any other ESP investigator. All other subjects in each series will be "inexperienced." Thus an "inexperienced" subject in Series A was, if he continued through Series B, "experienced" in the latter. A particular subject is considered as maintaining throughout a given series the status of "inexperienced" or "experienced" with which he began that series.

"New" materials for a given subject in a given series are those symbol sizes with which he had not been tested previous to the series in question. Other material will be designated as "old" material. Thus "new" material in Series A becomes "old" material when used by the same subject in Series B.

RESULTS

I. AS EVIDENCE OF ESP

The Evaluation of the Results in Relation to the Hypothesis of Chance Coincidences

The Experiment as a Whole. This report deals with a total of 3,868 runs with ESP cards, or 96,700 trials, each trial with a probability of success of 1/5. The successes, or hits, observed were 20,310. This number represents a deviation from mean chance expectation of 970 hits, or an average of 5.25 hits per run. The standard deviation (S. D.) of expected hits for this number of trials is 124.38, and the critical ratio (C. R.) of the result is 7.80.

Series A. Forty-two subjects participated in Series A with a total of 1,468 runs in which they scored 7,821 hits. This is 481 hits in excess of mean chance expectation or an average number per run of 5.33. The S. D. for this number of runs is 76.63. This gives a C. R. of 6.28 with an equivalent probability value (P) of 10^{-10}. The total of 169 runs witnessed by another staff member in addition to

Woodruff gave a positive deviation of 73, with an average of 5.43 and a C. R. of 2.81.

TABLE I

SUMMARY AND COMPARISON OF GENERAL EXPERIMENTAL RESULTS OF SERIES A AND SERIES B AND THE RESULTS OF THE CROSS-CHECK ON SERIES B

Series	Runs and Dev.	Av. Hits per Run	S.D.	C.R.
A	1,468 + 481	5.33	76.63	6.28
B	2,400 + 489	5.20	97.98	4.99
Total	3,868 + 970	5.25	124.38	7.80
	C.R. of diff. = 2.00			
Witnessed Tests in Series A	169 + 73	5.43	26.00	2.81
Cross-Check Series B	2,400 + 56	5.02	97.98	.57

Series B. Series B consisted of a total of 2,400 runs, or 60,000 trials, of which 12,489 were hits. This is a deviation of 489 in excess of mean chance expectation, or an average of 5.20 hits per run. The S. D. is 97.98, which gives a C. R. of 4.99 with the associated probability of 3×10^{-7}. An analysis was made of the 2,400 runs of this series by the chi-square method. This analysis, based upon the frequency of run scores for all the subjects as shown in Table II, gave a chi-square of 34.30, with 9 degrees of freedom and a probability of .000,078. Thus the deviation ratio method of evaluation and the method of chi-square both support the conclusions that results reliably different from chance expectation were obtained in Series B.

It is evident from the general summary in Table I that the results of the research as a whole and of the two series taken individually can not reasonably be attributed to chance factors. Because of the more elaborate precautions against sensory cues and experimental error which were taken in Series B, more interest attaches to the results of this series as regards the question of the interpretation of the deviations. It seems important to reach some kind of conclusion as to whether these results were due to ESP before proceeding to a consideration of further problems dealing with the nature of ESP.

The Results in Relation to the Problem of the Occurrence of ESP

Series B. Series B was planned shortly after the symposium on experimental methods in ESP research at the Columbus, Ohio, meeting of the A.P.A., in September, 1938. In planning their research, the investigators made every effort to take fully into account all the criticisms of methods made at the symposium, as well as those in

TABLE II

FREQUENCY DISTRIBUTIONS OF RUN SCORES FOR SERIES B BY INDIVIDUAL SUBJECTS AND WITH THE TOTAL DISTRIBUTIONS FOR THE EXPERIMENTAL SERIES AND THE CROSS-CHECK

	Frequency of Run Scores															*Runs and Dev.*
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Subjects																
H.G.	1	1	8	18	37	36	34	27	13	7	4	0	1	0	0	187: + 76
M.B.	1	1	11	26	27	30	28	14	13	3	2	2	0	0	0	158: + 8
B.Y.	1	1	6	12	18	14	7	5	8	4	2	0	0	0	0	78: — 2
J.Bd.	0	1	3	7	13	13	9	9	2	1	1	0	0	0	0	59: + 2
D.L.	0	1	11	10	17	29	19	14	8	4	4	1	1	0	0	119: + 46
O.M.	0	0	1	2	7	3	0	0	0	2	0	0	0	0	0	15: — 6
A.B.	0	7	6	13	23	21	17	17	4	2	0	0	1	0	0	111: — 17
A.M.	2	9	18	45	47	68	41	34	28	14	4	0	3	0	0	313: + 53
B.M.	1	2	8	11	22	20	17	11	12	3	1	1	0	0	0	109: + 17
L.D.	0	1	7	11	19	12	10	13	3	5	0	0	1	0	0	82: + 6
J.Br.	0	0	0	3	0	0	1	2	0	0	0	0	0	0	0	6: — 1
M.E.	0	1	0	0	5	3	1	3	2	0	0	0	0	0	0	15: + 4
R.K.	0	2	0	6	6	6	5	4	3	2	0	0	0	0	0	34: + 4
C.W.	2	6	5	9	13	14	14	10	3	2	1	0	0	0	0	79: — 24
E.G.	0	1	1	2	0	1	3	1	1	0	0	0	0	0	0	10: — 3
P.M.	0	2	7	18	25	26	25	24	9	10	11	4	1	0	0	162: + 136
B.Br.	0	1	2	9	10	7	8	9	3	0	0	0	0	0	0	49: — 3
C.H.	0	0	2	0	4	2	2	4	1	2	0	0	0	0	0	17: + 11
J.A.	0	2	0	5	3	10	3	2	3	2	0	0	0	0	0	30: + 3
B.J.	0	0	2	2	5	8	5	3	1	2	0	0	0	0	0	28: + 7
D.S.	0	0	1	1	4	0	3	3	1	0	0	0	0	0	0	13: + 3
J.B.	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	2: — 3
G.E.	0	1	2	4	6	7	7	3	0	0	0	0	0	0	0	30: — 11
B.B.	1	4	7	17	32	33	24	22	9	4	1	1	0	0	1	156: + 23
C.C.	0	3	10	23	34	38	33	25	11	8	5	4	0	0	0	194: + 75
T.E.	0	0	0	1	0	4	3	1	0	1	0	0	0	0	0	10: + 7
D.A.	0	0	2	6	2	12	10	9	3	4	2	0	0	0	0	50: + 43
M.W.	1	0	2	9	6	10	13	7	1	0	0	1	0	0	0	50: + 1
N.A.	0	0	8	17	24	24	17	20	10	3	2	1	0	0	0	126: + 33
N.S.	1	2	6	8	9	12	8	5	9	5	0	1	0	0	0	66: + 15
D.C.	0	0	0	4	4	6	6	0	1	1	0	0	0	0	0	22: + 1
C.K.	0	2	2	1	9	2	1	1	1	1	0	0	0	0	0	20: — 15
Total	11	51	138	301	432	471	374	302	163	92	40	16	8	0	1	2,400: + 489
$X^2 = 34.30$; P = .000,078																
Cross-check frequency	4	64	165	323	454	464	379	268	158	80	33	4	3	1	0	
$X^2 = 6.20$; P = .86																

the literature of the critical writers. In addition, efforts were made to anticipate criticisms which had never been made and which might never be seriously advocated.

In general, the criticisms directed against the published ESP reports have been classified as those pointing to the possibility of sensory cues in the experiment; those concerned with the occurrence of experimental errors in the observation of responses, recording results, and reporting the data; and finally those dealing with the methods of evaluation. The conditions which obtained in Series B may be considered in relation to these three general aspects of the research.

TABLE II B

CHI-SQUARE EVALUATIONS FOR INDIVIDUAL SUBJECTS (AS SUGGESTED BY THE REVIEW COMMITTEE) AND INDIVIDUAL CRITICAL RATIO EVALUATIONS

Subject	X^2	d.f.	P	C.R.	P
H.G.	9.51	8	.30	2.79	.0026
M.B.	3.49	7	.83	.32	.38
B.Y.	8.81	7	.27	— .11	.46
J.Ba.	2.62	4	.63	.13	.45
D.L.	8.61	7	.28	2.11	.018
A.B.	12.65	7	.08	— .81	.21
A.M.	13.13	9	.16	1.50	.067
B.M.	5.35	7	.62	.81	.21
L.D.	4.46	6	.62	.33	.37
C.W.	17.51	7	.015	—1.35	.089
P.M.	75.63	9	.000000	5.34	4.6×10^{-8}
B.Br.	0.85	4	.92	— .21	.42
B.Be.	4.34	8	.82	.92	.18
C.C.	13.48	8	.10	2.69	.0036
D.A.	7.43	4	.12	3.04	.0012
M.W.	4.60	4	.33	.07	.47
N.A.	5.04	7	.66	1.47	.071
N.S.	12.01	7	.10	.92	.18
Misc.	5.77	8	.67	.00	.50
Total	215.29	128	.00000083*		

*Derived by the use of the formula:

$$C.R. = 2X^2 - 2(d.f. - 1)$$

The Question of Sensory Cues. In Series B there was no direct sensory contact between the subject and the cards to be guessed. This aim was simply and effectively accomplished by removing the cards from the hands of the subject and by interposing the opaque screen between the subject's eyes and the shuffled pack. If sensory cues affected the subject's scores, therefore, this would have had to come about by some indirect or more subtle means. The possibilities may be examined in the light of the actual conditions.

a. Visual. Visual perception must obviously be controlled if the conditions are to be adequate for testing ESP, since the stimuli and cards are characteristically visual objects. One thinks first of possible reflecting surfaces in which the subject might have seen the card symbol. The top of the table on which the subject worked was covered by a blotter which would have prevented reflection even without the screen; and with the small shield back of the aperture of the screen, there could be no possibility of reflected visual cues. The walls of the room in which the experiment was conducted were of soft composition material and were equally poor as reflectors. The subject could not have made a practice of looking over the screen into Woodruff's eyes or glasses without having had his actions detected by one or both of the experimenters, one of whom sat behind the subject with the latter in full perspective.

One critic has suggested that in the usual form of the STM procedure in which the key cards are visible to both the subject and the experimenter, the experimenter may wishfully misplace a few cards in such a manner as to get the scores which he anticipates. This assumes that the experimenter may use sensory cues to produce spurious results. The conditions of Series B explicitly prevented this danger by having the subject alone know the order of the key cards. Any prejudice or will-to-produce on the part of the experimenter who handled the cards was effectively controlled.

b. Auditory and visual. If cues occurred, therefore, they must have been partly of an auditory character, effected with the following assumed steps: The experimenter either deliberately looked at the cards in the deck or unwittingly observed cues that identified them. He unconsciously or deliberately gave cues to the subject that could be heard by him but not by the other experimenter seated a few feet behind him. The subject would follow the cue and point to the key card indicated. Or the subject might give the experimenter auditory cues concerning the order of the key cards which would permit the experimenter to misplace some cards to increase the score. This is to assume, again, that the experimenter either looked at the faces of the cards or identified them through visual or tactual cues.

The facts as they bear upon these possibilities are these: (a) The experimenter gave no signal to the subject throughout the run other than that of the time to start. This is one of several points in which the methods of this study exceed the requirements for control against sensory cues and experimental error as laid down by Knight Dunlap and others (2) in their description for the condtions of an adequate experimental testing of the ESP hypothesis by means of card sorting. They suggested that the experimenter should give some vocal signal when he was ready for each trial throughout the run. (b) The rate at which the subjects proceeded in their indications in the present research averaged for some of the best scorers as fast as two cards per second. This in itself would appear to be an effective block against the interchange of auditory signals between experimenter and subject. (c) As later analyses of the data will show, subjects tended to decline in their ability to score above chance the longer they used a particular kind of stimulus material. This is a fact which is difficult to account for on the basis of the use of sensory cues, either visual or auditory. On the other hand, a decline in ability to demonstrate ESP has frequently been reported, even by investigators in experiments done at such a distance that the question of auditory and visual cues could not enter.

In certain respects, however, our conditions failed to meet the requirements laid down in the paper mentioned: (1) One specification was that the scores be withheld from the subject throughout the entire research. In this experiment, the subject knew his score at the end of each run of 25 trials. (2) The order of the key cards was to remain unchanged throughout. In our tests the order of the key cards was changed from run to run. (3) No computation of scores was to be made until the end of the experiment. They were frequently made in this experiment. (4) Each subject was required, as far as possible, to have the same number of tests as every other subject. No effort was made in the present work to obtain the same number from each subject. (5) Work periods were to be of uniform time length, were to consist of the same number of runs, and were to have the same distribution throughout the week. While the work periods in this experiment were roughly uniform as a matter of convenience, there was no effort made to keep the other points uniform. (6) Age range was to be restricted, for example, to two adjacent college years. Our subjects varied more widely than this and no effort was made to restrict age. (7) Subjects were to be requested not to use ESP cards in any other connection during the course of the experiment. No such request was made in this research. (8) A number of statistical requirements were made that were not carried out in this particular study. For example, the scores of each subject were to be totalled (a) for each set of five successive runs, (b) for each successive set of 25 runs, (c) for all the tests of each work period, (d) for the total tests of the experiment. This was done only for (c) and (d). The requirement stated that there should be tabulation of the total hits and misses for each of the stacks separately; that is, for each key card. This was not done. There was required also the average number of successes per subject, average score for experimental session, and percentage of successes. (9) It was required that the experiment be set up under the superintendence of three psychologists, each from a different university. It was, in addition, to be under the direction and control of two or more psychologists who are regarded by members of the profession generally as competent in the experimental field, one of whom was to be on duty during every work period. In this experiment there was no superintendence from psychologists of other universities, and since in the less objective professions, competence of the experimenter is mostly determined by the *a priori* acceptability of his findings, it is conceded that in this point, too, the requirement is not met.

While the majority of the requirements which have not been met can be recognized to have a certain value for experimental objectives not concerned here—objectives such as the comparative study of ESP test performance under certain conditions—we are unable to discover any reason warranting their general adoption. The proponents themselves gave no grounds for their being regarded as essential to a crucial test of the ESP hypothesis. The ninth requirement, regarding superintendence, is based upon the assumption that "competent" experimenters will remain "competent" should they become associated with an investigation in which the findings are favorable to the ESP hypothesis. So far as is known, there was never any question of the competence of the now considerable number of psychologists who have obtained results favorable to the hypothesis prior to their publication of these findings.

(d) The series of 400 trials made with the key cards facing inward, so that only their backs were visible to the subject, constitutes in a peculiar way a control upon the possibilities that sensory cues of visual and auditory characters may have been combined in producing the results. While it can not be denied that the subject may have identified some of the key cards from their backs, there will be no question that the subjects knew far less about the order of key cards in this series than in the others in which the symbols were fully visible. Consequently, if the subject was using information obtained visually concerning the key cards and was conveying such information to the experimenter as would permit the experimenter to misplace certain cards as visually perceived through cues on their backs, the results of this series should have been appreciably lower than those of the remaining 2,000 runs. Actually, the scores with blind STM were at the same average level as for the rest of Series B, an average of 5.2 per 25.

c. Faulty cards and shuffling defects. Any effect of inadequate shuffling or faulty card materials which permitted the subject to use inference to score above chance would indirectly involve sensory perception. This is true in the sense that the subject would have to apply knowledge which had previously obtained through the senses to infer something of the actual order of cards after they had been shuffled. This possibility was explicitly controlled during more than a third of Series B, during which Pratt shuffled the cards and Woodruff cut them out of sight behind the screen with a paper knife. The 830 runs done under this condition gave a deviation beyond mean chance expectation of 177 hits, an average of 5.21, almost exactly the same as that for the series as a whole.

Safeguards Against General Experimental Errors: a. Recording and checking. The possibility of errors in recording the results was avoided by the simple expediency of having Woodruff record the cards as distributed, and Pratt the key cards, without either one having any knowledge of the observations of the other till his record was fully made. The third person who checked the record sheets later did so without any knowledge of the scores as obtained from the cards after the run and recorded by each of the experimenters. Some light is thrown on the question of the accuracy of the two methods of checking, either from the cards or from written records of the symbols, by the following facts: When the scores as obtained from the record sheets were compared with those of the two experimenters as obtained from the actual cards, several discrepancies were found. In most instances in which the record sheets were again consulted, it was immediately evident that the error in checking had been made by the person working from the written records. In three instances it was evident from the record sheets that an error in recording had been made. One of these consisted in recording two key symbols of the same kind in the series of five where all were known always to be different. The other two were evident from a study of the symbol frequencies in the card distributions which showed that six of one symbol and four of another had been recorded for one run, when only balanced packs of five symbols of each kind were used. All told, three *recording* errors were discovered; that is, errors in which one of the experimenters had made a mistake in writing down the symbols. No errors in counting the scores from the cards at the end of the run were detected. Two of the recording errors lowered the run score by one hit in each instance, and the other raised it by one hit. The net result upon the total deviation as represented by the experimenters' scores was, therefore, to lower it by one hit for the entire 2,400 runs.

b. Deception. Experimental conditions which would make it impossible for one investigator wilfully to deceive his colleagues might not be attainable. However, it is worth pointing out that the conditions of Series B accomplished something in this direction, inasmuch as they made it difficult, if not out of the question, for one experimenter to practice deception upon the other even if he had wished to do so. The serially numbered record sheets which were obtained from the secretary for the purposes of the experiment were stamped with a seal which was always locked in the secretary's keeping and which could not have been duplicated or "borrowed" by either experimenter without considerable difficulty. The presence of the locked box into which the record sheets were deposited at the end of each run would have made it necessary, even if an experimenter had succeeded in duplicating the blank record sheets, for him to recover the legitimate record before substituting a faked one, or to recover the legitimate record before the check-up by the secretary if he intended to change it in a way to improve the scores. Finally, each experimenter kept his own complete record of the run scores as counted. If either one had wished to change the results, it would have been necessary for him to secure the record of the other and change this as well. These, it would appear, may not be insurmountable difficulties, but they are real psychological barriers to dishonest practices which those who wish to consider the question of fraud on the part of the experimenters may want to take into account.

The Question of Proper Statistical Evaluation

a. Sampling. The point most generally raised here is that of whether the data as evaluated were properly selected. In particular, were all the scores observed included in the final evaluation of the experiment? A positive answer to this question for Series B is particularly easy and emphatic because of the use of serially numbered record sheets. As stated previously, the length of each sub-series was definitely fixed in advance; the general procedure to be followed was outlined and the descriptive statement given to the secretary at the time the required number of serially numbered sheets was requested. In this way the investigators put themselves on record at each new stage of the research as to the additional number of runs that would be made. Each run was recorded on the sheet provided for the purpose before anything was known as to the actual score. In the final check-up from the written record, all of the blank sheets, duly filled in and signed, were accounted for. Any omissions would have been immediately obvious. There can be little question, therefore, that the 2,400 runs reported represent a consecutive series which was all the work done under the conditions described by the two experimenters within the time limits stated.

b. Optional stopping. Actually, the point at which the present experiment was arbitrarily terminated for the purposes of making this report did not represent an end of joint investigation by the two writers. The stopping point was determined actually by the occasion of presenting a report to the Southern Society for Philosophy and Psychology. The further work was subjected to alterations of conditions not relevant here. This raises a question as to whether the experimenters simply selected a favorable point at which to close the investigation, a point for which the only statistically reliable factor was their exercise of that right of optional stopping. The effect of optional stopping as related to ESP data has been emphasized by Leuba (6). The mathematical aspects of the problem have received particular attention from Greenwood, whose method was described in general terms earlier in this report. The optional stopping correction was discovered as the present research was nearing completion. The length of each sub-series as well as the maximum number of sub-series which would be done had not been stated before starting Series B. Before applying the optional stopping correction, therefore, it was necessary to make sure whether these characteristics were affected by the experimenter's knowledge of the scores throughout the experiment.

The lengths of the six sub-series in Series B were 300 runs for the first four and 1,200 and 400 runs for the last two, respectively. Actually, (a) the average remained fairly constant throughout the experiment; and (b) the experimenters were not aware of the manner in which either shortening or lengthening the sub-series would bear upon the evaluation of the results. These facts satisfy the writers that they were not influenced by the scores in fixing the length of sub-series. However, it was agreed that for the purposes of correcting for a possible effect of optional stopping, Series B would be treated as though it had consisted of 8 sub-series, each of 300 runs. In this manner the possibility of favorable variations in length are completely ruled out.

Likewise, no maximum number of sub-series for the experiment as a whole had been set at the start. Under the circumstances, the investigators, in order to make a fair correction for the effect of optional stopping, arbitrarily set the outside limit, beyond which the experiment would not have proceeded under any circumstances, as twenty sub-series. The two experimenters would need to work together intensively for one and one-half additional years in order to reach this maximum. This limit was deliberately placed high in order that the fullest allowance might be made for the effect of optional stopping. When Greenwood's correction is applied to the results of Series B, the probability is increased from 3×10^{-7} to 5×10^{-6}.

c. Cross-checks. In order to see what results would be yielded by the actual card distributions and by the key card orders of Series B if extra-sensory perception were ruled out, a cross-check between the card distribution of a particular run and the key card order for the third run in advance was made. The data were cross-checked within groups of 100 runs, following the system of checking the first distribution of card symbols against the order of key cards for the fourth run, the second distribution against the fifth order of key cards, etc., and finishing up each group by checking the ninety-eighth distribution against the order of key cards for the first run, the ninety-ninth against the second, and the one hundredth against the third. These 2,400 empirical scores gave 12,056 correspondences, or a positive deviation of 56 from mean chance expectation and an average of 5.02 hits per run.

A chi-square evaluation of the frequency distribution of scores obtained on the cross-check shows no significant departures from chance expectation. The chi-square was 6.20, with 11 degrees of freedom, with equal probability that 86 in 100 such empirical series would on the average give a worse fit to the theoretical binomial curve.

The results of the deviation ratio evaluation of the cross-check scores are shown in Table I, and those of the chi-square treatment of the frequency distribution of cross-check scores in Table II.

d. Care in statistical treatment of the data. Related both to the question of the general trustworthiness of the investigators and to that of the accuracy of statistical evaluations is the amount of care used in compiling and making computations from the data. The general summary of the results of Series B were calculated independently by two persons. Likewise, the cross-check with the written records of Series B was made independently by two persons, with a run-by-run comparison of their sets of 2,400 empirical scores and a final joint examination of the original records in cases of disagreement. The actual computations for Series A were the responsibility of one of the experimenters (Woodruff); but in making the further analyses of the data soon to be presented there was ample opportunity to check upon the accuracy of the figures. That is to say, the consideration of the data along the lines of various testing conditions offered a check both upon the general totals and upon the accuracy of the analyses themselves, inasmuch as the records were always retotaled and compared, after making any particular study of the results, with the general totals from which the divisions started.

e. "Stacking error." A conceivable source or error which will be called the "stacking error" may be described as follows: Woodruff, in laying off the cards of the pack, may have used either sensory cues or wilful deception to group the cards, laying an unusual number of like symbols in each pile. This is to suppose, of course, that he accomplished this by failing to follow exactly the subject's pointer. On the assumption that the experimenter actually did group the symbol suits as he laid down the cards, the element of chance in that step of the test is removed. Each run then reduces, in effect, to *five* trials in which the piles are compared with the key cards. Because of the fact that the experimenter is ignorant of the actual order of key cards, such a grouping of the symbols would not affect the average score expected. It would, however, increase the variability of the run scores, so that the probability of both high and low scores would be increased. The question at issue is whether the average of 5.2 per run for Series B would be significant in the face of the hypothesis that just such illegitimate groupings of the card symbols occurred with an indeterminate, equivalent reduction in "trials." (If this interpretation were preferred and the result were nevertheless shown to be significant, we would have to suppose that Woodruff had demonstrated ESP in favorably locating the piles.)

The results of the chi-square evaluation of the score frequencies obtained in the cross-check permit the definite conclusion that no unusual grouping of the symbols as they were distributed occurred. For if this had been the case, these arrangements would have affected the scores of the cross-check in the very manner which the hypothesis in question supposes was the case for the experimental series. The absence of symbol groupings is demonstrated by the fact that the chi-square evaluation for the cross-check gave a P of .80 (see Table II).

It is recognized that this statistical control is one which might break down in tests in which much higher averages, necessitating some degree of grouping to produce the observed scores, are obtained. However, in the present investigation it was effective because of the fact that the average rate of scoring over a relatively long series was not high enough to be associated with noticeable symbol groupings.

From the consideration of the results of Series B in the light of all the experimental conditions, the writers are unable to offer any explanation of the findings except to say the subjects demonstrated a degree of success in identification of the concealed cards and that this knowledge was not obtained through any recognized sensory channel. On the basis of the joint investigation in particular, a conclusion is reached that ESP occurred in this investigation.

The Occurrence of ESP—Series A

The question arises, of course, as to whether the introduction of the advances in methodology in Series B means that the writers intend to minimize the importance of the results of Series A or of earlier experiments in general. The answer must be that the following considerations appear to make such a course unnecessary.

In the first place, Series B effectively substantiates Series A. In each case, the results are shown to be highly significant—Series A giving a C.R. of 6.28; and Series B a C.R. of 4.99. The average of Series A is .13 higher than that of Series B, but, as shown in Table I, this is not a reliable difference (C.R. of the diff. = 2.00). Having been forced, for want of any other explanation, to conclude that ESP occurred in Series B, the writers prefer on the principle of parsimony of hypotheses to extend this conclusion to cover Series A as well. Actually, the only differences in conditions between the two parts of the investigation are the absence of a second observer throughout Series A and a reliance therein upon counting and re-counting the scores from the cards without making independent written records. However, 169 runs of Series A with an average of 5.43 were witnessed by a second experimenter. The absence of written records in Series A does not seem so serious in the light of the comparative study of the relative efficiency of methods of scoring in Series B which showed counting from the cards to be more accurate than the use of written records.

As long as the question of the occurrence of ESP was primarily at issue, Series B was rightly considered to represent a higher plane of evidence because of the more advanced experimental conditions. Now that the evidence on that first problem seems to justify further study of the results to see how varying the conditions of the experiment affected this phenomenon, the results of the entire investigation are admissible as evidence bearing upon possible relations of ESP.

II. RELATIONS SHOWN BETWEEN CONDITIONS AND RESULTS

The Relation Between Rate of Scoring and Size of Symbols

It was stated above that the primary purpose with which Woodruff undertook the tests described as Series A was that of making further observations upon the comparison of level of scoring and the size of symbols. The results bear out the earlier finding of L. E. Rhine (12); namely, that no significant relation is indicated.

The analysis of the *total* results of the entire research giving comparison of the scores for the four sizes of symbols used (Table III) shows no significant difference in the level of scoring for different sizes of symbols. It may be seen from the table that marked differences in averages resulted *in Series A*; total of 576 runs with the regular-sized 1½ inch symbols averaged 5.15 hits per run, while 691

TABLE III

RESULTS OF SERIES A AND B ACCORDING TO SYMBOL SIZES

Symbol Size	Series A Runs and Dev.	Av.	Series B Runs and Dev.	Av.	Total Runs and Dev.	Av.
1/16	201 + 121	5.60	282 + 19	5.07	483 + 140	5.29
1/4	691 + 273	5.40	321 + 51	5.16	1,012 + 324	5.32
1 1/2	576 + 87	5.15	1,197 + 286	5.24	1,773 + 373	5.21
2 1/4			600 + 133	5.22	600 + 133	5.22

runs with the ¼ inch symbols averaged 5.40, and 201 runs with the 1/16 inch symbols gave an average of 5.60. Not only is this difference consistently in the direction of a higher score upon the smaller symbols, but the difference between the ¼ and the 1½ inch sizes has the suggestive C.R. of 2.27 and that between the 1/16 and the 1½ inch ones, the significant C.R. of 2.81. (The smaller number of runs with the 1/16 inch symbols is accounted for by the fact that these were not introduced until relatively late in Series A.) But as far as the relation between scores and symbol size is concerned, the results of Series B do not follow the earlier pattern. With the 1/16 inch symbols, which had yielded the highest scores in Series A, 282 runs in Series B averaged only 5.07. With the ¼ inch size, 321 runs gave an average of 5.16. The 1½ inch symbols averaged 5.24 for a total of 1,197 runs, and the largest, 2¼ inch symbols, 5.22 for 600 runs. The differences in this series are not as striking statistically as those for Series A. As a consequence of the tendency toward a reversal of results in Series B the combined scores for the total research do not show any effect of size of stimuli upon ESP scores within the limits investigated. This conclusion is indicated by the last three columns of Table III, which show the totals for Series A and Series B combined.

The Relation Between Experience of Subjects and Rate of Scoring

Reference has frequently been made in the literature to a tendency for subjects to decline in scoring ability as they become more experienced with the usual laboratory tests. A consideration of the results of Series A in relation to the previous experience of the subjects with tests of this character led to the suggestion that this phenomenon of declining scoring ability might have produced the differences in results for the various stimulus sizes. This suggestion seemed to offer a possible explanation because of the fact that a larger percentage of experienced subjects participated in the first tests of Series A, when the two larger sizes of symbols were used exclusively, than in the last part when the smallest size was introduced. Therefore, the data of the entire research were analyzed from the point of view of the amount of experience of the subjects to discover whether there was a general tendency of subjects to decline in scoring ability.

For the purposes of this analysis, those subjects who had taken part in any formal ESP tests prior to their participation in Series A were classified for that series as "experienced" subjects. Others in Series A were classed as "inexperienced." In Series B all subjects who had already taken part in Series A or in any other formal ESP tests were classified as "experienced" subjects and all others as "inexperienced" subjects. A subject's classification as to experience was considered to remain unchanged throughout a given series, but the same subject might be inexperienced in Series A and experienced in Series B.

TABLE IV
RESULTS OF SERIES A AND B ACCORDING TO EXPERIENCE OF SUBJECTS

	Series A		Series B		Total	
Subjects	Runs and Dev.	Av.	Runs and Dev.	Av.	Runs and Dev.	Av.
Experienced	793 + 215	5.27	1,575 + 290	5.18	2,368 + 505	5.21
Inexper.	675 + 266	5.39	825 + 199	5.24	1,500 + 465	5.31

Table IV shows the analysis of the results along these lines for both Series A and Series B separately and for the research as a whole. The slight differences in favor of higher scores for the inexperienced subjects are statistically insignificant.

The Relation Between Newness of Stimulus Sizes and Rate of Scoring

The Problem. However, even a slight difference might lead to the discovery of an important principle. The proposition can logically be formulated in the following manner: All stimulus material used by inexperienced subjects was new to them. On the other hand, only part of the stimulus material used by experienced subjects was new. The difference in favor of the inexperienced subjects might have been caused by the use of a greater preponderance of new material. The problem, then, may be stated: Did stimulus material, when used by a given subject over a period of time, lose its effectiveness as indicated by a falling-off of ESP scores?

The General Evidence. With all the results in hand, it was necessary to set up certain arbitrary criteria of newness of material in order to make a general analysis of the data for a possible effect of a novelty factor. For this purpose, material was classified as "old" for a particular subject in Series A if that size of symbol had been used in formal tests by that subject before he took part in the present investigations. As none of the subjects had worked with any symbols except those of the regular 1½ inch size, the only use of old material in Series A was in those runs by experienced subjects with the regular 1½ inch ESP symbols. All other tests in Series A were considered to be made with "new" material. For Series B, all tests made with any symbol size by subjects who had previously used that particular size of stimulus, either in Series A or in other formal tests before entering upon Series B, were classified as tests with "old" material. All other tests in Series B (including those with all sizes of stimulus for inexperienced subjects and those with sizes used for the first time by experienced subjects) were made with "new" material.

Table V shows the results of the general analysis of the data into the old and new material categories. The difference between these two groups in Series A is statistically significant (C.R. of the diff. = 3.83) with the higher rate of scoring in the tests with new material. In Series B, a similar division of the work with the four stimulus sizes gives a difference in the same direction, though the rate of scoring with the new material is not significantly higher than that for the old (C.R. of the diff. = 1.50). When both series are combined and the distinction between new and old material is maintained, a significant difference in favor of higher scores with new stimulus material (C.R. of the diff. = 3.43) is obtained for the experiment as a whole.

TABLE V
RESULTS OF SERIES A AND B ACCORDING TO THE NEWNESS OF MATERIAL

Material (Symbol size)	Series A Runs and Dev.	Av.	Series B Runs and Dev.	Av.	Total Runs and Dev.	Av.
Old	331 — 10	4.97	1,493 + 367	5.25	1,238 + 112	5.09
New	1.137 + 491	5.43	907 + 122	5.13	2,630 + 858	5.33
C.R. of diff.	3.83		1.50		3.43	

Further Analyses. The strong suggestion that the sizes of symbols with which the subjects had had less experience were more effective as ESP "objects of perception" raises a number of questions as to the possible nature and origin of this newness factor. Some of these questions can be answered, tentatively at least, by further study of the data from the present investigation. Others can only be stated and considered speculatively. In any event, definite conclusions both as to the actual occurrence of the novelty effect and as to its nature must await further independent experimental confirmation. For what they may be worth, these questions may be raised and considered as far as the results of this investigation will allow.

a. Loss of newness effect. If stimuli lose their effectiveness for ESP scoring with use, the question arises as to when and how the loss occurs. Do the scores with a particular type of symbol drop off gradually, or is there a rather sudden decline after a period of optimum success for each subject? If a point-for-point consideration of the rate of scoring in the present research in relation to the amount of experience of subjects is made, some light might be thrown upon this question. The results of such a study are shown in Table VI, and the same data are represented graphically in Fig. 5.

For the purposes of this analysis, the distinction between Series A

TABLE VI

RELATION OF LEVEL OF SCORING TO THE AMOUNT OF EXPERIENCE WITH THE DIFFERENT SIZES OF STIMULUS MATERIAL

Successive Times of Using (Sessions)	Unclassifiable Runs and Dev.	Unclassifiable Av. Hits per Run	Classifiable Runs and Dev.	Classifiable Av. Hits per Run
1	304 + 17	5.06	1,041 + 409	5.39
2	106 + 0	5.00	571 + 208	5.36
3	44 + 22	5.50	432 + 158	5.37
4	32 + 9	5.28	303 + 6	5.02
5	28 + 7	5.25	240 + 56	5.23
6	52 — 10	4.81	147 + 62	5.42
7	39 + 8	5.21	121 + 22	5.18
8	28 — 11	4.61	75 — 26	4.65
9	19 + 2	5.11	73 + 21	5.29
10	28 + 10	5.36	46 + 9	5.20
11	26 + 14	5.54	27 + 4	5.15
12	17 + 2	5.12	10 — 1	4.90
13	13 — 12	4.08	12 + 3	5.25
14	9 — 8	4.11	11 — 5	4.55
15	14 — 6	4.57		
Total	759 + 44	5.05	3,109 + 926	5.30
		C.R. of diff. = 3.00		

and Series B was disregarded. All tests were divided into two classes. In one, designated as "unclassifiable," was placed all the work *by experienced subjects done with the regular 1½ inch symbols, which they had used in tests prior to first starting in the present investigation.* For these subjects, it was not possible to determine the amount of experience with the standard symbols before their participation in this experiment. In the second class, designated as "classifiable," was included the work of all subjects with those sizes which were used only in this investigation. For these tests, the rate of scoring in relation to the novelty of symbol sizes could be traced from the very first session of using any new material through successive occasions of being tested with that same material.

In Table VI all the results of the experiment are presented as they belong under these two groups. The results for all symbol sizes from the first experimental session in which each was used in this research are brought together as Session 1. The results of the second occasion of using each size are combined as Session 2, and so on for the entire experiment. The smaller numbers of runs for later sessions are due to the fact that not all subjects used each symbol size equally often and that the subjects served for an unequal number of sessions.

An examination of the unclassifiable column shows no significant trend in scoring throughout successive periods of working. As we

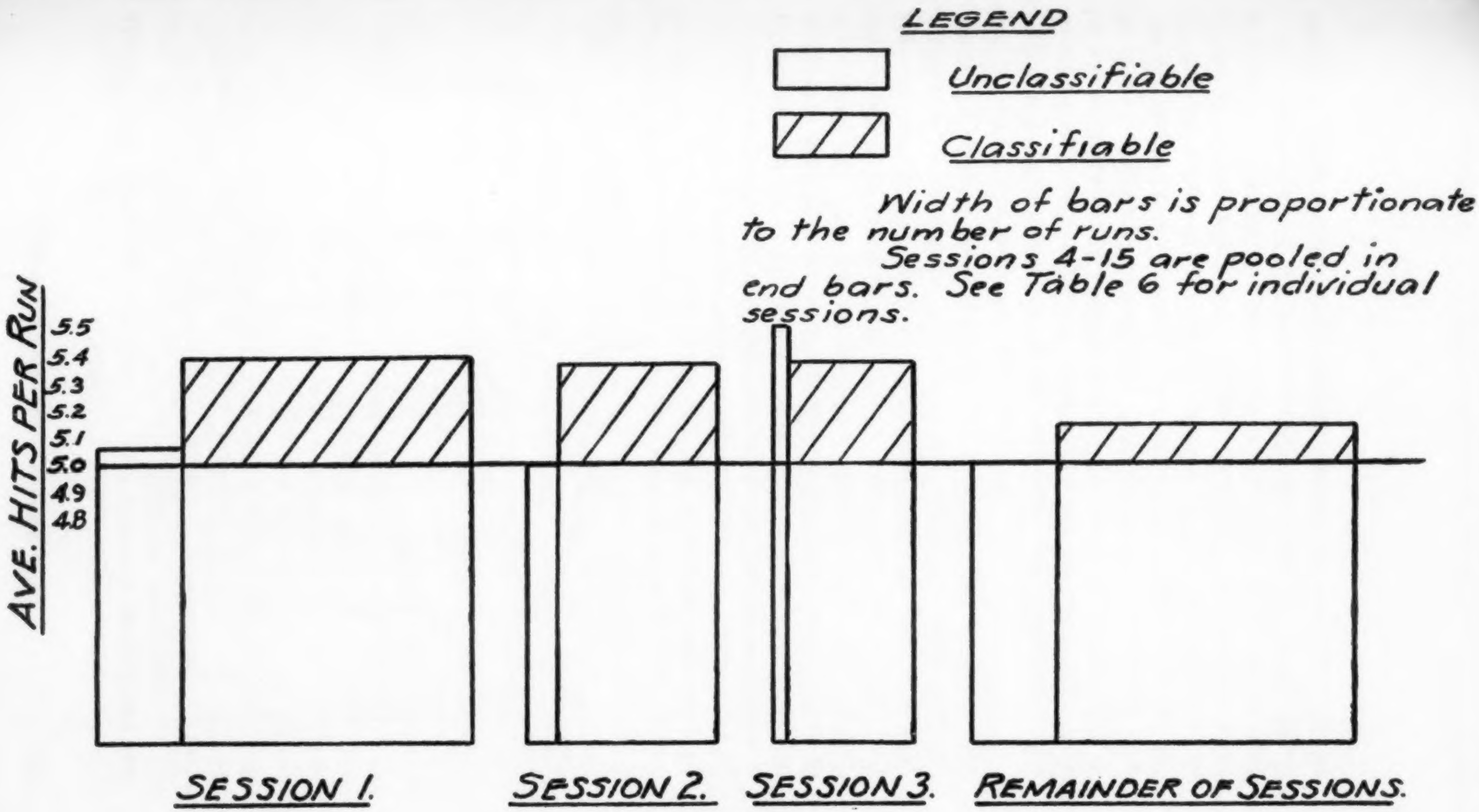

LEGEND
Unclassifiable
Classifiable
Width of bars is proportionate to the number of runs.
Sessions 4-15 are pooled in end bars. See Table 6 for individual sessions.
AVE. HITS PER RUN
5.5
5.4
5.3
5.2
5.1
5.0
4.9
4.8
SESSION 1.
SESSION 2.
SESSION 3.
REMAINDER OF SESSIONS.

Figure 5

should expect from the previous indication of the favorable effect of novelty, the results for this classification of symbols, with which the subjects had had an indeterminate amount of experience before the first session with them in this investigation, were low in general rate of scoring (an average of 5.05 for 759 runs).

On the other hand, the classifiable column, in which the subjects' experience with the various sizes of symbols can be traced, session by session, from the very first use of each size, seems to tell a different story. The first three sessions with new material give a fairly uniformly high average. Thereafter, fluctuations in average from session to session appear in a manner suggestive of those of the unclassifiable column from the start. If all the sessions in the classifiable column from 4 to 14 inclusive are combined, the average is found to be 5.14 for 1,065 runs. As contrasted with the average of 5.39 for the 1,041 runs of the first session and that of 5.36 for the 1,003 runs of the second and third sessions combined, this suggests that the subjects tended to obtain lower scores with a particular new size of stimulus symbol sometime after they had used it for three experimental sessions. No generalization to other experiments is possible and none is intended. In connection with the general evidence that novelty favors ESP scoring, the data of Table VI and Figure 5 are important in that they show an actual decline did occur for the successive uses of sizes of stimuli to which the subjects were introduced for the first time in this experiment.

b. Relation between experience and effect of newness. The question arises as to how experienced and inexperienced subjects compared in their ESP performance on new material. Did subjects who were experienced on old material before a given series and who got lower scores upon the old material, do worse with new material than the inexperienced subjects, for whom all material was new? The data in Table VII clearly indicate that new material was equally effective for experienced and inexperienced subjects.

TABLE VII

A COMPARISON OF "EXPERIENCED" AND "INEXPERIENCED" SUBJECTS WITH "NEW" MATERIAL

Subjects	Series A Runs and Dev.	Av.	Series B Runs and Dev.	Av.	Total Runs and Dev.	Av.
Experienced	462 + 225	5.49	668 + 168	5.25	1,130 + 393	5.35
Inexper.	675 + 266	5.39	825 + 199	5.24	1,500 + 465	5.31

c. Relation of subject's knowledge of material to effect of newness. It will be recalled that different methods were used for determining the order of presenting symbol sizes within an experimental session in which two or more sizes were used. An analysis distinguishing among the various methods of selection (regular alternation or rotation in which the subject knew when each size was to be used; or the experimenter's subjective determination of the order or following the cast of a die in which the size was not known to the subject until the end of the run) showed no significant differences among them for the general results. However, a question arises in connection with the effect of novelty of sizes and the fact that the subjects sometimes did not know during a run what size stimuli were being used: namely, what was the relation between the effect of new material upon scoring and the subject's knowledge of the kind of material being used?

In other words, the analysis presented in Table V showed that, in general, subjects scored better with new than with old material. The question now raised is this: Did that relation hold both when the subjects knew and when they did not know what size of stimuli was being used? Table VIII shows that the difference in scoring rate in favor of the new material was much greater when the subjects knew before each run what stimuli were to be used than when they did not know. Indeed, the difference in averages is statistically significant for the tests in which they knew about size of the stimuli (C. R. of the diff. = 3.75), which is not the case for those tests in which they did not know until after the run with which size of stimuli they were being tested (C. R. of the diff. = 1.09). However, when the subjects were kept ignorant of the size of the stimuli, the average with the old material was slightly higher and that of the new material lower than the general averages for each (see Table V). Consequently, the general average for all tests in which subjects did not know the size of symbols during the run is insignificantly below that of the total results of tests in which they knew.

TABLE VIII

A COMPARISON OF "NEW" AND "OLD" MATERIAL WHEN THE SUBJECTS KNEW AND DID NOT KNOW WHICH STIMULI WERE BEING USED

Classification of Stimuli	Subjects not Knowing Size Runs and Dev.	Av.	Subjects Knowing Size Runs and Dev.	Av.
Old	549 + 83	5.15	689 + 29	5.04
New	606 + 169	5.28	2,024 + 689	5.34
Total	1,155 + 252	5.22	2,713 + 718	5.26

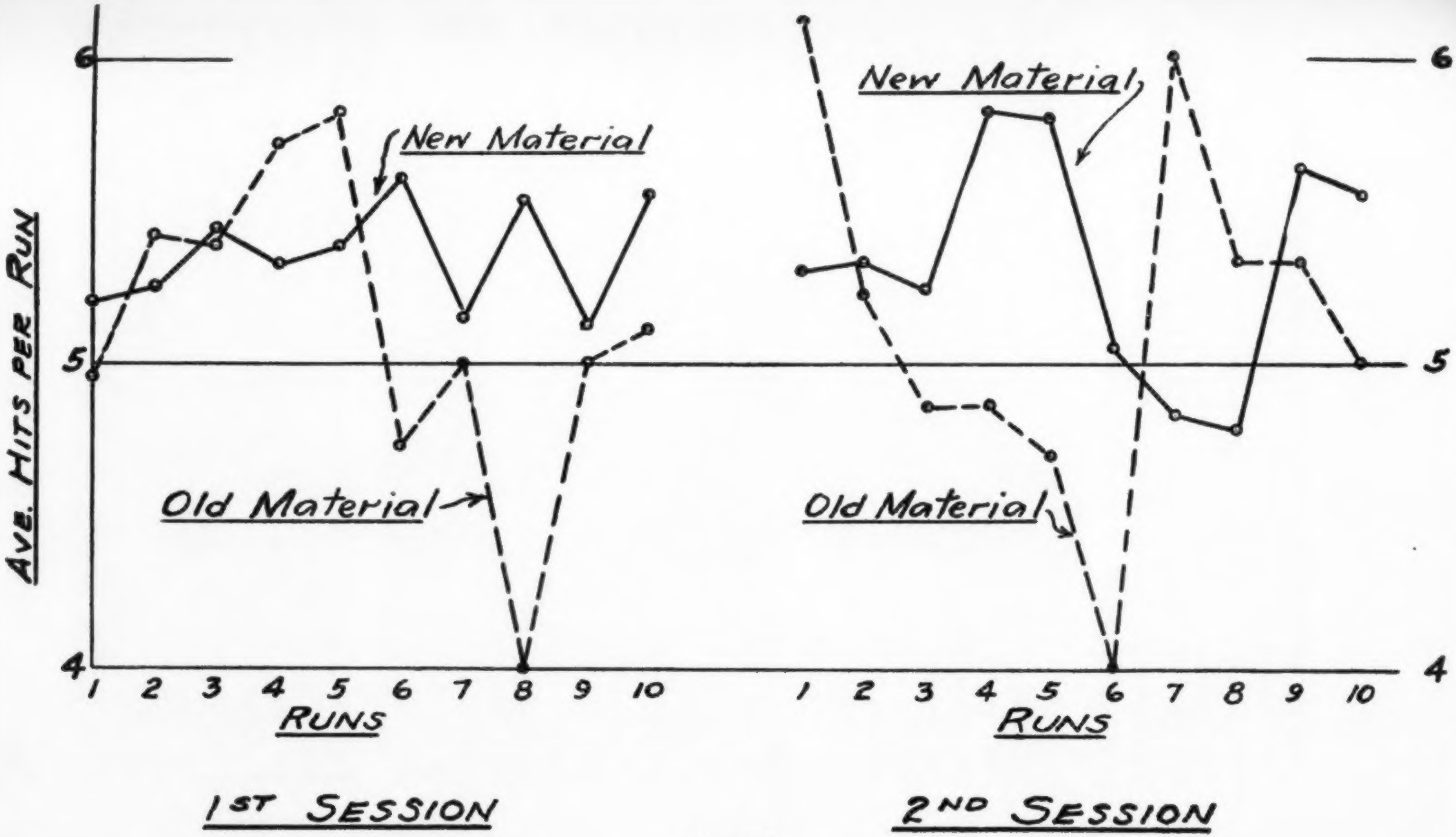
d. Scoring trends within experimental sessions. A point of interest in relation to Table VI is that of how the run-by-run performance curve of the new (classifiable) material compared with that of the old (unclassifiable) material. One question has to do with the trend of the scores during the *first few runs of the session*—or of Session 1 in particular. Rhine (11) reported a characteristic period of adjustment to a new condition reflected in the scores by a rising level of scoring during the first few runs. Another question is that of how the performance with the new and old material compared throughout the experimental session.

In Fig. 6 the results of the first two sessions of Table VI are shown graphically. The curves suggest that there was an adjustment period in the first session with both old and new material. Also, the average difference in favor of the "new" material seems to have resulted from a more consistent level of scoring. Because of the fact that sessions for various subjects were not of equal length, the points on the curves toward the end of each session do not represent as many runs as those toward the beginning. Each line of evidence, however, is only suggestive in character and takes on real significance only if compared with other similar lines. The evidence of an adjustment period may therefore be said to be stronger because it confirms the observations of earlier investigators, while the suggestion of a steadier, more sustained, rate of scoring upon new material remains to be confirmed or refuted by subsequent investigations.

DISCUSSION

In general, the evidence on the relation between the kind of stimulus material and the rate of scoring shows that some characteristic or characteristics of the symbol material affect the degree of success in identifying the symbols. The foregoing analyses point to the newness of stimuli as the most important factor. It is appropriate to inquire whether this was the *only* factor and to discuss how the effect of newness upon ESP scoring is to be interpreted.

There is more than a superficial relation in the results between size of stimuli and newness. The fact that the rate of scoring in Series A was inversely related to the size of stimuli was logically interpretable as due to either factor, size or newness. However, there would appear to be no basis for expecting the effect of size, if any occurs, to be in an *inverse* relation to the scoring level. This fact, in itself, strongly suggested that the explanation lay either in the new-



AVE. HITS PER RUN
6
5
4
New Material
Old Material
1 2 3 4 5 6 7 8 9 10
RUNS
1ST SESSION
2ND SESSION

Figure 6

ness factor or in still another one. It has been made clear from the presentation of the data that this suggestion was supported by Series B and the combined results of the entire experiment. The results from this investigation must therefore be taken as strengthening the general evidence that the known physical characteristics of the stimuli do not affect ESP. Size was not a factor within the limits of variation introduced in this study.

The factors of experience of subjects and newness of material are obviously closely related. In ESP tests in which the same material is used throughout, it is not possible to distinguish between the general experience of the subject and his use of the standard ESP symbols, as they both increase together. The present investigation involved variations in conditions such as to permit an analysis of the data distinguishing between these two factors.

The evidence here points strongly to a direct relation between the level of scoring and the number of runs for which each size of symbol had been used. This relationship is one which may offer further explanation of the generalization made by other investigators (see p. 123) that scoring ability decreases with experience. Subjects did decline with experience in this study as well, but this effect tended to be specific to each size of stimulus rather than generalized. Subjects who had used the standard ESP symbols before entering upon tests with new sizes did just as well with the new symbols as did completely naive subjects.

An important implication of the results is the suggestion that the scoring rate was kept at about the same level throughout the experiment by the introduction of new material when, presumably, it would have declined much sooner if the same size of stimuli had been used throughout. It is not yet possible to generalize for these subjects as to whether they can be made to score above expectation for an indefinite period simply by introducing new material at strategic points. Nor is it possible to state whether changes other than those involving size would affect the results in the same way. There has been a general clinical impression abroad among ESP workers that a change of conditions helps to keep the subject interested in the tests in a way that favors scoring. Further direct experimental evidence to define the conditions under which this generalization is applicable is obviously needed.

Unfortunately, the quantitative results of the research do not point the way to a definite interpretation of the psychological difference between new and old material. General observations suggest, however, that the experimental situation was such as to elicit more favorable motivation in the subjects when the new stimuli were used. One possible interpretation would be that the effect is related to general psychological satiation for the various kinds of material. If this were true, the decline in scores with the use of a particular size of symbol might be considered to be a psychological aspect of the relation between the subject and the stimulus, and the changing of other features of the experimental conditions might not have the same effect as changing the stimuli.

On the other hand, there are indications that the newness effect is partly, at least, a function of the relation between the experimenters and the subject; or, in other words, that the experimenters unintentionally used the new sizes of symbols as an opportunity for eliciting a more favorable attitude in the subjects. The experimenters always showed the subjects the new sizes before beginning the daily session in which they were first used. This was frequently done in a manner which challenged the subject to do better with the new sizes before beginning the daily session in which they were first used. During experimental sessions the experimenters, particularly Woodruff, frequently encouraged the subject between runs in the same challenging manner. The introduction of a new size of stimulus was made a special "talking point" between the experimenters and the subject. This probably resulted in a greater interest of the subjects in the new material for a few sessions, after which the challenge to do better either lost its effect or was shifted by the introduction of another new size. These speculations serve chiefly to emphasize the need for further research.

SUMMARY

(1) The results of Series B appear to bear in a crucial manner upon the problem of the occurrence of ESP. In that part of the research the conditions were carefully planned to control against the effects of hypothetical sources of error by explicit steps in the experimental procedure. These safeguarding conditions have already been summarized on p. 126 ff. The results of this period of joint investigation included 2,400 runs with a deviation of 489 hits beyond mean chance expectation, an average of 5.20 hits per run. The S.D. for 2,400 runs is 97.98, and the C.R. of the observed result is 4.99. The P-value for the result is 3×10^{-7}; when allowance is made for the

possible effect of optional stopping, this is increased to $P = 5 \times 10^{-6}$. The conclusion was reached that "perception without the use of recognized sensory channels" is the only principle which can reasonably account for the results of Series B.

Because of the essential similarity in results between the two main divisions of the research, this conclusion was extended—for purposes of further analysis of the data as they bear upon the nature of ESP—to the results of Series A as well.

(2) Analysis of the data according to the size of stimulus symbols used showed that size *per se* did not affect the results of the investigation as a whole (Table III). Suggestive differences among the sizes used in Series A in favor of the smaller stimuli proved on further analysis to be related to the newness of testing material.

(3) The amount of experience of subjects with ESP tests was not, in itself, directly associated with trends in scoring (Table IV). Again, there were suggestive differences in the direction of higher average scores by the inexperienced subjects, but these also proved to be related to the newness of testing material.

(4) When the results were considered in relation to the amount of subjects' experience with the different kinds (sizes) of stimulus material, subjects were found to score significantly better with new than with old material (Table V). The advantage of working with new material was found to decline after a time as the subjects became more experienced with the new material (Table VI). Experienced and inexperienced subjects scored equally well with new material (Table VII). When subjects did not know before or during a run what size of stimuli were used, the advantage in favor of higher scores with new material was not statistically significant (Table VIII). The evidence suggests that there was a period of adjustment during the first runs in the first sessions of using any material, whether new or old (Fig. 6).

REFERENCES

1. Carpenter, C. R., and Phalen, H. R. An experiment in card guessing. *J. Parapsychol.*, 1937, I, 31-41.
2. Dunlap, K., and others Adequate experimental testing of "extra-sensory perception" based on card sorting. *J. Parapsychol.*, 1939, III, 29-37.
3. Gibson, E. P. A study of comparative performances in several ESP procedures. *J. Parapsychol.*, I, 26-275.
4. Greenwood, J. A. Analysis of a large chance control series of ESP data.

J. Parapsychol., 1938, II, 138-146.
5. Greenwood, J. A., and Stuart, C. E. The mathematical techniques used in ESP research. *J. Parapsychol.*, 1937, I, 206-225.
6. Leuba, C. An experiment to test the role of chance in ESP research. *J. Parapsychol.*, 1938, II, 217-221.
7. MacFarland, J. D., and George, R. W. Extra-sensory perception of normal and distored symbols. *J. Parapsychol.*, I, 93-101.
8. Murphy, Gardner, and Taves, Ernest Covariance methods in the comparison of extra-sensory tasks. *J. Parapsychol.*, 1939, III, 38-78.
9. Pratt, J. G. Clairvoyant blind matching. *J. Parapsychol.*, 1937, I, 10-17.
10. Price, M. M., and Pegram, M. H. Extra-sensory perception among the blind. *J. Parapsychol.*, 1937, I, 143-155.
11. Rhine, J. B. *Extra-Sensory Perception*. Boston: Boston Society for Psychic Research, 1934.
12. Rhine, L. E. Some stimulus variations in extra-sensory perception with child subjects. *J. Parapsychol.*, 1937, I, 102-113.
13. Riess, B. F. A case of high scores in card guessing at a distance. *J. Parapsychol.*, 1937, I, 260-263.

A METHOD FOR ESP TESTING

J. H. MANLEY

Department of Physics, University of Illinois

Urbana, Illinois

The experimental methods of ESP have been subjected to sufficient criticism[1] to warrant serious consideration of a different technique. It is the purpose of this paper to outline a method which would remove most of the serious objections raised against present methods. Elaboration of the method is limited only by the ingenuity of the designer and the ultimate cost. Such elaboration, however, is beyond the scope of this paper which must necessarily be limited to suggesting the means of securing the features desired. These features will be summarized according to the outline of Rhine.[2]

The first requirement of any method is, to quote Henlein and Henlein[3] to: " . . . determine the normalcy and fitness of the experimental distribution from which the estimates are obtained." The ideal distribution is certainly a purely random one essentially infinite in extent, for all probabilities will then be completely independent. As has been pointed out, the methods of Rhine or any limited card series do not give independent probabilities unless made cyclical. The lack of randomness is even more serious for it can always be argued that any mechanical means of securing a distribution is subject to the definite error due to imperfections in the physical objects used. Pegs may not be perfectly round or accurately spaced, coins and dice not symmetrical, cards irregular, and so on. Although tests can be used to check distribution, every piece of equipment shoud be so tested, resulting in at least doubling the number of observations required.

It is indeed surprising that no one has used in statistical investigations what is undoubtedly the best available source of a random distribution, the processes of disintegration of atomic nuclei or the

---

[1] Kennedy, John L., *Psychological Bulletin*, 1939, **36**, 59.

[2] *Journal of Parapsychology*, 1939, **3**, 3.

[3] Henlein & Henlein, *J. Psychol.*, 1938, **5**, 135.

occurrence of cosmic rays. That these processes are purely statistical has been amply demonstrated by experiment. Their use would therefore constitute a source which should satisfy all possible objections as to "normalcy and fitness." It is necessary, of course, to transform such a process into a more usable form, but this can be done simply with the aid of a Geiger-Mueller (G-M) tube and any of the many quenching and amplifying circuits which have been developed for use with such a tube.[4] The output of any of these circuits is then electrical pulses, randomly distributed in time. The method of application of these electrical pulses for the investigation of statistical phenomena must be determined by the test desired. Let us consider the application to the usual post-selection ESP test.

For this type of test, the apparatus must make a random choice of one of five symbols, the result of which is unknown to the subject. To accomplish this the operator presses a start button which connects the counter amplifier to a five contact rotary switch which is operated by the counter pulses. The start button is arranged to release automatically after one second. The choice is then complete, for the rotary switch will have stopped on a certain contact determined by the number of pulses in the time allowed, a purely random number. (It is naturally necessary that the number of impulses per second from the counter be much greater than 5, say 100. Because of this condition, a few micrograms of radium near the G-M tube would be more satisfactory than the use of cosmic rays which would require a tube of excessive volume.) The output of the rotary switch may be used to operate lights in separate compartments containing the ESP cards, or to operate for suitable presentation of the choice of the ESP ability of the subject.

With the random choice complete, the subject is free to make his selection. (Selection before completion of the above operation can be prevented by a time-delay circuit.) Again, there are many possibilities for recording the subject's choice and the number of hits, but a tape recorder seems to offer the most fool-proof method.

Suppose the tape is furnished with six inking pens, five for the possible choices and one for hits. A simple series connection with the subject's choice button and the rotary switch will operate the sixth pen only when a hit is made. A light can be used to inform the

---

[4] *Review of Scientific Instruments,* 1938, **9**, 218.

[5] *Review of Scientific Instruments,* 1938, **9**, 83; 1939, **10**, 21.

subject immediately since the permanent record showing both the random and subject's choice should be concealed. Extension to record which choice the subject made as well as the hits is easily accomplished.

Although the system described involves the use of a few mechanical parts, any clues can be eliminated with care. Operation of the rotary switch at high speed should in itself eliminate clues from this element. For more rigorous elimination of clues, all mechanical parts may be replaced by vacuum tubes with a slight increase in complexity. In this connection it may be pointed out that a scaling circuit such as that reported by Lifschutz[5] provides a cyclical sequence of 2n possible choices and is therefore immediately adaptable to testing employing 2, 4, 8................2^n choices.

The general features of a device employing these suggestions may be summarized in accordance with the requirements given by Rhine.

1. Careful design and construction would assure all the dependability of a modern radio receiver.
2. The principle is essentially simple, the cost of the apparatus being determined entirely by the elaborateness of the recording and safeguarding attachments.
3. Unfortunately, any apparatus designed for more than laboratory use requires many precautions which increase the cost and therefore limit marketability, but that described could be made to fulfill very stringent requirements.
4. The recording method suggested provides a continuous self-check on the proper functioning of the apparatus.
5. The size of the apparatus would be such as to be readily portable.
6. The enclosure of all circuits in tamper-proof cabinets and the electrical nature of the apparatus preclude production of extra-chance results by external manipulation.
7. As suggested, a complete trial would be made in not more than two seconds.
8. The use of standard radio parts in the apparatus suggested would automatically require operation from 110 volt, 60 cycle supply.

From these considerations it can be concluded that this method of employing the random processes of nuclear disintegration will rigorously satisfy all criteria for research in which a reliable random distribution is required.

Furthermore, this distribution in conjunction with a reliable recording system such as can be constructed along the lines suggested, will provide data on ESP tests which are not subject to the criticisms directed at present methods.

The author wishes to acknowledge the assistance of Dr. R. G. Barker of the University of Illinois and Dr. J. L. Kennedy of Tufts College in the preparation of this paper.

A COVARIATION STATISTIC

JOSEPH A. GREENWOOD

Duke University

Murphy and Taves have used a statistic to measure the amount of co-variation of the results of two joint tasks by an ESP subject (1). It is a measure of the tendency of the subject to score high or low simultaneously in both tasks for the entire experiment, or, on the other hand, to vary in opposite directions. This statistic is the first product-moment of the two variables representing the scores of the two tasks, stated in theoretical standard units and measured about their theoretical means.

Although they termed this the "covariance" we note that the usual definition of "covariance" is identical with the preceding definition with the word "theoretical" replaced by "observed." Since these statistics are generally different in value it would avoid ambiguity and detailed explanation in each case to label the statistic they used by some other name, for example "covariation."

While in given (perhaps most) instances covariance and covariation are nearly enough equal to warrant evaluation of an observed covariation as though it were a covariance (that is, as a correlation coefficient is evaluated), for the sake of exactitude in evaluation the first four moments of the covariation statistic, C, are here presented.

Let $x_1, x_2, \ldots, x_n$ be variables for which the standard deviation of x_i is σ_i, and similarly let $y_1, y_2, \ldots, y_n$ be variables for which the standard deviation of y_i is σ_i'. Assume $E(x_i) = E(y_i) = 0$[1] and that the 2n variables are independent. ($i = 1, 2, \ldots, n$). Form

$$C = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i y_i}{\sigma_i \sigma_i'}, \text{ the covariation.}$$

[1] E operating on a variable is defined to be the mathematical expectation or average value of the variable.

Due to the independence of the 2n variables $E(C) = 0$. By definition, variance of C is $\mu_{2:}C = \sigma C^2 = E(C^2)$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \left\{ E \frac{x_i^2 y_i^2}{\sigma_i^2 \sigma_i'^2} \right\} = \frac{1}{n^2} \Sigma 1 = \frac{1}{n}.$$

$$\mu_{3:}C = E(C^3) = \frac{1}{n^3} \sum_{i=1}^{n} \left\{ E \frac{x_i^3 y_i^3}{\sigma_i^3 \sigma_i'^3} \right\} = \frac{1}{n^3} \sum_{i=1}^{n} \alpha_{3:x_i} \alpha_{3:y_i}.$$ [2]

$$\mu_{4:}C = E(C^4) = \frac{1}{n^4} \left\{ \sum_{i} E \frac{x_i^4 y_i^4}{\sigma_i^4 \sigma_i'^4} + 6 \sum_{i<j}\sum E \frac{x_i^2 y_i^2 x_j^2 y_j^2}{\sigma_i^2 \sigma_i'^2 \sigma_j^2 \sigma_j'^2} \right\}$$

$$= \frac{1}{n^4} \left\{ \sum_{i=1}^{n} \alpha_{4:x_i} \alpha_{4:y_i} + 3n(n-1) \right\}.$$

The skewness and kurtosis of C are

$$\alpha_{3:}C = \frac{\mu_{3:}C}{\sigma C^3} = \frac{1}{n^{3/2}} \sum_{i=1}^{n} \alpha_{3:x_i} \alpha_{3:y_i},$$

$$\alpha_{4:}C = \frac{\mu_{4:}C}{\sigma C^4} = \frac{1}{n^2} \left\{ \sum_{i=1}^{n} \alpha_{4:x_i} \alpha_{4:y_i} + 3n(n-1) \right\}.$$

In the case in which we are immediately interested, the variables x_i have identical distribution functions as do also the variables y_i. Then the above parameters reduce to

$$C = \frac{1}{n\sigma\sigma'} \sum_{i=1}^{n} x_i y_i,$$

$$E(C) = \text{Mean} = 0,$$

$$\mu_{2:}C = 1/n,$$

$$\mu_{3:}C = \frac{1}{n^2} (\alpha_{3:x} \alpha_{3:y})$$

$$\mu_{4:}C = \frac{1}{n^3} \left\{ \alpha_{4:x} \alpha_{4:y} + 3(n-1) \right\},$$ [3]

[2] $\alpha_{3:x_i}$ or $\sqrt{\beta_{1:x_i}}$ is defined to be skewness of the x_i distribution. In general $\alpha_{r:x_i}$ is the rth standard moment of x_i.

[3] The 4th standard moment or α_4 is called kurtosis. It is also written β_2.

$$\alpha_{3:C} = \frac{1}{\sqrt{n}} \alpha_{3:x}\, \alpha_{3:y}\,,$$

$$\alpha_{4:C} = \frac{1}{n} \left\{ \alpha_{4:x}\, \alpha_{4:y} + 3\,(n-1) \right\}.$$

It is obvious that the two latter parameters have the normal limiting values, 0 and 3, respectively, as n approaches infinity.

Under the assumption that $\frac{|x_i|}{\sigma_i}, \frac{|y_i|}{\sigma_i'} < \sqrt{N}$, then $|x_i\, y_i|/\sigma_i\, \sigma_i' < M$.

Setting $z_i = x_i\; y_i/\sigma_i\; \sigma_i'$, the variables z_i are bounded. Further,

$$\mu_{2:\, z_1 + \ldots + z_n} = \mu_{2:\, z_1} + \ldots + \mu_{2:z_n} = n$$

since the variance of standard variables equals one.

The conditions of the Laplace-Liapounoff Theorem (2, Second Case, p. 294) are fulfilled and we may conclude that the distribution of the mean of the z's or C has a limiting normal distribution. It is evident that the above conditions are fulfilled in all practical situations likely to be encountered in an applicaiton of the covariation method.

In the interest of convenience and with due regard for ordinary requirements of accuracy a Type III Pearson curve seems to be a satisfactory approximation to the distribution of C.

The following illustrates the C statistic: Suppose a first task of guessing 25 selections of an ESP card drawn from a deck is paired with guessing cards of a bridge deck with two known cards deleted. Again assume replacement after each guess so the probability of success is constantly and independently 1/50 from trial to trial in the second task. Let the deviations from expected values 5 and 1, respectively, be denoted by x and y.

The assumed set of twenty pairs of observed scores is as follows:

1st	6	4	3	5	8	9	5	3	3	5	7	10	4	2	0	5	3	1	7	8
2nd	2	1	0	0	1	3	1	0	0	1	2	4	1	0	0	1	2	2	3	1

This becomes

x	1	−1	−2	0	3	4	0	−2	−2	0	2	5	−1	−3	−5	0	−2	−4	2	3
y	1	0	−1	−1	0	2	0	−1	−1	0	1	3	0	−1	−1	0	1	1	2	0

$\Sigma xy = +38$, $n = 20$, $\sigma_x = 2$, $\sigma_y = \sqrt{50\,(.02)\,(.98)} = .990$,

$\alpha_{3:x} = .300$, $\alpha_{3:y} = .971$ since binomial $\alpha_3 = \frac{q-p}{\sigma}$.

$C = 38/(20)\ (2)\ (.990) = .960,\ \sigma C = 1/\sqrt{20} = .224.$

Critical ratio $= C/\sigma C = +4.29.$

$\alpha_{3:}C = .065.$

Upon referring to Salvosa's tables of areas for the Type III Curve (3) under $t = 4.29$ and interpolating between tabular skewnesses of 0 and .1, one obtains $p = .000021$. This assumes that a positive C was expected. If the sign of C were immaterial then this probability must be doubled.

REFERENCES

1. Murphy, Gardner and Taves, Ernest. Covariance methods in the comparison of extra-sensory tasks. *J. Parapsychology* (1939), **3**, 38-78.

2. Uspensky, J. V. Introduction to Mathematical Probability. New York: McGraw-Hill Book Company, Inc., pp. ix and 411 (1937).

3. Salvosa, L. R. Tables of Pearson's Type III function, *Annals of Mathematical Statistics* (1930), **1**, 191-198.

A THEORY OF EXTRA-SENSORY PERCEPTION

OLIVER L. REISER
Department of Philosophy, University of Pittsburgh

INTRODUCTION

In recent years we have been told to the point of weariness that the modern world is sick, afflicted with a mortal disease—a "sickness unto death" as Kierkegaard terms it. Again and again it has been stated that the world we live in is growing more chaotic and is headed for disaster. Everything is in a state of crisis, or is passing from one crisis to another, each succeeding crisis worse than its predecessor. We are told that socially we are geared too high for our own sanity; that our ethics is out of step with our scientific advance. . . . And so the diagnoses go.

It is interesting to note that the remedies proposed to cure the ills in many cases rest on a reversion to religious and philosophical primitivism. The modern Thomistic movement offers neo-Scholasticism as its cure for the aimlessness and lack of depth of modern society. The "crisis" theologians, Karl Barth and others, emphasize man's complete inability and utter dependence, and offer a return to supernaturalism as our human salvation. Nicholas Berdyaev tells us that "the world is entering an epoch of Caesarism," and argues that the present state of the world calls for a moral and spiritual revolution. In all these cases, the reader will observe, there is an obvious lack of faith in man's intelligence, in his ability to solve rationally the problems he has created. As opposed to such intellectual primitivism and cultural atavism the group known as Humanists maintain and assert their faith in the powers of human intelligence. Among moderns the Humanists stand alone in telling us that we must go forward rather than backward, that we need not less science but more. The Humanist may agree with the Oxford Group movement and the crisis theologians that we need a spiritual revolution, but he insists that it will have to be one inspired by a scientific understanding of nature; it must be guided by intelligence.

But how shall such a new scientific world view as I have hinted at, a vision which will provide an emotional outlet guided by intelligence, be attained? What new philosophical synthesis can again inspire man in this despairing age? Let us here try to picture for you such a vision, and then let you judge whether it may serve to again unite and inspire man to look forward with hope towards an uncertain future. The view I here present is one which has been in process of development for many years. The broader undertaking of which the present essay is a fragment involves a two-fold task: first, to demonstrate that a radically new mode of human thought and orientation will be operative in the future, and second, to indicate what the world will appear to be when it is understood in terms of these new principles. I have experienced great difficulty in trying to outline in brief form the broad features of a view which is so comprehensive and yet so technically intricate, but if the reader will follow with patience the following exposition I believe he will secure a good idea of the main features of this proposed synthesis.

The outlines of this theory were first presented in an article on "Emergence, Dimensionality, and Extra-Sensory Perception," which appeared in *Psyche* (London) in 1937 (Vol. 17). The present essay is an expanded version of that argument.

I.—FACTS AND THEORIES

As everyone knows, the belief in what is termed "supernormal" phenomena is very old. In recent years investigators in the field of psychical research have attempted to study such phenomena—"clairvoyance," "telepathy," and the like—under laboratory conditions. The results, real or spurious, have provided the occasion for much controversy. Among the recent investigations along this line we find the tests made by Dr. J. B. Rhine at Duke University. The results of these experiments seem to Dr. Rhine to justify the belief in what he terms *extra-sensory perception* (ESP). Dr. Rhine and his coworkers believe that they have considerable data supporting the genuine reality of clairvoyance and telepathy, and they maintain that the conditions of the experiments and the mathematical handling of the results are sufficiently beyond criticism to convince anyone who is reasonably open-minded. It goes without saying, however, that the critics of this work do not agree with this thesis.

Whatever may be the final conclusion, these investigations in the field of extra-sensory perception have attracted much attention. Attacks upon and defenses of this work in parapsychology (as it is called) have been frequent and lively. The present writer has witnessed some of these experiments and has been in touch with some of the individuals doing this work and has finally arrived at the conclusion that there may be "something in it." In trying to understand the unwillingness of Dr. Rhine's critics to go along with him in his statement of "facts" and theories, the conclusion has forced itself upon me that the opposition to his work is due in a large measure to our present inability to explain such results. This of course is a familiar mode of reasoning. It is the old "argument from inconceivability" set in a new context.

In the present paper I shall outline a theory which attempts to provide an explanatory foundation for extra-sensory perception. The aim of this presentation is not to try to prejudice the reader in favor of Dr. Rhine's conclusions, but merely to remove one difficulty which has made it impossible for some to approach this field with an open mind. The facts must be judged quite independently of our ability to explain them, but if a theory can be invented which will show how such facts might be understood, this may contribute towards the development of a more objective attitude in a highly controversial field.

In the present view we are not defending a dualistic theory of psychology. Our theory rigidly excludes supernaturalistic ideas; it excludes the miraculous and the inexplicable. The elements which enter into the construction of the present theory have already been defended by the writer on other occasions, quite apart from their relevance to the field of psychical research. The constituent elements of the theory which is here offered are these: (1) the theory of *emergent evolution;* (2) the system of reasoning termed *non-Aristotelian logic*; (3) the notion of *psychic levels*; and (4) the doctrine known as *religious humanism*. Now we aim to bring all ideas to a focus on the phenomena of extra-sensory perception.

It will be noticed that the last constituent element of our theory takes us into the field of religion. It will be obvious to the discerning that the ultimate aim of our enterprise is nothing less than a new theory of biological and mental evolution. Contrary to the view of those who hold that religious and philosophical considerations have no place in the development of a scientific hypothesis, I hold that what the world needs today is a synthesis which will bring together the interests represented by science, art, religion and philosophy. Before entering into the technical details of this new outlook, let us glance at the philosophical presuppositions of our theory.

II.—THE NEED FOR A NEW PHILOSOPHY

It is obvious that the world is undergoing a profound reorganization in thought and in social relations. To some students of society these fundamental readjustments are evidence of a general disintegration of European-born civilization as it passes into a new Dark Age, while to others, less pessimistic in outlook, these rapid and disturbing changes appear as the prelude to the emergence of a new type of culture. For the moment the question of which view is correct need not concern us. It is sufficient to note that in this environment of dislocations, where problems are so numerous and difficult that their solution seems to demand almost superhuman effort and intelligence, there is always danger that perplexed nations, like human individuals, will seek to dispose of apparently insuperable difficulties by adopting the devices of the mentally immature. Confronting an environment too perplexing for mastery, some individuals succumb to atavistic tendencies and return to simpler and more primitive modes of adjustment, such as are natural to a child. These mental regressions represent an escape from reality; here problems are solved by being ignored.

At the present time the nations of the world are faced by the necessity of evolving a new machinery of international understanding and cooperation. In the presence of this unprecedented demand there is the constant temptation to resort to social atavism. It may appear far-fetched to compare nationalism and political isolationism to the withdrawing reactions of an insane man, and yet the unwillingness to enter into a new level of integration does resemble the infantilism of an "adult" who seeks to return to the world of the child because he cannot adjust himself to the world of "grown-ups." That is to say, fascism, economic autarchy, and the like appear as regressions to earlier forms of political-economic organization which ought to be outmoded in the present world of interdependent units. These obsolete forms linger on primarily because we haven't yet discovered a method of creating positive techniques of international living. The disturbing thing is that even though we recognize the antiquated nature of these survivals, we can't put them where they belong—in the archaeological museum of social fossils.

If we think this situation through, it begins to appear that the fundamental difficulty here is largely a result of the failure of philosophy. Excessive nationalism is rampant because we have deep-seated urges which must canalize themselves through emotional outlets, but lacking any higher modes of expression they are forced to manifest themselves through the older forms of expression embodied in our inherited institutions. The political state, collapsing through the inherent frustrations of economic maladjustment, preserves itself by supplying the justification and occasions for such emotional orgies, paying the piper for the song by mortgaging the uncreated wealth of future generations. In an earlier age religion provided the needed emotional outlet, but modern science has all but destroyed the authority of the older religious appeal, without supplying any substitute. Men cry for a goal, for a purpose in life giving meaning to action. Finding nothing in the domain of contemporary science and religion, men allow themselves to be led back to the ancient flesh pots, the age-old outlets of chauvinism and social egotism (nationalism). Thus mass feelings find energy-escapements through atavistic forms of culture, and men worship the old tribal gods and pray at the shrine of hollow nationalistic personifications, knowing in their hearts that they have moved into a new world where men must find the "Unknown God"—else they perish.

Here and there we see evidence of a breaking with the old forms and techniques. But as we rise above the past, we experience its power to drag us back. None the less—and in spite of set-backs—we do glimpse the form of a new world order emerging: a world-state guided by a world-awareness and animated by a social consciousness born of science. Through radio, telegraph, rapid transit and newspapers a consciousness-of-the-world is being transformed into a world-consciousness. The new organ of integration is gradually crystallizing its own skeleton. Eventually the political, economic and religious motives will again unite, producing a philosophical synthesis quite radical and startling in character.

Let us see how this will come about.

On the physical side the world is growing into a new unity through manifold processes of integration. When the problem of wireless communication was solved, the range of man's auditory environment was enlarged to the point where we can now hear sounds at practically the same moment they are produced in any part of the world. The radio has thrown a girdle 'round the world which greatly extends man's environment. In a similar way the moving picture has projected man's visual world beyond the ordinary limitations of time. Events which happened in distant places and at former times can now be reproduced at will. Thus the instruments which enable us to transcend the normal restrictions of time and space are not only changing the content of our thought, but are also *intensifying our awareness.* The instruments of publicity—radio, newspapers, television, etc.—are now accepted institutions of society, and these are making us more aware of each other and of ourselves. By enlarging the sensory environment we are changing the inner life of the organism.

But how does this undoubted unification of the world through science contribute to the formation of this emerging synthesis which we have described as a new world religion? This we now propose to investigate.

III.—THE FUTURE OF MAN

We have referred to the consciousness-of-the-world which is developing. This phrase implies that we are all growing more *aware,* that our sensitiveness to reality is becoming richer and deeper and more intense. That this increased capacity for experience is related to the increasing complexity of man's physical and social environment is an idea we have already suggested. This must be so, if man's consciousness is social in reference. But on the biological side we have equally good reason for believing that as we enter into new environmental (social) relations our inner life expands. Thus the view that mind has reached its apex, that the wave of consciousness has finally and for all time reached its culmination in man *as now constituted,* is neither good sociology nor good biology.

To those persons whose thinking starts and stops with the "special creation" theory the idea that man is still in the process of being created may come as a shock. We might suggest in passing, however, that if one is looking for religious sanctions for this view we can always refer to the utterance of that ancient voice of hope—*it doth not yet appear what man shall be!* Whether we like it or not, that fact is that we are living in a world which is still incomplete. The philosopher Nietzsche, sometimes miscalled the Anti-Christ, saw clearly that *man is a bridge;* like the ape, he exists for what is to come after. Man must go beyond himself; he will be superseded by the superman. This coming man, however, will not be the creature imagined by Nietzsche. Instead of raising certain of our present human attributes to the *n*th degree, he will possess new psychic capacities not manifested at the present time in any considerable proportion of the human race. To show that such psychic functions as we are attributing to our future humanity are within the range of scientific possibilities, let us begin by looking at evolving man from a biological point of view. Here we have recourse to the views expressed by several eminent students in the field.

In his address before the American Philosophical Society, meeting at Philadelphia in 1929, Dr. Ales Hrdlicka argued that man, a product of biological evolution, is still evolving, and that there is practical certainty that this future evolution, as in the past, will be mainly in the direction of intellectual development. It is quite true that some biologists have argued the contrary view, assuming apparently that the limit of man's physical evolution has been reached, so that the next step lies in taking advantage of the vast potentialities of social evolution. While this is a debatable matter, Dr. Hrdlicka maintains that the further mental development he has postulated may be expected to be attended by an additional increase in brain size, although this gross increase will be of moderate proportions. The main changes, he thinks, will be in the internal organization of the brain, in a greater blood supply, and in an increased effectiveness in the use of the brain. Along somewhat the same lines, we find Dr. Frederick Tilney holding that even though it is true at the present time we make use of only one-fifth of our brain, nevertheless the brain of modern man is not a finished product. Remembering that the first known man made his appearance hundreds of thousands of years ago, and that since that time man's brain has increased in volume and acquired greater refinements in structural detail, it seems likely to Dr. Tilney that the present brain represents an intermediate stage in its ultimate development. In his treatise, *The Brain from Ape to Man,* Dr. Tilney reviews this steady advance, and then puts the question: "Is there still a possibility of further evolving in the development process so clearly seen in the brain of primates, so obviously reaching its present culmination in the brain of man—is there still a latent power in the human brain for the expression of yet unsuspected potentialities and beneficial progress?" This question was recently answered in positive terms by another student of living organisms, Dr. Alexis Carrel, in his book, *Man the Unknown.* Here these "unsuspected potentialities" turn out to be psychic powers such as have apparently been possessed by those who claim the "occult" gifts of clairvoyance and telepathy.

IV—THE EVOLUTION OF REASONING

The view that is here being advanced as a tentative hypothesis is in some respects similar to that put forth by Dr. Carrel, except that I wish to add that the changes which may take place may be so fundamental and far reaching as to involve the substitution of a new logic for the older logic which the human race has been employing for thousands of years. This new mode of thought may call for a revision of the ancient "laws of thought" which have regulated thinking ever since the time of Aristotle. In order to be concise, let me state at once that I am proceeding on the supposition that the evolution of human mentality during historical times may be summed up under three stages, as follows: (1) the pre-Aristotelian period; (2) the Aristotelian period; and (3) the non-Aristotelian period. We hold that primitive man functions on the first level of human mentality; that the human mind of today (of civilized nations) is functioning on the second level of Aristotelian logic; and that in the future the human mind will move on to the third level, the level of the non-Aristotelian mode of understanding.

Now let us briefly consider the characteristics, the "axioms," of each of the above three levels of orientation.

(1) *The Pre-Aristotelian Mentality.*—Here we have the stage of primitive man, who deals with nature in terms of wholes. The researches of Lévy-Bruhl have revealed that the primitive mind is "prelogical" in the sense that it does not conform to the categories which the reasoning of classical European science has established. Lévy-Bruhl is convinced that primitive man does not observe the fundamental canon of Aristotelian logic, the *law of contradiction,* but follows an entirely different principle which he designates by the term *participation.* On the first level, the pre-logical mode of adjustment, the axiom is: *everything is everything else.* Thus primitive man's *personifications* of nature are based on what have been called false identifications—"I am other things." The "animistic" system is an expression of mystical participation in the sense that it does not distinguish between the *self* and the *not-self.* There are no sharp dichotomies in nature because the Aristotelian "laws" of *identity, contradiction* and *excluded middle* are not respected.

(2) *The Aristotelian Mentality.*—On this next level of mental-social evolution we get sharp distinctions in nature. The reasoning of this level of orientation is based on the familiar "law of identity," that *A is A*: everything is identical with itself and distinct from the "other." Here the axiom is *this is this,* and *that is that,* and *this is not that.* This logic involves a sharp distinction between an "object"

and its "environment" and dichotomizes the *self* and the *not-self*. This is the logic of modern science, which undoubtedly took over the presuppositions of Greek (primarily Aristotelian) logic. Here, unlike primitive man's orientation, there is a separation of reasoning and emotion into distinct faculties, and the activities of science are connected with man's rational life while the affective responses are excluded from science (reasoning) and left to the domains of religion, poetry and "metaphysics."

(3) *The Non-Aristotelian Mentality.* — In proposing that the third stage of mental evolution is, or will be, the non-Aristotelian mode of orientation we mean that after the present age of specialization in science has passed, or has been supplemented by an era of coordination and synthesis of knowledge, we will attain an insight into the interconnectedness of things which will resemble primitive man's sense of "participation" in the sense that here, on a higher level, we again realize the limitations of the classical laws of thought. On this coming third level we return to the idea that *everything is everything else,* except that this non-Aristotelian principle (unlike the pre-logical principle of primitive mentality) will be based on an understanding of an underlying unity, provided by a sub-universe of continuity, so that the distinction between "object" and "environment" becomes relative. Individuality (identity) is to some extent illusory. In its ethical application this means that it is really true that we are our brother's keepers, and that he who would save his life must lose it.

We have already noted that one significant feature of the science and philosophy which develops in connection with Aristotelian logic is the separation of intellect and feeling, reason and emotion. The present emphasis on the part of positivistic philosophy on the study of cognitive meanings and the rigid exclusion of emotional elements, allegedly because of the affinity of feeling with poetry and religion, is only the latest consequence of this schism. The present impasse between sterile intellectualism and irrational emotionalism, running through the whole of modern life and separating religion and politics from the life of reason, is the unfortunate social consequence of this elementalistic psychology and cultural atomism. In the present organismic (non-elementalistic) view this dualism and consequent mental conflict is resolved.

The essential unity of nature and life which we have suggested is easily recognized in *mystical pantheism*, and this attitude is difficult to "understand" precisely because of its super-logical nature. The intuitions of such a view are finely portrayed in Emerson's tantalizing poem, *Brahma,* from which the following two stanzas are quoted:

> If the red slayer think he slays,
> Or if the slain think he is slain,
> They know not well the subtle ways
> I keep, and pass, and turn again.
>
> They reckon ill who leave me out;
> When me they fly, I am the wings;
> I am the doubter and the doubt,
> And I the hymn the Brahmin sings.

Such "mystical participation" is taken for granted in poetry. But that the present difficulties in science are due to the use of a faulty logic, and that their solution calls for a new mode of understanding in any way analogous to Emersonian pantheism, are things that require considerable proof. Nevertheless there is evidence indicating that the traditional "laws" dating back to Aristotle will have to be limited in their applications. The statement of Dr. A. N. Whitehead, that the world is in the midst of a most profound scientific revolution, only hints at what is coming. Not only have the new and revolutionary discoveries in physics upset our traditional ideas about the fundamentals of nature—*space, time* and *matter*—but the reconstruction in our thinking which physics necessitates goes much deeper. Following Alfred Korzybski's thesis, we have for a number of years argued that the Newtonian world-picture was based fundamentally on Euclidian geometry and the traditional Aristotelian laws of thought, and that this Aristotelian-Euclidian-Newtonian scheme of nature forms one coherent pattern. But now relativity physics and wave mechanics compel a modification of this classical world-view, and the new picture will be non-Aristotelian, non-Euclidian, and non-Newtonian.

We hold that the development of this modified view calls for a rejection of the time-honored "laws of thought," which will be replaced by new principles of orientation. We shall not attempt here to show this in detail.[1] But since I propose to show that Dr. Rhine's

[1] This has been done in the following articles: "Physics and the Laws of Thought," *Psyche,* 1931, Vol. 11, 70-78, "Non-Aristotelian Logics," *Monist,* 1935, Vol. 45, 100-

results in ESP also call for a rejection of classical science and the creation of a new scientific world-view, it is necessary at least to indicate roughly what, from our point of view, is wrong with classical physics. To be brief it is necessary to be dogmatic, and I therefore merely sketch in outline what seem to be the fundamental assumptions of the classical theory of nature. These are as follows:

ASSUMPTIONS OF CLASSICAL SCIENCE

(1) *Whatever is, is.* (This is the "law of identity.")

(2) *A thing is what it is.*

(3) *A thing is where it is.*

(4) *The same thing cannot be in two different places at the same time.*

(5) *Two different things cannot be in the same place at the same time.*

(6) *In order that any thing can get from one place to another it must move through the intervening space, and it must take some time to do this.*

(7) *The same thing, or event, can be observed from two different points of view at the same time.*

(8) *Two different events can happen simultaneously, and they can be observed as simultaneous from the same point of view.*

It is my contention that the discoveries of present-day science discredit the universal validity of these once-universally accepted axioms. For example, it is known to the experts that axioms (7) and (8) are upset by relativity physics, which rejects "simultaneity" of events in different frames of reference. That is to say, Einstein was led to the special theory of relativity by challenging the traditional idea that two events can happen in *different* places at the *same* time. Again, to pass on to a simpler situation, the absolute truth of axiom (4) is challenged by evidence showing that, in a sense, *two bodies may occupy the same space at the same time* (this appears in quantum mechanics). Later on I shall indicate the evidence disproving axiom (6), *i.e.*, evidence shows that, in a sense, *the same body may be in two different places at the same time.* This means that certain supposed fundamental relations between objects (or "matter") in space

117; "Non-Aristotelian Logic and the Crisis in Science," *Scientia*, 1937, Vol. 61, 137-150; and "Aristotelian, Galilean and Non-Aristotelian Modes of Thinking," *Psychological Review*, 1939, Vol. 46, 151-162.

and time (relations which classical physics took as axiomatic) are now discovered to be valid *only within certain limits.* Thus we now find that physics and logic must revise their ideas of what is "possible" in nature. Logic cannot escape this revision because the "laws of thought" have historically been interpreted as laws of reality.

The significance of this development for ESP research can readily be seen by turning for a moment to Dr. Rhine's results. In experiments where subjects were set to the task of calling cards at a distance of several hundred miles the results which Dr. Rhine amassed, and reported in his book on *Extra-Sensory Perception,* indicate that the ordinary laws of radiation do not hold, and this suggests that a non-radiant energy is at work in ESP. These facts of distance clairvoyance and telepathy therefore bring us face to face with the circumstance that space relations and possibly time relations also are not binding for the mind as they were supposed to be for the physical world in classical physical science. If Dr. Rhine's results are valid, they necessitate the acceptance of a kind of energetics not limited by the customary inverse-square law, *i.e.,* there is no decrease of effectiveness of extra-sensory perception with increase of distance, as is the case for known energies. *Since this physical law is a consequence of the geometrical properties of Euclidian space, and is necessitated by the Newtonian law of force, the results of Dr. Rhine really seem to suggest the need for a non-Aristotelian logic in this field.* Of course the validity of this argument rests to a considerable extent on the soundness of our prior thesis that *Newtonian physics is indeed an exfoliation of the presuppositions of Aristotelian logic and metaphysics,* as that synthesis was passed over the historical bridge of Euclidian geometry to become the conceptual framework of the Cartesian-Newtonian mechanistic physics of modern science. The defence of this thesis was presented in our *Scientia* article (previously referred to) and cannot be repeated here.

V.—EMERGENT DIMENSIONALITY

We have said that the theory of emergent evolution forms an integral part of the new philosophy of nature on the basis of which we shall attempt to erect a humanistic religion for mankind. Let us see how this is to be accomplished.

Various writers have different interpretations of the meaning of "emergent evolution." For us it is a name for the process whereby the "ultimate particles" out of which all things are made (possibly positive and negative electricity) combine and recombine in ever-increasing degrees of complexity to produce new and higher syntheses or "organisms." (On this view, which here agrees with Whitehead's philosophy, even atoms are organisms.) Our own development of this idea of the emergent evolution of progressively more intricate behavior-complexes is connected with the notion of a historically new or emergent dimension, a concept which was deliberately framed to provide a reconciliation of the *relativity* of motion (as Einstein treats it) and the *absolutivity* of motion.[2]

We recognize that the type of motion in which the science of mechanics is primarily interested is subject to all the principles of Einstein's theory of the relativity of motion. But *growth* and *evolution* (biological and psychological), types of motion (change) in which physics has hitherto not been interested, are *not* relative. This is a form of change to which present Einsteinian relativity does not apply. Motion as represented by the fourth coordinate of the space-time continuum is relative; but evolution, we insist, calls for a new dimension of time (a new form of temporal organization). This historically new dimension of growth is the $n + 1$ dimension, where "n" is any lower and earlier spatial dimension of "materiality" out of which this higher temporal organization of growth appears. Thus on our conception emergence adds a "degree of reality" to any "lower" plane of being.

Whenever we can refer to a system as a whole, with its spatial coordinates and its own "local" time, this time is transposable across the parts, as Professor Wheeler says.[3] If now this system (K_1) enters into dynamical interaction with another system (K_2), the two together may form a new system, and this system, so long as it is treated as a whole, will have its own (emergent) time transposable across the whole. *This new ("public") time is what we mean by the emergent dimension.* The "social" order which brings an emergent public time out of the "local" times of the individual constituents may even have its own "emergent mass," as George H. Mead puts it, and this, for us, represents the *field* or *gestalt* property of the family of subordinate systems. The significance of this idea is that it allows us to utilize the notion of an "absolute" time in what we call organismic

[2] This idea was first expounded in my volume, *Philosophy and the Concepts of Modern Science*, 1935, Chapters 1 and 8.

[3] Cf. "Organismic Logic in the History of Science," by R. H. Wheeler, *Philosophy of Science*, 1936, Vol. 3, 26-61.

or non-elementalistic situations. That is, in such cases we can determine whether one event is "simultaneous" with another when they can be "experienced" together by the "consciousness" of the "organism" which spans the local time of its own atomic (or "cellular") constituents. Aside from its importance for psychic phenomena (to be discussed later), it is interesting to note that this idea can be used to resolve the famous wave-particle difficulty in physics.

It is now generally known that in one set of experiments light acts *as if* it were a wave-phenomenon, and yet in another set of experiments light clearly acts *as if* it were a corpuscular-phenomenon. But what this situation really means, I think, is this: we must now recognize that the former separation in physics of the "observer" and the "observed"—in this case the "sink" and the "source" of radiation—is artificial; they both play correlative rôles. Light is a manifestation of a non-elementalistic or wholeness situation, and "particle" and "wave" concepts taken alone and in isolation give only part of the story. This is our reinterpretation of what Niels Bohr terms the principle of *complementarity*.[4]

At this point we pause for a moment to exhibit how this notion of "organic" time as an emergent coordinate associated with the unique (absolute) dimension of growth and evolution fits in with our non-Aristotelian approach. The process of emergence, by means of which a thing changes (ceases to be what it was and becomes what it is), defies the laws of Aristotelian logic in the sense that it is unintelligible in terms of the traditional "laws" of thought. This unintelligibility in terms of Aristotelian habits of thinking is curiously reminiscent of the difficulties inherent in Zeno's paradoxes of motion. To see this let us turn to the ancient Greeks for a moment.

In order to show that Zeno's paradoxes of motion are indeed related to the "laws" of traditional logic which Aristotle stabilized, let us note first of all that the "law of excluded middle" would be strictly applicable in a universe of discontinuous movement, but it does not hold in a temporally continuous process. This is illustrated by the first premise of one of Zeno's arguments: *A thing must either*

[4] A statement of this reinterpretation and its significance for non-Aristotelian logic is given by the writer in an article on "Physics, Probability and Multi-valued Logic," to appear in a forthcoming article in the *Philosophical Review*. In this article we also try to show that our theory of the relativity of the observed to the observer (in a non-elementalistic situation) is in harmony with A. S. Eddington's theory of the conjugate rôle of the "thing" and its "comparison object." For the electron this is the universe as a whole.

move where it is or where it isn't. This is the law of *tertium non datur,* or excluded middle, that *A is either B or non-B,* but not both. Zeno then continues: But a thing cannot move where it is; neither can it move where it is not; therefore, motion is impossible! Or putting the argument in symbolic form:

$$m < w + w'$$
$$m \nless w$$
$$m \nless w'$$
$$\therefore m = 0$$

Now the difficulty here is that motion is precisely the process whereby a thing gets from where it is to where it wasn't: a third possibility which the law of excluded middle completely overlooks. In reality, therefore, $m < w + w' + (w \rightarrow w')$. Thus we agree with Brouwer that Aristotelian logic was derived from an abstraction from the mathematics of *finite* classes (and *discontinuous* processes), which was then universalized. Brouwer goes on to argue that the law of excluded middle is inapplicable to (cannot be shown to hold for) the domain of the *transfinite.* But the idea of *infinity* (along with that of *continuity*) underlies the whole modern mathematical analysis of motion (in differential calculus). And so Brouwer, like Hegel, must reject the modern handling of the problem of motion and change, and like Hegel, though for somewhat different reasons, he is forced to deny the applicability of one of the classical laws of thought of Aristotelian logic. (Those versed in the technical details of philosophy will note here that Brouwer agrees with Bergson's view that continuity cannot be handled in the classical manner as a completed aggregate of points.) In our own view we try to bring these two views (of Hegel and Brouwer) together. We hold that the symbol $(w \rightarrow w')$ represents neither logical addition nor logical multiplication, nor any other operation of traditional or modern symbolic logic. This is what in mathematics introduces continuity and infinity into the analysis of motion and change, but for us this is what symbolizes the passage from the "is" to the "is not" of any changing or evolving entity. But especially the symbol $(\rightarrow)$ designates the change or growth whereby the new time-dimension emerges. *At this point the particle-aspect, associated with "identity," is lost in the emergence of a phenomenally new whole, with its own time-system transposable across the parts.*

VI—AN OBJECTION CONSIDERED AND A COMPARISON

Of course the author is aware of the difficulties inherent in this theory. One objection to our conception arises in connection with our unorthodox use of the term "dimension." It will be pointed out that the use of the notion of higher dimensions (or hyper-space) in relativity physics, or the multi-dimensional phase space of wave mechanics, has no physical significance. These tricks of non-Euclidian geometry, it will be stated, imply nothing physical beyond three dimensions: the *n*-dimensional manifold is a dodge which must be interpreted to refer to (a) the number of *independent variables* of some physical system, or (b) the number of *degrees of freedom* of a configuration.

In replying to this point I can only say that the Euclidian-Newtonian assumption that the only "real" spatial dimensions of the physical world are the three coordinates of classical physics (*i.e.,* Cartesian coordinates) is a naive view which came, perhaps, from the acceptance of the Pythagorean-Platonic-Aristotelian doctrine that "God geometrizes," and does so only in accordance with the scheme of Greek logic and mathematics. Now we should know better.[5] The detailed exposition of our broader definition of dimension as the emergent time-axis cannot be undertaken here. On another occasion I shall try to show how the current conception of the "spin" of the electron, which turns space into time and is connected with the principle of the indistinguishability of electrons and the "inter-change energy" (or *resonance* between the elements) of newly-forming aggregates, may find its place in this theory. This, however, is for the future. Now I am concerned to differentiate this view from a somewhat similar conception presented by a British investigator,

Those who are acquainted with the view of Mr. J. W. Dunne, first expounded in his book, *An Experiment with Time* (1927), and more recently set forth in his volume, *The Serial Universe,* may seem to detect a similarity between that view and the one here advocated. In the serial universe time has a "regressive" character; for example, the time-dimension for a three-dimensional observer is merely the direction in which his field of presentation is traveling in a four-

[5] It is interesting to note that in his latest attempt at a theory to link gravitation and electricity into one unified field theory, which will explain all physical happenings in one broad concept, Einstein has found it necessary to introduce a *fifth* dimension. Thus Einstein takes the idea of Professor Theodor Kaluza, who used the idea of a fifth dimension as a mathematical notion without physical meaning, and ascribes physical reality to the fifth dimension. For us these additional dimensions beyond the bare space-time-matter level (e.g., the levels of life, mind, social organisms, etc.) are not given as antecedent realities through all "eternity," but they "emerge" as nature evolves.

dimensional manifold. Thus every time-traveling field of perception is contained within a field one dimension higher. The symbol, $\sqrt{-1}$, by means of which orthodox relativity transforms time into space, in Mr. Dunne's view represents the rotation of an axis of time until its features coincide with those of the time of the next lowest geometrical map. Mr. Dunne believes that, in terms of this theory, he is able to show how the "perception" of events, for example in dreams, might precede the actual happening of these events in our familiar physical world.

One serious objection to this view arises in connection with the postulation of an "infinite regress of time dimensions." This leads Mr. Dunne into the difficult concept of the "Observer at Infinity." In our own view, as in Mr. Dunne's, we accept the idea that in the physical application of multi-dimensional geometry time plays the rôle of the next highest dimension, but for us this is an "emergent," something historically new, which is generated by the aggregation of matter in the process of producing a whole which has associated with it a public time, binding the parts into a dynamic synthesis. Even though, on our theory, the number of such dimensions which may emerge may be unlimited, there is no "Observer at Infinity" who can look back upon the serial order and resolve it into a spatial manifold of a lower order. For us the emergent time dimension is real; for Mr. Dunne it appears to be illusory.

VII.—ORGANISMIC TIME

As an illustration of the macroscopic (public) time transposable across the microscopic parts and binding the source and sink of radiation together we may cite the example given by C. G. Darwin in his book, *The New Conceptions of Matter* (1931). In connection with his exposition of quantum mechanics (p. 90) Darwin refers to the experiment in which it appears that electrons will not pass through one small hole in a shutter unless another hole is made close beside it. (The statement of the details of this experiment is much clearer in J. H. Jeans' book, *The New Background of Science*, 1933, p. 159.) In Darwin's interpretation "the only possible way of explaining this is to say that each of the electrons knows all about both holes, or has gone through *both* holes at the same time, because only thus could we get the cancelling effect characteristic of interference." But only five pages later (p. 95) in the same volume Darwin points out that Einstein, in upsetting the idea of absolute time in nature, showed experimentally that "in fact it is really impossible to determine whether two events in different places occur at the same instant." This seems to contradict the statement on page 90. But the contradiction is only apparent.

In the first case (the electrons going through both holes simultaneously) the result is a part of one total experiment (or wholeness-situation) and no problem of relativity of time-measurements is involved. That is, the various parts of the instrumental set-up are surveyed as parts of one common public time; whereas in time-measurements across two independent coordinate systems (frames of reference) there is no measure of absolute simultaneity, *unless both systems become part of a more inclusive system which in turn is treated as a whole.* The transition in organisms from intra-cellular to cellular, or from cellular to inter-cellular synthesis, is an illustration of this gestalt (field) property in which the "local" times of the atomic constituents are incorporated into a public time transposable across the parts. Later on we shall indicate the analogous linkage of the "sender" and "receiver" which occur when extra-sensory processes act "across" space and time to unite the mind and its object through what we shall term a "psychic level."

VIII.—SPACE, TIME AND EXTRA-SENSORY PERCEPTION

The reader may well wonder what all this has to do with ESP and our new philosophy of nature. In replying to this query let us first go back for a moment. We have already made the point that the new world picture of science is modifying our conceptions of what is "possible" in nature. Following this up in more detail, we now urge that the advent of non-Aristotelian principles of reasoning teaches us (a) that certain things (or events), such as ESP phenomena, may appear possible in nature when our minds are freed from slavery to traditional habits of thought, and (b) certain phenomena of a "psychic" nature may become "understandable" and even more easily manifested when human minds begin to function uniformly on the coming non-Aristotelian level of orientation.

Next we point out that the notions of telepathy and clairvoyance (both of which involve communication at a distance appearing to violate familiar space and time relations) can be made rational (if they can be made intelligible at all) only through the notion of an "absolute" time. For example, if a person receives telepathic messages about events which happen at some distant part of the world at the same instant at which they happen, this means that somehow distant events can be "simultaneous" and can be experienced as such. "Premonitions" and "precognition" probably also presuppose an ability to place events in an absolute time-scale.

After the writer formulated the foregoing theory of emergent organismic time it was then discovered that at least one other author had speculated along similar lines, and I quote the following passage of Professor A. P. Ushenko's volume, *The Philosophy of Relativity* (London, 1937, p. 49), as indicative of this parallel conception:

> The assertion that there is no physical interaction between distant events must not shut the door on the possibility of metaphysical instantaneous transactions at a distance. Even the ordinary functions within a single organism, the organic relationships, might easily happen to be on a level which is to a certain extent free from purely physical restrictions. For it seems to be a fact that the whole volume of one's body may be sensed at the same instant; and one may speculate whether this togetherness of all parts within an organism is capable of extension to its environment, as, for example, when a fencer learns to feel with the tip of his rapier. Also there are believers in telepathy and in the instantaneous propagation of emotional influences. All such opinions could be allowed for, if one conceives the world as a hierarchy of ontological levels, of which the physical level gives the basic framework of temporarily unrelated events (contemporaries) as a field of potentiality for their various interrelations on the higher levels, the organic transactions being, perhaps, the simplest mode of such inter-relation, beyond which there may be other as yet unexplored modes. This is a fertile source for metaphysical conjectures.

The main difference between the above view as stated by Dr. Ushenko and my own theory lies in the fact that for me there is (as in the experiment cited by Darwin) room even in physics for a dynamic unity of source and sink which makes possible a public time transposable across the parts of the physical situation.

It is now clear to the reader that our own theory of extra-sensory perception involves the notion of an organismic situation binding the members of the human race together. In the present view we are forced to assume that the "local" time of each human individual is now, through a process of "mutual aggregation" (to borrow a phrase from Josiah Royce), beginning to cohere into such a group time. Telepathy, clairvoyance and the like may turn out to be indications of this dynamic unity whereby a new social whole is emerging. But what is this emerging organism which is producing a public time as a new time-axis? And how shall we understand and explain such a remarkable event in biological evolution? Before attempting to answer these questions, which take us into the field of biology, let us restate the physical basis of our theory.

We have argued that the advances in contemporary physical science serve to make us more open-minded about possibilities in the field of extra-sensory perception. In time the recent revolutions in physics will help create a new type of theory in psychology. Already, under the influence of gestalt theory, based on *field* physics, modern psychology is being led step by step closer to the idea of consciousness as a pulsating electromagnetic field in and around the brain and the central nervous system. This idea provides a mechanism for the instantaneousness and richness of content of conscious experience and helps us to understand some of the newly-discovered facts of *electroencephalograms* (cortical rhythms due to changes of electrical potential in the brain). It may also help us to explain the evolutionary intensification of consciousness previously discussed. In the human brain the movement of liquid ions to and from colloid interfaces cannot give a sufficient degree of speed and flexibility for psychic life.

But even this modification of traditional brain physiology is not sufficient to explain Rhine's results. For if consciousness were some form of familiar physical radiation it would obey the usual inverse-square law. But in Rhine's experiments on distance-telepathy—as we have already observed—the results do not fall off with increase of distance. This ability of mind to triumph over what the older physics would regard as the normal limitations of time and space is one of the most interesting features of Rhine's work. Actually the "new" physics also has followed this tendency to transcend the older limiting conditions of nature, and this is true whether one is thinking in terms of relativity theory or of quantum theory, in either their earlier or later forms. Let us consider this for a moment in more detail, before passing on to an examination of the biological-social organism.

Turning first to relativity theory, it is very interesting to note that while Einstein's theory states that the velocity of any form of radiation cannot exceed the limiting velocity of light, this does not exclude the idea (presented in the de Broglie-Schrödinger theory) that *certain kinds of group waves can travel at any velocity*, and this does not contradict the teaching of relativity physics concerning the constancy of the velocity of light. In the more recent theorizing of Dirac there is another departure from the ordinary ideas of relativity, in the sense that in the interior of the electron it is possible for a signal to be transmitted faster than light. Here there is a region of failure of the elementary properties of space-time. Now since in the physical world the amount of space to be traversed, or involved in the transmission of influences, is inversely proportional to the velocity of transmission of such influences, we can say that space progressively loses its reality as a limiting condition in nature as we increase the velocity. In the above case, in which a source and sink of radiation are linked together into a dynamic whole, we can say that the space-time interval of separation is zero, or we can introduce the notion of "virtual contact," or we can try to cover the situation by saying that the influence in its transmission approaches an "infinite velocity." But in our own language we would say that this is a wholeness-situation, not explicable in elementalistic-atomistic terms.

It is interesting to note that in quantum phenomena similar problems arise. The original photoelectric effect is still with us: light of a certain *frequency* will knock out an electron from an atom, and this is quite independent of the *intensity* of the light; in this case it makes no difference how far distant the atom is from the source of the radiation. In a sense this is similar to distance-telepathy results, except that the transmission of the symbol on Dr. Rhine's cards is a more complicated affair. This "mystery" of physics (the photoelectric effect) was one of the origins of quantum theory, which later was transformed into "wave mechanics." The latest speculations in this field commit the physicist to the doctrine (altogether inexplicable on the older Newtonian-particle physics) that when a "particle" passes through a slit it may be considered as a group of waves and the frequency of each harmonic train in the group is changed by the modulation due to the shutter of the slit.

In this fashion we see that no matter how we "take" our modern physics, it still remains true that mutual influences and transactions are possible which the older physics, with its antiquated ideas of "space," "time" and "matter," would have been forced to declare impossible. About the only outstanding scientist who realizes the significance of these new physical ideas for psychology and has the courage to declare it is Professor J. B. S. Haldane. In his recent book, *The Marxist Philosophy and the Sciences* (1939), Professor Haldane states (p. 169): "I do not see why a dialectical materialist should reject *a priori* the possibility of such alleged phenomena as telepathy and clairvoyance . . . if their occurrence should be proved, I do not think this would disprove materialism, or even revolutionize science; though it would open up an important new field, and very probably facilitate the study of the human mind as a natural phenomenon." Haldane is led to this supposition as a result of considering the discovery that elementary particles will leak through a "potential barrier" which they could never cross if classical physics were true. As Haldane says, "the fact is that, whether or not we take the wave system as a reality, the electron is influenced by surrounding objects in a manner not contemplated by physics up till the last twelve years" (p. 168).

Surely it is something of a triumph that the old dilemma of "action at a distance" *versus* "action by contact" is now solved in a non-elementalistic logic. This "organismic" situation is unintelligible in a completely "atomistic" view of nature, and it provides the physical homologue for such "short-circuiting" processes as we shall later assume to occur in psychical processes. Before leaving this matter let us once again emphasize that on our theory space and time are not antecedent realities (like vessels or containers) into which things are put. They emerge as simpler entities of a lower order interact to produce more complex aggregates. On the human level the space and time of a *psychic continuum* (or a *psychic level*) are conditions for mental interactions, as we shall now try to show.

IX.—WORLD AWARENESS

Just as in previous pages I have made use of the results of Dr. Rhine, so now I am going to incorporate into our new world religion the views of another investigator, Dr. C. Hilton Rice, whose approach to the study of man is from the side of medicine.[6] Dr. Rice's central insight into the approaching unity of mankind is based on the fundamental thesis that *the organic kingdom as a whole is literally and in fact an organism, with the human race taking the place of the developing nervous system (the neuroblasts) of this organism.* That is, the organic kingdom as a whole is the body of a single embryonic and

[6] The untimely death of Dr. Rice in 1937 occurred before he could publish his treatise on *The Visible Organism*, though an abstract of the theory appeared in *Psyche* (London), January, 1929. Dr. Rice was one of Dr. Rhine's best subjects in ESP research, and some of his results, edited by Dr. J. G. Pratt, were presented in the *Journal of Parapsychology*, December, 1937.

developing being that is feeding upon the substance of a gigantic egg, the earth. According to this picture of the world organism the plant and animal kingdoms form (functionally) the endoderm and ectoderm of a super-organism, and the human race serves as the nervous system of the embryo. This evolving system of life is operated by the energy from the sun, which fabricates the essential substances which the earth-yolk feeds to its embryo. As we have said, the nervous system of this super-organism consists of the sum total of the nervous systems, with the human race functioning as the cerebrum, the whole held together, not by "material" continuity as in the case of the cellular structures of the component parts, but by the "herd instinct," the "group mind," etc.

The way in which this sun-planet-organism hook-up is maintained, so that the outer layer of air and the inner layer of rock enclose between them a layer of water, through which the rock-layer protrudes to form continents and islands, is a matter for science to investigate. For Dr. Rice the most interesting phase of this developing embryonic being is the manner in which a great composite mind is beginning to dawn and reveal its form and potentialities in the social consciousness. Thus man's deep religious sense finds its confirmation in the coming into existence of a being in whose image man fancies he has been created. And just as the unborn babe cannot know and communicate with its parents until it has developed the ears to hear and the eyes to see, so this huge embryo, the earth-organism, cannot know itself until it has developed organs of sight and hearing, faint anticipations of which we now see in radio and radio-vision. These are the precursors of the extra-sensory perception of Humanity, the brain of the embryonic earth-organism.

In his theory of extra-sensory perception Dr. Rice puts forth the suggestion that our sense centers are two-way mechanisms that register impressions both from the sense organs and from the cortex. How, he asks, does a child in night terrors see objects that have no reality? His wide staring eyes show plainly that the object is registered in the sight center and projected outward by a reversal of the mechanism of distance reception. Apparently a part of the cortex (in the sleeping state) is sending impulses to the cells of the visual center and these impulses are transmitted into images of things "seen." In what is known as "eidetic imagery" the phenomenon occurs in the waking state. In short, it looks to Dr. Rice very much as though the visual center may respond to both sensory and non-sensory stimuli, the one type coming from without and the other from "within." In extrasensory perception the cortical cells may act as receptors and transmit impulses to be interpreted visually.

Such is the reasoning of Dr. Rice. Now let me integrate this with some speculation of my own.

The "physical" basis of the new psychic unity of mankind, if it is ever to be attained and understood, must be pictured first of all in its most general terms. As we have indicated in our volume, *Philosophy and the Concept of Modern Science* (Ch. VII), we make the general assumption that the activity of any entity of nature (electron, atom, cell, organism) always takes place within a field or environment. The entity itself is a behavioral unity of its constituents, and any entity, plus its field, yields an entity of a higher order. This field, or "level," is a result of a compounding of the microscopic fields to produce a macroscopic field. Thus the synthesis of residual atomic fields produces a molecular field, and the compounding of molecular fields produces a molar field. As we have already noted, the explanation of telepathy and clairvoyance seems to demand some sort of psychic level or continuum, and this, we have surmised, may arise out of a compounding of biological (cortical) fields to produce a super-organic field. Thus, just as a molecular field is created, by the synthesis of the electromagnetic fields of the atomic constituents, so the mental fields of each human brain, under appropriate conditions, might be responsible for the creation of a psychic level.

We hold that just as each synapse levies a minute toll on each nerve process to build up a psychic field which forms the basis of the "consciousness" of each individual human being, so each human consciousness makes its contribution to a collective consciousness, a psychic continuum or level which is the medium of interaction in telepathic and clairvoyant rapport. Just as molecular fields utilize (and are created out of) the residual electromagnetic fields (or unsaturated bonds) of atoms, so the residual fields of human brain fields combine to produce a super-individual field. But even though we suppose that in this fashion a collective human consciousness is being born, the theory that the human race constitutes the neuroblasts (embryonic nerve cells) of the developing earth-organism places restrictions upon the theory of emergent evolution. In other words the form of humanity already exists as the potential framework guiding the whole course of biological evolution, random and haphazard as that may appear to be. The potential form of humanity acts as a morphogenetic field of force controlling neuroblasts of the embryonic organism so that the "mutations" behind man's evolving psychic faculties are not due completely to "chance." The emerging world-organism helps create the inter-personal continuity which its gradual synthesis heralds and foreshadows.

X.—Some Unsolved Problems

It is true that this view still requires further development. Many questions can be raised which are difficult to answer. We have already discussed the physicist's possible difficulty with our conception of "dimensionality." And here are some additional problems: Why has the existence of a medium such as we have postulated—a psychic field—never been experimentally demonstrated through the use of physical apparatus? Why are those who are gifted with ESP so relatively rare in our population? And why do these psychic powers appear to run in families? And why are ESP faculties so flickering and fitful, so readily fatigued, and so uncertain in manifestation? These and other interesting questions still remain to be answered. I cannot reply to all these questions—even if I knew the answers!—but we can throw out a few suggestions.

With reference to the first question, we may suppose that, aside from the possibility that the psychic medium or continuum may not be susceptible of investigation by physical means, negative results may be due to the fact that the psychic ether is still in the process of being created. Or again, failure to detect the presence of a "psychic field" might be explained on the theory that the phenomena of a higher dimension cannot be trapped in the instruments of a lower dimension. In connection with this suggestion we may point out that our theory of the emergent dimension as a new form of spatio-temporal organization results in a theory of a "level" or field not subject to investigation by such experiments as the famous Michelson-Morley experiment. This experiment was performed to decide whether or not the luminiferous ether was dragged along by the earth in its onward motion through space. The fact that all experiments on the motion of material bodies relative to the ether have led to negative results (except the controversial results of Dayton C. Miller) does not discredit the notion of a "field" as we employ it. The Michelson-Morley experiment was performed on and about the earth only, and no other body played a part, hence it might be argued that the result is what should have been expected: the earth is at rest relative to itself; while relative to the sun the earth moves and relative to the earth the sun moves. In so far as both the sun and the earth (plus the other bodies of our solar system) enter into a dynamical configuration which makes it possible to treat the sun-planet system as a whole there is a "cosmic ether," but this cannot be detected by experiments *within* the system.

Returning to the suggestion that a psychic continuum for the human race is still in the process of creation, we surmise that perhaps a true social mind is being generated by the gradual synthesis of a super-mental field, the physical basis of which is the sum total of all nervous systems. If it is true, as Sir James H. Jeans has imagined, that each individual consciousness is a brain cell in the universal mind, then the present inter-communication between disparate areas *within* the individual human brain will be paralleled in the world-mind by direct communication *between* human minds.

The remaining questions cannot be answered satisfactorily, but it does appear that there is some hereditary basis for the presence of psychic faculties in certain individuals. This may be due to a mutation which, once it occurs, is biologically established and continues to reappear in subsequent generations so long as they are not eliminated in the "struggle for existence." This aptitude may rest upon some change within the brain or in the body generally. Since the faculty of ESP seems to be related to unusual powers of "concentration" (or possibly "integration," accompanied by a corresponding detachment so far as the immediate environment—"distracting stimuli"—is concerned), this may arise out of some change in cerebral chemistry which permits an unusual type of orientation of molecules at the biological interfaces which give rise to electromotive forces (bioelectric potentials). In our volume, previously referred to, we have pointed out that potassium, the only (spontaneously) radioactive substance in the body, has the power to facilitate such molecular orientation, and we have proposed that in this fact may be found one clue to the interaction of mental fields and biochemical processes in the brain. A final possibility which we must mention is that possibly the explanation of these unusual powers will be found in the new ideas which chemists are taking over from wave mechanics to explain "chemical affinity." The notions of "resonance energy," "electron interchange," the "spin" of atomic particles, and the like we are bound to hear much of in the future.

These are a few of the suggestions which can be brought forth to explain the body of facts which investigators of things psychic have turned up. They all indicate that it is at least possible to conceive of some sort of "mechanism" whereby the space-time intervals which normally isolate individuals from each other may be overcome. If we are indeed moving towards the creation of a world-mind where direct inter-personal continuity is established, we get a new insight into human motivation. Thus the normal human craving for fellowship (the "herd instinct") appears not merely as a *vis a tergo,* a psychic regression to the group mind of primitive man; it is also a *vis a fronté,* a striving towards a higher unity—the next emergent level of nature.

If all this be true, as time goes on the "law of identity" will become even less satisfactory as a description of human individuality. And thus we are confirmed in our conclusion that extra-sensory perception, defying the time-honored laws of Aristotelian logic in their scientific applications, is but a feeble and uncertain intimation of psychic powers yet to be evolved and perhaps eventually to become universal in the human species. *Evolution is not yet through with the human organism, for still higher functions remain to be developed.* Humanity thus appears as a god in embryo, a developing being with the psychic powers—omniscience and omnipresence—which man has hitherto assigned to his God. Perhaps man will eventually find that he is made in the image of God because God is being made in the image of Humanity.

XI.—EVOLUTION AND THE NEW HUMANISM

This doctrine that man is a potential god is of course one form of humanism. As Charles Francis Potter has stated, when the radical nature of humanism is recognized, its truly revolutionary possibilities will become manifest. This is the only non-supernaturalistic religion which can recapture the moral idealism and emotional drive of the ancient and obsolete forms of religious expression. The way in which the direct realization of the unity of mankind may help to create a new technique of political and economic living is a matter beyond the scope of the present essay. But if the picture I have here tried to paint for you is correct in its mean features, you may be sure that a new type of social science will eventually be on its way to realization.

A REVIEW OF RECENT CRITICISMS OF ESP RESEARCH, II

C. E. Stuart

Duke University

Abstract: Critical discussions of ESP research, published since mid-1938, are reviewed. Those examined are Kellogg's criticism of mathematical techniques; Kennedy's application of alternative hypotheses to the whole field of research; Leuba's and Lemmon's selection criticisms; Goodfellow's, Lund's, and Fernberger's proposal of the effect of preferences and suggestions; and Rogosin's social evaluation of the research. It is noted that "optional stopping" presents a difficult problem experimentally, although mathematically solutions are available.

Since the publication of the last article under this title in 1938, critical statements regarding ESP research have continued to appear in various journals. In the main, there has been no addition to the points of criticism previously raised, but the fact that many continue to be raised indicates the need of keeping them under discussion. There is little to be gained by mere repetition of argument, however, so the following selection of points for discussion is determined more by this writer's judgment of interesting, or important, or typical criticisms rather than by an attempt to be exhaustively complete.

Criticisms of Mathematical Techniques

C. E. Kellogg's "A Note in Reply to Mr. Charles E. Stuart" stands alone among recent criticism as being primarily directed to the mathematical evaluation of ESP results. A complete rebuttal to this "Reply" would require an equally lengthy note and would be of interest to very few persons other than the principals. Certain repetitions, however, may be of general interest.

Chance Hypotheses. Dr. Kellogg repeated once more his view that "the discrepancies between the methods [Normal, Binomial, and Matching] increase for high scores with increase in the number of runs . . . , " but since he does not refer to the opposing view discussed by Stuart and Greenwood (18), it is important to clarify the difference in such a way as to close controversy or reopen it upon some common ground.

The Normal, Binomial, and Pure Matching hypotheses are mathematical approximations to the various statistics to be expected in an ESP test if none other than chance factors produce the scores. In any test in which the subject is free to call as he pleases, no one of these hypotheses fits the experimental conditions exactly. Greenwood (3) found that of the three hypotheses when fitted to empirical chance data the Binomial was the best fit.

(a) For a *single run* score of nine or more, the Normal underestimates the Binomial which, in turn, is lower than the Pure Matching approximation to the true probability. These discrepancies are greater the higher the score.

(b) Although no proof is at present available, the following statement seems permissible: If a given *average score* results from a number of runs, any discrepancies between probability values on the three hypotheses will be increased if further runs *maintaining that average* are added.

(c) If, however, a series of runs yields a given *critical ratio,* and further runs are made maintaining that critical ratio, then any discrepancy between hypotheses will be *decreased* with additional runs. This statement is illustrated clearly in a table given by Stuart and Greenwood (18: p. 303).

On the above three statements, certain practical comments are relevant: (a) The probability of an individual run score has never been an important statistic in ESP research. No crucial conclusion has depended upon an exact computation. (b) The maintenance of an average score has been of great interest, but observed results have been characterized more frequently by variability or change of average. Furthermore, the exact chance probability of a maintained high average is of no crucial bearing, since the maintenance of a high extra-chance average is in itself a guarantee of the non-chance character of the scores. (c) The security of a given critical ratio as lying beyond an agreed probability criterion has been of first importance in ESP research. The third statement above means, in practice, that a fairly large critical ratio from a reasonably large number of runs will have a significantly chance probability upon any one of the three approximations. And the larger the number of runs, the more certain we may be that the hypothesis used will not introduce an important error.

Variance. Kellogg again insists upon the use of the observed variance of the scores as the proper statistic to use in finding significance of an observed mean from the expected chance mean. He clarifies his position as follows: "Stuart's remarks concerning this section are obviously quite irrelevant to my argument, which refers not to the relative frequency of chance scores, but to the trustworthiness of experimental data, as supposedly evidential of something other than chance. Perhaps I am mistaken in believing this to be the question to which the answer is sought by all this research. The estimates of the chances, discussed above, provide no answer; they only show, in any case, whether the question is worth asking."

The ambiguous usage of "trustworthiness" above again befogs the issue. Of course, the research has been concerned with trustworthy data. But the statistical reliability of a given observed average has seldom been of crucial concern.* The major problem has been whether an observed score deviated significantly enough from chance expectation to exclude the chance hypothesis. The reliability of the scores as representing some expected theoretical characteristic of ESP has not been of primary interest and has, no doubt, been inadequately dealt with heretofore.

CRITICISMS OF EXPERIMENTAL METHODS

In "A Methodological Review of Extra-Sensory Perception" (5) John L. Kennedy examines a great many of the published reports of research to see if they can be explained by a number of alternative hypotheses—notably sensory cues and recording errors. Any experiment which does not, in its reported conditions, explicitly exclude the counter-hypothesis is excluded as ESP evidence on the ground that the results are explainable by the alternative. All the research considered is so excluded by Kennedy except Warner's (20), Riess's (15), and Rhine's Pearce-Pratt distance series (14). These are classified as "inexplicable."

From one point of view, it might be agreed that, since there is no difference except a linguistic one between "inexplicable" and "crucially supporting the ESP hypothesis," there can be little objection to the Kennedy study. But Kennedy's procedure contains certain presuppositions, the nature of which must be clearly understood if much of the recent critical literature is to be intelligible. The presuppositions concern the appropriate *classification* of research results.

---

* A case in which it was of concern occurs in a report by the present writer (17). Scores were obtained under two sets of conditions. In order to show that the observed averages were significantly different, the observed standard deviations were used, as Dr. Kellogg proposes. But when the chance or non-chance nature of the averages was examined, the theoretical standard deviation was used as the best measure of expected variation. Note that the first problem was one of reliability of difference between averages, neither one of which had a theoretical expectation within the problem; and the second problem dealt with testing whether an observed average might be significantly deviant from a population the statistics of which were exactly theoretically determinable.

Let us take, for example, an Open Matching series conducted with commercial ESP cards, in which the subject works alone and records his own results. He offers his results in good faith as evidence for ESP.

Kennedy would say that the experiment did not constitute ESP evidence because he could set up experiments involving the use of sensory cues and recording errors that would reproduce the results. That is, the results should be classified as belonging to the class of experiments upon observational error, and constitute evidence supporting the occurrence of such errors.

Rhine would say that the experiment constituted an ESP test because the experimenter intelligently and in good faith intended it to be so. But he would exclude it from consideration as crucial evidence for ESP since there was no adequate objective exclusion of alternative hypotheses. That is, the experiment belongs in the class of ESP tests, but does not belong in the class of crucial ESP tests.

Both of these types of inferential classification follow from the "law of parsimony." Kennedy provides a simpler hypothesis to explain each experiment, and excludes the intent and subjective opinions of the experimenter as of little value. Since there is thus no common factor to be considered, the findings of the crucial ESP tests have no bearing upon the ones he explains on simple grounds. Rhine, on the other hand, finds in the crucial tests a common factor which might be attributable to all the experiments, and he would hold that all are properly classifiable under the heading of tests of that common ESP factor, but that each must be evaluated in terms of the degree to which alternative hypotheses are excluded.

The difference of interpretation is of first importance in understanding recent discussion. Kennedy (6), for example, examines the tests of the *Handbook for Testing Extra-Sensory Perception* and shows how the results of each of the tests up to a certain point may be open to interpretation by alternative hypotheses. He thereby ignores the major point of that publication, namely, that the succession of tests described is, in the opinion of the majority of successful ESP experimenters, the best way of reaching the point of competently safeguarded experimental techniques.

The exclusion of the implied aims of an experiment, and inferential evidence of suggestive bearing are often necessary for critical evaluation of a given piece of work. But if it is stringently applied throughout an experimental field, it becomes merely a set of arbitrary standards whose only justification is whether they allow a consistent evaluation. The fact that Kennedy had to classify work as "inexplicable" shows that his standards did not produce a consistent set of explanations for the field of work. The standards themselves are, therefore, open to question.*

A further complication of the same problem involves the nature of the background population of a given research. In his monograph Rhine (13) published all his research on the ground that the experimenters' judgment was of weight in considering the possibility of alternative explanation. The background population of this research might therefore be characterized as "All card-guessing tests seriously aimed at producing positive evidence bearing upon the ESP hypothesis." Pratt, on the other hand, took the position that he would present as evidence only that work in which the experimental conditions were such as to exclude given alternatives. The background population of Pratt's research is therefore more limited than that for Rhine's original report. Actually, however, as the standards of research change, the basic populations from which a particular experiment must be considered a sample are limited more and more. The Pratt and Woodruff experiment (12) contains so many novel restrictions that it can be pertinently classified only as the total of work of its kind.

These matters are important when somewhat vague rules of permissibility of selection are suggested by Leuba and Kellogg who tacitly assume that the background population of any research report is "all card guessing tests." Such an assumption simply denies the fact of crucial differences in experimental procedures.

Selection. Clarence Leuba (8), in an article in the *Journal of Parapsychology,* reported an experiment from which he induced a number of suggestions regarding non-permissible selection. In a later article (9), he amplifies his views to reach the following summary: "The large number of correct responses obtained in the early Duke experiments and reported in two recent books, fail to appear when sensory cues and recording errors are eliminated; the more modest successes of the later experiments, under better controlled conditions, can be obtained by the selection of lucky subjects and through the influence

* For example, Kennedy excludes the Stuart tempo work on the ground of lack of independent records and therefore infers that the results are due to recording errors. But to do so he had to ignore the argument presented in that paper against the motivational direction of possible errors, and the fact that his own experiments, loaded for recording errors, gave results inadequate to account for several research results pointed out by Murphy (11). One of the findings of the Pratt and Woodruff study was that the counting method in an STM procedure was very accurate.

of the experimenter's attitude in determining the point at which the testing will be discontinued."

(These conclusions do not occur in Leuba's first article, and the only additional evidence presented is a repetition of his experiment under two new conditions. No results are given with the statement: "In neither experiment, however, were the results as significant as in our first experiment. . . . " Since the total results of the first experiment were not significant, and the significance of results of the individual "subjects" never fully examined, the comparison throws little light upon the issues involved.)

Leuba's summary does not concur with the facts. The highest ESP scoring was reported three years after the publication of the early Duke experiments. Significant results have been obtained without the selection of "lucky" subjects (however that selection is to be accomplished). The Pratt and Woodruff report of "more modest successes" in this number of the *Journal* is fully controlled to deal with the optional stopping hypothesis.

V. W. Lemmon (7) repeats the "optional stopping" hypothesis in his critical statement at the Southern Society for Philosophy and Psychology symposium. It is possible, he points out, to continue testing with each subject until a positive score is obtained by chance. Ultimately a significant positive result will be obtained. Theoretically, as recently proved by Greenwood and Greville, this procedure is possible. But in practice, it involves the same fallacy as occurs in gambling "systems." A gambling system will work if an unlimited capital is available and an infinite number of plays are possible. The studied use of optional stopping would give consistently significant ESP test results if experimental time were unlimited. But such time is limited by hours, and days, and even lifetimes.

The optional stopping hypothesis is countered best by its experimental exclusion or control, the latter along such lines as proposed by Greenwood (4). But Greenwood's application of systematic selection of stopping point to his 20,000-run chance series, which *never* obtained a significant positive average, displays clearly its inadequacy as a general counter-explanation of all ESP results.

Dr. Kellogg calls for a "full and frank explanation" of Rhine's statement of procedure: "If, during the performance for record, the score drops below a 6 in 25, it is legitimate to quit scoring for a time," and asks if the procedure were really adopted. The answer is that it has been used (mainly by Rhine), but wholly at the will of the experimenter. Rhine's statement appears both full and frank, but an example might clarify it. Suppose a subject was known to be capable of consistent above-chance scoring, and the experimenter wants a series which exemplifies that capacity. The subject then proceeds to get the following run scores: 6, 8, 7, 9, 7, 4. At this point the experimenter can make a more or less intelligent judgment about the last run score. He may decide that the 4 is merely a chance deviant from the subject's usual scoring level and proceed with the experiment. Or he may decide that the 4 is a qualitatively different score from those preceding. If the latter decision is made, the experimenter may proceed with the experiment immediately, upon the belief that the low score marks a very temporary break in the series. Or he may decide that it marks a break that may continue for some time. If this last decision is made, he may "quit scoring for a time." He may stop the testing for a minute, or an hour, or a day. He may give the subject a wholly different experimental task. Or he may suggest continuing the same procedure "off the record," with the understanding that whatever scores are made, they are not to be part of the experiment. He may then watch for any trend of the off-record scores that may indicate a return to the previous scoring level. If such a trend appears, he may decide to continue the experimental tests for record.

The assumption of this procedure must be kept clearly in mind. First, it is supposed that ESP capacity may give sporadic test results. Second, it is supposed that the subject is capable of periods of consistency of scoring in spite of the variable nature of the process. And third, it is supposed that the experimenter can acquire various judgmental skills in dealing with the subject to predict certain scoring trends. If any one of these assumptions is not valid in a particular case, then, although the score results are an experimentally proper sample, the experimenter has wasted his time in using the procedure.

The question arises whether the procedure does not permit of nonpermissible sampling on a chance hypothesis. The answer is that in order that extra-chance scores result from the procedure, the second assumption above must be valid. The procedure would not result in extra-chance results from a chance series and would, indeed, be a valid technique for choosing a random sample from such a series.

It is probably relevant to point out that in the more crucial tests reported, other requirements made this procedure impracticable so that, although theoretically valid and interesting, it has not been widely used.

Patterns of Preference. From September 1937 to January 1938, the Zenith Radio Corporation sponsored a radio program in which was incorporated a series of short "telepathic" tests for audience response. These tests consisted typically of five responses to a randomly determined two-choice situation; for example, the third test required a "black" or "white" response to five signals.

An analysis of the results by L. D. Goodfellow (2) begins: "Neither coincidence nor telepathy, but the natural response of an audience to secondary cues caused the 'highly successful' results in the Zenith Radio experiments in telepathy. Approximately three-fourths of the audience's seventy-six attempts to receive an impression telepathically yielded results significantly different from chance expectation. Apparently, the audience was responding to definite factors—not chance. The most significant result of this study is the discovery of these factors." And, in conclusion, he says: "The two most important of these factors are (1) the pattern or sequence used by individuals in recording this process, and (2) the set or predisposing influence of subtle suggestions found in the test instructions."

The first point of concern here is that the Zenith tests did not constitute adequate crucial tests of ESP (or telepathy) in the first place. They were obviously open to the type of alternative explanation Goodfellow set out to prove. But since Goodfellow proposes a similar type of analysis for ESP tests, it is pertinent to examine his proof critically.

(a) His Tables I and II give two analyses of the same data. The totals of successes do not agree in the two tables.*

(b) By a separate experiment, it was found that an "audience aggregate pattern was 11211, 1 representing the favorite or suggested symbol and 2 representing the alternative one (2: p. 627)." Upon this principle, the results of six succeeding Zenith tests were predicted. But this prediction requires the second response to be similar to the first on the average. Yet (on p. 610) it is noted that in the Zenith results "the data show that for approximately fifty-three per cent of people it was different."

(c) The predisposition of the audience to respond in a given way to the first trial is attributed to suggestions in the instructions. These "suggestions" were discovered by presenting recordings of the instruc-

* The present writer noted this discrepancy in the original Zenith data in October, 1937. It was corrected in later tests but apparently the original data were not reexamined.

tions to a number of judges. But, since it is not explicitly denied, it appears clear that the judges knew the results of the tests. The finding of "suggestions" was therefore a simple task of rationalization. The control of the effect of suggestion upon the judges could have been easily accomplished and is crucial to the whole conclusion.

F. H. Lund, by entitling his report "Extra-Sensory Perception Another Name for Free Association?" and by stating the problem as though a crucial distinction were being exercised, obscures a very real problem within a tangle of merely linguistic difficulties. First, ESP is self-evidently not another name for free association since any perception is assumed to be limited by the nature of the object perceived. Second, the term "free association" is customarily used in a deliberately paradoxical sense to indicate a response which is not limited by instructions but which is assumed to be more or less strictly determined by causes not explicit in the consciousness of the subject. The customary ESP test is more properly a multiple choice situation. The response is limited strictly to a specified range, usually the naming of one of five symbols.

Lund gave six trials of guessing ESP cards to each of 596 subjects in 15 sub-groups. The results were close to chance and he remarks appropriately that "there is no evidence that ESP in any way affected the subject's responses." He noted that there was a marked preference for the star and a greater preference for the middle card of the order of symbols displayed on the blackboard. He found the subjects avoided repetition of a symbol known to have been drawn on the previous trial. These observations were, of course, independent of any consideration of the success of response. All of these are what might be expected to occur in a multiple-choice situation, and they would tend to support the obvious contention that the ESP test is such a multiple choice situation.

Lund makes a further observation, however, in the section entitled "Influence of Repetition," in which it is noted that when a symbol was drawn from the deck, returned, and then a similar symbol drawn, a 50 per cent above chance success was attained on "first trials." This correlation of success with some aspect of the stimulus situation might easily have suggested reexamination of the results in terms of real problems. For example: How did the degree of symbol preference, the degree of position influence, the degree of non-repetition, relate to degree of success? The customary measure of ESP is the degree of success. If it is hypothecated that the same causes producing variations in association produce the variation in success that characterize ESP, it seems only reasonable to test the hypothesis in respect to that customary measure.

S. W. Fernberger, under another interrogation, "Extra-Sensory Perception or Instruction?", reports giving instructions "loaded" for a "black" response in five "black or white" free choices made by college-student subjects. He found a pattern of response similar to that found in one of the Zenith experiments. That is, "black" occurred as a response with greater-than-chance frequency on the first and third trials, and close to a 50 per cent chance expectation on the others. The supposition is that had the test been an ESP test and had "black" occurred as the stimulus on the first and third trial purely by chance, then there would have been an extra-chance deviation appearing to indicate an ESP relation, but which could be wholly attributed to the loaded instructions and a natural tendency to pattern the responses. This conclusion is the same as that proposed by Goodfellow.

Neither Goodfellow nor Fernberger distinguish between these tests and the customary ESP tests. The five trial, two choice test differs from the 25 trial, five choice test particularly in its amenability to effects from habitual patterns of response. In everyday life, two choice situations are vastly more common than five choice situations, and likewise more liable to habitual organization. Furthermore, whereas the effect of instructions upon a single five-trial series may be considerable, the customary ESP test, in which the subject makes repeated runs of 25 trials without instructions after his first visit to the laboratory, does not provide a situation for even loaded instructions to be effective.

General Criticism. H. Rogosin (16) "evaluates" ESP research in a recent article by discussing what has been said about it and may be said about it. He mentions no experiment, method, or conclusion of a research report. He finds Warner and Raible's suggestion for ESP controls in psycho-physical techniques "a direct slam at psychology." He finds the concept "sixth sense," as used by popular writers, an evidence of cultural lag, since psychologists do not talk about "senses" any more. He points out that "the great advances that have been made in spreading ideas based on experimental evidence of the working of the brain, have been wiped out by the popularization of the Duke experiments"; thus raising again the mind-body problem, which had been settled. He finds that there has been controversy concerning the mathematical evaluation of ESP which he feels is unsettled, that most of the ESP investigations must be considered as unreliable, due to a neglect of history and standardization. He deplores the propagation of incorrect ideas about psychological work, these incorrect ideas being accepted widely because (a) the work was done under university sponsorship; (b) it was "pushed" by science writers O'Neill and Kaempffert; and (c) it fits in with a flight-from-reality trend in contemporary culture. He charges McDougall and Rhine with disagreeing with most psychologists and possessing unorthodox beliefs, charges O'Neill and Kaempffert with bias in favor of psychical belief, and deplores idealism.

No doubt a program of research may be evaluated in this way. But when every assertion is open to challenge both as to factual basis and strictness of implication, the usefulness of such evaluation is in question. ESP investigators are concerned with methods, results, and conclusions of research. Rogosin is not so concerned. And there the matter rests.

Published criticisms of the past year have been characterized by a marked decrease in purely mathematical discussion. The general approval of mathematicians, the excellent original work of Greenwood and Greville, and the fact that of all features of ESP research, the mathematical problems have had the most ready and generous assistance from specialists in the field, all tend to suggest definite advance in this realm of the research.

The most interesting of current criticism is the optional stopping hypothesis. In its mathematical aspect, important techniques have been developed by Greenwood. But in respect to ESP theory, it may continue long to be a difficult problem, since the length of an experiment is bound up with the "scoring level" of the subjects. Yet that scoring level seems to be variable in respect to so many factors so far uncontrolled that it may be necessary to develop the universally repeatable experiment before this difficult hypothesis can be fully dealt with.

REFERENCES

1. Fernberger, S. W. 'Extra-sensory perceptions' or instructions? *J. Exper. Psychol.*, 1938, **22**, 602-607.

2. Goodfellow, L. D. A psychological interpretation of the results of the Zenith radio experiment in telepathy. *J. Exper. Psychol.*, 1938, **23**, 601-632.

3. Greenwood, J. A. Analysis of a large chance control series of ESP data. *J. Parapsychol.*, 1938, **2**, 138-146.

4. Greenwood, J. A. An empirical investigation of some sampling problems. *J. Parapsychol.*, 1938, **2**, 222-230.

5. Kennedy, J. L. A methodological review of extra-sensory perception. *Psychol. Bull.,* 1939, **36**, 59-103.

6. Kennedy, J. L. Suggestions concerning the nature and production of "extra-sensory perception" [Manuscript].

7. Lemmon, V. W. The role of selection in ESP data. *J. Parapsychol.,* 1939, **3**, 104-106.

8. Leuba, C. An experiment to test the role of chance in ESP research. *J. Parapsychol.,* 1938, **2**, 217-221.

9. Leuba, C. Has recent research undermined the evidence for extra-sensory perception? *J. Appl. Psychol.,* 1938, **22**, 549-553.

10. Lund, F. H. Extra-sensory perception another name for free association. *J. Gen. Psychol.,* 1939, **20**, 235-238.

11. Murphy, G. [Limits of recording errors.] *J. Parapsychol.,* 1938, **2**, 262-266.

12. Pratt, J. G., and Woodruff, J. L. Size of stimulus symbols in extra-sensory perception. *J. Parapsychol.* [this number].

13. Rhine, J. B. Extra-Sensory Perception. Boston: Bruce-Humphries, 1934.

14. Rhine, J. B. Some basic experiments in extra-sensory perception. *J. Parapsychol.,* 1937, **1**, 70-80.

15. Riess, B. F. A case of high scores in card guessing at a distance. *J. Parapsychol.* 1937, **1**, 260-263.

16. Rogosin, H. An evaluation of extra-sensory perception. *J. Gen. Psychol.,* 1939, **21**, 203-217.

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

18. Stuart, C. E., and Greenwood, J. A. A review of criticisms of the mathematical evaluation of ESP data. *J. Parapsychol.,* 1937, **1**, 295-304.

19. Stuart, C. E., and Pratt, J. G. Handbook for Testing Extra-Sensory Perception. New York: Farrar and Rinehart, 1937.

20. Warner, L. A test case. *J. Parapsychol.,* 1937, **1**, 234-238.

EXPERIMENTS ON THE NATURE OF EXTRA-SENSORY PERCEPTION

I.—Repetitions of the Rhine Experiments[1]

John L. Kennedy

Fellow in Psychical Research, Stanford University

Introduction

Acceptance or rejection of a new scientific discovery, assuming of course that the data presented in experimental reports are not inconsistent, is usually contingent upon successful repetitions of the phenomenon under the conditions stated in the experimental reports. The generalizations based upon the conditions of the experiments in question may themselves be open to criticism, but a scientific experiment at least should allow repetition of results with reasonable precision.

Inconsistencies in the experimental reports concerning the discovery and further elaboration of Extra-Sensory Perception have already been pointed out (2) (3) (4) (11). However, the criterion of repeatability of the ESP results has also been surrounded with qualifications to the extent that, if the qualifications are accepted, the publication of non-confirmatory results at this time might truly be considered superfluous. Nevertheless, the present writer wishes to present as the first paper in a series devoted to criticism of ESP: (1) the results of an extensive search for subjects with this hypothetical ability and (2) an evaluation of some of the qualifications to the criterion of repeatability which have been seriously proposed to account for results such as are presented here.

Experimental Methods

The general methods as presented in "A Handbook for Testing Extra-Sensory Perception" were followed in the present experiments. The specific methods utilized were: (1) Open Matching (OM), (2)

---

[1] Communication No. 3, new series, from the Psychical Research Laboratory at Stanford University. This paper has been read and approved by the Stanford Committee on Psychical Research.

General or Undifferentiated ESP method (GESP), (3) Down Through (DT), and (4) Pure Telepathy (PT).[2]

The commercial variety of ESP cards was used with (OM), (GESP) and (DT) methods. New packs were used whenever the cards became soiled or bent. With the (OM) method, the cards were shuffled twice through an I-Deal mechanical shuffler and cut by the experimenter just before their use by the subject. For the (GESP) and (DT) methods, 10 separate decks of ESP cards were run twice through an I-Deal shuffler before the experimental session; the order of the cards in the deck was separately recorded and the decks were used in order. No cards were used in the (PT) work.

In the experiment with the (OM) method, all cards matched to each key symbol were recorded by frequency in each suit. The cards were turned over by the experimenter (the writer) after the subject had matched them to the key symbols, arranged in suits and recorded on a plain sheet of paper. In the other experiments, the commercial ESP record pad was used. In the case of the (GESP) work, the calls were separately recorded by an observer (the writer) as well as by the experimenters. In (DT) work, the experimenter (the writer) recorded the subject's calls and the subject recorded the card series. In the (PT) procedure, the subject's calls were recorded by an observer (the writer) and by the experimenter. The symbols chosen by the experimenter for sending were not recorded.

TABLE I
THE DISTRIBUTION OF SUBJECTS, EXPERIMENTERS, AND DECKS ACCORDING TO ESP METHOD USED

Method	*Subjects*	*Experimenters*	*Total Decks*
Open Matching (OM)	100	1	1600
General ESP (GESP)	68	16	982
Down Through (DT)	33	3	382
Pure Telepathy (PT)	3	12	130
Totals	204	32	3094

The great majority of subjects and experimenters in these experiments were students in the elementary course in psychology, although anyone willing to devote the time to this experiment was tested. Data on the number of subjects, experimenters, and decks of cards guessed are presented in Table I.

[2] The version of the (PT) method described in Rhine's monograph (8) was repeated in this experiment. The (PT) method described in the Handbook (10) differs from the earlier method in that it advises independent recording.

RESULTS

1. *Open Matching Method.* In the total of 1,600 packs matched in the Open Matching procedure, 7,936 hits were scored, 8,000 hits were expected by chance and the obtained deviation from chance expectancy was —64. According to the Table of Appendix A in (10), the standard deviation of the theoretical distribution is 81.6 for 1,600 runs, which yields a critical ratio (D/σ) of .78. The odds, as given in the Table on Page 59 in (10), against getting this result by chance alone are 6-1, hence it may be concluded that the deviation was produced by chance factors.

The 1,600 packs reported here were collected in two sessions. At one experimental session, the subject matched eight packs of cards and repeated this procedure at a second session at a later time. By means of a correlational technique, it is possible to decide whether or not subjects demonstrate an extra-chance consistency in scoring within the total chance scoring of the whole experiment. Figure I presents the scatter plot for this correlation. Each dot represents the score of a single subject on Tests I and II. The Pearsonian r is $-.0004 \pm .0675$, indicating nothing but chance consistency. The extreme cases encircled in the figure have been reported in a previous paper (5).

2. *General Extra-Sensory Perception Method.* For the total of 982 packs collected by this method, the obtained number of hits was 4,987, the expected number 4,910 and the difference was + 77. The standard deviation computed was 63.91 and the critical ratio, 1.20. Of the 68 subjects and 16 experimenters tested with this method, not one demonstrated anything but chance performance.

3. *Down Through Method.* The 382 packs guessed by 33 subjects with the Down Through method yielded a total of 1,852 hits where the expected number was 1,910. The standard deviation is 39.78 and the critical ratio is 1.46, indicating odds of 15-1 against the chance explanation of the results. No individual subject or experimenter obtained extra-chance results.

4. *Pure Telepathy Method.* In the 130 packs guessed with the Pure Telepathy methods, a total of 767 hits was scored. The expected number by chance was 650 and the difference was + 117. The standard deviation for 130 packs is 23.26 and the critical ratio is 5.02, yielding odds of approximately 3,383,000 to 1 against the chance explanation of these results. Since these are the only extrachance data in the present experiment, they merit further analysis.

SCATTER-DIAGRAM OF TOTAL HITS PER TEST

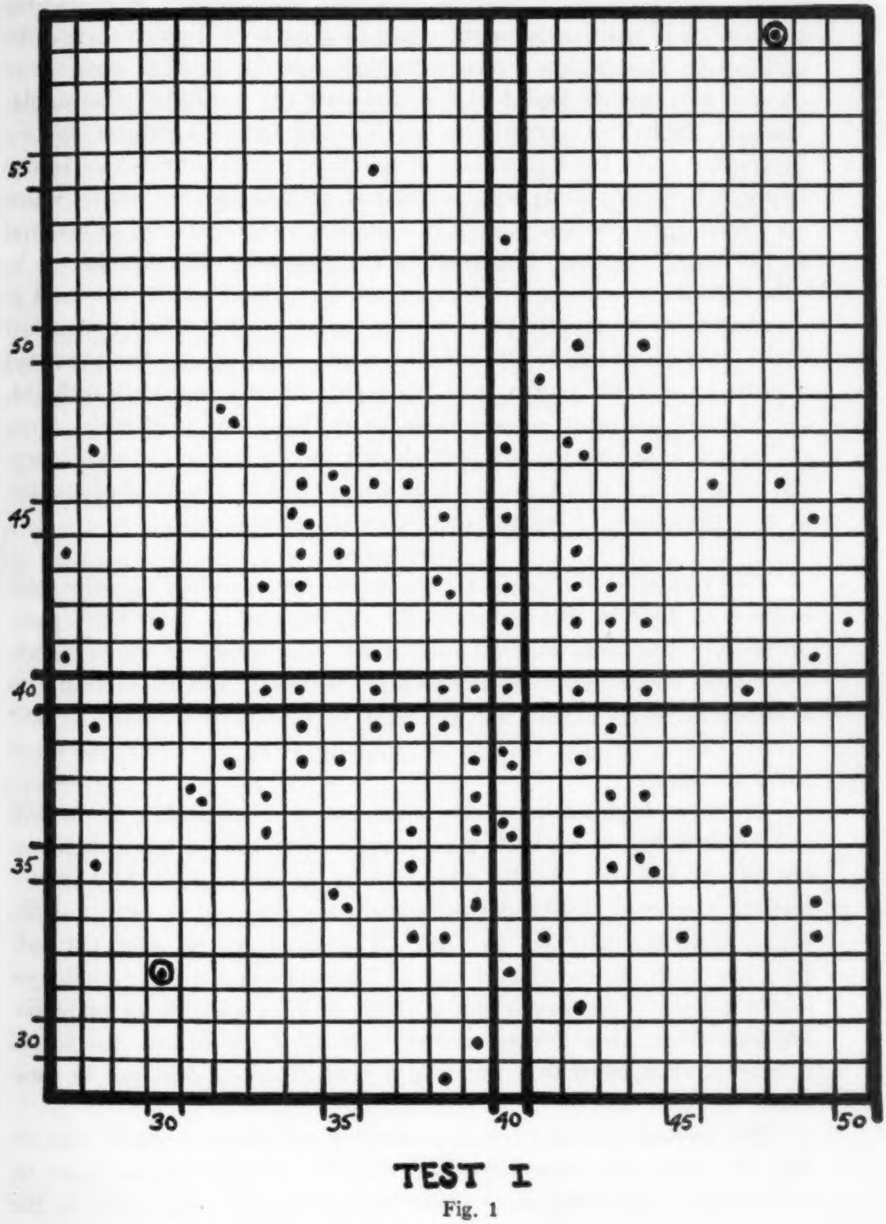


Fig. 1

The PT method as used in the present experiment introduced several unsatisfactory points of methodology in conducting ESP experiments. First, the experimenter or sender was allowed to choose the symbols to be sent without reference to a pack of shuffled cards. According to Goodfellow's recent findings with respect to patterns of choices (1), this method does not eliminate the possibility that similar "mental habits" or preferences in choosing between subject and experimenter may have produced a spuriously large number of coincidences when the results were compared with theoretical values which are based upon the assumption of complete randomization of material to be "sent." Second, in recording the results of the experiment, he (the "sender") was required to record the subject's calls only and to indicate a hit by merely checking the called symbol when it matched the symbol on which he was concentrating. Kellogg (4) has presented a critique of both points indicating that results obtained with this method are certainly questionable as evidence for telepathy. Since the actual order of the symbols chosen by the sender is not known, the data, as they stand, do not lend themselves to any kind of analysis or interpretation. They are merely experimental curiosities.

Discussion

The results of the present experiment are certainly negative with respect to ESP in 204 subjects, 32 experimenters, and 3,094 packs of cards. The only extra-chance result was obtained under poorly controlled conditions. Some attempt, however, should be made to evaluate negative results in the light of certain hypotheses in ESP research which would, if accepted, explain away non-confirmations of the ESP theory.

The first qualification to the criterion of repeatability of the ESP experiments has to do with the assertion that only a few people can obtain extra-chance results when acting as subjects in ESP experiments. Certainly individual differences are the rule in other psychological abilities; why not in ESP? The incidence of good ESP subjects has been estimated at 1 out of 5 persons tested (8, p. 106), yet in the present experiment not a single subject capable of maintaining an extra-chance average under controlled conditions was found. It may be concluded that the 1 in 5 generalization does not fit these facts.

The second qualification is embodied in recent research findings (6) (7) that only some experimenters can produce extra-chance results with some subjects. This finding applies not only to the telepathy experiments in which the experimenter takes an active part as "sender" but also to the Down Through and other clairvoyance methods in which he acts only as shuffler and recorder. None of our 32 experimenters were able to obtain extra-chance scores with the subjects they tested when conditions were adequately controlled.

Finally, the hypothesis of impermanence of ESP ability (7, p. 158) is perhaps the most important barrier to meaningful repetition of the ESP results. According to this hypothesis, even if a subject who at one time had shown evidence of ESP ability by his extra-sensory scoring had been tested in the present group and had obtained results at the chance level, the failure of the ability to manifest itself should be looked upon as a temporary loss of the ability.

The facts on which these hypotheses are based seem to the present writer to be amenable to other interpretations. Specifically, the role of the experimenter and the experimental conditions needs to be more fully elaborated before accepting the above hypotheses as anything other than easy rationalizations for "chance" results obtained under what appear objectively to be the same conditions as those which produced "extra-chance" results. Non-ESP factors capable of producing spurious evidence for ESP will be discussed more fully in following papers.

CONCLUSIONS

It is difficult to draw conclusions from negative results such as are reported in this paper. The writer believes, however, that the following summary statements may be made:

(1) The results of a relatively extensive search, conducted with several ESP methods, for individuals who are capable of obtaining the extra-chance scores in card-guessing were entirely negative when these methods were completely controlled for source of error.

(2) The Pure Telepathy method, as used originally by Rhine, yielded extra-chance results in this experiment. The conditions under which the data were collected, however, preclude analysis of the nature of these extra-chance scores.

REFERENCES

1. Goodfellow, L. D. A psychological interpretation of the results of the Zenith radio experiments in telepathy. *J. Exper. Psychol.*, 1938, **23**:601-632.

2. Gulliksen, H. O. Extra-sensory perception—what is it? *Am. J. Sociol.*, 1938, **43**:623-632.
3. Kellog, C. E. Dr. J. B. Rhine and extra-sensory perception. *J. Abnorm. (soc.) Psychol.*, 1930, **31**:190-193.
4. Kellogg, C. E. New evidence (?) for extra-sensory perception. *Sci. Mon.*, N. Y., 1937, **45**:331-341.
5. Kennedy, J. L. The visual cues from the backs of the ESP cards. *J. Psychol.*, 1938, **6**:149-153.
6. MacFarland, J. D. Discrimination shown between experimenters by subjects. *J. Parapsychol.*, 1938, **2**:160-170.
7. Pratt, J. G., and Price, M.M. The experimenter-subject relationship in tests for ESP. *J. Parapsychol.*, 1938, **2**:84-94.
8. Rhine, J. B. Extra-sensory perception. Boston: Boston Soc. Psychic. Res., 1934.
9. Rhine, J. B. New frontiers of the mind. New York: Farrar and Rinehart, 1937.
10. Stuart, C. E., and Pratt, J. G. A handbook for testing extra-sensory perception. New York: Farrar and Rinehart, 1937.
11. Wolfle, D. L. A review of the work on extra-sensory perception. *Am. J. Psychiat.*, 1938, **94**:943-955.

A Critical Review of "Discrimination Shown Between Experimenters by Subjects," By J. D. MacFarland[1]

John L. Kennedy

Fellow in Psychical Research, Stanford University

I. Introduction

In the September, 1938 number of the *Journal of Parapsychology*, MacFarland (4) published an experimental paper in which evidence was presented that the experimenter is an important variable in ESP scoring. The following quotation from the abstract of the article gives the general conditions and the results:

"Five selected subjects made 15,300 calls equally divided between the GESP and DT test procedures. Each call was aimed at two target decks and counted twice, making 30,600 trials. An average of 5.48 was obtained, and a critical ratio of 8.37. Two experimenters, one with each target deck, were present throughout; all recording was doubly witnessed, and the results were doubly rechecked. Sensory cues were eliminated by having the subjects located several rooms away from the cards and experimenter, with a signalling device used by the subject to communicate his calls to the experimenter.

"In the research here reported, two experimenters were compared, one of whom had been consistently successful in previous work and the other 'unsuccessful.'[2] The subjects and the external conditions were the same for both. But the two experimenters still continued at approximately their previous level of success, one at chance and the other consistently above.

"The GESP procedure gave higher score averages than DT for all five subjects. Notably higher scoring was achieved in GESP when it happened that both experimenters were looking at the same card. No U-curves in this adaptation of the DT procedure were discovered, contrary to earlier experience with DT" (4, p. 160).

The appearance of MacFarland's paper should be looked upon as

---

[1] This paper has been accepted and approved by the Stanford Committee on Psychical Research as Communication No. 6 from the Psychical Research Laboratory. The writer acknowledges with thanks the assistance of J. V. Uspensky and Quinn McNemar in the mathematical portion of this paper.

[2] In the following report, the initials J.M. designate the successful experimenter, D.H., the unsuccessful experimenter.

an important event in the course of ESP work since it represents the first experiment, to the present writer's knowledge, in which two experimenters made independent records of the subject's guesses.

The specific conduct of the experiment may best be described by quoting from the experimental report:

"Two standard ESP techniques were used. In the GESP technique, experimenters D.H. and J.M. sat side by side in front of the receiving set. Each experimenter held a deck of cards which he shuffled thoroughly. The button on the platform was pushed three times to give the subject his cue to begin calling. Each experimenter looked at the top card of the deck momentarily, placed the card face down on the deck on which the set was located, and each recorded independently the call made by the subject as it was indicated by the lighting of one of the compartments. This continued through a standard run of 25 cards, at which time the experimenters signalled with their push button that the run was completed. No other signalling occurred; the pace was set by the subject. The double record of the subject's calls secured by the independent recording of the two experimenters was checked by placing the two columns parallel to each other and both experimenters inspecting each record. Both experimenters then checked the two decks of cards against the doubly-recorded order of the calls. In this way there was secured double witnessing of all checking. . . .

"The DT technique was used with the same precautions. In this technique, however, the experimenters did not disturb their decks of ESP cards till after the completion of the run of 25 cards . . ." (4, pp. 162-163).

In Table I, pertinent data concerning the experimenters, number of runs and deviations from chance expectancy are presented. These values are taken from MacFarland's Table III (p. 165), corrected by the present writer for errors in counting hits. Three hundred and five complete runs through a deck of ESP cards (commercial variety) were collected by J.M. with the GESP technique, three hundred and six with the DT method and three hundred and six by D.H. for both methods. Significant deviations from chance expectancy occur in J.M.'s work with both methods. The deviations of plus 395 for the GESP method and plus 174 for the DT method

TABLE I
MACFARLAND ESP RESULTS

Experimenter	*Runs*	*Hits*	*Expected*	*Deviation*	X^2
GESP METHOD					
J.M.	305†	1920	1525	395	127.89*
D.H.	306	1574	1530	44	1.58
DT METHOD					
J.M.	306	1704	1530	174	24.74*
D.H.	306	1517	1530	–13	.14

* indicates significant X^2's

† one deck was thrown out since only 24 cards and calls were recorded

will be considered more fully in this report with respect to certain patterns of successes and errors in recording.

II. Errors in Recording in the Card Column

Through the courtesy of Dr. J. B. Rhine and Mr. MacFarland, photostatic copies of the original records of this experiment were obtained by the present writer and subjected to further analysis. When the calls were checked again for errors in recording, a surprising number was found in which the same call as recorded by two separate experimenters did not agree in the two records. Since the experimental conditions were such that disagreements of this character indicate that one of the experimenters recorded erroneously, the hits produced by these calls appear to the present writer to be open to question. It should be noted further that these cases of disagreement provide no information concerning which one of the experimenters made the error. The following table, in which the possible effects of these errors are presented, was kindly provided by Mr. J. L. Woodruff of the Parapsychology Laboratory at Duke University.

TABLE II
Effect of Discrepancies in Calls in MacFarland's GESP Records

(1)	Increase of 1 hit in M's score, or no effect	28
(2)	Increase of 1 hit in H's score, or no effect	4
(3)	Increase of 1 hit in M's score, or decrease of 1 hit in H's	8
(4)	Increase of 1 hit in H's score, or decrease of 1 hit in M's	3
(5)	Increase of 1 hit in either M's or H's score	13
(6)	No effect in either	10
(7)	No effect, or decrease of 1 in M's score	4
(8)	No effect, or decrease of 1 in H's score	5
	Total	75

A total of 75 discrepancies in the calls was found. As stated above, an accurate evaluation of them in terms of their effect on the score is well-nigh impossible since it is impossible to determine from the data in which record the error was made. The assignment of errors in the card recording, however, is more easily accomplished. Errors here consist of failures of the 5:5:5:5:5 frequencies of the five symbols in the ESP deck to occur. When one of the symbols is recorded at a frequency of 6, for example, it is obvious that some kind of error has been committed. In 9 decks of J.M.'s GESP work, the equal frequencies of the five symbols have been disturbed. For one of these cases, 7 squares were recorded instead of 5. Again it is impossible to assess the effect of card column errors on the hit scores in these decks.

In concluding this section, one can only assert that the presence of these readily noticeable errors casts doubt on the objectivity of the recording procedure of the whole experiment. The remainder of the present paper will be devoted to the question of recording errors in the card columns, errors which are not readily noticeable but which may be hypothecated on the basis of the ways in which the hits are put together to produce the significant positive deviations in Table I.

III. PATTERNS OF HITS IN THE MACFARLAND DATA

From inspection of the photostatic records, it appeared that, in the case of J.M., the successful experimenter, hits occurred in pairs, *i.e.*, a hit appeared to generate another in the next adjacent position down through the deck. The frequency of "isolated" hits, *i.e.*, (1) a hit surrounded by failures in the adjacent positions, or (2) success-failure and failure-success respectively for the top and bottom positions, also appeared to be high. Reference to Figure 1 and examples given there will clarify the meaning of the term isolated hit. Accordingly, the frequencies of (1) pairs of hits, (2) same-symbol pairs of hits, (3) unlike-symbol pairs of hits and (4) isolated hits, were counted for each experimenter and each method. These frequencies were compared with expected chance frequencies as predicted according to the Matching Hypothesis.[3] The expected frequencies for pairs of hits and like and unlike symbol pairs of hits were calculated according to the recently presented method of Stevens (5). The formulae for these calculations are presented in the Appendix at the end of this paper.

Table III presents the results of the above analysis and gives chi-squares computed for each category on the basis of a 2 x 2 table with 1 degree of freedom (3, pp. 3-4). In the pairs of hits category, J.M.'s GESP work is far above the level attributable to chance variation, while D.H.'s pairs of hits are at the chance level. When the pairs of hits for both experimenters are distributed over the two categories of same symbol and different symbol pairs, however, an interesting fact becomes apparent. J.M.'s deviation in pairs of hits may be adequately accounted for by his deviation in unlike symbol pairs. The frequency for like symbol pairs in both J.M. and D.H. is identical and insignificantly different from the frequencies predicted by chance. In the case of D.H., the unsuccessful experimenter, the frequencies of both like and unlike symbol pairs are at the chance level. The

[3] The Matching Hypothesis, suggested by Greenwood & Stuart (2) for evaluation of the mean expectancy of ESP scores, assumes that the subject's calls are equivalent to another deck of ESP cards which is matched to the target deck.

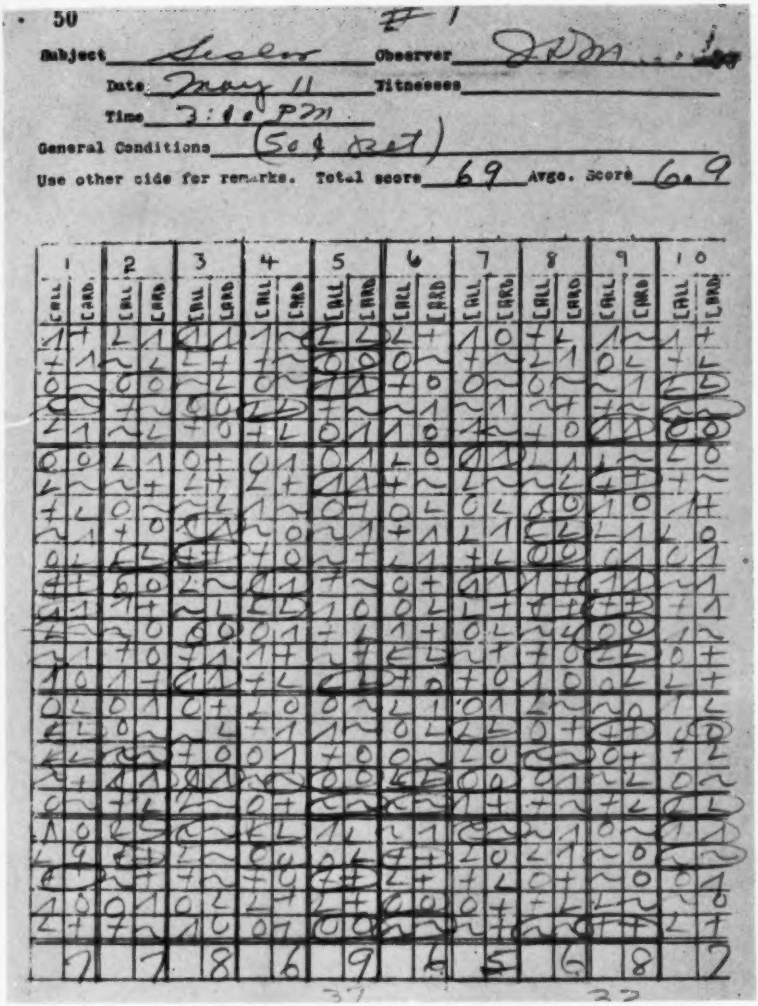


FIGURE 1. PHOTOSTATIC COPY OF ORIGINAL RECORD IN MACFARLAND DT EXPERIMENT
(a) *Pairs of hits.* A pair of hits was counted if the next adjacent position down through the record also contained a hit. Thus, in the case of a run of three hits, 2 pairs were counted; for runs of 4, three pairs, etc. In column 9 of the above record, then, 3 pairs of hits were counted. A pair of same symbol hits may be found in column 1 in positions 17 and 18, a pair of unlike symbol hits in column 1 in positions 11 and 12.
(b) *Isolated hits.* Type 1. Isolated hits were counted when the order no hit—hit—no hit was encountered. Thus, in column 9, 3 isolated hits were counted. Type 2. Isolated hits were counted when either the first or last position in the deck contained a hit followed or preceded by no hit. In column 9, one such case was counted, the final hit. Both types were included in the category "isolated hits."

isol
cha
wo

Pai
Sam

Unl

Isol

Pai
Sam

Unl

Isol
• in

me
car
syn
doc
cor
of
cau
nex
IV.

of
the
Th
pre
hav
bol
by
J.M

abo
Hy

isolated-hits category in J.M.'s GESP work also exhibits an extra-chance influence, while no such influence is present in D.H.'s GESP work.

TABLE III

THE LOCI OF EXTRA-CHANCE SCORING IN THE MACFARLAND EXPERIMENT

	J.M.				*D.H.*			
	Obt.	*Exp.*	*D*	X^2	*Obt.*	*Exp.*	*D*	X^2
GESP METHOD								
Pairs of hits	444	295.8	147.2	76.32*	283	295.8	–12.8	
Same Symbol Pairs	35	40.8	–5.8		34	40.8	–6.8	
Unlike Symbol Pairs	409	255.0	154.0	96.51*	249	255.0	–5.8	
Isolated Hits	1131	995.6	135.4	21.44*	1017	995.6	21.4	
DT METHOD								
Pairs of hits	329	295.8	33.2	3.88	291	295.8	–4.8	
Same Symbol Pairs	31	40.8	–9.8		15	40.8	–25.7	16.87*
Unlike Symbol Pairs	298	255.0	43.0	7.55*	276	255.0	21.0	
Isolated Hits	1084	995.6	88.4	9.02*	987	995.6	–8.06	

* indicates significant X^2's

The second part of Table III presents the same data for the DT method. Here the pairs-of-hits category is on the borderline of significance for J.M. and at the chance level for D.H. Again the unlike symbol pairs of hits for J.M. exhibit an extra-chance influence, as does the isolated-hits category. In D.H.'s record, however, a discordant result is found in the obtained frequency of like-symbol pairs of hits, which shows a highly significant negative deviation. The causation of this significant negative deviation will be discussed in the next section.

IV. THE EFFECT OF SUBJECT PREFERENCES IN GUESSING ON THE APPLICATION OF THE MATCHING HYPOTHESIS TO AN ANALYSIS BASED ON PAIRS OF HITS.

The applicability of the Matching Hypothesis to the above analysis of pairs of hits depends upon the correctness of the assumption that the subject's calls may be considered as another shuffled deck of cards. This assumption is questionable and certainly is not justified in the present case. The calls in this experiment indicate that the subjects have a definite tendency to avoid calling adjacent like pairs of symbols. Accordingly, the calls (as recorded by D.H. for DT work and by each experimenter for GESP work) and card series of D.H. and J.M. were separately searched for same symbol pairs.

Table IV shows that the subjects in both methods called only about half the number of same symbol pairs expected if the Matching Hypothesis were valid for these data. Furthermore, D.H. in the

TABLE IV
FREQUENCIES OF SAME-SYMBOL PAIRS IN MACFARLAND'S EXPERIMENT

	J.M. Calls	*J.M.* Cards	*D.H.* Calls	*D.H.* Cards
GESP				
Circle	130	253	126	309
Square	118	269	121	297
Star	101	255	99	298
Cross	85	213	84	263
Waves	133	232	139	260
Totals	567	1222	572	1427
DT				
Circle	138	274	138	220
Square	138	275	138	258
Star	115	230	115	266
Cross	88	256	88	237
Waves	152	245	152	285
Totals	631	1280	631	1266

GESP procedure either did not shuffle his cards adequately or made errors in recording since a highly significant positive deviation on same symbol pairs in the cards is found ($X^2 = 40.40$). These considerations make obvious the need for evaluation of pairs of hits of like and unlike symbols by means of empirical rather than theoretical probability.

V. CORRECTIONS IN EXPECTED NUMBERS OF SAME AND UNLIKE SYMBOL PAIRS OF HITS BASED UPON EMPIRICAL PROBABILITY.

Certain peculiarities in the calls of the subjects make the calculation of empirical probabilities for the categories of Table III a difficult and questionable procedure. On the face of it, the problem appears to be simple. The data in Table IV provide the necessary information for such approximations. If the assumption is made that same symbol pairs in the calls and cards are randomly distributed, the frequency of circles, for example, in J.M.'s GESP work may be treated as a probability fraction 130/7320 for the occurrence of circle pairs in the subject's calls. Similarly, for the cards, the probability for circle pairs is 253/7320. Multiplication of these probabilities will yield the probability for circle pairs of hits, from which the expected number of such hits may be easily determined. Treating the other four symbols in the same way and summing the expected numbers, the total expected frequency for pairs of same-symbol hits in J.M.'s GESP work is 19.0, for DT, 22.0. However, the data of Table V appear to call in question the assumption of random distribution of the subjects' calls.

Accordingly, the empirical expected number was calculated according to an assumption which more nearly fits these facts, namely, that the subjects' calls of same symbol pairs are distributed randomly in each single deck. Applying the above method to the 305 decks of J.M.'s GESP work, an empirical expected number of same symbol hits

TABLE V

DISTRIBUTION OF NUMBER OF SAME SYMBOL PAIRS CALLED IN EACH DECK BY SUBJECTS IN J.M.'S GESP AND DT WORK

Number of Pairs	*Number of Decks*	
	GESP	*DT*
0	50	64
1	96	57
2	82	72
3	47	51
4	14	42
5	8	14
6	3	3
7	1	1
8	0	0
9	0	0
10	1	0
11	0	0
12	0	1
13	1	1
14	1	
15	1	
	305	306

of 20.6 is obtained, for the 306 decks of DT, 21.7. Table VI shows, however, that this last assumption does not fit the facts. When the same-symbol calls of the subjects are distributed according to position in the deck it appears that certain positions are preferred and others avoided. The avoidance of the first three pairs of positions is especially striking.

The writer does not know how to allow for the preferences exhibited in Table VI. Certainly they act to increase the variance of the expected number of same-symbol pairs of hits. The problem of the exact solution of the expected number and variance of same symbol pairs of hits must be left to someone more capable at handling the subtleties of empirical probability than the present writer.

The frequencies of 35 pairs of same symbol hits in J.M.'s GESP work and 31 such hits in J.M.'s DT work (Table III) are also questionable. Four of the GESP pairs have members in which the calls do not agree between the two experimenters or there are more than 5 such symbols recorded in the cards. One case of disagreement in the DT work was discovered. If these questionable hits are subtracted from the totals, the obtained numbers become 30 for the GESP work and 30 for the DT work. The presence of erasures, and

TABLE VI
THE RELATION BETWEEN SAME-SYMBOL PAIR CALLING AND POSITION IN THE DECK IN J.M.'s GESP AND DT WORK

Positions	*Number of Same Symbol Pairs* *GESP*	*DT*
1 & 2	6	5
2 & 3	15	17
3 & 4	11	15
4 & 5	21	20
5 & 6	36	50
6 & 7	24	23
7 & 8	15	25
8 & 9	25	24
9 & 10	22	25
10 & 11	40	42
11 & 12	25	30
12 & 13	18	29
13 & 14	23	30
14 & 15	34	25
15 & 16	33	38
16 & 17	28	31
17 & 18	20	16
18 & 19	19	22
19 & 20	35	28
20 & 21	27	26
21 & 22	22	29
22 & 23	21	24
23 & 24	24	26
24 & 25	22	30

writing over of a first recorded symbol by a second one seems to indicate a certain amount of dependence of one experimenter on the recording of the other. How much importance to attribute to these facts is still another unanswerable question on the basis of the data as they stand.

The present writer is inclined to the point of view that the deviations obtained in the category of like symbol pairs of hits may be treated as sampling deviations, *i.e.,* as indicating only chance variations. The reasons for such a treatment may be summarized as follows:

(1) The expected numbers of same symbol pairs of hits calculated by empirical probability methods should be looked upon as lower limits of the true expected numbers. Tables V and VI and VII also show that the variance of these values is quite high.

(2) The presence of approximately the same number of like symbol pairs of hits in both successful and non-successful experimenter's records. Comparison is rendered dubious here since the unsuccessful experimenter has a non-chance frequency of same-symbol pairs in his cards.

(3) The large negative deviation in the non-successful experimenter's DT same symbol pairs of hits which is produced by

the same number of pairs in the calls and approximately the same number in the cards as the large positive deviation of the successful experimenter in DT.

If the above arguments are accepted as approximations until the exact solutions are reached, certain interesting hypotheses may be made. The writer is fully aware that conclusions reached are tentative and will depend upon further analysis for their corroboration. However, if the lower limits of the expected numbers of the categories of Table III are used, the error will be on the side of underestimation rather than overestimation of the results. In Table VII approximations to the correct expected values are used in computing chi-squares. It is assumed for the present that only the relative frequencies of same and unlike symbol pairs of hits are affected by the calling preferences. Hence, the expected number of isolated hits and pairs of hits remains the same as in the Matching Hypothesis. How correct this assumption will be will depend upon further analysis.

TABLE VII

CORRECTIONS OF EXPECTED FREQUENCIES AND ACTUAL FREQUENCIES OF TABLE III FOR THE SUCCESSFUL EXPERIMENTER

	Obt.	*Exp.*	*D*	X^2	*P*
GESP					
Pairs of hits	440	295.8	144.2	73.26	<.001
Same Symbol Pairs	31	20.6	10.4	4.37*	.02
Unlike Symbol Pairs	409	275.2	133.8	67.89	<.001
Isolated hits	1127	995.6	132.6	20.28	<.001
DT					
Pairs of hits	328	295.8	32.2	3.51	.05
Same Symbol Pairs	30	21.7	8.3	2.90*	.10
Unlike Symbol Pairs	298	274.1	23.9	2.17	.10
Isolated hits	1084	995.6	88.4	9.05	<.001

* corrected for continuity (3, p. 3)

VI. DISCUSSION AND INTERPRETATION

Examination of Table VII indicates that the loci of extra-chance scoring in J.M.'s GESP work are in the unlike symbol pairs of hits and isolated hits category while for the DT work, the important locus is in the isolated hits category. If the hypothesis is accepted that the extra-chance influence does not extend significantly beyond the chance level to the same symbol pairs of hits category, an important hint as to possible interpretations of these results is obtained.

So far as the present writer can see now, the results are open to three possible interpretations. (1) According to one possible point of view, this analysis is invalid since it is based upon statistical artifacts caused by the discrepancy in the number of cases in the like and unlike symbol pairs categories. It should be pointed out, however, that this discrepancy is a function of the probability set-up in the ESP deck. (2) The ESP interpretation of J.M.'s GESP deviation on pairs of hits might be that the subjects "saw" the symbols in pairs. This point of view, however, seems to be questionable when the categories of same and unlike symbol pairs are considered. It would seem that once having hit a given symbol, it should be easier to hit the next one if it were a symbol of the same kind. This is certainly not the case, however, since almost all the deviation of pairs of hits is provided by the unlike symbol category. Similarly, the present writer cannot see any good ESP reason why the isolated-hits category should be at an extra-chance level in both GESP and DT for J.M. Does this show that the subject prefers to miss before and after he has called a hit?

(3) An hypothesis that seems to fit the facts better than the ESP interpretation, is based upon the assumption of inversion errors in recording the cards. Reference to Figure 1 will help to clarify the mechanism of these hypothetical errors. In column 1, positions 1 and 2, the calls were star and cross while the cards appeared as cross and star respectively. If the positions of the star and the cross in the card column were reversed (an inversion error in recording) (3, p. 78), a pair of unlike symbol hits would be produced which would automatically raise the score. In column 1, positions 17 and 18, two squares were guessed in the calls and two squares appeared in the cards. In the case of same symbol pairs, inversion adds nothing to the hit score. Inversion of two symbols may also at times produce a spurious same symbol pair of hits. In column 1, positions 17 and 18, the same symbol pair of hits could have been produced by inversion of the symbols in positions 18 and 19 or in positions 16 and 17 in the cards. Spurious isolated hits may be produced by inversion errors in which only one member of the pair produces a hit; thus, in column 2, position 7 in the calls is a wave and position 8 in the cards is also a wave. Inversion of positions 7 and 8 in the cards would produce an isolated hit, *i.e.,* no hit—hit—no hit. The present hypothesis assumes that inversions of both types occurred in J.M.'s GESP work and that the DT work contains inversions mainly of the second type. However, some evidence that the small positive deviation in J.M.'s DT unlike symbol pairs of hits may be due to inversions of the first type is found in considering the raw data in Figure 1. This set of 10 trials was collected by the DT method with J.M. as experimenter. It is headed "50c bet" and has the highest scoring rate in J.M.'s DT work. In the calls, only 10 cases of pairs of the same symbol are found

while the cards contain 47 such pairs. The expected number of same symbol pairs of hits is approximately .4. Actually, 1 pair of same symbol hits was scored. In the case of unlike symbol pairs of hits, the expected number is approximately 9.3. Twenty-one cases of unlike symbol pairs of hits were scored. If the assumption is correct that the large deviation on unlike symbol pairs is due to inversion errors, subtracting 22 (11 x 2) questionable hits (the unlike symbol deviation x 2) from the total of 69 would yield 47 hits, a deviation of —3 from the expected total number of 50 for 10 runs.

Finally, the question arises as to whether or not the large deviations given in Table I can be accounted for in terms of inversion errors. If the assumption is made that the deviations on unlike symbol pairs and isolated hits are products of recording errors, subtractions based on suspected errors would include for J.M.'s GESP deviation (1) approximately 260 questionable hits from the unlike-symbol-pairs category and (2) approximately 130 hits from the isolated-hits category, or a total of 390 hits. Since the total deviation for J.M.'s GESP work was 394, such a subtraction would reduce his results to the chance level. Similarly, for the DT work of J.M., (1) subtract 48 hits for unlike symbol pairs deviation and (2) 88 for isolated deviation, or a total of 136. The total deviation for J.M.'s DT work was 174, subtracting 136 yields a deviation of 38, a result well within the limits of chance variation.

VII. CONCLUSIONS

On the basis of the analysis presented in this paper, the writer believes that the following conclusions may be drawn:

1. When certain facts are made evident, it appears that the MacFarland results are questionable, even as to their extra-chance nature. The loci of extra-chance scoring in these data lend themselves to interpretation in terms of inversion errors in recording the cards.

2. The indicated numbers of hypothetical recording errors are sufficient to explain the extra-chance deviations in this experiment.

3. The relatively large number of recording errors actually found in checking the call columns lends weight to the card column errors explanation of MacFarland's results.

4. The presence of errors in recording in the call columns and the presumptive evidence for inversion errors in the card columns indicates the necessity for truly independent recording in which the recorder does not have knowledge of the subject's guesses while recording.

BIBLIOGRAPHY

1. Fisher, R. A. & Yates, R. Statistical tables for biological, agricultural and medical research. London: Oliver & Boyd, 1938. Pp. viii + 90.
2. Greenwood, J. A. & Stuart, C. E. A review of criticisms of the mathematical evaluation of ESP data. *J. Parapsychol.,* 1937, **1**, 295-305.
3. Kennedy, J. L. A methodological review of extra-sensory perception. *Psychol. Bull.,* 1939, **36**, 59-103.
4. MacFarland, J. D. Discrimination shown between experimenters by subjects. *J. Parapsychol.,* 1938, **2**, 160-170.
5. Stevens, W. L. Tests of significance for extra-sensory perception data. *Psychol. Rev.,* 1939, **46**, 142-150.

APPENDIX

I. GENERAL FORMULAE FOR THE EXPECTED FREQUENCIES OF TABLE II ACCORDING TO THE MATCHING HYPOTHESIS

Two identical decks of cards, Deck I and Deck II, are shuffled and matched.

Let N = number of cards in the deck

m = number of cards in a suit

n = number of suits

A. Same-symbol adjacent pairs of hits (E_s).

According to the method of Stevens, the probability for the occurrence of two same symbols in the positions *x* and *y* of Deck I and *x'* and *y'* in Deck II, thus producing a pair of same-symbol hits is

$$P_s = nm^2(m-1)^2/N^2(N-1)^2$$

Since there are (N—1) adjacent pairs of places in which the event can happen, the expected number of same-symbol adjacent pairs of hits is

$$E_s = nm^2(m-1)^2/N^2(N-1)$$

Evaluating for the ESP deck, (N = 25; m = n = 5)

$$E_s = 2/15 \text{ per deck}$$

B. Unlike-symbol adjacent pairs of hits (E_u)

Similarly,

$$P_u = n(n-1)m^4/N^2(N-1)^2$$

and

$$E_u = n(n-1)m^4/N^2(N-1)$$

Evaluating for the ESP deck,

$$E_u = 5/6 \text{ per deck}$$

C. Adjacent pairs of hits (E_p)

Since the two cases above represent the possible ways in which the event of adjacent pairs of hits can occur, the probability is obtained by adding the probabilities for A and B. Thus,

$$P_p = \frac{nm^2[(m-1)^2 + m^2(n-1)]}{N^2(N-1)^2}$$

and

$$E_p = \frac{nm^2[(m-1)^2 + m^2(n-1)]}{N^2(N-1)}$$

Evaluating for the ESP case,

$$E_p = 29/30 \text{ per deck}$$

D. Isolated single hits, i.e. Type (1) the pattern failure-success-failure and ($E_{I(1)}$), Type (2) successes in the end positions with failures adjacent ($E_{I(2)}$).

(1) failure-success-failure

Let *a* represent any particular symbol, *b*, any other symbol and *c* any symbol neither *a* nor *b*. Then five separate patterns can be distinguished with respect to the occurrence of symbols in any three adjacent positions, the *x*th, *y*th, and *z*th in Deck I. They are:

(A) aaa
(B) aab
(C) baa
(D) aba
(E) abc

The problem then can be solved by defining the number of patterns in Deck II which will satisfy the pattern failure-success-failure for each pattern in Deck I. Since the probabilities for (B) and (C) will be identical, the expression

$$P_1(_1) = P(_A) + 2P(_B) + P(_D) + P(_E)$$

will yield the required probability for Type (1) above.

It can be shown that $P(_A)$, the probability for Type (1) isolated hits in which the pattern aaa occurs in Deck I, is

$$P(_A) = \frac{(n-1)\,m^3(m-1)\ (m-2)\ [\,(m-1)\ (n-2)\,m]}{N^2(N-1)^2(N-2)^2}$$

Similarly,

$$P(_B) = \frac{n\,(n-1)\,m^4(m-1)\ [\,(2n-3)\ (m-1)\ +\ (2n-1)\,m]}{N^2(N-1)^2(N-2)^2}$$

$$P(_D) = \frac{n\,(n-1)\,m^3(m-1)\,[\,(n-4)\ (m-1)\ (m-2) + (2n-1)\,m\,(m-1) + (2n-4)m^2]}{N^2(N-1)^2(N-2)^2}$$

$$P(_E) = \frac{n\,(n-1)\ (n-2)\,m^4[\,(n-4)\ (m-1)\ (m-2) + (2n-2)\ (m-1) + (2n-3)m^4]}{N^2(N-1)^2(N-2)^2}$$

Evaluating for the ESP case,

$$P_1(_1) = .1276$$

Since there are (N—2) adjacent triads of places in which the event could happen

$E_1(_1) = 2.93457$ isolated hits of Type (1) per pack.

(2) successes in the end positions with failures adjacent.

Two different patterns can be distinguished in Deck I. They are

(A′) aa
(B′) ab

By the same analysis as above, it can be shown that

$$P(_{A'}) = n\,(n-1)\,m^3(m-1)\,/N^2(N-1)^2$$

and

$$P(_{B'}) = \frac{n\,(n-1)\,m^3[\,(n-4\,(m-1)\ +\ (n-2)\,m]}{N^2(N-1)^2}$$

Solving for $P_1(_2) = P(_{A'}) + P(_{B'})$ and evaluating for the ESP case,

$$P_1(_2) = .1597$$

Since there are only two pairs of places in the deck in which the event can occur

$E_1(_2) = .3194$ isolated hits of Type (2) per pack.

For the total expected number of isolated hits as defined above,

$$E_1 = E_1(_1) + E_1(_2)$$

$$E_1 = 3.2540 \text{ per pack}$$

EXPERIMENTS ON THE NATURE OF EXTRA-SENSORY PERCEPTION

III. The Recording Error Criticism of Extra-Chance Scores[1]

John L. Kennedy and Howard F. Uphoff
Stanford University

I. Introduction

In the first paper of this series, the senior author (5) presented the results obtained from an unsuccessful attempt to reproduce the phenomena of extra-sensory perception under the conditions recommended by those who have been successful in obtaining extra-chance scores in card-guessing. The results were negative with regard to the ESP hypothesis, but certain parts of the experiments were planned in such a way that they illustrate possible sources of error in ESP methodology which may help in understanding positive extra-chance results. The recording in three of the methods (GESP, DT and PT) was controlled by additional records, taken at the time of the experiment, by means of which the number and direction of recording errors could be determined accurately. Corrections in the critical ratios reported in the first experiment, necessitated by elimination of the effects of recording errors, are presented in the first section of the present paper. The second section will be devoted to a series of trials, obtained with a believer in ESP as recorder, in which the same controls were applied. Finally, since this work led us to suspect a correlation between belief in parapsychological phenomena and the tendency to make recording errors, an extensive study of this relationship, carried out with the aid of modern attitude testing techniques, is presented.

II. Recording Errors in Several ESP Methods

In the previously reported experiments with the GESP, DT and PT methods (5), recording of the card series was controlled by shuffling and recording 10 decks of cards before the experimental session. These decks were kept in order and the recorder was instructed to use them without cutting or shuffling. The call series was controlled by placing an observer in the experimental room who made an independent record of the subject's guesses. Errors in recording were discovered

---

[1] This paper has been read and approved by the Stanford Committee on Psychical Research as Communication No. 5 from the Psychical Research laboratory.

by applying these independent records to the card and call recording. Table I presents the number and effect of errors made by student recorders when the GESP method was used.

TABLE I

RECORDING ERRORS WITH THE GESP METHOD

	Calls			Cards		
Effect:	*Raise Score*	*No Effect*	*Lower Score*	*Raise Score*	*No Effect*	*Lower Score*
Number:	53	29	8	27	49	9

These errors increase the "true" score by 63 hits and spuriously raise the critical ratio by approximately 1 point. When this correction is applied to the total of hits for GESP work reported in the previous paper, it changes from 4987 to 4915, and the critical ratio (D/σ) from 1.06 to .08.

In the DT work reported in the previous paper, controls on recording errors were effected by applying the previously-recorded card orders to the subject's recording of the cards in the experiment. Errors in recording the call symbols were not considered since this recording was carried out without knowledge of the card symbols. Table II shows the same tabulation as Table I for the DT errors.

TABLE II

RECORDING ERRORS WITH THE DT METHOD

	Cards		
Effect:	*Raise Score*	*No Effect*	*Lower Score*
Number:	30	74	52

The DT errors spuriously lower the total score for this method by 22 and the critical ratio by approximately .6. When the correction is applied to the DT total of 1852, it becomes 1874, the expected number by chance is 1910 and the corrected critical ratio is .90.

For the PT work an independent record of the subject's call made at the time of the experiment was later compared with the agent's recording of calls. Table III shows the errors in the agent's recording.

TABLE III

RECORDING ERRORS WITH THE PT METHOD

	Calls		
Effect:	*Raise Score*	*No Effect*	*Lower Score*
Number:	4	54	0

The errors change the reported critical ratio so slightly that they may be ignored.

III. SPECIAL EXPERIMENTS WITH A BELIEVER IN ESP AS RECORDER

In this experiment a series of trials is reported based on the use of a modification of the Rhine Undifferentiated ESP [General ESP or Telepathy-Card] Test. The conditions were as follows: (1) The receiver was placed in one room, the agent in an adjoining room with an open door between; (2) instead of cards, the agent used a record sheet in which the order of symbols from 10 shuffled decks of cards had been recorded before the experiment; (3) an independent recorder was stationed in the receiver's room, recording the calls separately without knowledge of the card symbol. We believe that the agent was unaware of this check on her recording during the first 40 packs. The agent, in beginning a series of 25 guesses, concentrated on the first symbol in the record book and signaled this fact to the receiver by pressing a telegraph key. The receiver called out the name of the guessed symbol, which was recorded both by the agent and the independent recorder. The agent recorded the called symbol in the space in the record book adjacent to the symbol on which she was concentrating and went on to the next symbol. At the conclusion of the run of 25 the agent went back over the record and circled the hits.

Before discussing the unusual results obtained with this method, the character and beliefs of the participants should be given due regard. Both were elderly ladies. The agent believed strongly in the existence of thought-currents, was favorably disposed toward mediumistic phenomena, and reported personal telepathic and clairvoyant experiences. She had read reports of the work on telepathy and clairvoyance at Duke University and was extremely interested in attempting to produce equally good results in the Stanford laboratory. The percipient was interested in the experiments but remained sceptical about ESP even though the high scores were enthusiastically accepted by the agent.

Two groups of trials under these conditions are reported here. The first group of 40 packs or 1000 trials was gathered under conditions of excitement and expectancy of good scores without knowledge of the presence of the independent recorder in the receiver's room, the second group of 30 packs or 750 trials under conditions in which the agent knew that the recorder was keeping an independent record of the calls.

The total hits in the first set of 1000 trials was handed in as 229 with the expected number 200, yielding a critical ratio of 2.26. Although this difference is not statistically significant according to the generally accepted criterion, it is large enough to suggest an extra-chance cause. The scoring in successive sets of 10 decks was 71, 70, 47, and 41 respectively, showing that the positive deviation for the total 1000 trials was provided by large deviations on the first two sets of 10 packs.

When the recorder's independent record of the subject's calls was checked against the agent's record of the same calls, it was found that errors in recording had been made by the agent, errors which accounted for all the positive deviation in the experiment. Now and then the agent, instead of recording the subject's calls with strict accuracy, involuntarily (we believe) recorded the same symbol as appeared in the record book, the symbol she was attempting to send. In the check-up for hits after the run was completed, these erroneous hits were counted with the correct ones. In the first group of 10 packs, 21 errors in recording were found, 20 of which increased the score automatically and one had no effect on the score. The number of authentic hits was thus reduced to 51, a deviation of plus 1 from chance expectancy. In the second group of 10 packs, 23 errors were made, all in the direction of increasing the score. For the total set of 1000 trials, the effect of recording errors was to increase spuriously the number of hits by 46. With these errors removed, the critical ratio for the first group of 1000 trials became 1.33 based upon a negative deviation of 17 from chance expectancy.

In the second group of 750 trials, gathered when the agent knew that a check on the calls was being recorded, only three errors were made, two of which increased the score. With errors left in, the critical ratio, based upon a positive deviation of 10, was .89; with errors removed, .72.

These preliminary findings concerning the possibility of recording errors as an explanation of some ESP results and the apparent effect of a strong motivation of belief in increasing the frequency of such errors in the one case reported above led us to carry out further experiments in which the factor of belief could be more adequately measured and the experimental conditions for errors in recording more fully understood.

IV. Recording Errors as a Function of the Attitude of the Recorder

1. *The Measurement of Belief*

A survey of the field of modern attitude measurement led to the choice of the Likert (6) rather than the Thurstone (15) method for the construction of an attitude scale with two comparable forms for the measurement of belief in telepathy. Details regarding the Likert method of attitude scale construction may be found in the appendix of "Public Opinion and the Individual" by Murphy and Likert (7), as well as in an earlier monograph by Likert (6).

A brief characterization of the method may be useful here. Starting from the assumption that an opinion is a verbal expression of an attitude, 60 statements expressing opinions relating to belief or disbelief in telepathy were culled from various sources. After discarding the most obviously defective of these statements, the remaining 49 were arranged in random order with five possible responses beneath each statement as illustrated below.

1. Some people may be able to use telepathy.

Strongly Approve	Approve	Undecided	Disapprove	Strongly Disapprove

The preliminary test was then given to a group of 102 subjects taken from various psychology classes in the summer of 1938. The first analysis of the responses to these statements by the method of internal consistency (see below) showed that six more were comparatively non-differentiating. These six statements were also discarded.

The scoring of the scale was done in two ways, both of which are described by Likert (6). First, the sigma values for each response were calculated, *i.e.,* the frequency of response was reduced to a percentage of total responses to that statement, and from this the average sigma value of that percentage was determined. As shown in Table IV, the negative statements, *i.e.,* those which expressed disbelief in telepathy, had the largest negative value given to the "strongly approve" responses; the positive statements had the largest positive value given to "strongly approve" responses.

TABLE IV

SIGMA VALUES FOR TWO CHARACTERISTIC STATEMENTS

Telepathy has no scientific basis. (Negative statement)

	Strongly Approve	*Approve*	*Undecided*	*Disapprove*	*Strongly Disapprove*
Frequency of response	14	25	38	20	5
Percentages	13.7	24.5	37.2	19.6	4.9
Sigma values	—1.59	—.66	.17	1.09	2.10
Corresponding 1-5 values	1	2	3	4	5

Telepathy is possible between persons who are intimately acquainted. (positive statement)

	Strongly Approve	*Approve*	*Undecided*	*Disapprove*	*Strongly Disapprove*
Frequency of response	4	35	36	19	8
Percentages	3.9	34.3	36.3	18.6	7.8
Sigma values	2.16	.86	—.14	—.96	—1.86
Corresponding 1-5 values	5	4	3	2	1

The other method of scoring was much simpler insofar as it involved neither calculation of sigma values nor positive and negative values. The values 1, 2, 3, 4, and 5 were arbitrarily assigned to the various responses with the 1 given to the response which indicated strongest disbelief in telepathy and the 5 to the response which indicated strongest belief in telepathy. Table IV shows how these values were assigned.

Likert has found that the two methods of scoring yielded highly comparable results ($r = .987$ to $.997$). The similar correlation obtained in this study of .993, S.E. = .001 verifies Likert's findings and justifies the use of the simpler method as the only one employed in the subsequent work reported here.

The remaining 43 items of the original scale were given to another group of 182 subjects in an elementary psychology class in the fall of 1938. Since this group was larger and more homogeneous than the summer group, it was employed in the final selection of items for the scale. Two possible ways in which differentiating items could be selected for the final form of the scale were (1) by the criterion of internal consistency and (2) by the use of criterion groups. The criterion of internal consistency employs the two extremes of the experimental group itself. The proportion of cases at the extremes which have been used by various workers has not been constant. The Thurstones (15) in constructing their Neurotic Inventory used the highest and lowest 7%, Likert (6) and Hall (3) both used highest and lowest 9%, Allport and Vernon (1) and Rundquist and Sletto (11) used extreme quartiles, and Smith (12) in an inventory for the measurement of inferiority feelings used the 27% on the extremes. Extreme quartiles were selected in the present study since it was decided that with groups of this size the effects of errors or inconsistent responses on the part of a few subjects would be minimized. Critical ratios were obtained for each statement on the basis of the responses of these extreme quartiles.

Because of certain objections to the use of the criterion of internal consistency, criterion groups were also employed and the critical ratios between the responses of these groups also calculated. Fortunately, it was possible to obtain the coöperation of a Psychical Research Society,[2] the members of which were used as the criterion group of believers. Ten blanks were obtained from this group. For the criterion group of non-believers, graduate students in the Psychology Department of Stanford University provided 18 blanks. The distributions for these two small groups are given later with similar

[2] We wish to thank the members of the Oakland Society for Psychical Research who filled out the attitude scales.

data for the other groups studied.

It is interesting to note in Table V that, in general, the two methods select essentially the same group of items as most-differentiating. In Table V the item numbers refer to the items on the scale of 43 statements, while numbers in the adjoining columns refer to the sizes and ranks of the critical ratios which that particular statement yielded. Thus, Item I yielded a critical ratio of 5.2, which ranked 26th in order of magnitude when calculated using the responses of the criterion groups, and a critical ratio of 7.1 which ranked 15.5 when calculated using the responses of the extreme quartiles of the experimental group itself. It will be noted that in Table V only the rankings of the 33 most-differentiating items as selected by each method are given. It is evident that 31 items are thus selected by both methods. In other words, it would appear that the end result will be approximately the same whichever method is employed. The data given in Table V are calculated from the responses of the summer group (N = 102) and the criterion groups. A similar analysis was made on the basis of responses of the fall group (N = 182) and the criterion groups with almost identical results. In the latter case, 30 instead of 31 items were selected.

Considering the small size of the criterion groups, it seems remarkable that such close correspondence was obtained. Possibly, if larger criterion groups had been available, even closer correspondence would have resulted. The greatest inconsistency, which can be traced directly to the small size of the "Believer" group, occurs in the case of Item 6.

The items were arranged in rank order according to the size of the critical ratios obtained by each method and rank-difference correlations were computed and transformed into Pearsonian *r's*. The results of this analysis are presented in Table VI.

Utilizing the data obtained by both of the above methods, but relying chiefly upon the data from the fall group, the 38 most differentiating items were selected and divided into two groups. The items were arranged in rank order according to the size of the critical ratios obtained from the responses of the extreme quartiles of the fall group and then separated on the basis of odd and even numbered ranks. Some adjusting was done so that similar statements would not occur in the same scale and so that the sums of the critical ratios of each group were about equal. None of the statements used yielded critical ratios of below 3.6, and all but seven of them had critical ratios of 6.0 or greater. The nineteen statements in each form of the scale were

TABLE V

COMPARISON OF THE MOST DIFFERENTIATING ITEMS AS SELECTED BY BOTH THE CRITERION GROUPS TECHNIQUE AND BY THE CRITERION OF INTERNAL CONSISTENCY

Item Number	*Size and Rank of Critical Ratio Using Criterion Groups*		*Size and Rank of Critical Ratio Using Internal Consistency*	
	Size	*Rank*	*Size*	*Rank*
1	5.2	26	7.1	15.5
2	8.4	10	9.0	9
3	9.0	9	11.3	2
4	3.7	—	4.4	—
5	1.8	—	4.6	—
6	11.5	1	5.2	31.5
7	10.8	3.5	13.6	1
8	7.6	14	8.7	11
9	3.0	—	2.8	—
10	2.4	—	6.5	22
11	10.9	2	7.0	17
12	3.7	—	8.8	10
13	10.8	3.5	8.0	12.5
14	4.3	31	6.1	24.5
15	4.0	32.5	6.5	22
16	5.3	25	3.0	—
17	7.8	13	5.3	19
18	9.8	6	7.1	15.5
19	5.7	22	5.3	30
20	8.0	12	9.5	6
21	5.6	23.5	2.1	—
22	3.8	—	4.3	—
23	1.5	—	3.4	—
24	8.1	11	10.7	3
25	3.7	—	3.5	—
26	10.3	5	8.0	12.5
27	1.8	—	3.2	—
28	6.3	18	5.9	27.5
29	7.2	15	10.2	4
30	5.9	21	5.9	27.5
31	9.2	7	5.2	31.5
32	6.1	20	6.5	22
33	1.8	—	3.9	—
34	4.7	28	4.9	33
35	6.5	17	9.4	7
36	5.6	23.5	5.9	27.5
37	9.1	8	9.9	5
38	4.6	29	6.9	18
39	5.1	27	6.1	25.5
40	6.2	19	9.1	8
41	4.0	32.5	7.3	14
42	4.4	30	5.9	27.5
43	7.0	16	6.6	20

TABLE VI

CORRELATIONS BETWEEN RANKS OF CRITICAL RATIOS OBTAINED IN THE VARIOUS GROUPS OF SUBJECTS

Groups	*r.*	*S.E.*
Summer Group and Criterion Group	.59	.10
Fall Group and Criterion Group	.53	.11
Summer Group and Fall Group	.79	.06

then arranged in random order. The two forms are given below. The designations "Positive" and "Negative" were not placed after the statements on the scale. A positive statement is scored as shown in 1, with the 5 given for the "Strongly Approve" response, 4 for "Approve," 3 for "Undecided," 2 for "Disapprove," and 1 for "Strongly Disapprove." Similarly, the negative statements are scored as shown for statement 2. The total score is obtained by summing the numerical score for the statements. By this method of scoring, a high total score (range, 19-95) would indicate belief in telepathy while a low score would indicate disbelief. Needless to say, the numbers 1, 2, 3, 4, and 5 were not placed under the responses as shown here. The number in parentheses after each statement is the critical ratio obtained from the responses of the extreme quartile of the fall group.

FORM A

1. Some people may be able to use telepathy. (Positive) (9.5)

Strongly Approve	Approve	Undecided	Disapprove	Strongly Disapprove
(5)	(4)	(3)	(2)	(1)

2. I don't believe telepathy is possible. (Negative) (12.8)

Strongly Approve	Approve	Undecided	Disapprove	Strongly Disapprove
(1)	(2)	(3)	(4)	(5)

3. People who believe in telepathy are superstitious. (Negative) (6.7)
4. All so-called "proofs" of telepathy are merely matters of coincidence. (Negative) (10.8)
5. The next step in the evolution of the brain will be the development of telepathy between human beings. (Positive) (5.9)
6. It has been proved to me that there is no such thing as telepathy. (Negative) (5.5)
7. I have never seen nor heard of any telepathic results which could not be adequately explained by other factors. (Negative) (8.4)
8. Identical twins can use telepathy better than people not so closely related. (Positive) (4.2)
9. I believe that general impressions can be sent by telepathy. (Positive) (6.8)
10. I believe in telepathy because of experiences I have had. (Positive) (10.9)
11. It is conceivable that telepathy is possible. (Positive) (7.4)

12. Telepathy is "all bunk." (Negative) (12.2)
13. I would like to believe that telepathy is possible. (Positive) (6.5)
14. Telepathy is present to some extent in all of us. (Positive) (7.7)
15. I have witnessed phenomena which can only be explained by telepathy. (Positive) (9.3)
16. No intelligent person now denies that there is something to telepathy. (Positive) (3.6)
17. Telepathy has now been scientifically proved. (Positive) (6.2)
18. I am convinced telepathy is impossible. (Negative) (11.4)
19. People who believe in telepathy are more emotional than rational. (Negative) (8.0)

FORM B

1. I am convinced telepathy is possible. (Positive) (12.9)
2. It is silly to believe in telepathy. (Negative) (12.1)
3. I have been able to use telepathy at times. (Positive) (9.2)
4. The time will come when everyone will be able to use telepathy. (Positive) (7.9)
5. I believe more effort should be exerted in an attempt to discover more about telepathy. (Positive) (7.1)
6. People believe in telepathy only because they want to believe that man has some mystical power. (Negative) (7.6)
7. Telepathy is possible between persons who are intimately acquainted. (Positive) (10.6)
8. I do not understand how telepathy operates but I think there is something to it. (Positive) (12.2)
9. I believe in telepathy because of things I have witnessed. (Positive) (9.4)
10. Telepathy may occur when the receiver is hypnotized by the sender. (Positive) (4.1)
11. Only ignorant people believe there is anything to telepathy. (Negative) (5.1)
12. Telepathy has no scientific basis. (Negative) (6.7)
13. Most of my friends doubt the reality of telepathy. (Negative) (6.4)
14. Results of modern experimentation with brain waves indicate that there is a real basis to telepathy. (Positive) (5.6)
15. I believe that precisely-worded messages can be sent by telepathy. (Positive) (6.0)
16. The study of telepathy is a waste of time. (Negative) (8.3)
17. People who can use telepathy are probably further advanced

along the evolutionary scale than the rest of mankind. (Positive) (6.5)

18. Modern experiments prove beyond a doubt that messages can be sent by telepathy. (Positive) (10.0)
19. I believe in telepathy because I have had it proved to me. (Positive) (9.8)

The responses of the fall group were used in the calculation of the reliability of the scale. When this group was scored on Form A and Form B (19 items in each), the correlation obtained between forms was r = .92, S.E. = .011. No retests were given.

In the preliminary work which led to this scale it was found that some subjects had difficulty in defining telepathy. Therefore, on the first page of the scale itself, the following definition was given:

> Telepathy may be defined as the communication of feelings, impulses, ideas or more complex experiences from one mind to another, which is assumed to be effected without the use of sense organs.[8]

The definition was followed by this request:

> Check the position on the following line where you think your attitude toward telepathy would fall. (√)

Extreme Disbelief	Neutrality	Extreme Belief

The purpose of this rating was to discover the relationship existing between such estimates and the scores obtained on the scales. The line upon which the estimate was to be placed was 150 mm. long. The estimate of belief was reduced to a numerical figure simply by measuring the distance in mm. between the check mark and the "disbelief" end of the line. The range of estimates was from 3 to 136 with the mean at 65 or somewhat toward the disbelief end of the line for the fall group. The median estimate was found to be at 75.5, or practically at the midpoint. Table VII gives the data for both estimates and scores on the scale for this group.

TABLE VII
ESTIMATES OF BELIEF AND SCALE SCORES FOR THE FALL GROUP

	Range	*Mean*	*Median*	*Standard Deviation*
Estimates of Belief	3-136	65.00	75.50	27.71
Scores on Form A	29-82	57.16	58.40	9.74

It is interesting to note that both the median and the mean scale scores for the group fall very close to the figure which would result

[8] Taken from "Dictionary of Psychology" by H. C. Warren (16).

from checking the "Undecided" response to each statement.

The correlation between the estimates of belief and the scores obtained on Form A of the scale was found to be r = .73, S.E. = .034. These results agree with a similar comparison made by Thurstone and Chave (14), who report a correlation of r = .67 between attitude scores and self-ratings.

In Table VIII, distributions of Form A attitude scale results in various sample populations are given. The winter group was measured for the purpose of selecting subjects for an experiment to be discussed in the last section of this article.

TABLE VIII
ATTITUDE SCALE RESULTS IN SOME SAMPLE POPULATIONS

Group	Range[4]	Mean	Median	Standard Deviation
Fall Group, N = 182	29-82	57.16	58.40	9.74
Winter Group, N = 128	27.74	53.54	55.44	11.36
Graduate Group, N = 18	26-56	36.17	36.00	8.75
Psychical Research Society Group, N = 10	60-89	73.10	71.50	8.47

Since attitude scales deal with verbal behavior, their validity is usually difficult to determine. However, from Table IX it is possible to obtain some information about the validity of this scale. The graduate group was known to be composed of disbelievers, although one member *characterized himself as an agnostic* regarding telepathy. His score of 56, indicating a neutral attitude, was the highest of the group. The Psychical Research Society, on the other hand, was known to be composed of believers. Those who claimed to be neutral in their belief also obtained scores which approached neutrality. These facts, together with the fact that the scores of these criterion groups show no overlap, indicate that the scale measures belief in telepathy with considerable accuracy.

2.—*The Effect of Belief on ESP Recording by the GESP Method.*

The attitude scale makes possible the investigation of a number of problems relating to the dynamic action of belief or disbelief on the objectivity of response to situations involving responses which may be colored by such belief. The ESP recording situation as explained in the first section of this article and in a previous paper (4) appears to be of this type. Accordingly, the scale was used to select subjects for a controlled and "fixed" telepathy experiment in order to discover the effect of extremes of belief or disbelief on error commission using the GESP method.

In the winter quarter of 1938, Form A of the scale was given to

[4] On both forms of the scale the possible range of scores is from 19-95, 19 indicating extreme disbelief, 95 indicating extreme belief.

the elementary psychology class, and 14 subjects at each extreme of the attitude distribution were selected as recorders for the experiment. The score of each of these subjects was more than one standard deviation removed from the mean of the class. The mean score of the believers was 69.71; median 69; the mean of the disbelievers 34.86; median 35.00. These scores, it will be noted from Table VIII, closely approach the mean scores of the criterion group.

The subjects used in the experiment did not know its purpose. Nothing was said at the time of the experimental sessions concerning the attitude scale, which had been given to the class several weeks earlier, on the basis of which they had been selected as subjects. Identical instructions were read to each subject. The experimental conditions, rooms and equipment were the same for all subjects. The following instructions[5] were read to the subjects:

> In order to be sure that all of the subjects receive the same instructions I am going to read them to you.
>
> This is an experiment to test telepathic sending ability under conditions in which no sensory cues are available. Now, obviously it is impossible to test all the members of the class; therefore a random selection of subjects was made from the class roll and you were chosen as one of the subjects.
>
> In this experiment you will be the *sender* and I will be the *receiver*. You will remain seated at this table while I will go into the adjoining room and sit at a table at the opposite wall with my back toward this room.
>
> You will see before you a set of 10 decks of cards, the decks being numbered from 1 to 10 consecutively. You are asked to use the decks in order and in the manner I will describe. Also notice that in front of you is a record pad and a signaling key, the sounder for which is in the next room.
>
> Here is a sample deck of cards. It is composed of 25 cards, five each of 5 symbols; cross, star, square, waves and circle. (At this point, a card of each type was indicated.) For convenience in recording, use the symbols as shown at the top of this record sheet. (The commercial ESP record book was used in the experiment. The abbreviations O, L, $\wedge$, +, $\sim$, are recommended for recording circle, square, star, cross, and waves respectively. These abbreviations were pointed out to the subject.)
>
> When I go into the next room and say that I am ready, you are to take the first deck of cards and, without either shuffling or cutting, place it on the holder[6] before you in this manner with the bottom card revealed. It has been found that hand

[5] These instructions are based upon descriptions of the GESP method in Rhine (8).

[6] In order to minimize the possibility of the decks becoming disarranged, a small bookstand was used to hold the cards.

shuffling and cutting does not produce a random distribution, hence these decks have been machine shuffled. Concentrate on the first symbol revealed and at this time touch the telegraph key to let me know that you are ready. I will then make my call which you will record in the first square under "call" in the first column.

Immediately after you have recorded this call, take the card off the deck and lay it face down on the table before you. Concentrate on the next symbol, signal to me, record my call, remove the card, and follow the same procedure through the deck. (Here the procedure was demonstrated.) When I have given the last call, say "That's all" so that I will not be trying to get a further message.

Now, place the deck again on the holder and this time record the correct symbols for the cards in the second column under the figure I entitled "Cards." Do this through the deck. Each time you come to a row in which the two symbols—*i.e.,* the one in the call column and the one in the card column—coincide, place a circle around the two symbols as shown here. Total the number of hits in the pack, put the deck aside, and go on to deck number 2.

Most of the subjects are curious as to the results we get. So that you may be able to interpret our results let me say that by chance alone we would expect to get an average of about five hits out of every deck. Roughly speaking an average of anything over five is better than chance expectancy. It is not necessary to get a majority of hits, say 15, in order to get results above chance.[7]

In order to provide identical conditions for each subject, to control the experiment adequately, and to facilitate the search for errors in recording, eight decks of the ten used in the experimental sessions were arranged in the same order for each subject. The order of the cards was taken at random from decks in a previous experiment. The last two of the ten decks which each subject used were not prearranged and the results of these were not used in the experiment. They were injected merely to afford some variability in the total results in order to allay suspicion on the part of subjects who might have compared results.

Two separate sets of calls were arranged and recorded. One set yielded results substantially above the chance level when they were compared with the prearranged cards (average of eight hits per deck). The other sets of calls were arranged to give an average of approximately three hits per deck. Every attempt was made to eliminate error in making the calls. For this purpose a typewritten card was

[7] This information was provided to make sure that each subject knew the chance expectancy.

prepared for the calls which were to be given for each deck. Thus, two sets of eight cards were prepared, each of which was headed for the number of the deck for which it was destined and the series of calls to which it belonged, *i.e.,* series giving results above or below chance. The symbols were not used, but the words for the symbols (star, cross, waves, square, and circle) were spelled out on the cards, thus eliminating the confusion which often occurs in translating the symbols into verbal form. A sliding guide was used on these cards for the purpose of keeping place accurately.

Each subject served as recorder in two experimental sessions on different days. At one of the sessions, the subject received results in line with his expectations as predicted by the attitude scale; in the other, results contrary to his expectations. Thus, a believer would receive results above the chance level during one session and below during the other. The temporal presentation of the two series of calls was reversed for some subjects.

Since the card orders were the same in all cases and since there were only two sets of calls, the checking of records for errors was quite easily and accurately carried out. All the records were checked twice and errors in both call and card columns were counted.

In a total of 445 decks[8] of cards or 11,125 separate trials, 126 recording errors were made. All the subjects made some recording errors; four made only one error while one made fourteen and another sixteen errors. Table IX shows the distribution of error commission. When the errors are classified according to their effect on the score, it was found that 57 increased the score, 27 decreased the score, and 42 errors had no effect on the score.

TABLE IX
INDIVIDUAL DIFFERENCES IN ERROR COMMISSION

Errors per Experiment	*Number of Subjects*	*Total Errors*
0	0	0
1	4	4
3	8	24
4	8	32
5	4	20
8	2	16
14	1	14
16	1	16
	28	126

Table X shows the classification of errors according to the belief of the subject.

[8] Because of time limitations, one subject could stay for only 6 decks and another for 7.

TABLE X

COMPARISON OF ERRORS MADE BY BELIEVERS AND DISBELIEVERS

Subjects	*Increased Score*	*Decreased Score*	*No Effect*	*Total*
Believers	36	9	22	67
Disbelievers	21	18	20	59
Totals	57	27	42	126

It should be noted in Table X that both groups made approximately the same number of errors which had no effect on the score but that the number and particularly the percentage of errors increasing the score was considerably greater for the believers than for the disbelievers, while the number of errors that decreased the score was twice as large for the disbelievers as for the believers.

The errors were also analyzed according to whether the results were in line with the expectations of the subjects or contrary to expectations. Table XI presents the results of this analysis, together with the effects of the errors (+ = increased score, — = decreased score, 0 = no effect on score).

TABLE XI

EXPECTATION AND ERROR COMMISSION

Subject	*Results Contrary to Expectation*				*Results in Line with Expectation*			
	+	—	0		+	—	0	
Believers	18	3	19		18	6	3	
Disbelievers	14	8	10		7	10	10	
Total	32	11	29	72	25	16	13	54

The data in Tables X and XI offer some substantiation of the hypothesis that the direction of preconceived belief is an important factor in the production and direction of recording errors with the ESP method used in this experiment. The fact remains, however, that even the disbelievers made a considerable number of errors which increased the score although the results they were getting were opposed to their assumed expectation, and the believers made the same number of errors which increased the score under both conditions of results contrary and results in line with expectation (Table XI). These facts may be understood in the light of the experimental situation. Over and above the assumed effect of belief, the nature of the task set for the recorder was such as to load the error-making with errors that increase the score. Thus, the recorder is concentrating on the card symbol at the time he is recording the call symbol. Lapses in attention and consequent automatisms will at times allow the card symbol to be recorded in the call column. When the cards are recorded next to the calls, the same symbol is again recorded in the card column and a hit is automatically scored.

The 27 errors that decreased the score may be divided into six involving the recording of a wrong symbol and 21 errors of oversight. The type of oversight error is presented in Table XII.

TABLE XII

Errors of Oversight

Subject	*Failure to Encircle Hit*	*Error in Addition*	*Total*
Believers	6	1	7
Disbelievers	9	5	14
Total	15	6	21

There is evidence that the tendency to make errors diminished as practice advanced for the group as a whole. Table XIII presents the findings with respect to this question.

TABLE XIII

Errors Made During the First and Second Hours of Experimentation

Subjects	*First Hour*			*Second Hour*		
Believers	28	6	19	8	3	3
Disbelievers	13	9	12	8	9	8

However, when the records of individual subjects are studied, their rank in error commission generally remains approximately the same. Table XIV gives a detailed report of the error commission of the four subjects who made the greatest number of errors.

TABLE XIV

Protocol of Errors with Four Subjects

Deck Number	*Subjects* 26	19	11	12
1	1	4	0	0
2	0	1	0	2
3	5	0	0	0
4	6	0	1	0
5	0	1	0	0
6	0	0	1	2
7	1	1	2	0
8	0	0	0	0
9	1	0	0	0
10	0	4	2	4
11	0	0	0	0
12	0	0	1	0
13	1	0	0	0
14	1	0	1	0
15	0	3	0	0
16	0	0	0	0

Subject 26 improved considerably during the second session. However, in the case of the next three subjects whose total scores ranked next in magnitude, no improvement is evident. This finding agrees with that of Crosland (2) who reported that large individual differences with respect to errors in proof-reading continued in spite of practice.

V.—Summary

This paper was devoted to an experimental demonstration of the

recording error criticism of ESP methods. In the first experiment, student recorders with the GESP and DT methods committed enough errors to produce a marked effect on the total hit score and on the critical ratios for these methods. In one case, reported in more detail, recording errors when left in the data raised the hit score above the chance level; the results were slightly below chance when errors were removed. The fact that this individual was also a believer in ESP led to an experiment with college students as subjects on the effect of extremes of belief and disbelief, as measured by an attitude scale, on accuracy of recording with the GESP method.

Using the Likert method, two comparable forms of a scale, each consisting of 19 items, were constructed. The reliability of the individual forms was found to be r = .92. For the selection of most-differentiating items, both the method of internal consistency and the method of criterion groups were tried. It was found that essentially the same items were selected by the two methods. Estimates of belief, as measured on a rating scale, correlated .73 with scale scores. The validity of the scale was indicated by the fact that the criterion groups (which were not used in the construction of the scale) showed no overlap of scores.

Two groups of recorders were selected by means of this attitude scale, a group of believers on the belief extreme and a group of disbelievers whose scores fell in the disbelief extreme of the scale scores. These two groups of subjects served as recorders in a pseudo-telepathic experiment in which the tendency to make recording errors could be studied. Although the bulk of the evidence in this experiment is not striking, the direction of the errors appears to lend some support to the hypothesis that the direction of ESP extra-chance scoring may be influenced to some extent by the expectancy of the recorder who tends to make errors in the direction of his expectancy.

Some evidence was found that individuals differ greatly in their susceptibility to error commission. The groups of believers and disbelievers made fewer errors during the second hour than during the first, thus indicating a general improvement with practice. However, when the records of individual subjects were studied, it was found that they retained approximately the same rank in number of errors committed. Three of the four subjects with the highest error scores showed no improvement during the second hour of experimentation.

A consideration of recording errors and their possible effects on extra-chance scores reported in the ESP literature again shows the importance of truly objective recording devices in psychological work of this character. In the summer of 1938, Rhine (9) demonstrated an automatic recording machine for ESP experimentation. If such a machine is constructed to eliminate adequately the suspected errors of the bulk of previous ESP experimentation and if it provides a truly random selection of the material-to-be-guessed, questions as to the objectivity of extra-chance scores in ESP experiments will not be raised. Although two recent papers [Rhine (19) and Taves (13)] have been devoted to mechanical methods for recording and selecting material-to-be-guessed, we have not yet seen nor heard of extra-chance scores produced by an ESP machine which satisfy the above requirements.

VI. Conclusions

The following conclusions may be drawn from the results presented in the present paper:

(1) The GESP and DT methods as customarily used do not eliminate the possibility that recording errors may, in large measure, be responsible for extra-chance scores. With college students as recorders, it has been demonstrated that large discrepancies can be found between controlled and uncontrolled hit totals.

(2) The tendency toward error commission may be magnified by the presence of positive belief in the phenomena of parapsychology on the part of the recorder. One case is cited in which recording errors spuriously increased the hit score from a true chance performance to an extra-chance performance.

(3) When student recorders are selected who have extremes of either belief or disbelief in telepathy, their errors tend to be in line with their beliefs, *i.e.,* those who disbelieve tend to make errors which decrease the score.

(4) Serious attempts should be made to objectify the recording procedures in order to present crucial data on the ESP hypothesis.

BIBLIOGRAPHY

1. Allport, G. W. & Vernon, P. E. Studies in expressive movement. N. Y.: Macmillan, 1933.
2. Crosland, H. R. An investigation of proofreader's illusions. *Oregon Univ. Pub.,* 1924, **2**, pp. 168.
3. Hall, O. M. Attitudes and unemployment. A comparison of the attitudes of employed and unemployed men. *Arch. Psychol.,* 1934, 165 pp.
4. Kennedy, J. L. A methodological review of extra-sensory perception. *Psychol. Bull.,* 1939, **36**, 59-103.
5. ——— Experiments on the nature of extra-sensory perception. I. Repetitions of the Rhine experiments. [This issue.]

6. Likert, R. A technique for the measurement of attitude. *Arch. Psychol.*, 1932, 140.
7. Murphy, G. & Likert, R. Public opinion and the individual. N. Y.: Harper, 1938.
8. Rhine, J. B. Extra-sensory perception. Boston: Boston Soc. Psychic Res., 1934.
9. ——— The ESP symposium at the A.P.A. *J. Parapsychol.*, 1938, 2, 247-272.
10. ——— Requirements and suggestions for an ESP test machine. *J. Parapsychol.*, 1939, 3, 3-10.
11. Rundquist, E. A. & Sletto, R. F. Personality and the depression. Minneapolis: Univ. of Minn. Press, 1936.
12. Smith, R. B. The development of an inventory for the measurement of inferiority feelings at the high school level. *Arch. Psychol.*, 1932, 144.
13. Taves, E. A machine for research in extra-sensory perception. *J. Parapsychol.*, 1939, 3, 11-16.
14. Thurstone, L. L. & Chave, E. J. The measurement of attitude. Chicago: Univ. Chicago Press, 1929.
15. Thurstone, L. L. & Thurstone, T. G. A neurotic inventory. *J. Soc. Psychol.*, 1930, 1, 3-30.
16. Warren, H. C. Dictionary of psychology. N. Y.: Houghton Mifflin, 1934.

LETTERS AND NOTES

To the Editors:

December 27, 1939

Subject: Review—Size of Stimulus Symbols in Extra-Sensory Perception—J. G. Pratt and J. L. Woodruff

The Pratt-Woodruff paper has been reviewed by the following members of the Committee: L. Dick, J. J. Gibson, E. R. Hilgard, J. L. Kennedy, L. Long, R. R. Willoughby, and the writer. The specific statements are credited to the respective reviewers in the body of the report. Statements not credited to another member of the Committee are the responsibility of the writer.

The members of the Committee have been impressed with the thoroughness with which the experimental work has been conducted and the report written up. From the standpoint of "repeatability" the report is very satisfactory. The procedure has been described in complete detail. Every step is explicitly treated.

Several suggestions concerning experimental techniques and statistical procedures have been made by the Committee.

1. From the statement on Page 124 "With the pack face down in his hand in readiness for dealing, the E signalled to the S to begin by saying 'All right';" and from the description of procedure on Page 130, it is apparent that the technique used was *clairvoyance*. It seems desirable to state this explicitly (Dick and Long).

2. Although it might conflict with Rhine's theories concerning the relationship between E and S, it seems that one way in which causal data might be secured concerning the factor of "newness" would be to inform S that the sizes of the stimuli are to be changed, but to continue the use of the standard cards. This would also eliminate the checking of scores after each run; although E might check them without letting S observe and then give S the results (Dick and Long).

3. There is a question in our minds concerning the value of the attempt to prove the honesty of the E's. All efforts to eliminate errors in recording are obviously desirable, but actually it is impossible to establish one's integrity by the method used in this paper. Anyone might still doubt the honesty of the E by charging collusion. The burden of the proof rests with the "critics" and it would be more dignified to ignore the unscientific charges (Dick, Long, Sells).

In relation to this point, Dr. Kennedy has raised the following question: "Why not eliminate the human element in recording entirely by the use of a suitable mechanical system?" He has also asked: "Why didn't the second E watch the first E instead of the S alone?"

4. The recording of card symbols in Series B should be checked to see whether or not each symbol is presented at a frequency of five (Kennedy).

5. Series B and Table II appear to be a real contribution. The experimental conditions seem adequate and adequately described and the publication of hit distributions, subject by subject, is a minimum essential that has been long overdue (Willoughby).

The publication of individual records in Table II should set an important example for future writers on ESP. In this connection, Dr. Willoughby's request for individual Chi-Squares and the evaluation of the empirical Chi-Square distribution is extremely important. "One computational matter has been omitted. The reader can supply this for himself since the necessary data are given, but it would save time and effort if it could be printed. I refer to the absence of individual Chi-Squares and the evaluation of the distribution. These should measure the deviation of each individual performance from the theoretical distribution published by Huntington for this case (the binomial distribution might be used). Since in any distribution of Chi-Squares some Chi-Squares are expected to be high and some low by chance and since Chi-Squares have certain convenient properties, the individual Chi-Square coefficients should be distributed and a test of significance made between the obtained distribution and the theoretical Chi-Square distribution with appropriate mean and standard deviation. Presumably this comparison will show that the distribution of Chi-Squares is unlikely to have arisen by chance. However, this analysis ought to be made and reported in the text rather than left to the imagination or labor of the reader" (Willoughby).

It would be highly desirable either to follow this suggestion or to use an equivalent procedure which would report the probability values for individual subjects.

6. An effort should be made to determine that the card orders provided a chance distribution. It is likely that this information cannot now be obtained since the original order of the cards before matching was not recorded. I think that it would be interesting in the light of the "stacking" hypothesis (*i.e.* the presence of more

cards of a certain symbol in one of the five positions, than in the others), to carry out a Chi-Square analysis of the frequencies of the five symbols as they appeared in the five key card positions, irrespective of which key cards were attached to that position. If a certain symbol tended to show up more frequently in the right hand position than in the other positions and if the subject had a preference for placing the symbol there, extra-chance scores would be made. Such an analysis would check this hypothesis (Kennedy).

7. Dr. Hilgard offered the following criticism of the study:

"In studies in which the mean deviation from chance is slight, controls necessarily must be more meticulous than if deviations are large. In this study an extra hit is scored once in five runs through the pack. While this yields satisfactory critical ratios, and cannot be attributed to chance, very slight cues might make a difference *once* in each 125 calls. While the authors are quite aware of this, they have failed to control several cues, among which may be mentioned: (1) proximity of subject and experimenter, permitting auditory cues, (2) low screen, permitting observation by the subject of the experimenter's head. The differences between series B, with better controls, and series A, with poorer controls, conforms to the usual finding that the better the control, the less the excess over chance."

In view of the great care with which this paper has unquestionably been prepared, it is suggested that as many of the minor revisions recommended in this report as possible be made prior to publication. It is hoped that future writers on ESP will use the Pratt-Woodruff manuscript as a model for careful reporting of experimental results.

SAUL B. SELLS
Chairman
Board of Review

* * * *

In response to the Board's comments that Dr. Kennedy's article entitled "Repetitions of the Rhine Experiments" (p. 206 this issue) seemed biased in tone, Dr. Kennedy made important changes in it. No purpose seems served by full printing of criticisms based on another form of the paper.

January 19, 1940

To the Editors:
Subject: Review—Experiments on the Nature of ESP III. The Re-

cording Error Criticism of Extra-Chance Scores. Kennedy, J. L. and Uphoff, H. F.

The above paper has been reviewed by the following members of the Committee: Lillian Dick, Louis Long, R. R. Willoughby, and the writer. Specific criticisms submitted by members of the Committee are so designated. All other statements are the responsibility of the writer.

1. In view of the many references to recording errors in the literature on ESP a formal determination of its extent was necessary.

"As a reader I am pleased that the recording error has now been evaluated and a good estimate of its magnitude obtained" (RRW).

2. The reviewers all point out the significant facts, first, that the findings of this paper are general and not related to the interpretation or criticism of any specific research, and second, that the large extra-chance deviations obtained in such experiments as Rhine's could not be explained by recording errors of such slight magnitude as those found in these experiments. Drs. Dick and Long also direct attention to the finding that recording errors decrease with practice.

". . . its contribution to an evaluation of the possibility of the existence of the ESP phenomenon is negligible" (LD and LL).

". . . the finding that errors decrease with practice detracts from the importance of this experiment as a criticism of Rhine's work, since Rhine's recorders were all highly experienced" (LD and LL).

"The results, however, are both essentially negative and irrelevant to the criticism of the Duke results and others like them; and it obscures issues to conclude as if this were not the case" (RRW).

"We know now, not that 'recording errors may, in large measure be responsible for extra-chance scores', but that recording errors are negligible in the production of large scores, *even where they were allowed to occur*" (RRW).

3. A statistical test of the significance of the data presented in Tables X and XI should be made and included in the discussion (LD and LL, SBS).

4. The paper, in common with the preceding articles of Dr. Kennedy which have been reviewed, betrays an unscientific bias in favor of a particular result. This is demonstrated most noticeably on page 244: "Although two recent papers [Rhine (19) and Taves (13)] have been devoted to mechanical methods for recording and selecting material-to-be-guessed, we have not yet seen nor heard of extra-chance scores produced by an ESP machine which satisfy the above requirements."

Inasmuch as both papers referred to were published in 1939 and dealt with the development of apparatus in preparation for experimental work, it would be reasonable to be somewhat more patient. Nevertheless, the implication of the author's statement is that the phenomenon is frightened away by machines.

5. The writers have inadvertently (we hope) exempted themselves from certain rigid procedures which they have prescribed for so-called pro-ESP authors.

"The experiment reported in pages 237 ff., for example, might have been strengthened by the application of one of Dr. Kennedy's own suggestions with regard to Rhine's failure to use a machine for recording. A really convincing check on the agent would have been supplied by a machine. As long as there is a possibility of auditory cues (to say nothing of ESP) and as long as 'the attitude of the observer might influence his recordings', the authors should have made it impossible for such factors to operate by substituting a machine for the independent recorder" (LD and LL).

"In any case, the use of one elderly woman as an agent can scarcely constitute a check on Rhine's work" (LD and LL).

6. "The elaboration of a new attitude scale is a competent job, although it appears that here too, the results are essentially negative, *i.e.,* the simple cross on line device is for practical purposes just as good. The main experiment may perhaps be brought under the level of aspiration concept and is also interesting in spite of the slimness of results" (RRW).

Aside from whatever bearing it may have on ESP, the main experiment has much interest in relation to "frame of reference."

Dr. Willoughby's comment "that nothing is known about the factors causing the more adequate Duke results, though much effort has by this time gone into demolishing the obviously demolishable ones" epitomizes the reaction of the Committee to this paper.

Saul B. Sells
Chairman
Board of Review

January 8, 1940

To the Editors:

We have read the comments of the Review Committee with interest and have been glad to act upon their specific suggestions as fully as seemed to us consistent with the logic of our experimental comments upon the suggestion of the corresponding number in the Committee's report.

1. The procedure *was* clairvoyance, and we have been glad to state this explicitly at the proper point in the description of the procedure.

2. The research suggested here by two members of the Committee would be intended to give results bearing upon the question of whether the newness effect depends entirely upon the subject's attitude toward the testing material. We would like to call attention to Table VIII, which presents data relevant to this question, data which were obtained under conditions that did not necessitate the use of deception. In short, in a part of our experiment we told the subjects that different sizes of symbols would be used but that the subject would not know during a particular run which cards he was attempting to match. Under these conditions, the advantage of using "new" material largely disappeared.

3. We must confess that, personally, we share the Committee's doubt of the value of attempting to control against deliberate dishonesty on the part of an investigator. However, the question of the effect of bias on the part of the investigator has certainly been raised in the criticisms of past work in ESP. Some critics have gone so far as to charge individuals with deliberate dishonesty. What we aimed to accomplish in this experiment was to make it impossible (as far as was humanly possible) for the bias of any single individual to affect the scores either unintenionally or deliberately. Actually, we considered that certain features of the procedure bore upon the question of the trustworthiness of the investigators; but we did not intend to draw a sharp line between general carelessness in the handling and reporting of data on the one hand and deliberate dishonesty on the other. In fact, our conditions would have been exactly the same even if we had explicitly stated that we were not at all concerned with the deliberate intention to deceive. But since the procedure did bear upon this possibility as well, it seemed best to point out that fact for the benefit of those few who choose to raise the "unscientific" issue. Of course, no control against collusion to deceive was intended; nor, in our opinion, is any possible.

The suggested use of a suitable mechanical system is appropriate enough, if only sufficient financial resources were at hand to develop apparatus to this end, which inventive ingenuity has already advanced to the point of demonstrated practicality. The further suggestion from Dr. Kennedy that the second experimenter should have watched the first seems beside the point, as the primary necessity for completely independent recording of the key cards and the distributed pack required that the E's *not* watch each other from the beginning of a run until after the symbols were recorded. And at all other times in the investigation the experimenters did watch each other.

4. The reason for this suggestion to count the frequency at which each symbol was recorded in each run is not clear. With the use of independent recording, the only circumstance under which a lack of balance in the symbols could affect the scores in a consistent direction is the highly unlikely one of using an unbalanced deck with the subject knowing or preferring the symbols of higher frequency and favoring those key cards in his pointing. We believed at the time of the experiment that we were always using balanced packs, and at no time did we observe that this was not the case. In the absence of any definite proof of this fact at the present time, however, we decided to count the distribution of symbols in a representative sample of the recorded runs. The symbols in each run with a serial number ending in zero were counted. In this way, the distribution of symbols was checked in each of 240 runs, or one-tenth of the total of Series B. This count was first made by Woodruff, and then repeated by Pratt without any knowledge on his part of the results found by the first E. No departures from five of each symbol in any of the 240 runs counted were found by either experimenter.

5. We have added a new table (II B) to meet the suggestion of Dr. Willoughby that the work of individual subjects be evaluated independently. In this table we have presented independent chi-squares for each subject who did at least 40 runs and have combined those of less than that number in a miscellaneous classification. This gave 19 sub-series, for each of which a chi-square evaluation was worked out. We were unable to comply with the second part of Dr. Willoughby's suggestion, namely, that a chi-square of the empirical distribution of individual chi-squares be obtained, for the obvious reason that 19 cases are not enough for the proposed evaluation. We did, however, make a direct combination of individual chi-squares and obtained the associated P-value for the total number of degrees of freedom. We have included the critical ratio evaluation for individual subjects for good measure.

6. The suggestion by Dr. Kennedy to make a chi-square analysis of the frequency of appearance of symbols in the various positions

irrespective of key cards is appreciated for the fact that, unlike the statement in paragraph 4, he offers a definite hypothesis upon which the chi-square results might bear. But even if such a study were made with significant results, there is a further assumption in his argument that such extra-chance consistency could produce spurious results; viz.: "If a certain symbol tended to show up more frequently in the right hand position than in the other positions and *if the subject had a preference for placing the symbol there* (italics ours), extra-chance scores would be made. Such an analysis would check this hypothesis." The analysis suggested goes only half way, as it concerns only one of the assumptions: that of the consistency of placing the various symbols of the pack in the five possible positions. For an analysis relevant to the hypothesis in question, it would be necessary to make a study of the consistency of distribution of the key cards as well.

But knowing what hypothesis is in question, it is possible for us to state that we have taken this specific point into account and have checked upon the possibility of making this interpretation of the results by means of the cross-checking of the data as described on page 142 of the report. In other words, if the significant average obtained in the experimental series was attributable to a general tendency for the symbols of both the ESP pack and the key cards to fall in the same positions from run to run, significant scores on the cross-check should have resulted. The fact that the cross-check average for Series B was only 5.02 is clear evidence that such an explanation of the significant experimental results could not be offered.

7. We are unable to evaluate Dr. Hilgard's criticism as he fails to show wherein the discussion of the question of sensory cues in Series B on pages 135-9 is inadequate to support our conclusion that sensory cues were not the basis of the results.

J. G. PRATT
J. L. WOODRUFF

[illegible]