[illegible] December, 1940 Number 2

The Journal of Parapsychology

[illegible] and Bernard F. Riess, Editors
[illegible], Managing Editor

CONTENTS

The [illegible]

QUARTER[illegible] WITH THE[illegible]

[illegible]VEYANCE[illegible]

[illegible] Journal is pu[illegible] all [illegible] Murphy. O[illegible] Riess, Hunter C[illegible] Correspond[illegible] [illegible] is advised [illegible] of [illegible]mentation. Since th[illegible] mail, it [illegible] will not [illegible] and contrib[illegible] copies of their [illegible] should [illegible] in [illegible]

[illegible] year. [illegible] issue[illegible] Correspondence [illegible] subscriptio[illegible] back numbers, and other [illegible] communicatio[illegible] to Edward F. Riess, [illegible] College, New Y[illegible]

[illegible] subscription pric[illegible] [illegible] annual issue [illegible] back [illegible] Volumes 1 a[illegible] numbers will [illegible] when [illegible] is given [illegible] of [illegible] should [illegible] Edward F. Riess, Hunter [illegible] York City.

[illegible] of R[illegible] includes [illegible] D. Selig, Louis Lo[illegible] Lillian D[illegible] J. L. R[illegible]

[illegible] have [illegible] mathematical co[illegible] are [illegible] involved.

[illegible] through the [illegible] of [illegible] will be [illegible]

STUDIES IN EXTRA-SENSORY PERCEPTION:

III. A REVIEW OF ALL UNIVERSITY OF COLORADO EXPERIMENTS

By

Dorothy R. Martin and Frances P. Stribic

ABSTRACT

All data, with the exception of a short series designed to test for telepathy, obtained by the experimenters over a three year period in the University of Colorado laboratory are here reported. The data embrace a total of 12470 runs (311,750 trials) performed by 322 subjects. Four primary experimental series are included: 1. a preliminary series of 3171 runs performed by 39 subjects in 1937; 2. a group series of 2840 runs performed by 284 subjects in 1938-1939; 3. a main experimental series involving 4180 runs performed by 13 subjects in 1938-1939; 4. a series of 2279 runs involving several variations in experimental technique and performed by a single subject in 1938-1939. In all except the preliminary series, tests were screened. Unless otherwise indicated the DT or 'down through' technique was used exclusively.

A control series obtained by matching the subject's guesses against the reverse order of the pack was employed throughout. This series is termed the *reverse* series. An additional control was supplied by chance expectancies as computed upon the binomial hypothesis.

Results for the four series yield *direct* scoring averages of 5.80, 5.00, 5.79, 6.90, with critical ratios of 24.09, -.12, 25.19, 43.35, respectively, as against *reverse* scoring averages of 4.98, 4.95, 4.97, 5.04, with critical ratios of -.40, -1.39, -.90, 1.06, respectively. The aggregate data yield a *direct* scoring average of 5.83 with C.R. 45.70 as compared with a *reverse* average of 4.98 with C.R. 1.00.

Detailed analysis of the data reveals: 1. pronounced variation in the scoring success achieved by different subjects, 16 subjects attaining averages which yield critical ratios above 2.5, 315 subjects attaining

averages within the -2.5 to +2.5 C.R. range. A single subject (who performed only 30 runs) achieved a score with a critical ratio below -2.5. 2. marked daily fluctuation in scoring for both high and low-scoring groups, with notably greater consistency among the former. Certain unusual scores suggest possible correlation between scoring success and physical fitness. 3. a significant excess for consecutive hits in the *direct* scoring of the high-scoring group. Scoring advantage would appear to reside primarily in this excess. 4. inconclusive results with respect to the relation between scoring success and card position. For a single subject, however, there is maintained a sufficiently direct relationship between these factors to suggest the possibility that there may be a general functional relationship. 5. no evidence in support of the implication that records of high-scoring 'subjects' may have arisen from fortuitous confluence.

INTRODUCTION

In December, 1936, experimental investigations were undertaken at the University of Colorado with a view to examining the hypothesis of Extra-Sensory Perception. It is the purpose of this paper to present in rather great detail a complete resumé of all work relating to the clairvoyance aspect of the hypothesis. Much of this work has been exploratory in nature and has little of interest to offer. Certain of the more significant results have already been reported in this journal (7, 8, 9). It seems, nevertheless, desirable that the entire collection of data be presented at this time in order that a background may be offered against which the more significant portions of the work may be evaluated.

EXPERIMENTAL MATERIALS

The materials used throughout all tests consisted of so-called standard ESP cards. These cards were arranged in packs of twenty-five, there being in each pack five cards of each of five designs (star, cross, waves, circle, rectangle). Prior to the fall of 1938, and throughout *all* of the experimental work with our special subject Mr. C. J., heavy hand-stamped cards supplied through the courtesy of Dr. J. B. Rhine were used exclusively. These cards present few if any visual or tactual cues. Subsequent to September, 1938, when an adequate supply of these cards was no longer available, the regular commercial cards were substituted for all subjects except, as previously noted, for Mr. C.J. These cards, which are by no means free from visual and tactual cues, were used only under screened conditions.

GENERAL PROCEDURES

Orientation

Prior to the undertaking of the initial test, all subjects were informed as to the purpose of the tests, were given a description of test procedure, and were invited to examine a sample pack of ESP cards. Brief comment was made on the results to be expected on the basis of a chance hypothesis and on reported deviations from this hypothesis. Skepticism toward the ESP postulate was given generous expression. The need for extensive and rigorously controlled experimental investigation was emphasized.

Procedure for individual tests.

The general procedure for individual tests is recorded below. Any exceptions to this procedure will be treated separately in subsequent sections of this report.

The subject was seated at a table and was separated from the experimenter by a 15"x18" opaque wooden screen without apertures. The DT or *down through* method was employed throughout. The experimenter shuffled and cut the pack, and placed it face down before her, taking care to avoid seeing the bottom card. The subject then recorded twenty-five guesses for the *down through* order of the pack. After the completion of each run, and in the presence of both subject and experimenter, the pack sequence was recorded and checked against the guessed order and the coincidences noted. The data obtained by the matching of the guessed order against the card order for which it was intended will be termed the *direct series* (in contrast to the *reverse series*---a control series derived from the matching of the subject's calls against the reverse order of the pack).

Recording Techniques for Individual Tests

Prior to the fall of 1938, all recording in the individual experiments was done by the subject under the careful surveillance of the experimenter. At the conclusion of the recording of guesses for each run, the subject brought his record to the experimenter, seated himself at her left, and recorded the card symbols as the experimenter called them from the target pack which lay face up on the table before her. Only after the recording was completed and this recording checked by the experimenter was the card removed (and turned). A correspondence or 'hit' was indicated (by a horizontal line through call and card symbols) immediately upon being encountered and this hit was checked by both subject and experimenter against the card which lay face up on the table. At the end of each recording, the hits for the entire run were summed and the total score entered on the record sheets. It would thus appear that a genuine hit would rarely fail to be recorded except in the event of a miscall of the card by the experimenter. On the other hand, a recorded hit which was not genuine would require that the experimenter not only fail to check a particular recording but also fail to observe and check the recording of the hit line. This double failure is conceivable. To the authors, however, it would seem to constitute extreme experimental negligence.[1]

That errors occur despite precautions is evident from the fact that in 1000 runs of the UT series (p. 175) which have been subjected to an intensive error search, there occurred 58 runs in which 64 recording errors were caught and corrected at the time of recording (these were primarily of the slip of tongue or of pen variety) and 5 runs which were discarded due to violation of the 55555 check.[2] Of the 64 detected errors, 4 would have affected the score favorably, 5 adversely, the remainder leaving it unaffected, thus indicating that the preponderance of error is confined to the less-interesting and probably less attended non-hit regions. This suggests a slight relaxation of vigilance on the part of both experimenter and subject in these more monotonous stretches. That such 'relaxation' might have led to increased scores in the form of unmarked hits is, of course, possible. Actually, 2 such hits were found, both in recordings in which the 55555 check was violated.

Inversion errors induced by the subject's call series would appear to be excluded by the technique of recording since these errors require the retarded recording of a pair or sequence which has been called as a pair

1. One subject, (in the 1937 series) assuming such negligence, made repeated attempts to fraudulently achieve high scores. Failing this, the subject discovered 'time pressure', and withdrew from the experiment. The records were discarded.
2. This check has, as a matter of course, been applied to all data. The 55555 check is a determination of the frequency of each symbol in the deck. There should be, in this case, five of each.

or sequence. The remote chance that there occur consecutive recording errors equivalent to the inversion error and thus immune to the 55555 check is present but unimpressive as a scoring factor. The hypothesis that error of inversion might play an important role in extra-chance scoring has been entertained by the authors since the beginning and was, perhaps, decisive in the choice of a recording technique.

Since September 1938, the recording has been done by the experimenter, the subject's guesses being shielded during the process and the 55555 check being applied upon the conclusion of the recording of each run. The incidence of recording error as evidenced by detected error has shown marked increase under this procedure, 76 instances of misrecording of some form having been noted and corrected during the experimental period. No errors have been found through application of the 55555 check subsequent to the experimental period. Since the added vigilance at the hit position was here missing, it would be anticipated that the incidence of error affecting the score would display an increase. Thirty-three such instances, 12 favorable to increased scoring, 21 unfavorable, occurred.

Test Supervision

All experimental work was conducted under the supervision of at least one of the authors. This does not, however, imply that one of the authors was present during the administration of every individual test. Several student assistants participated in the test administration. These assistants, though not of graduate standing, were mature students whose major interests were in the fields of mathematics or of psychology. Integrity and intelligence, together with demonstrated reliability and established accuracy in performance, formed the basis for their selection. Compensation, supplied by grants from the National Youth Administration, was in no wise dependent upon the continuance or upon the outcome of the experiments. These student assistants were carefully trained, first as witnesses, later as experimenters under observation, before being entrusted with the independent administration of tests. They were assigned only to exploratory tests or to the continued testing of subjects who had long maintained a relatively consistent scoring record. All test administration was subject to supervision in the form of witness or taking over by author D.M. without warning. No assistant was himself able to achieve extra-chance scores either in standard screened tests or in such 'off-the-record' tests as were self administered. No evidence in support of the hypothesis that scoring success is a function of the individual experimenter was obtained.[2a]

Procedure for Group Tests

In the group tests, which were purely exploratory in nature, the following procedure was observed. Each member of the group was provided with a standard record sheet and was asked to record his guesses for the DT order of the pack which had been shuffled, cut, and placed by the experimenter on the deck behind a reading stand which served as screen both for the group and for herself. When all had finished recording guesses, the order of the target pack was slowly called to the subjects and recorded by them. The experimenter herself kept an independent record of the card order. Each subject was asked to indicate hits as noted and to sum his score before proceding to the next test. When, for purposes of completeness, it seemed imperative to include these tests in this report, the record of the experimenter was compared with the guessed order of each subject and the official score (and the incidence of recording error on the part of the subject) noted. It is, of course, at once obvious that under this procedure an impressive scoring record could be fraudulently achieved by any individual subject through the simple device of deferred recording of guesses. This flaw in the procedure was exposed to each group at the end of the experimental period, and the reaction of each was such as to suggest that the thought had failed to occur to the members. The essentially chance nature of the results would tend to indicate that this form of potential fraud was negligible in incidence, if extant at all.

Official scoring of this series of 2840 runs reveals in the unofficial student recording 607 instances of recording error, of which 149 affected the score positively and 86 affected the score negatively. In Table 0.1 these errors have been tabulated according to type. From this tabulation it appears that greatest error incidence results from the 'chain error' in which one misrecording in the run causes displacement of all succeeding card recordings. This type of error would, of course, be subject to immediate detection in individual tests. The relatively high incidence of miscellaneous error

2a. Table 0.1A (Appendix) presents scoring averages obtained by each of several experimenters in a series of 1000 DT runs performed by a single subject.

which would have affected the score positively, suggests either deliberate misrecording or a tendency to record the symbol seen rather than the symbol called by the experimenter. The low incidence for the error of inversion is of peculiar interest to the authors since only this type of error or its equivalent could escape detection on application of the 55555 check. Scoring errors which do not involve errors in recording are of little significance in an error study since they would, of course, be eliminated in any recheck of the data.

Since recording conditions which obtained for these tests were as conducive to error as any which are conceivable, it would seem that the error data here presented might offer something of an upper bound or, more conservatively, at least an average measure for such error incidence as might be induced by the subject's unshielded calls.

CONTROL SERIES: THEORETICAL AND EMPIRICAL

Theoretical Series

Perhaps the most interesting controls are those theoretical ones deriving from the hypothesis that chance alone is operative in scoring success. On the assumption that the subject's calls are truly random--i.e., that each call is independent of all others--the distribution of scores is given by the binomial $(4/5 + 1/5)^{25}$, with mean = 5, and variance = 4. This distribution for 1000 runs is given in Table 0.2. If, however, it be assumed that knowledge of the pack composition will constrain the subject to call exactly five calls of each symbol--a situation obviously realized in the matching of two ESP packs against each other--the frequency distribution becomes that of the matching hypothesis. Exact frequencies for this case have been supplied by Greville (4). For purposes of ready comparison with the frequencies of the binomial hypothesis, these frequencies for 1000 runs are included in Table 0.2. Greenwood (3) has shown that the variance for this case = 25/6. That either of these hypotheses will obtain in practice is, of course, unlikely. Regardless, however, of the nature of the call series, it is established: 1. that the expected score per run is always 5; 2. that the maximum possible value for the variance (25/6) is that of the matching hypothesis; and 3. that either the binomial or the matching hypothesis is adequate in supplying theoretical values for the distribution of successes (2, 3). It is of interest to note that Greenwood (2), in an extensive empirical investigation of a free calling chance series, has found the binomial hypothesis to offer the best description of the data actually assembled by him.[3]

Matching Series

In order to test empirically the adequacy of the mathematical predictions, several empirical controls were introduced. The first was derived from 1000 matchings of two well-shuffled packs. For purposes of checking and analysis, these pack sequences were recorded. Results ar are presented in Table 0.2. This series will be termed Matching Series I.

A second matching series of 1000 runs was derived from the recorded pack sequences arising in a previously reported study (9). In this series, each recorded pack sequence was compared with the pack sequence consecutive to it and the coincidences noted, thereby subjecting each target pack to a mechanical or 'robot' calling. In order to achieve 1000 comparisons, the last sequence was compared with the first. This series, which is also presented in Table 0.2, will be termed Matching Series II. This series supplies not only an empirical control on theoretical frequencies, but also a check on the adequacy of shuffling for the 1000 runs involved.

Reverse Series

In order to provide a more adequate empirical control, each run of the entire experimental series has been subjected to an additional and, we believe, valuable check. In it, the subject's guesses for each run were compared with the reverse (in general, the *up through*) order of the pack and the coincidences noted. The series thus resulting will be termed the *reverse series*. This series supplies a control which for each subject exhibits the same symbol preferences and pattern idiosyncracies as does the *direct* experimental series. Should the scoring effects of these preferences or pattern idiosyncracies be negligible (as theoretical considerations would indicate) this series should yield an essentially chance distribution. This it does (subject to such variation as may be introduced through extra-chance scoring on card 13, the middle card, which is common to the two series).

3. The binomial hypothesis will accordingly, with one exception, be used for theoretical predictions throughout this paper. In all references to critical ratio, however, the more conservative values derived from the matching hypothesis will be recorded.

I. PRELIMINARY SERIES

During the fall of 1936 and the spring of 1937, exploratory experiments in the field of Extra-Sensory Perception were initiated. The procedure employed in this series was that of the DT or *down through* technique described in a previous section, with the exception that in this instance the tests were unscreened. Thirty-nine subjects participated in these experiments. All were university students; most of them were sophomores taking the required course in General Psychology. The group was composed of individual volunteers. They were almost equally divided as to sex. With one subject, Miss E.W.[4], the G.E.S.P. technique[5] was substituted. With another subject, Miss D.W., distance[6] was introduced as an experimental factor. With this subject 110 additional runs performed under distance conditions have been added to those already reported (7).

For convenience of reference, those subjects who completed 100 or more runs will be termed *major subjects;* those who completed less than 100 runs will be termed *minor subjects*. Since a detailed report of results achieved by each subject is already available (7), the results gathered from *minor subjects* will here be grouped.[7]

Results of Preliminary Series

A summary of results (both *direct* and *reverse*) for this preliminary series is recorded in Table 1.1. It will be seen that the *direct* series yields an average of 5.80 hits per 25, with the highly significant critical ratio of 24.09, whereas the *reverse* series yields an average of 4.98 hits per 25 and the clearly non-significant critical ratio of -.40.

4. A detailed account of the procedure with this subject is given in a previous article (7).
5. A technique in which either clairvoyance or telepathy or both may be operative.
6. A total of 200 were performed under 'distance' conditions. These include 60 runs (Av. 7.60, C.R. 9.87) in which the subject and experimenter were separated as far as a room 20'x20' would permit; and 140 runs (Av. 7.84, C.R. 16.49) in which subject and experimenter were in adjoining rooms.
7. For readiness of reference, scoring summaries for each of these *minor* subjects are included in Table 1.1A (Appendix).

Frequency data for the two series are recorded in Table 1.2 and are given graphic representation in Fig. 1.2. For purposes of comparison, frequency predictions for the entire series based upon the binomial hypothesis are included in both table and figure. It will be observed that modes for both *direct* and *reverse* distributions occur, in accordance with chance expectation, at 5. It is to be noted, however, that the *reverse* frequencies correspond closely throughout the entire range with those predicted by chance ($\chi^2 = 8.58$, $n = 12$, $P = .75$), whereas the *direct* series exhibits a significant excess of frequencies above 6 and a corresponding deficiency of frequencies below 6 ($\chi^2 = 1673.10$, $n = 12$, $P = < 10^{-18}$). This discrepancy between chance prediction and the *direct* series must be largely attributed to the performance of the six *major* subjects who achieved average scores which on the basis of a 2.5 critical ratio criterion[8], may be regarded as statistically significant. Individual frequency distributions for these subjects reveal that for Miss E.W., the mode occurs at 8; for Miss D.W., for Mr. C.J., and for Miss B.C., at 7; for Mr. C. B. at 6; and for Mr. E.A. (whose average score of 5.60 with C.R. of 3.94 is not impressive), it appears at 4. It is interesting to note, however, that it is this latter subject who has scored the highest number of hits per run ever to be obtained in this laboratory. For readiness of comparison, an aggregate distribution table for these high-scoring subjects and for all other subjects[9] has been presented in Table 1.2. The contrast between the performance of those subjects who score significantly and those who do not becomes increasingly evident on further detailed analysis. The records of all subjects have, accordingly, for this and for all subsequent series, been subjected to analysis with respect to: 1. daily fluctuation in scoring, 2. frequency of consecutive hits, and 3. frequency of hits with respect to card position.

Daily scoring averages for each subject in this series appear in Table 1.3A[10]. It is of interest to note that on July 27, Miss D.W. for the first time approached

8. For purposes of classification, this basis for significance has been arbitrarily adopted throughout this paper. The authors should, however, hesitate to accept this criterion as an adequate basis for crucial decision.
9. Of these low-scoring subjects only one achieved a C.R. below -2.5. This subject in 30 runs attained an average score of 4.00 with a C.R. of -2.60.
10. Tables which bear the label 'A' appear in the Appendix.

a chance level of scoring. On inquiry, it was learned that she had neglected lunch and had devoted the two previous days---except for a few hours stolen for sleep--to the typing of a thesis. Paradoxically, she here displayed for the first time, interest in her scoring ability, expressing confidence that when rested she could again achieve her previous scoring record. Her acceptance of a teaching position and departure shortly thereafter from the university made it impossible to put this to the test.

In Table 1.4 is presented the tabulation of consecutive hits (*direct* and *reverse*) for both *major* and *minor* subjects together with the predictions (on the basis of the binomial hypothesis) for these consecutive hits. The table reveals consistent differences both between the *direct* and the *reverse* series and between the *direct* series and chance predictions as against a close parallelism between the latter and the *reverse* series. From this table it would appear that for those subjects who score significantly, the discrepancy between *direct* and *reverse* series results not only from an increased frequency of isolated and paired hits, but also from the greater frequency of consecutives of higher order.

Summaries of hit frequency with respect to card position for both *direct* and *reverse* series are presented in Table 1.5A. This table contains a tabulation of the frequency of scoring for the top card, the second card, the third, etc. In Table 1.5, average scores (per 25) for the first, second, and succeeding five-card groups are recorded. These analyses, undertaken as a result of experiments previously published and later to be reviewed in this paper (p. 29), reveal for this series a slightly W-shaped pattern for scoring success. This would seem to confirm findings of Rhine, Pegram and others (11, 10, 1) who report for the DT method a W or U-shaped curve of operation. Comparison with the *reverse* data, together with statistical treatment (maximum critical ratio of the difference between *direct* scores for the 5-card groups being 2.02) fails, however, to establish for our aggregate results clear statistical significance.

When, however, records for high-scoring and for low-scoring subjects are isolated, it is noted that a somewhat more pronounced W-shaped curve obtains for the former (with maximum critical ratio of the difference equal to 2.47), whereas for the latter a U-shaped curve is obtained (with maximum critical ratio of the difference equal to 1.57) thus suggesting a possible association between the W-shaped curve and above chance scoring success. Examination of records for individual subjects tends somewhat to confirm this position. For only two subjects, however, are the scoring trends sufficiently pronounced to command statistical support. For one of these subjects, Miss E.W., the W-shaped curve fails to obtain, and scoring success appears to vary directly with the distance of the card from the top of the pack. This might seem impressive were it not for the fact that in this single instance the standard DT procedure was replaced by the G.E.S.P. procedure. Again it would appear that a clearly significant relationship between scoring success and card position fails to emerge.

More striking, however, in this connection, is the rather consistent incidence of high scoring on the first and last cards of the pack. This at once suggests that, despite experimental precautions, adequate protection against sensory cues may not have been given these cards. In many instances, the subject gave no attention to the shuffling process or sat with back toward the experimenter, thus rendering inapplicable the hypothesis of sensory cues. In the case of Miss D.W., who exhibits in most marked degree this tendency toward high scoring on cards 1 and 25, it will be recalled that 200 runs were performed under 'distance' conditions. These runs yield 77 hits (with an average of 9.64 per 25) for the first card, and 107 hits (with an average of 13.33 per 25) for the last card of the pack. There are instances, however, in which the hypothesis of sensory cues might apply. For this reason, adjustment has been made for scoring on the first and last cards of the pack, by assigning to these cards a score derived from the average scoring on the adjacent four cards. These corrections are presented parenthetically in Table 1.5. It is noteworthy that this adjustment does not essentially alter the scoring pattern. It does, however, reduce the statistical significance which may be assigned to the observed variations. That the superiority of *direct* over *reverse* scores for high-scoring subjects does not derive from this high scoring on the first and last card positions is at once obvious. It is, moreover, to be noted, (Table 1.5A) that for each of these subjects (with the single exception of Mr. E.A.), *direct* scores exceed chance expectancy for all card positions.

II. GROUP SERIES

The group tests, which were quite exploratory in design, were conducted primarily as class room demonstrations. Two hundred and eighty-four students participated. The procedure for these tests has been described in detail above (p. 164). It will be recalled that it was such as to admit these tests to the screened category. Ten runs were performed by each subject.

Results of Group Tests

In Table 2.1 is presented a summary of the results both *direct* and *reverse* for this group series. The frequency distribution for the entire series appears in Table 2.2. This distribution is given graphic representation in Fig. 2.2. Summaries for daily fluctuations in scoring, for consecutive and for card position frequencies are contained in Tables 2.3A, 2.4, 2.5 and 2.5A. A distribution of total scores attained by individual subjects is contained in Table 2.6 and is presented graphically in Fig. 2.6.

From Table 2.1 it appears that the results here obtained (average score = 4.995, with C.R. = -.12) are in rather perfect accord with chance expectation. That this chance character of the results obtains not only with respect to score, but also with respect to frequency distribution and to consecutive and card position frequencies, is evident from Tables 2.2, 2.4, and 2.5. From Table 2.6 it appears that no subject achieved notable scoring success.

Of interest is the character of the distribution of individual totals in Fig. 2.6. Whether additional trials (or increase in the number of subjects tested) would have accentuated the apparent trimodality of the direct distribution, thus suggesting the presence of individuals from three diverse populations, is a matter for speculation. The authors, having discovered no subjects who consistently maintain scores below the chance level, should anticipate the disappearance of at least one of these apparent 'modes'.

III. MAIN EXPERIMENTAL SERIES

The Main Experimental Series was initiated in the fall of 1937. All tests were performed under screened DT conditions described in the above section on <u>General Procedure for Individual Subjects</u>.

The subjects participating in this series will again be classed as *major* or *minor*, the term *major* being reserved for those who completed the 500 runs set as an arbitrary minimum for this series. The *minor* subjects were those who were unable to allow sufficient time for this arbitrary minimum. The *major* subjects were, with one exception, interested students, most of whom had read Harper's account of the work at Duke University (12), and who had something of a bias for or against the ESP hypothesis. The group tests administered in the classes were singularly unfruitful in supplying subjects who were ready to submit to sustained testing.[11] Tests with each subject in this series were continued as long as his convenience and interest--permitted. In no case was a subject dropped because of low scoring or pressed to go on if he scored high.[12] The number of runs performed during a given work period were determined by the time available and by the subject's tempo. The minimum number of runs for any experimental period was ten.

Some comment with respect to *major* subjects may be of interest. Mr. C.B. was an upper class student with a major in psychology. His attitude was that of the interested experimentalist, curious and non-committal. Mr. H.D., a sophomore with a flair for the unusual, was fascinated with the idea of the experiment, and offered himself as a subject with a rather certain feeling of success. Despite his unimpressive scoring, he retains both interest and 'faith'. Mr. R.S., a sophomore premedic student, with a wide range of interests and of abilities, proposed himself as a subject in order that he might assist in 'blowing the hypothesis to bits'. He tried various guessing techniques which he hoped would result in extra-chance scores, but failed to achieve his objective. It was then suggested by the experimenter that he suspend his rational attack in favor of recording the first symbol that occurred to him. To his intense amazement, he began immediately to achieve extra-chance scores. He persisted with the firm conviction that his 'run of luck' would break. His present attitude is one of puzzled acceptance of extra-chance scoring as a phenomenon which demands further study. Miss D.M. is one of the authors.[13] She submits to the tests only when she is too

11. They did, however, contribute several *minor* subjects.
12. It must be noted, however, that high scoring is conducive to maintenance of interest in the experiment and that this factor of interest often has direct bearing upon the willingness of the subject to submit to sustained experimentation. In this sense, individual experiments tend to be weighted somewhat in favor of the high-scoring subject.
13. Miss F.P.S., the other author, had performed numerous self-administered tests prior to the initiation of our experimental work. No records of these tests were kept. Scoring was somewhat above chance.

weary for other exertion. Mr. C.J., our subject extraordinary, who has continued experimental work throughout his last three years of college, presented himself as a subject, eager to try the tests, and confident that he would be able to score above chance. This confidence seemed to spring from several 'notable coincidences' in his own experience and from his trust in his mother's reliable 'intuitions'. During the first two years of experimentation, he was not only definitely interested in the ESP hypothesis, but seemed to derive genuine satisfaction from the actual testing procedure. During the past year, however, the fun has apparently vanished from the actual process of card guessing, and he has continued primarily because of the sense of responsibility which he holds toward the work. This very natural ennui has been accentuated by the nervous tension and fatigue resulting from excessively heavy academic and activity programs. A slower tempo of performance has resulted. In consequence of this slower tempo and of the subject's limited availability, it has been impossible to achieve the completion of the entire schedule of experiments which had been planned for the year.

Results of Main DT Series

In Table 3.1 is presented a detailed summary of the results (both *direct* and *reverse*) for this series. It will be seen that the entire series of 4180 runs yields a *direct* average of 5.79 hits per 25, with the highly significant critical ratio of 25.19, whereas the *reverse* average is 4.97 hits per 25, with the clearly non-significant critical ratio of -.90.

Examination of individual results for both *major* and *minor* subjects reveals that these subjects present marked variation in scoring success. For the high-scoring *major* subjects, Messrs. C.J., R.S., and C.B., the average scores of 6.89, 6.03, 5.50, and the critical ratios 29.35, 11.25, and 6.52, respectively, are such as to indicate either the occurrence of several extremely unusual chance events, or high probability that the results are extra-chance. On the other hand, scores of 5.17, and 5.06, with critical ratios of 1.84, and .66, for Miss D.M., and for Mr. H.D., respectively, reveal the essentially chance character of these records. Similar variations in scoring success are exhibited by the *minor* subjects, of whom four obtain scoring averages significantly above chance. For readiness of comparison, a summary for high and for low-scoring subjects is presented in Table 3.1. This contrast between subjects becomes even more striking when their records are subjected to the further customary analysis with respect to distribution of hits, frequency of consecutive hits, daily fluctuation in scoring, and frequency of hits with respect to card position.

In Table 3.2 appear the frequency distributions (*direct* and *reverse*) for both *major* and *minor* subjects. Frequency distribution summaries for high and for low scoring subjects are presented in Table 3.2 and are given graphic representation in Fig. 3.2. The marked contrast between the two groups of subjects is again evident, both with respect to the location of the mode and with respect to the frequency of occurrence of scores above the mode.

Daily scoring averages are presented for each subject in Table 3.3A. Of primary interest is the consistency with which above-chance scoring is maintained by the majority of high scoring subjects. Noteworthy, however, is the prolonged decline and eventual recovery exhibited by the scoring of Mr. C.B. With Mr. C.J., on the other hand, it is interesting to note that on October 26th, and on that date only, scoring dropped to the chance level. This drop in scoring level occurred on the Monday following a vigorously celebrated Homecoming week-end at the University.

In Table 3.4 is presented a tabulation giving for each subject the frequencies for isolated and for consecutive hits, together with the predictions on the basis of the binomial hypothesis, for these frequencies. Again it appears that the advantage enjoyed by high-scoring subjects resides primarily in their achievement of consecutive hits of the second and of higher orders.

From Mr. C.J.'s records for this series it became at once apparent that for him the level of scoring throughout the pack was not constant, higher scoring success being achieved for those cards which were situated near the top of the pack. It was this observation which prompted the inclusion in all series of the analysis of scoring success with respect to card position. This analysis for all subjects in this series is presented in Tables 3.5 and 3.5A and is given graphical representation in Fig. 3.5. It may be noted that the aggregate data establish highest scoring for the first five cards in the pack, and lowest scoring for the last group of five, the maximum critical ratio of the difference between scores for successive 5-card groups being 4.34. When, however, the performance of various individual subjects is scrutinized, it is noted that wide variation occurs. (For no subject does the W-shaped curve found in the preliminary series appear.) Since the scoring variations exhibited by Mr. C.J. are so pronounced as to dominate the entire series, it has seemed desirable to present summaries for the high-scoring subjects and for the entire series both with and without the data for him. These summaries are included in Tables 3.5 and 3.5A. From these it appears that for both *major* and *minor* high-scoring groups, exclusive of Mr. C.J., maximum scoring success is achieved in the fourth 5-card position and minimum in the last 5-card position. Statistical treatment, together with considerations of consistency, would, however, indicate that only the latter variation may be regarded as significant. For the low-scoring subjects, variations in scoring are clearly non-significant.

IV. EXPERIMENTAL VARIATIONS WITH MR. C.J.

A. The UT Series

In consequence of the findings in the Main Experimental Series, there appeared a definite need for further intensive study of individual subjects. Extensive experimental variations were, accordingly, undertaken with the subject who furtunately combined high scoring ability with availability---Mr. C.J.

In the attempt to discover whether scoring ability might be subject to voluntary change in direction, and to further investigate the remarkable relationship previously found to obtain for this subject between success in scoring and the position of the card in the pack, a second series of 1000 runs was initiated in January of 1938. This series will be termed the UT series.

The experimental procedure was identical with that reported for the Main Experimental Series with the single exception that the subject was here invited to guess the *up through* rather than the *down through* order of the pack. It was in this series that the 'robot' check (*Matching Series II)* was introduced as an additional control. It becomes obvious that the control which has been called the *reverse series* would here be obtained by comparing the subject's guessed order (intended for the *up through* pack order) with the unintended or *down through* order of the pack.

Results for the UT Series

This series, together with an analysis of the data obtained, has been published in a previous report.

report. (9). For a detailed discussion of these results, the reader is, accordingly, referred to this report.

A summary of results for the series, presented in Table 4.1, reveals a *direct* average of 7.39 with the highly significant critical ratio of 37.01, in marked contrast to *reverse* and *matching* averages of 5.13 and 4.91, respectively, with non-significant critical ratios of 1.95 and -1.38, respectively.

Hit frequencies for this series appear in Table 4.2; daily fluctuations in scoring in Table 4.3A; consecutive hit frequencies in Table 4.4. Card position data are contained in Tables 4.5 and 4.5A and are given graphic representation in Fig. 4.5.

From Table 4.5A it appears that the striking relationship between success in scoring and card position found to obtain for this subject in the DT series, is here maintained despite the reversal in technique. Application of the Spearman Rank Difference Method of correlation here yields $\rho = .94 \pm .02$. The *reverse* and *matching* series, on the other hand, reveal an apparently random distribution of hit frequencies with respect to card position, $\rho = .30 \pm .13$, and $\rho = .07 \pm .14$, respectively. That this observed relationship between scoring success and card position is more than a statistical artifact is a postulate which appears worthy of further examination.

B. The EE Series

Series EE_{RL} and EE_{LR}

In order to further test the apparent relationship between scoring success and card position, additional experimental variations were introduced. Primary among these variations is that of the EE (equal exposure) series which was undertaken in the fall of 1938. In this series, a screen 57" long and 22" high was so placed between the experimenter and the subject as to shield completely both cards and experimenter from the vision of the subject. The experimenter shuffled and cut the pack and then dealt out all 25 cards face down in a single row parallel to the screen. The subject then recorded his guesses for the cards from right to left or from left to right as designated by the experimenter. As previously noted (p. 6) all card recording was done by the experimenter (and witnessed by the subject), the subject's calls being shielded during the recording process. In response to the request of the subject, who chafed at the slower tempo (resulting from inter-run delay) imposed by this experimental procedure, and who was eager to re-establish his previous scoring record, 140 DT runs (performed under the same screened conditions) were interspersed among the first 140 runs of the EE series. These DT runs, by establishing the subject's DT level of scoring for this particular period, offer a needed control on the EE series.

The EE series includes two principal sub-series. The first, EE_{RL}, consists of 400 runs in which the order of the subject's calls was from right to left; the second, EE_{LR}, consists of 140 runs in which the reverse or left to right order obtained. When it became evident that the results of this second series exhibited not only high internal consistency but also close parallelism with those of the RL series, the series was arbitrarily terminated, and the short time available for experimentation was devoted to additional experimental variation.

In Table 5.1 is presented a summary of results for these series (EE_{RL}, EE_{LR}, DT). The distributions for hit frequencies appear in Table 5.2 and Fig. 5.2; day-by-day averages for this period in Table 5.3A; summaries for consecutive hits and for card position data in Tables 5.4, 5.5 and 5.5A.

Series EE_{25-5}

An examination of the results of the two EE series prompted the division of the subject's task into small experimental units. It was, accordingly, decided to introduce EE_{25-5} in which only five cards were dealt from the pack at one time; the subject's guesses for the right-left order of each set of five (within the given pack) were then recorded. Scoring was deferred until guesses for the entire pack had been recorded. Owing to time pressure, only 110 runs were performed in this series.

The summary of results for this series is included in Table 5.1. Frequency distributions, daily fluctuations, consecutive hit and card position frequencies appear in Tables 5.2, 5.3A, 5.4, 5.5 and 5.5A, respectively.

Series EE_5

In order to create a more pronounced break between consecutive sets of five, a second and last variation was introduced in the form of EE_5. The procedure was here identical with that of EE_{25-5} except that here the cards were returned to the pack and the entire pack was again shuffled before the succeeding set of five cards was dealt. Recording and shuffling necessarily occurred at the end of each set of five calls. Since the recording was here subject neither to witness nor to the 55555 check, each recorded set of five was immediately subjected by the experimenter to a second check against the target cards. For convenience of reference, a set of five consecutive runs is treated as a run of 25. Sixty such runs were performed. Results and customary analyses are included in Tables 5.1, 5.2, 5.3A, 5.4, 5.5 and 5.5A.

Discussion of Results for the EE Series

The Summary Table 5.1 reveals that for these series, the subject's scoring level (average 6.16 and C.R. of 16.58) is, in all but the short EE_5 series, consistently below previous levels attained by him. That this lowering of the scoring level can scarcely be attributed to the shift to the EE technique seems apparent from inspection of the control DT series. In fact, the average score (6.23, with C.R. = 16.09) for EE exceeds that for DT (average 5.79 with C.R. = 4.60). While statistical treatment (C.R. diff. = 2.34) suggests that this difference may be significant, the authors feel that comparison between the two series should be confined to that data accumulated over the same experimental periods. When applied to the DT series and to those runs (also 140) of the EE series which were performed during the corresponding experimental periods, statistical treatment yields a critical ratio of 3.33 for the difference between the mean scores of 5.79 for the former and 6.60 for the latter. That the EE technique may be more conducive to extra-chance scoring than is the DT technique is indicated by these results. Similar observations to the effect that lower scoring seems associated with the DT technique have been recorded by others (1, 11). Again a possible relationship between scoring success and a physical factor (card exposure) is suggested. This postulated superiority of the EE technique as a basis for scoring success gains support if attention is directed to the scoring level maintained in the EE_{RL} series for the first five cards of the pack. During 110 consecutive runs of this series, there was obtained for these five cards an average score of 11.00 hits per 25---a

14. The series was terminated by the subject's departure from school. He returned only for the commencement exercises.

scoring level never maintained during the subject's periods of optimum scoring for a comparable number of consecutive runs of a DT or UT series. Even if the entire EE series (which includes 100 runs of approximately chance scoring in January of 1939) be included, the scoring average of 8.39 established for the first five cards of the pack is comparable to the scoring averages obtained for these cards at the peak of the subject's scoring success.

Worthy of note, moreover, is the average score of 13.17 maintained throughout the 60 runs of EE_5 for the last 5 cards in each set of 25. This average has only once (and then for 20 runs) been equaled by this subject for *any* 5-card group and has never been approached for the last 5 card positions in a DT or UT series (the maximum score previously recorded for these positions being one of 9.33 maintained for 30 consecutive runs).

Inspection of Table 5.3A reveals for these series marked departure from the consistency in scoring which has previously characterized the performance of this subject. Notable, moreover, is the improvement both in the level and in the consistency of scoring subsequent to March 3, 1939. Whether this improvement represents a return by the subject to his customary 'form', or a response to variation in technique (i.e., to division of the task into smaller units) is a matter for speculation. In the absence of adequate data and controls, the authors incline to the position that both factors were operative.

As would be anticipated in consequence of the reduced average scoring level, the excess of consecutive frequencies over predicted frequencies is, for these series, less marked than for previous ones.

Examination of card position data for EE series (exclusive of EE_5) reveal for these series high scoring for the first five cards of the pack, near chance scoring for the intermediate cards, followed by varying degrees of scoring recovery for the last five cards. For EE_{RL} and for EE_{LR} this 'recovery' fails to command statistical support (the maximum critical ratio of the difference between scores for the last and for the intermediate 5-card positions being .82 and .13, respectively); for EE_{25-5}, on the other hand, it receives strong statistical support (the maximum critical ratio of the difference being 8.95). Only in the arbitrary runs of 25 formed from the EE_5 series, is the approximately constant level of scoring which might have been anticipated for the EE series maintained. The DT series, which here serves as a control, exhibits a slight but statistically non-significant U (or W)-shaped trend line.

C. MINOR VARIATIONS

After the completion of the long DT series and prior to the initiation of the UT series, several minor experimental variations with Mr. C.J. as subject were initiated. In these short series, it was our purpose to so accustom the subject to experimental variations that these would entail neither loss of confidence nor disruption of scoring. Eventually it was hoped that variations unknown to the subject (e.g., blank cards) might be introduced.

<u>DT-Ten Packs</u> (DT_T)

The first variation introduced was that of using ten different packs for each set of ten runs. This variation serves the twofold purpose: 1. of providing a situation free from the inadequate shuffling criticism (in which it has been maintained that unexpected scores might be achieved by the subject through the conscious or unconscious recall of sequences which might persist in consecutive shuffles), and 2. of providing a situation which requires, on the part of the subject, the ability to focus his scoring efforts upon a particular pack. The series thus obtained will be termed DT_T.

Ten ESP packs were shuffled and placed face down in a row at right angles to the screen. Pack 1 was removed from its position, cut, and placed directly behind the center of the screen. The subject was then asked to record his guesses for the DT order of this particular pack. The pack was checked as usual; the subject recorded; the experimenter turned and called the cards and noted each recorded symbol and each hit as it was recorded. Pack 1 was then returned to its initial position, and pack 2 was moved to the central position. This practice was continued until all ten packs had been called. If time for more than ten runs was available, the subject left the room; the set-up was again prepared; and on the subject's return, the procedure was repeated. Since the ten target packs were still in position, a second check on the recorded card sequences was made at the end of each set of ten runs. Subsequent to the experiment, calls were checked against the nine pack sequences toward which they were not directed.

Results for TD_T

A summary of the results both *direct* and *reverse* for this series and for the control series derived from the nine pack check is presented in Table 6.1. Hit distributions for these series appear in Table 6.2 and Fig. 6.2. Tables 6.3A, 6.4, 6.5, and 6.5A present summaries for daily fluctuations, for consecutive and for card position frequencies. A more detailed analysis of the subject's success in focusing his scoring effort is offered in Table 6.6, which exhibits the number of hits obtained by matching of each call sequence against all packs on the table. In this Table it will be noted that the diagonal cells represent the 'voluntary' hit frequencies; all other cells contain the 'involuntary' or chance hit frequencies derived from the non-target packs. It is to be noted that the totals exclude the diagonal cell entries, and thus represent only the involuntary hit totals. Column totals exhibit all involuntary hits for call sequences 1, 2, 3, etc.

The results clearly indicate that high scoring (average 8.17, with C.R. 16.31) results from the matching of the subject's calls against those packs for which the calls were intended, whereas no extra-chance averages derive from the matching of the calls against those packs for which they were not intended. It would thus appear, contrary perhaps to expectation, that the scoring effort of this subject admits of a high degree of voluntary direction.

Worthy of note again in this series is the consistency in scoring, the excess in consecutive hit frequencies, and the observed direct relation between scoring success and position of the card in the pack. (ρ = .82 $\pm$ .02).

DT-Delayed Knowledge of Score (DT_D)

A second variation was introduced in order to test the motivating strength of the subject's knowledge of his score. It was felt by the subject that knowledge of the score on a given run exerted a positive motivating influence on the succeeding trial, that is to say, if the score were low, he would attempt to better it; if it were high, he would try to match it. Accordingly, a short series in which the subject's knowledge of his score was deferred until after the completion of ten runs was introduced. The procedure was the customary DT procedure with the following variation in recording. On the completion of the subject's calls, a record of the pack sequence was made by the experimenter on a separate record sheet. The subject's calls were then read to the experimenter who recorded them and checked the hits. Since the experimenter's recording of the pack sequence was here without witness, a re-check of the recorded order against the target pack was made by the experimenter before proceding to the next run. The subject's independent record of his own calls was available for a subsequent call check.

Results of DT_D

A summary of results together with the customary analyses with respect to frequency distribution, scoring fluctuation, consecutive and card position frequencies appears in Tables 6.1, 6.2, 6.3A, 6.4, 6.5 and 6.5A.

The results of this series, average 6.50 and C.R. 8.09, indicate that despite deferred knowledge of score the subject was able to maintain a scoring average well above the chance level. Comparison of the average for this series with that of the immediately preceding DT_T series would seem to suggest that deferred knowledge is associated with depression of the scoring level. Examination of the records reveals, however, exaggerated daily fluctuations in scoring (average 4.80 for the first experimental period of 30 runs, average 7.80 for the following period of 50 runs, and average 6.17 for the last period of 40 runs) rather than a consistently depressed scoring level. Since, however, the subject's reaction to this first major variation in experimental technique may represent an uncontrolled factor in the first thirty runs of the series, and since the long holiday intermission[15] (Dec. 1 to Jan. 5) may have a bearing upon scoring for the last experimental period, it is difficult to assign any clear significance to the observed depression in scoring.

More striking, however, is the shift in scoring pattern exhibited by this series, the linear trend of the previous series (DT, DT_T) being here replaced by a U (or W)-shaped trend line. While the elevation in scoring for the last 5-card position which produces this pattern shift does command a certain statistical support (the maximum critical ratio of the difference between the last and the intermediate 5-card groups being 2.69), this support is challenged by the *reverse* series which yields a corresponding maximum critical ratio of 2.94. The need for extreme caution in the interpretation of results

15. Depressed scoring immediately following long vacation periods seems rather typical for this subject.

based upon so small a number of runs is obviously indicated. Despite the ambiguity of these findings, it is however to be recalled that similar U (or W) curves appeared in the Preliminary (unscreened) series and in the 1938-1939 DT series, and have been found by others (1, 10, 11) to be associated with the DT technique.

DT-Ten Packs--Delayed Knowledge (DT_{TD})

When it became apparent that high scoring could be maintained despite delayed knowledge of score, the DT_D series was quickly terminated in order that a combination of the ten pack (DT_T) and the delayed knowledge (DT_D) techniques might be effected.

In this series, DT_{TD}, as in the DT_T series, the ten packs were shuffled and placed face down in a row at right angles to the screen. Pack 1 was removed from its position in the row, cut, and placed in the center position behind the screen. The subject then recorded his calls for the DT order of this pack. Upon receiving the subject's signal that he had completed the recording of his calls, the experimenter recorded the pack sequence on a separate record sheet. The subject then read his calls to the experimenter who recorded them and checked the hits. Pack 1 was then returned to its initial position and pack 2 placed in the center position and the experiment continued. As in the previous tests with ten packs, a second check of the recorded pack orders against the ten target packs was made by the experimenter at the conclusion of each set of ten runs. As before, the subject's calls for each pack were available for a subsequent call check and for checking against the nine non-target packs.

Results of DT_{TD}

Results for this series are presented in Tables 6.1, 6.2, 6.3A, 6.4, 6.5 and 6.5A.

That high scoring (average 7.30 with C.R. 9.43) should be maintained under this ten-pack-delayed-knowledge technique was a source of genuine satisfaction to the subject. Of psychological interest is the fact that knowledge of results plays a relatively unimportant role in the scoring success of this subject.

Card position frequencies again exhibit (though in less marked degree) the usual pattern of declining hit frequencies with position of the card in the pack.

Scoring on the non-target packs continues, as in DT_T, to be that of chance expectancy (average 4.99 with C.R. -.02). Here again is evidence that the subject's scoring efforts can be focused with remarkable success on a particular target pack.

UT Variations: UT_D, UT_{50}, and UT_{75}.

Subsequent to the completion of the UT series and during the few remaining days of the Spring quarter of 1938, three short series were undertaken in order to further test the relationship between scoring success and card position.

Since the DT_D series yielded data which failed to support the previously observed relationship between card position and scoring success, it seemed desirable that deferred knowledge again be introduced as an experimental factor. This was done in a short UT series--UT_D--of 90 runs. The procedure for this series was identical with that for the DT_D series except for the reversal in the order of calling the pack.

Results of UT_D

The results of this series are presented in Tables 6.1, 6.2, 6.3A, 6.4, 6.5 and 6.5A. It is to be noted: 1. that despite the excellent scoring here achieved, the average 6.74 is below that for the immediately preceding 100 runs of the UT series (average 7.09); 2. that there again occurs that shift in scoring pattern which was observed in DT_D, and which is characterized by depression in scoring for the central cards of the pack and elevation for the bottom ones. While this pattern variation commands only ambiguous statistical support (maximum critical ratio of the difference between scores for the last and for intermediate 5-card positions being 2.89 both for *direct* and for *reverse* series), its sudden reappearance with the re-introduction of the deferred knowledge factor cannot be entirely ignored.

As a further variation, it was determined progressively to increase the size of the pack in order to ascertain whether the apparently typical direct relationship between hit frequency and card position would be maintained, and if so, whether a depth of pack could be obtained for which scoring on the bottom cards would be approximately reduced to a chance level. The target packs were accordingly enlarged to 50 and 75 card size by the pooling of two or of three standard packs. The series resulting from the use of these packs will be designated as UT_{50} and UT_{75}, respectively. The former comprises 15 runs (of 50 calls), the latter only three runs (of 75 calls). These series were to have been continued in the fall of 1938. In view, however, of the very limited time which the subject could make available to us, variations which gave promise of more decisive results were at that time introduced. These series, accordingly, remain unfinished.

Results for these series (UT_{50}, UT_{75}) are presented in Tables 6.1, 6.2, 6.3A, 6.4, and 6.5. For purposes of the distribution table, scores are given in terms of 25-card runs.

While inadequacy of data renders conclusions unwarranted, it is of interest to note that a scoring trend not inconsistent with that established for the DT and UT series is here exhibited, highest scoring success being again achieved for the top cards of the pack and lowest for the bottom cards.

SUMMARY OF RESULTS

A summary of scores for all experiments performed in our laboratory is presented in Table 7.1. The frequency distribution for all data is presented in Table 7.2 and is given graphic representation in Fig. 7.2. The customary analyses with respect to daily fluctuations, consecutive and card position frequencies are contained in Tables 7.3A, 7.4, 7.5 and 7.5A. It will be noted that in 12470 runs performed by 332 subjects over a period of three years an average score of 5.83 per 25 with the highly significant critical ratio of 45.70 has been attained.

The frequency distribution for the aggregate data displays a single mode at 5 thus failing to reveal the diversity in scoring exhibited by the groups (high and low-scoring) of which it is composed. The form of the distribution does, however, diverge sufficiently from the theoretical binomial distribution ($\chi^2 = 3992.12$, $n = 13$) to establish high probability ($P = 10^{-19}$) that the distribution was not generated by the operation of chance alone. The *reverse* distribution, on the other hand, yields $\chi^2 = 8.92$, $n = 13$, $p = .79$ thus exhibiting an excellent fit to the theoretical chance distribution.

From Table 7.3A it appears that daily scoring fluctuates in somewhat random fashion. It is noteworthy, however, that in the *direct* series, daily averages drop below the chance level only thirty-five times while in the *reverse* series oscillations above and below the chance line occur with almost equal frequency, there being 104 above, 108 below.

Frequencies for consecutive hits again emphasize the deviation from chance expectancy exhibited by the *direct* data as contrasted with the close agreement with chance expectancy which characterizes the *reverse* data.

Since the relationship between card position and hit frequency exhibited by Mr. C.J. is so consistent and so pronounced as to mask the less marked forms of variation displayed by other subjects, the summary data for the individual tests have in this analysis been separated into several categories representing the performance: 1) of all low-scoring subjects; 2) of all high-scoring subjects exclusive of Mr. C.J.; 3) of Mr. C.J. exclusive of the EE series; 4) of Mr. C.J. in the EE series. Inspection of the various categories fails, however, to reveal a significant common tendency. The need for the further intensive study of individual high-scoring subjects is apparent.

DISCUSSION

Among the results of these studies, several seem to merit special emphasis. Primary, perhaps, among these is the marked contrast exhibited under all analyses between the aggregate data for the *direct* and for the *reverse* series. It will be recalled that these series are derived by subjecting identical experimental data to treatments which differ only in that one respects, the other violates the subject's intention that his calls shall be matched against a given order of the pack. If chance alone be credited with producing scoring success, indifference to order of operation must necessarily be postulated. Actually, however, the *direct* series yields for the 12470 runs (311,750 trials) a highly significant average score (5.83) whereas the *reverse* series yields a statistically non-significant average (4.98). It is at once obvious that the only experimental factor which could be associated with this difference is the subject's intention that the cards be called in a given order.

A second point of interest is offered by the contrast maintained under all analyses between the performance of individual subjects. It is at once evident that the discrepancy between *direct* and *reverse* series derives largely from the scoring of those subjects who have been termed 'high-scoring'. Of 332 subjects tested, 16 achieved performances which admit them to such classification. That the high scoring of these subjects may have resulted from the chance operation of the form of selection used by Leuba (6) in his generation of good 'subjects' is conceivable but difficult of acceptance. The presence of 316 subjects with essentially chance performance[16] suggests that scoring success may not be regarded primarily as a function of the experimenter or of the experimental technique.

Among the performances of the high-scoring subjects, that of Mr. C.J., our subject extra-ordinary, who, throughout 3659 runs covering a period of three years, maintained a scoring average of 6.85 with a critical ratio of 54.83, would seem to merit special comment. Characteristic aspects of the performance of this subject include:

1. marked consistency in daily scoring, with some evidence in support of the postulate that scoring success may be correlated with physical fitness.

2. definite improvement in scoring level over the initial two-year period with sharp decline followed by partial recovery for the third year of the experiment.

3. success in subjecting scoring effort to arbitrary direction as evidenced in successful shifts from the DT to the UT and to the EE technique and in the successful focusing of scoring effort upon a designated pack.

4. success in achievement of consecutive hits. In fact, the scoring advantage of this subject would seem to derive almost entirely from this excess of frequency in consecutive hits of the second and of higher orders.

5. pronounced variation in hit frequency with respect to card position. This variation assumes two distinct forms. The first, which approximates a linear relationship, occurs in the main DT and UT series, and in the special variations included under DT_T and DT_{TD}. The second, represented by a U or W-shaped curve,

---

16. The emergence in the aggregate data of an average *direct* score of 5.13 with a C.R. of 3.61 for all low-scoring individual subjects represents the generation of an apparently significant result from the pooling of records which were individually non-significant. Whether this final result represents a statistical artifact, the accumulation of small but persistent recording error, or the existence of whmewhat wide-spread but marginal ability to score above chance, is a matter for speculation. It should be noted that much of the scoring here involved derives from the unscreened Preliminary Series.

appears in the preliminary (unscreened) DT series, throughout the EE series, in the 1938-1939 DT control series, and under the deferred knowledge technique in both DT_D and UT_D. The appearance of a new curve of operation in the EE series is not disturbing. Rather it offers added support to the postulate (derived from the persistence in the UT series of the direct relationship established for the main DT series) that the basic factor here involved is physical. To reconcile the findings for the various DT and UT series is, however, difficult. The fact that the U-curve appears when scoring is relatively low whereas the linear relationship obtains when scoring is optimum invites speculation. Pattern instability associated with inadequacy of data, or the existence of several factors operating with varying effectiveness might be postulated. The dominance of the U-pattern throughout the preliminary and the EE series, and of the linear pattern throughout the main DT and UT series would suggest limited applicability for the first of these hypotheses. The multiple factor hypothesis seems, however, to merit further consideration. In simplest form, it might involve the postulate that some primary and relatively unstable psychological factor plays a dominant role when scoring is low, and that a secondary and physical factor, associated with card position, becomes apparent when this primary factor achieves increased stability (thus providing a certain constancy or uniformity of background[17]). This postulated dominance and instability of the primary factor might be invoked in the endeavor to account for the diverse performance curves of the various high-scoring subjects. Essentially, however, the fascinating problem of the relation between scoring success and card position still stands.

A final point of interest with reference to this topic is presented by the secondary rhythmic fluctuations which appear throughout the curves of operation of Mr. C.J. and of several high-scoring subjects. It is interesting to recall in this connection the findings of Guilford (5) with regard to liminal stimuli, and to speculate upon the possible bearing of these findings upon scoring pattern.

17. An electric circuit might offer a crude analogy. With sporadic (or violently fluctuating) electro-motive force, an intermittent or fluctuating current would result. The effect of the introduction into the circuit of variable resistence would be partially masked. With relatively steady electro-motive force, on the other hand, the relation between current and resistence is at once evident.

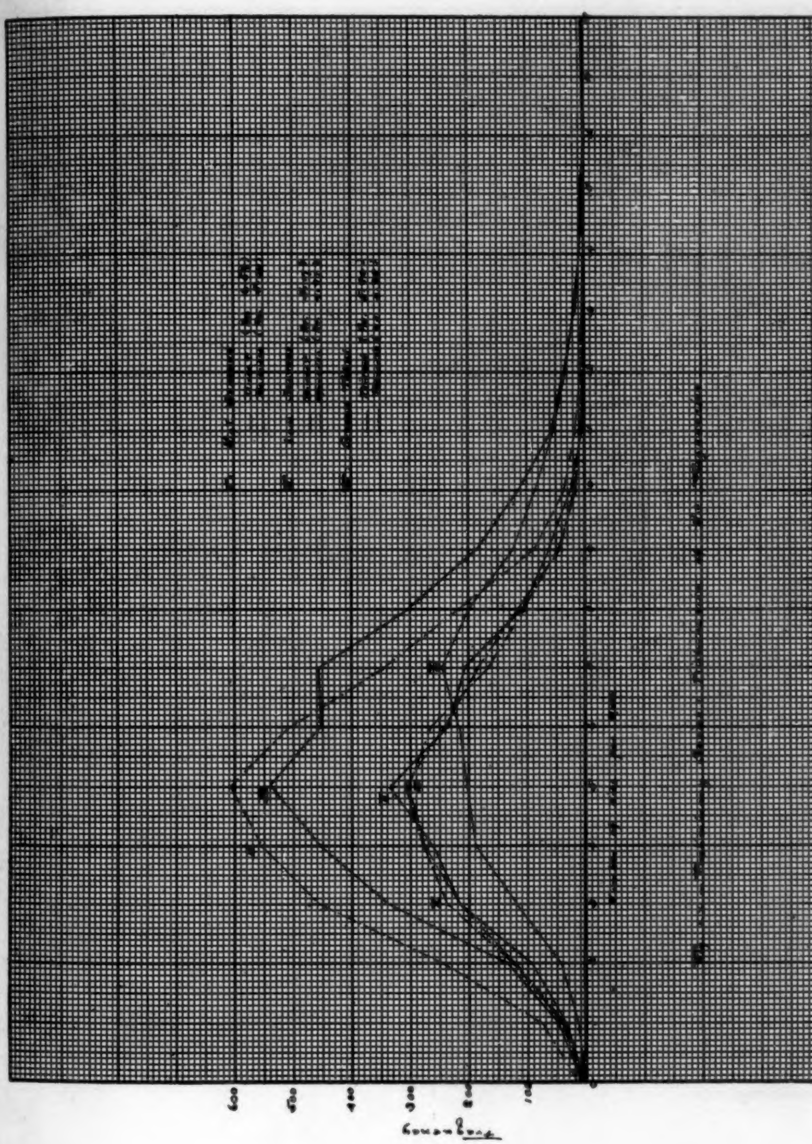
CONCLUSION

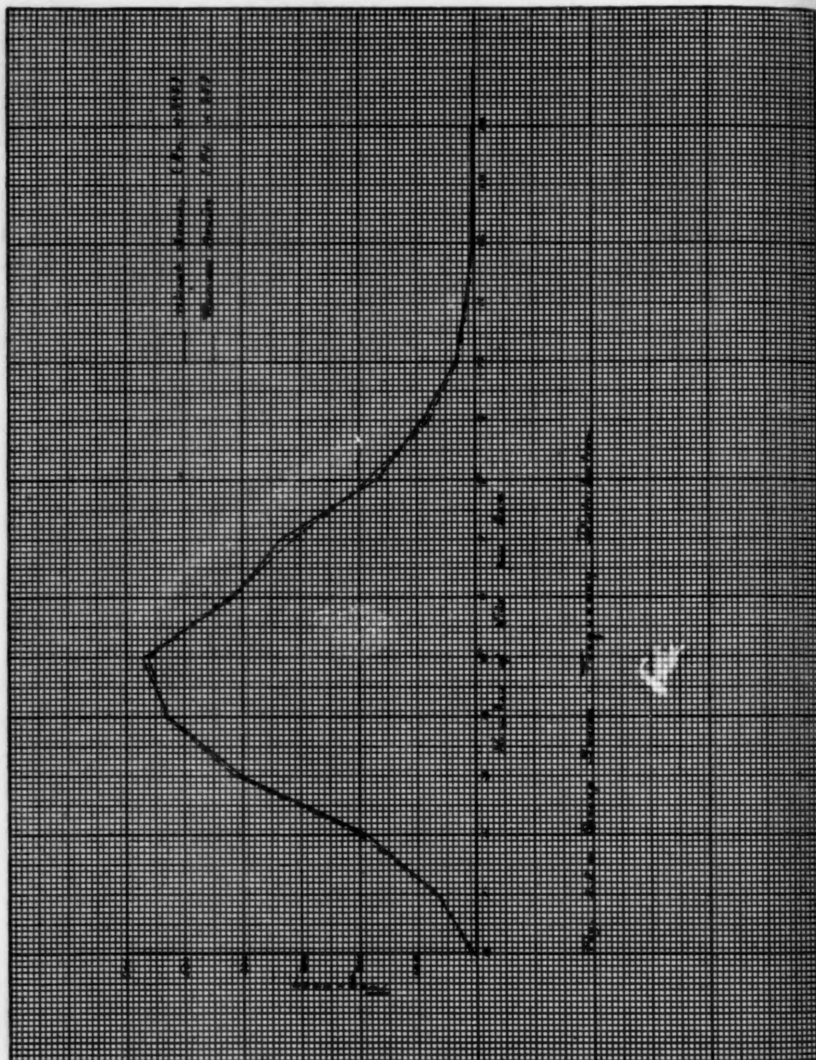
Throughout this study it has been the purpose of the authors to present and to evaluate their results against a chance hypothesis (as exemplified theoretically by the binomial predictions and empirically by the *reverse* series). Under all analyses there has been established for the *direct* series a unique statistical position with respect both to chance predictions and to the *reverse* series. With sensory cues excluded as a scoring factor for all screened data, three hypotheses with respect to the obtained results are suggested: 1. the results represent the chance occurrence of the highly improbable; 2. errors of recording imposed upon the operation of chance have produced apparently extra-chance results; 3. certain individual subjects possess an ability--admittedly unstable and imperfect--to perceive without known sensory means. That chance, operating in Boulder, Colorado during the years,1937, 1938, and 1939, should be fairly represented by these results, or that errors of recording should exhibit the frequency and the subject-preference necessary to produce these results are hypotheses which the individual reader may choose to adopt. To the authors, both positions seem untenable. The hypothesis that extra-sensory perception occurs seems inescapable.

BIBLIOGRAPHY

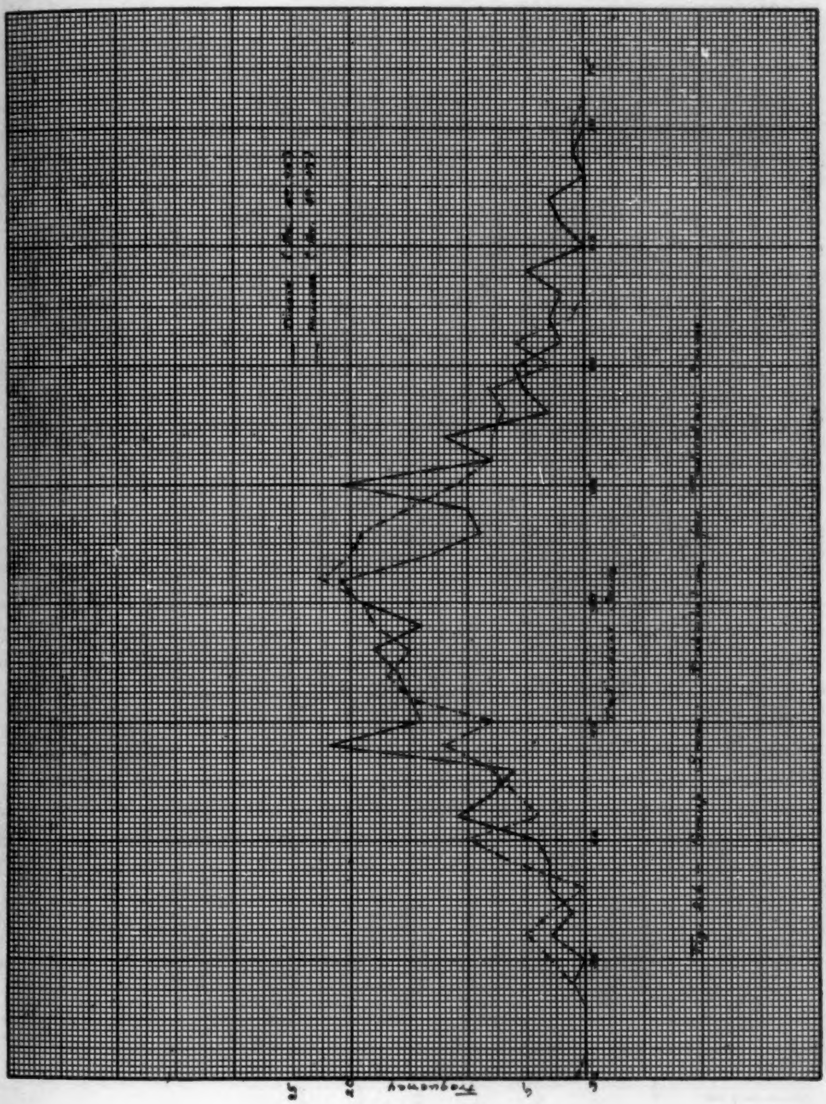
1. Gibson, E.P. A study of comparative performance in several ESP procedures. *J. Parapsychol.* 1937, 1, 264-275.
2. Greenwood, J.A. Analysis of a large chance series of ESP data. *J. Parapsychol.* 1938, 2, 138-146.
3. Greenwood, J.A. Variance of the ESP call series. *J. Parapsychol.* 1938, 2, 60-64.
4. Greville, T.N.E. Exact probabilities for the matching hypothesis. *J. Parapsychol.* 1938, 2, 55-59.
5. Guilford, J. P. 'Fluctuations of Attention' with weak visual stimuli. *The Amer. J. of Psychol.* 1927, 38, 534-583.
6. Leuba, C. An experiment to test the role of chance in ESP research. *J. Parapsychol.* 1938, 2, 217-221.
7. Martin, D.R. Chance and extra-chance results in card matching. *J. Parapsychol.* 1937, 1, 185-190.
8. Martin, D.R. and Stribic, F.P. Studies in extra-sensory perception: I. An analysis of 25,000 trials. *J. Parapsychol.* 1938, 2, 23-30.
9. Martin, D.R. and Stribic, F.P. Studies in extra-sensory perception: II. An analysis of a second series of 25,000 trials. *J. Parapsychol.* 1938, 2, 287-295.

10. Pegram, M.H. Some psychological relations of extra-sensory perception. *J. Parapsychol.* 1937, 1, 191-205.
11. Rhine, J.B. Extra-sensory perception. Boston-Bruce-Humphries, Inc. 1935.
12. Sharp, V. and Clark, CC. Group tests for extra-sensory perception. *J. Parapsychol.* 1937, 1, 123-142.
13. Wright, E.H. The case for telepathy. Harpers Magazine, Nov. 1936.
14. Wright, E.H. The nature of telepathy. Harpers Magazine, Dec. 1936.





Colorado Book Store



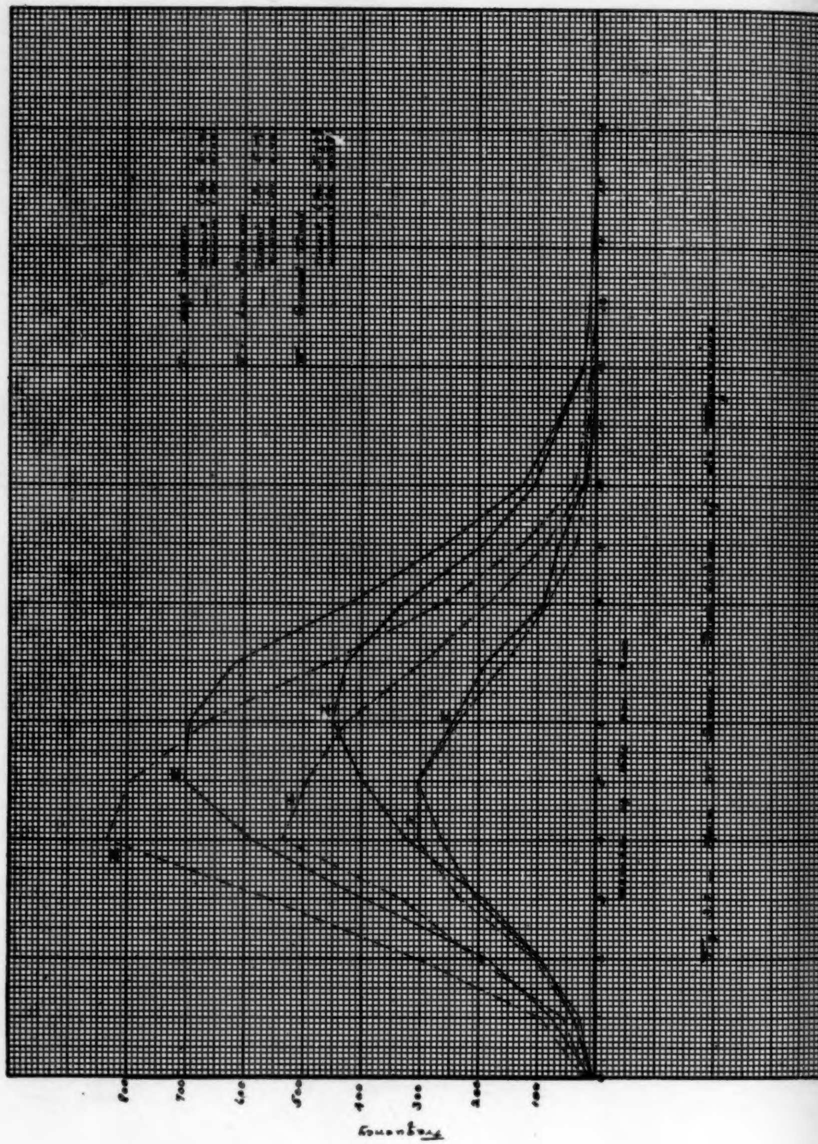
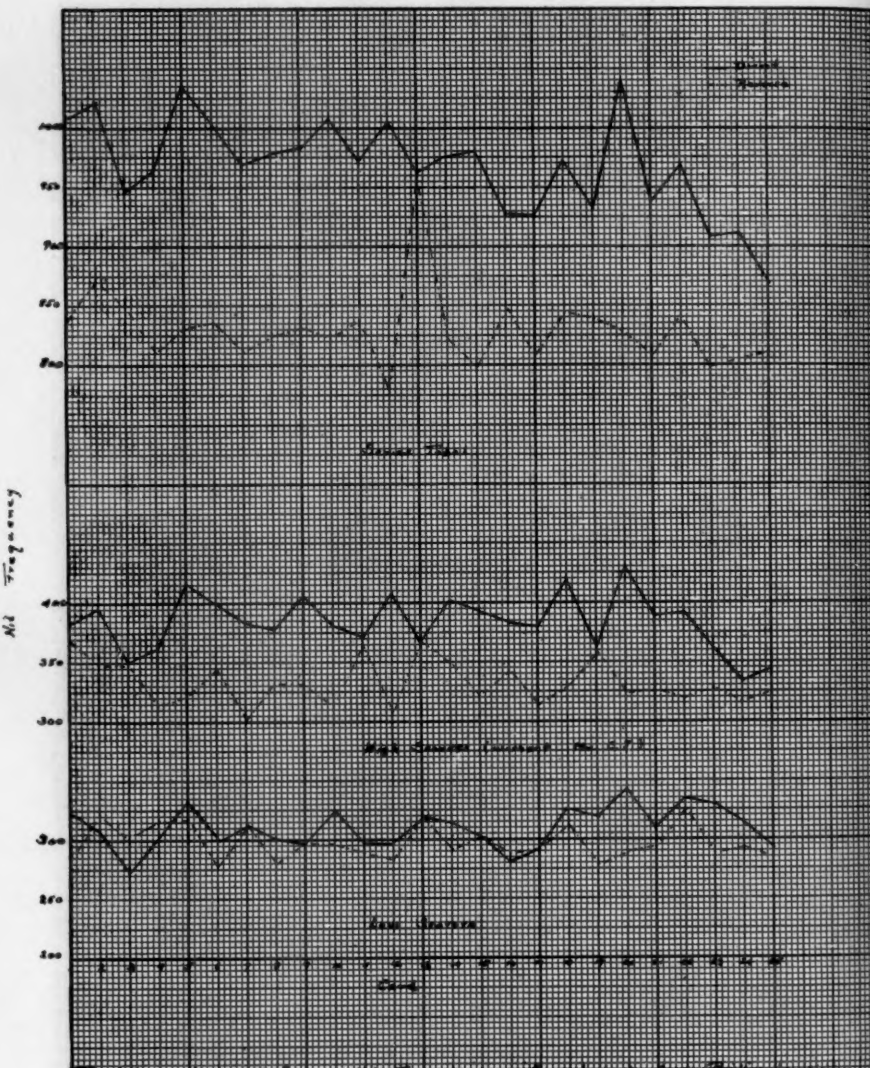
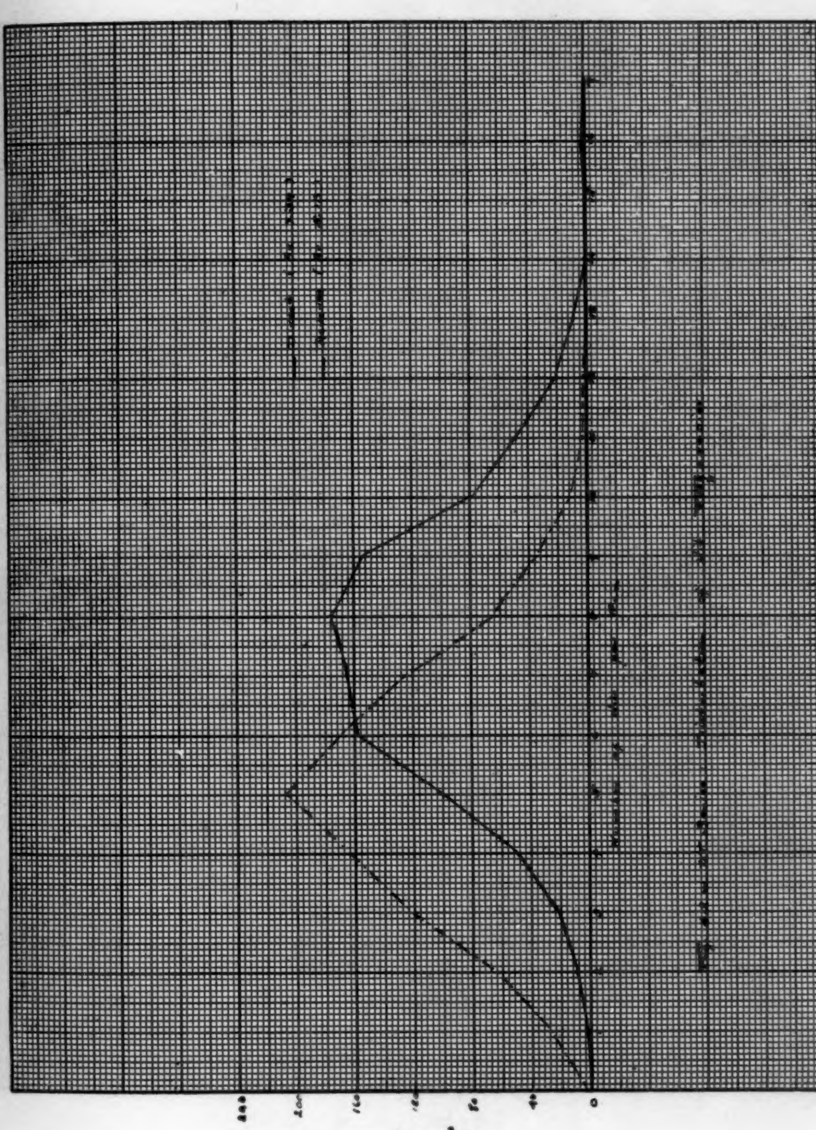
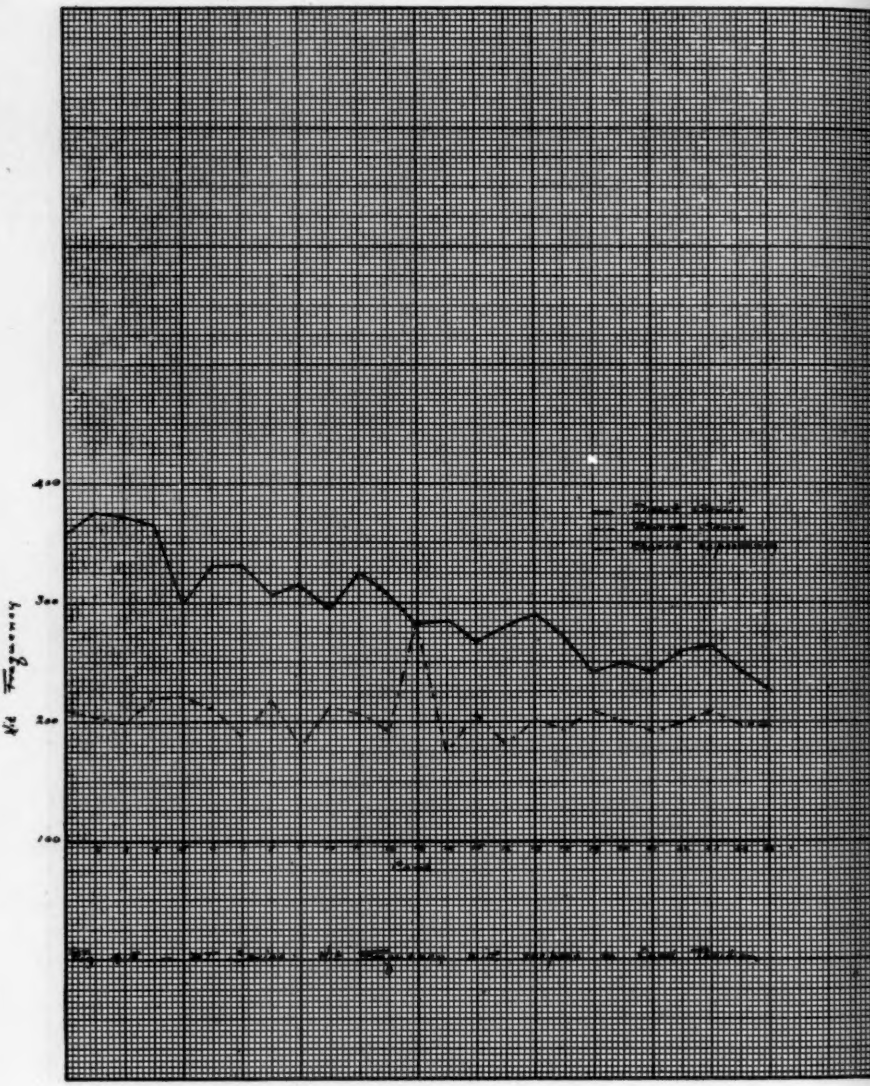


Fig. 3.5 - Main DT Series: Hit Frequency with respect to Card Position







400
300
200
100

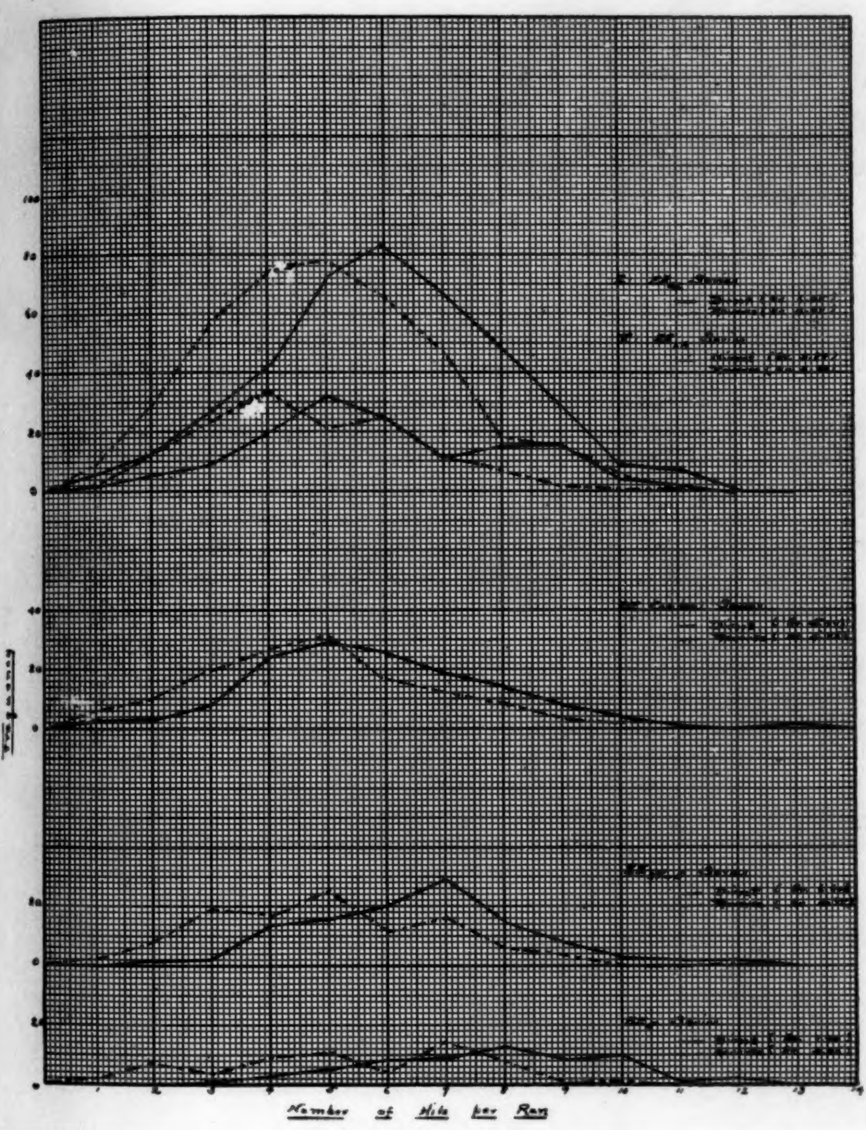
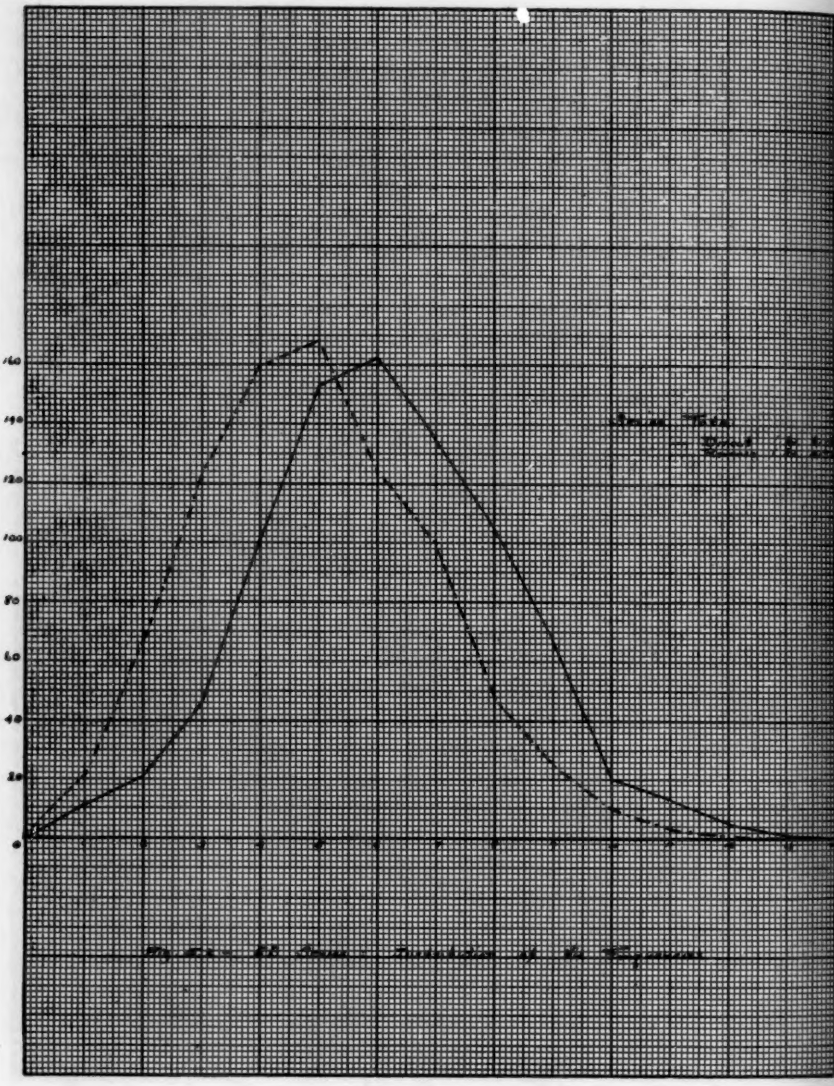


Fig. 5.2 — EE Series: Distribution of Hit Frequencies


160
140
120
100
80
60
40
20
0

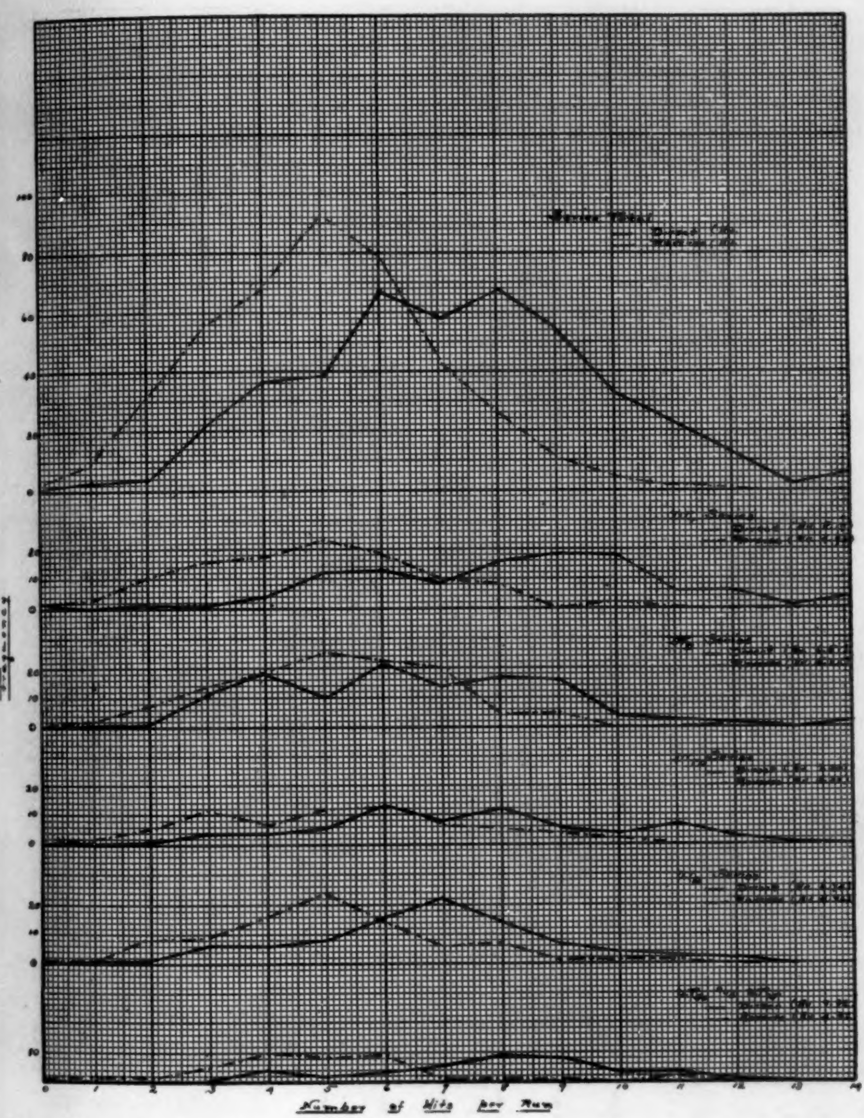


Fig. 6.2 — Minor Variations: Distribution of Hit Frequencies

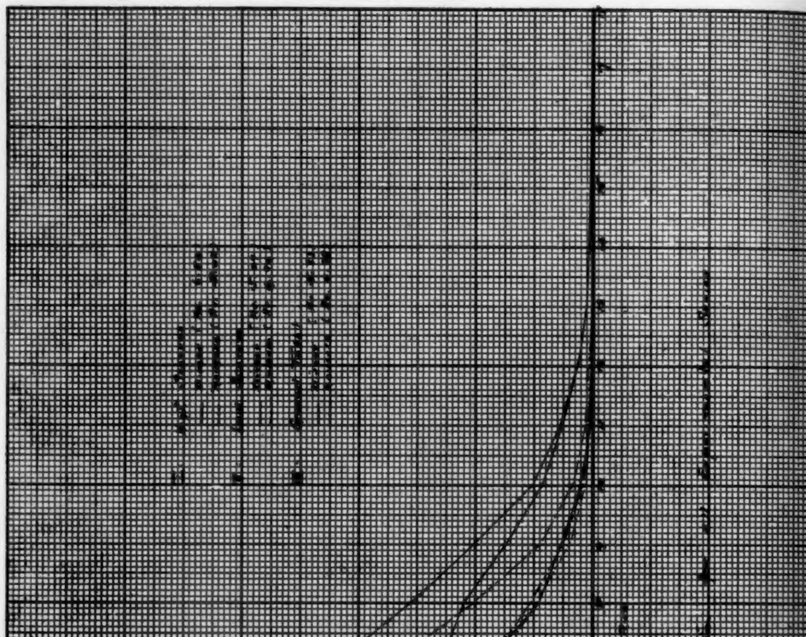

2400
2000
1600
1200
800
400
0

Fig. 7.5 – Hit Frequency with respect to Card Position for All Experimental Series

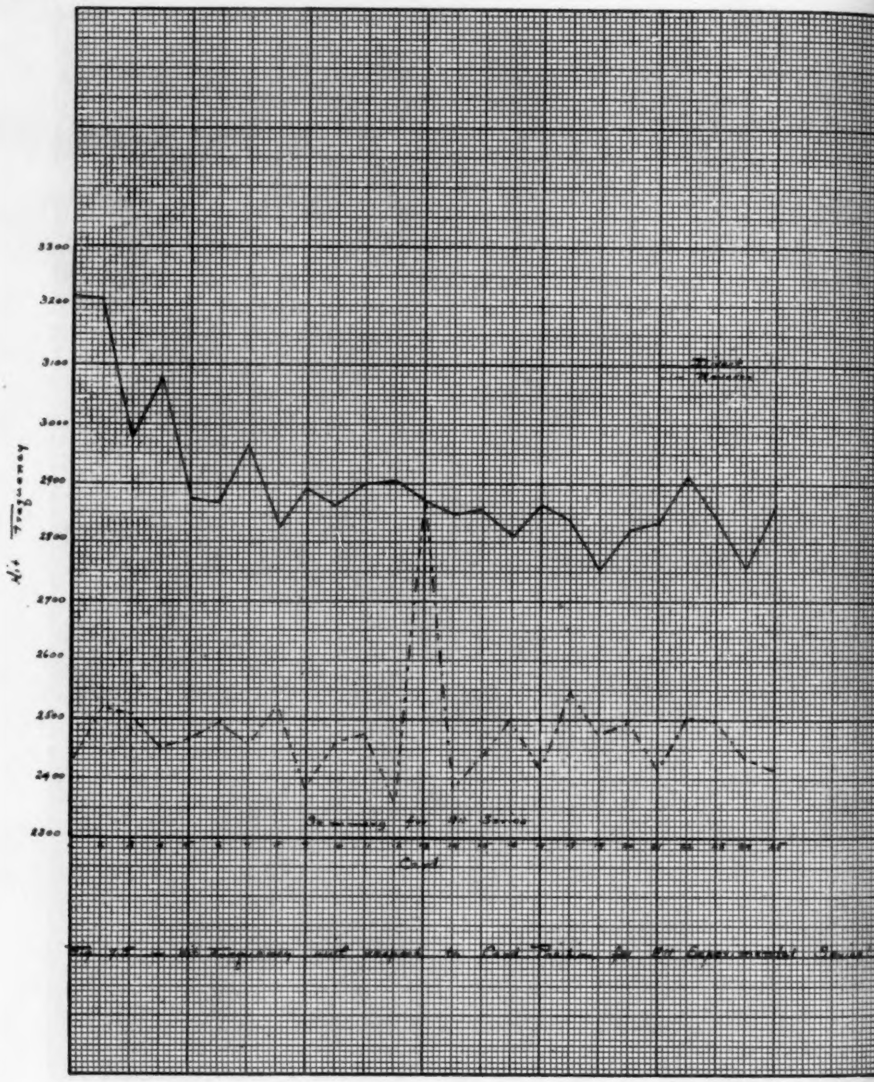

3300
3200
3100
3000
2900
2800
2700
2600
2500
2400
2300
Hit Frequency
Card

Table 0.1

ERROR FREQUENCIES FOR GROUP TESTS (2840 RUNS)

Type of Error	Total Error Abs.	Total Error Rel.	Positive Error Abs.	Positive Error Rel.	Negative Error Abs.	Negative Error Rel.	Net Effect on Scoring Average
Recording	316	.00445	56	.00079	69	.00097	-.0046
Chain	37	.00052	11	.00015	1	.00001	.0035
Inversion	254	.00358	82	.00116	16	.00022	.0232
Miscellaneous							
Scoring							
Marked hits	24	.00034	24	.00034			.0085
Unmarked hits	177	.00249			177	.00249	-.0623
Total	808	.01138	173	.00244	263	.00369	-.0317

Table 0.1A

SCORES OBTAINED BY VARIOUS EXPERIMENTERS
IN 1000 DT RUNS PERFORMED BY A SINGLE SUBJECT

Experimenters	No. of Runs	Total Score	Av./25
A.B. and J.S.	122	842	6.90
A.B. and K.M.*	6	39	6.50
D.M. and assistants A.B. and J.S.	217	1552	7.15
A.B.	301	2069	6.87
D.M.	124	754	6.08
L.L.**	10	69	6.90
J.S.	210	1508	7.18
L.S.**	10	60	6.00

* Professor Karl F. Muenzinger
** Assistant Instructor in Psychology

Table 0.2

FREQUENCY DISTRIBUTIONS FOR CONTROL SERIES (1000 RUNS)

Empirical Series																
Coincidences	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Matching Series I	8	28	73	117	198	195	156	112	68	29	13	2	0	1	0	0
Matching Series II	0	30	75	144	184	199	173	91	61	31	7	4	1	0	0	0
Theoretical Predictions*																
Matching Hypothesis	4	25	73	137	184	192	160	110	63	31	13	5	1	0	0	0
Binomial Hypothesis	4	24	71	136	187	196	163	111	62	29	12	4	1	0	0	0

*Accurate to nearest integer

Table 1.1

PRELIMINARY SERIES: SUMMARY OF RESULTS

	Direct Series						Reverse Series			
Subjects	No. of Runs	Score	Av./25	Dev. from n.p.	σ	C.R.*	Score	Av./25	Dev. from n.p.	C.R.
Major Subjects										
D.W.	301	2258	7.50	753	35.39	21.28	1507	5.01	2	.06
E.W.	180	1311	7.28	411	27.37	15.02	950	5.27	50	1.82
C.J.	380	2438	6.43	538	39.76	13.53	1889	4.97	-11	- .28
B.C.	190	1270	6.68	320	28.11	11.38	995	5.23	45	1.60
C.B.	340	2071	6.09	371	37.51	9.86	1668	4.90	-32	- .85
E.A.	180	1008	5.60	108	27.37	3.94	852	4.73	-48	-1.75
A.B.	120	637	5.30	37	22.35	1.66	576	4.80	-24	-1.07
J.H.	260	1353	5.20	53	32.90	1.61	1311	5.04	11	.33
D.M.	280	1451	5.18	51	34.13	1.49	1426	5.09	26	.76
E.J.	110	557	5.06	7	21.39	.33	580	5.27	30	1.40
Total for Major Subjects	2341	14354	6.13	2649	97.89	26.84	11754	5.02	+49	.50
Total for Minor Subjects	830	4269	5.10	119	58.77	2.02	4055	4.90	-95	-1.61
Grand Total	3171	18623	5.80	2768	114.87	24.09	15809	4.98	-46	- .40

SUMMARY FOR HIGH-SCORING AND FOR LOW-SCORING SUBJECTS

Subjects	No. of Runs	Score	Av./25	Dev. from n.p.	σ	C.R.	Score	Av./25	Dev. from n.p.	C.R.
High Scorers	1571	10356	6.59	2501	80.85	30.93	7861	5.00	+ 6	.07
Low Scorers	1600	8267	5.17	267	81.60	3.27	7948	4.97	-52	- .64

$$\text{*Critical Ratio} \equiv \frac{\text{deviation from chance expectancy}}{\text{standard deviation for chance}}$$

Table 1.1A

PRELIMINARY SERIES: SUMMARY OF RESULTS FOR INDIVIDUAL MINOR SUBJECTS

Direct Series

Subjects	No. of Runs	Score	Av./25	Dev. from n.p.		C.R. C.R.	Score	Av./25	Dev. from n.p.	C.R. C.
C.A.	40	244	6.1	44	12.90	3.41	223	5.5	23	1.77
C.D.	70	389	5.5	39	17.06	2.25	324	4.6	-26	-1.52
E.R.	10	62	6.2	12	6.45	1.86	47	4.7	- 3	- .43
L.F.	10	59	5.9	9	6.45	1.40	46	4.6	- 4	- .62
B.P.	60	318	5.3	18	15.80	1.14	274	4.6	-26	-1.64
L.DG.	40	215	5.3	15	12.90	1.16	200	5.0	0	0
R.LaF.	23	126	5.4	11	9.78	1.12	110	4.8	- 5	- .51
R.M.	20	110	5.5	10	9.12	1.10	89	4.4	-11	-1.21
B.N.	10	57	5.7	7	6.45	1.09	42	4.2	- 8	-1.24
M.W.	40	213	5.3	13	12.90	1.01	192	4.8	- 8	- .62
V.M.	40	212	5.3	12	12.90	.93	208	5.2	8	.62
J.W.	10	55	5.5	5	6.45	.75	48	4.8	- 2	- .30
C.McS.	56	290	5.1	10	15.26	.66	255	4.5	-25	- .66
M.DeB.	30	157	5.2	7	11.17	.63	143	4.8	- 7	- .62
G.B.	30	157	5.2	7	11.17	.63	164	5.4	14	1.24
L.L.	10	53	5.3	3	6.45	.47	45	4.5	- 5	- .77
R.K.	10	52	5.2	2	6.45	.31	56	5.6	6	.92
L.B.	30	152	5.0	2	11.17	.18	132	4.4	-18	- .63
A.L:	30	151	5.0	1	11.17	.09	159	5.3	9	.80
G.D.	30	149	4.7	- 1	11.17	- .09	140	4.7	-10	- .89
H.P.	25	122	4.8	- 3	10.20	- .29	120	5.0	- 5	- .49
F.G.	20	95	4.7	- 5	9.12	- .55	100	5.0	0	0
K.K.	20	94	4.7	- 6	9.12	- .66	98	4.9	- 2	- .21
A.Be.	10	45	4.5	- 5	6.45	- .77	57	5.7	7	1.10
W.G.	10	43	4.3	- 7	6.45	-1.09	45	4.5	- 5	- .77
W.K.	40	184	4.6	-16	12.90	-1.24	193	4.8	- 7	- .54
R.W.	36	163	4.5	-17	12.24	-1.39	181	5.0	1	.07
S.N.	40	181	4.5	-19	12.90	-1.47	204	5.1	4	.30
J.K.	30	121	4.0	-29	11.17	-2.59	159	5.2	9	.80

Table 1.2

PRELIMINARY SERIES:

Distribution of Hit Frequencies (Direct)

Subject	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
D.W.			4	9	14	37	40	58	47	33	27	11	12	3	3	2	0	1	
E.W.	1	2	1	17	15	22	19	20	26	17	7	14	8	5	4	2			
C.J.	2	2	15	42	46	41	54	59	40	22	23	16	6	10	2				
B.C.			4	10	24	28	21	35	29	17	11	4	5	2					
C.B.	1	4	8	20	57	47	60	53	42	26	16	5	1						
E.A.		4	10	21	33	29	23	24	16	9	4	3	1	2	0	0	0	0	1
A.B.		1	3	18	19	31	16	14	12	5	1								
J.H.		4	15	37	46	48	41	32	22	12	3								
D.M.	2	6	21	32	40	60	50	30	23	8	7	1							
E.J.		5	6	12	18	27	18	14	4	3	2	0	1						
Total Major Subjects	6	28	87	218	312	370	342	339	261	152	101	54	34	22	9	4	0	1	1
Total Minor Subjects	2	20	50	117	141	170	110	119	48	34	13	3	2	0	1	0	0	0	0
Grand Total	8	48	137	335	453	540	452	458	309	186	114	57	36	22	10	4	0	1	1
Binomial Predictions	12	75	225	431	592	622	518	351	198	93	36	13	4	1	0				
High Scorers	4	12	42	119	189	204	217	249	200	124	88	53	33	22	9	4	0	1	1
Low Scorers	4	36	95	216	264	336	235	209	109	62	26	4	3	0	1				

Distribution of Hit Frequencies (Reverse)

Subject	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
D.W..	2	10	19	39	58	58	48	33	16	10	5	3							
E.W.		4	10	24	31	34	34	19	10	6	5	1	0	1	0	1			
C.J.	1	11	29	50	68	74	67	36	26	15	2	1							
B.C.		2	12	24	30	44	31	23	11	10	1	2							
C.B.	1	5	30	56	56	58	62	42	23	2	3	1	1						
E.A.	2	4	21	24	36	34	24	12	17	4	2								
A.B.	2	3	12	20	15	22	20	13	9	4									
J.H.		6	20	30	55	55	38	26	18	7	6	1							
D.M.		6	22	42	49	46	42	34	25	8	4	2							
E.J.		1	10	13	20	15	20	20	4	2	2	1	2						
Total Major Subjects	8	52	185	322	416	440	386	258	159	68	30	12	3	1	0	1			
Total Minor Subjects	5	23	60	134	139	169	121	91	58	21	6	3	0	0	0	0			
Grand Total	13	75	245	456	555	609	507	349	217	89	36	15	3	1	0	1			
Binomial Predictions	12	75	225	431	592	622	518	351	198	93	36	13	4	1	0	0			
High Scorers	6	36	121	217	279	302	266	165	103	47	18	8	1	1	0	1			
Low Scorers	7	39	124	239	276	307	241	184	114	42	18	7	2	0	0	0			

Table 1.3A

PRELIMINARY SERIES: DAILY FLUCTUATIONS IN SCORING

Miss D.W.

Date	Number of Runs	Direct Average	Reverse Average
Feb. 6	10	7.30	5.40
Mar. 6	30	6.83	4.96
Mar. 11	25	6.28	5.84
Mar. 12	36	7.58	5.30
Mar. 14	60	7.58	4.61
Apr. 3	30	7.66	4.86
June 14	80	8.63	5.05
July 27	30	5.76	4.66

Mr. C.J.

Date	Number of Runs	Direct Average	Reverse Average
Feb. 17	20	5.95	4.45
Feb. 18	10	5.00	5.60
Feb. 19	20	5.00	5.05
Feb. 20	20	6.00	5.45
Feb. 23	20	5.25	4.65
Mar. 5	30	5.10	4.86
Apr. 5	30	6.80	4.73
Apr. 16	10	5.30	4.90
Apr. 21	30	5.26	4.80
Apr. 22	10	6.20	5.10
Apr. 27	10	5.70	6.20
Apr. 29	20	7.45	5.15
May 4	30	6.36	5.60
May 5	20	8.15	4.40
May 13	10	7.00	3.80
May 19	10	4.50	5.30
May 20	30	8.83	4.70
May 21	30	8.43	5.13
May 27	20	6.05	5.10

Mr. E.A.

Date	Number of Runs	Direct Average	Reverse Average
Jan. 10	50	5.60	4.44
Feb. 9	20	5.65	4.40
Feb. 16	10	7.40	4.00
Feb. 17	20	5.50	4.70
Feb. 18	40	5.63	4.95
Feb. 21	20	4.60	5.05
Feb. 25	10	4.50	5.60
Mar. 17	10	6.90	5.30

Miss E. W.

Date	Number of Runs	Direct Average	Reverse Average
Apr. 12	30	6.56	5.30
Apr. 29	20	7.45	5.00
Apr. 30	10	4.70	4.20
May 5	20	8.40	5.50
May 10	20	10.40	5.50
May 12	20	7.25	5.10
May 13	10	4.90	5.20
May 17	10	9.60	4.70
May 28	20	6.05	6.25
June 1	20	7.00	5.10

Miss B.C.

Date	Number of Runs	Direct Average	Reverse Average
Apr. 30	60	6.85	5.29
May 4	70	6.84	5.20
May 11	10	5.40	5.70
May 17	20	6.75	5.40
May 20	30	6.36	4.93

Mr. C.B.

Date	Number of Runs	Direct Average	Reverse Average
Apr. 26	10	6.90	4.00
Apr. 28	30	6.00	4.60
Apr. 29	50	6.46	4.86
May 3	40	6.05	4.72
May 5	30	6.53	5.03
May 10	30	7.26	5.00
May 12	60	5.56	5.11
May 17	20	5.70	5.15
May 19	20	4.75	4.08
May 25	50	6.00	5.00

Miss A.B.

Date	Number of Runs	Direct Average	Reverse Average
Apr. 8	20	5.64	4.74
Apr. 16	10	6.20	5.00
Apr. 19	10	5.40	5.20
May 13	10	4.60	4.20
May 19	20	4.80	5.14
May 21	10	4.90	3.60
May 22	10	5.00	5.00
May 22	10	5.00	5.10
June 1	20	5.84	4.84

Table 1.3A (cont'd)

Mr. J.H.			
Date	Number of Runs	Direct Average	Reverse Average
Feb. 17	10	4.50	5.30
Feb. 23	60	5.36	4.98
Mar. 2	40	5.42	5.05
Apr. 13	40	5.47	4.85
Apr. 14	20	5.05	4.90
Apr. 27	20	5.05	5.30
May 4	10	5.90	5.30
May 19	50	4.76	5.32
May 28	10	5.10	4.00
Miss D.M.			
Feb. 4	10	5.10	5.30
Mar. 2	10	4.40	5.20
Mar. 3	30	5.67	5.47
Mar. 28	20	4.90	6.15
Mar. 29	20	5.25	4.45
Apr. 12	10	5.20	4.30
Apr. 16	10	5.60	5.30
Apr. 19	20	4.35	4.95
Apr. 26	30	5.60	5.00
Apr. 27	10	4.20	5.70
May 1	10	5.10	4.80
May 5	10	3.80	5.10
May 9	10	4.50	5.40
July 2	40	5.40	5.05
July 3	20	5.95	5.60
July 6	10	5.40	4.00
Aug. 24	10	5.40	4.90
Miss E.J.			
May 18	50	5.16	5.24
May 19	10	4.80	4.60
May 20	50	5.02	5.48

Minor Subjects			
Date	Number of Runs	Direct Average	Reverse Average
Feb. 4	20	4.75	5.35
Feb. 5	20	4.30	4.85
Feb. 8	10	5.90	4.60
Feb. 9	10	4.30	4.50
Feb. 10	40	5.23	4.65
Feb. 12	9	6.11	5.44
Feb. 17	40	5.55	4.93
Feb. 18	20	5.65	4.50
Feb. 19	46	4.78	4.85
Feb. 21	20	5.10	5.40
Feb. 23	14	5.07	4.36
Mar. 1	10	5.50	4.80
Apr. 6	60	5.13	5.40
Apr. 8	80	5.11	4.65
Apr. 10	25	4.88	4.80
Apr. 12	80	5.50	4.76
Apr. 15	60	5.21	5.15
Apr. 16	40	5.35	4.63
Apr. 19	30	4.90	5.07
Apr. 20	30	4.03	5.30
Apr. 22	40	4.90	4.88
Apr. 27	29	4.93	4.45
Apr. 30	40	5.53	5.00
May 3	20	4.70	4.90
May 6	27	5.44	4.66
May 19	10	5.30	4.50

Table 1.4

PRELIMINARY SERIES:

Frequencies for Isolated and for Consecutive Hits

Subject		1	2	3	4	5	6	7	8	9
D.W.	Direct	1240	308	80	20	15	0	1	0	0
	Predicted	982	189	36	7	1	0	0	0	0
	Reverse	1040	184	26	4	1	0	0	0	0
E.W.	Direct	676	150	71	14	9	1	1	1	0
	Predicted	588	113	22	4	1	0	0	0	0
	Reverse	590	109	31	8	1	2	0	0	0
C.J.	Direct	1288	300	106	25	12	3	3	3	1
	Predicted	1240	238	46	9	2	0	0	0	0
	Reverse	1263	234	46	5	0	0	0	0	0
B.C.	Direct	647	179	60	15	5	0	0	0	0
	Predicted	620	119	23	4	1	0	0	0	0
	Reverse	600	116	43	7	0	1	0	0	0
C.B.	Direct	1110	307	82	19	5	0	0	0	0
	Predicted	1110	213	41	8	2	0	0	0	0
	Reverse	1104	190	49	8	1	0	0	0	0
E.A.	Direct	623	128	25	5	2	4	0	0	0
	Predicted	588	113	22	4	1	0	0	0	0
	Reverse	570	106	18	4	0	0	0	0	0
A.B.	Direct	416	72	23	2	0	0	0	0	0
	Predicted	392	75	14	3	1	0	0	0	0
	Reverse	377	70	17	2	0	0	0	0	0
J.H.	Direct	878	165	36	8	1	0	0	0	0
	Predicted	849	163	32	6	1	0	0	0	0
	Reverse	831	158	39	8	3	0	0	0	0
D.M.	Direct	911	200	36	8	2	0	0	0	0
	Predicted	914	176	34	6	1	0	0	0	0
	Reverse	902	181	35	13	1	0	0	0	0
E.J.	Direct	356	76	9	3	2	0	0	0	0
	Predicted	359	69	13	3	0	0	0	0	0
	Reverse	359	82	15	3	0	0	0	0	0
Total Majors	Direct	8135	1885	528	119	53	8	5	4	1
	Predicted	7641	1468	282	54	10	2	0	0	0
	Reverse	7636	1430	319	62	7	3	0	0	0
Total Minors	Direct	2769	543	103	21	3	1	0	0	0
	Predicted	2709	521	100	19	4	1	0	0	0
	Reverse	2661	492	94	21	5	2	1	0	0
Grand Total	Direct	10904	2428	631	140	56	9	5	4	1
	Predicted	10350	1989	382	73	14	3	1	0	0
	Reverse	10297	1922	413	83	12	5	1	0	0

Table 1.5

PRELIMINARY SERIES:

Average Scores (per 25)

For Successive Five Card Groups

Subject	Direct					Reverse				
	I	II	III	IV	V	I	II	III	IV	V
D.W.	7.34	6.91	7.52	7.61	8.12	5.27	4.97	5.02	4.80	4.98
E.W.	7.89	7.56	7.53	6.97	6.47	4.97	5.44	5.81	5.25	4.92
C.J.	6.34	6.20	6.46	6.30	6.77	4.78	4.87	5.40	5.12	4.70
B.C.	6.53	6.50	6.97	6.47	6.95	5.58	5.32	4.84	5.37	5.08
C.B.	6.04	6.13	5.92	6.04	6.31	4.78	4.97	5.04	4.80	4.94
E.A.	5.17	5.67	5.64	5.39	6.14	4.83	4.61	4.81	4.44	4.98
A.B.	5.17	4.67	5.54	5.75	5.42	4.46	4.63	4.75	4.83	5.33
J.H.	5.35	5.23	4.87	5.62	4.96	4.77	4.85	5.27	5.15	5.17
D.M.	5.00	5.07	5.43	4.98	5.43	5.23	4.50	5.21	5.29	5.23
E.J.	4.77	5.32	4.68	5.27	5.27	5.14	5.91	4.86	5.27	5.18
Total <u>major</u>	6.28	6.22	6.37	6.33	6.54	5.15	5.12	5.33	5.20	5.19
Total <u>minor</u>	5.39	5.20	4.87	5.06	5.20	4.80	4.88	4.80	5.14	4.81
Grand total	6.04 (5.94)	5.94	5.96	5.99	6.18 (6.05)	5.06	5.06	5.19	5.19	5.09

Subject	Direct					Reverse				
	I	II	III	IV	V	I	II	III	IV	V
High Scorers	6.54 (6.35)	6.46	6.64	6.49	6.85 (6.65)	5.00	5.00	5.16	4.96	4.91
Low Scorers	5.26 (5.11)	5.15	5.00	5.20	5.22 (5.17)	4.87	4.86	4.95	5.15	5.01

Table 1.5A

PRELIMINARY SERIES:

Hit Frequency with Respect to Card Position (Direct)

Card	D.W. (60)*	E.W. (56)	C.J. (76)	B.C. (38)	C.B. (68)	E.A. (36)	A.B. (24)	J.H. (52)	D.M. (56)	E.J. (22)	Total Major (488)	Total Minor (166)	Grand Total (654)
1	110	62	102	55	85	45	31	62	58	25	633	197	830
2	66	61	86	58	85	40	29	55	66	20	566	179	745
3	90	44	97	43	77	38	17	54	47	18	525	187	712
4	100	57	108	48	83	28	19	54	65	20	582	167	749
5	76	60	89	46	81	35	28	53	44	22	534	165	699
6	74	52	86	41	81	34	21	52	51	22	514	158	672
7	96	56	78	49	96	47	22	52	65	25	586	190	776
8	86	48	95	52	78	44	16	58	56	23	556	178	734
9	82	59	109	63	70	39	33	54	55	26	588	163	751
10	78	57	103	42	92	40	20	56	59	21	568	174	742
11	100	60	81	49	80	39	22	52	56	21	559	152	711
12	88	60	105	54	86	44	25	48	67	23	600	151	751
13	84	59	108	63	83	44	25	51	65	20	600	165	765
14	87	50	94	45	85	34	38	57	67	19	574	178	752
15	94	42	103	54	71	42	23	45	52	20	546	164	710
16	85	49	93	44	87	35	30	59	52	26	558	180	738
17	100	55	94	60	95	37	28	73	60	30	632	166	798
18	98	52	79	48	73	34	21	61	52	25	543	180	723
19	92	45	104	47	85	45	30	57	55	19	577	157	734
20	85	50	109	47	71	43	29	42	62	16	554	157	711
21	98	52	88	57	100	44	29	52	69	19	608	186	794
22	78	45	101	55	91	45	26	50	60	23	574	177	751
23	92	44	104	50	95	31	22	63	56	28	583	175	758
24	87	41	106	54	72	40	16	53	58	23	550	140	690
25	134	51	116	48	71	61	37	40	61	25	644	185	829

* Chance expectancy for each card position (.2N).

Table 1.5A

PRELIMINARY SERIES:

Hit Frequency with Respect to Card Position (Reverse)

Card	D.W. (60)*	E.W. (36)	C.J. (76)	B.C. (38)	C.B. (68)	E.A. (36)	A.B. (24)	J.H. (52)	D.M. (56)	E.J. (22)	Total Major (468)	Total Minor (166)	Grand Total (634)
1	71	45	65	28	60	35	18	54	55	20	451	157	608
2	65	35	79	36	78	29	24	44	56	30	474	159	633
3	54	34	80	55	63	44	21	42	57	22	470	162	632
4	62	33	69	41	68	28	24	63	63	19	470	157	627
5	67	32	70	54	56	38	20	45	62	22	466	161	627
6	45	38	69	39	64	43	24	53	60	28	463	176	639
7	69	41	72	47	66	21	21	53	48	27	465	164	629
8	61	42	85	47	82	46	22	58	43	27	513	155	668
9	62	36	79	35	56	24	20	42	51	27	432	158	590
10	62	39	65	34	70	32	24	46	50	21	443	157	600
11	41	37	72	26	80	38	20	61	55	18	451	174	622
12	66	40	68	31	57	31	23	47	64	25	452	157	609
13	84	59	108	63	83	44	25	51	63	20	600	163	763
14	53	38	80	24	55	31	16	63	54	25	439	145	584
15	58	35	82	40	68	29	30	52	56	19	469	158	627
16	62	37	74	44	78	36	24	58	65	24	500	176	676
17	47	36	83	42	58	27	16	49	57	26	441	172	613
18	71	46	77	44	54	54	30	53	63	19	491	176	667
19	47	32	83	28	71	32	27	57	51	21	449	164	613
20	62	38	72	46	65	31	19	51	62	26	472	165	637
21	59	33	63	37	66	35	23	61	57	26	460	160	620
22	54	31	77	39	60	38	24	48	60	22	453	152	605
23	57	41	80	42	77	40	31	59	61	21	509	176	676
24	67	32	72	44	53	35	28	53	52	26	462	171	633
25	63	40	65	31	80	31	22	48	62	19	462	140	602

* Chance expectancy for each card position (.2N).

Table 1.5A

PRELIMINARY SERIES:

Card Position Summary for High and for Low-Scoring Subjects:

Card	High Scorers (314)	Low Scorers (320)	High Scorers (314)	Low Scorers (320)
1	457	373	304	304
2	396	349	320	313
3	389	323	328	304
4	424	325	301	326
5	387	312	317	310
6	368	304	298	341
7	422	354	316	313
8	403	331	363	305
9	422	329	292	298
10	412	330	302	298
11	409	302	294	328
12	437	314	293	316
13	441	322	441	322
14	393	359	281	303
15	406	304	312	315
16	391	347	331	345
17	441	357	293	320
18	384	339	326	341
19	418	316	293	320
20	405	306	314	323
21	439	355	293	327
22	415	336	299	306
23	416	342	337	348
24	400	290	303	330
25	481	348	310	292

*Chance expectancy for each card position (.2N)

Table 2.1

GROUP SERIES: SUMMARY OF RESULTS

Direct Series						Reverse Series			
No. of Runs	Score	Av./25	Dev. from n.p.		C.R.	Score	Av./25	Dev. from n.p.	C.R.
2840	14187	4.995	-13	108.71	-.12	14049	4.95	-151	-1.39

Table 2.2

GROUP SERIES: DISTRIBUTION OF HIT FREQUENCIES

Hits per Run	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Direct	15	70	190	404	528	572	415	335	168	90	32	14	3	2	2
Reverse	17	69	200	414	532	562	421	314	186	80	30	12	3		

Table 2.3A

GROUP SERIES: DAILY FLUCTUATIONS IN SCORING

Date	Number of Runs	Direct Average	Reverse Average
Apr. 19	470	5.04	4.87
Apr. 20	590	4.91	5.01
Apr. 22	780	5.03	4.81
Oct. 25	450	4.79	5.08
Oct. 27	550	4.97	4.98

Table 2.4

GROUP SERIES: FREQUENCIES FOR ISOLATED AND FOR CONSECUTIVE HITS

	1	2	3	4	5	6
Direct	9341	1705	331	90	13	3
Predicted	9270	1781	342	65	13	2
Reverse	9195	1750	335	65	13	4

Table 2.5

GROUP SERIES: AVERAGE SCORES (PER 25) FOR SUCCESSIVE FIVE CARD GROUPS

Series	I	II	III	IV	V
Direct	4.97	4.94	5.06	4.94	5.07
Reverse	4.87	4.94	4.92	5.00	5.00

Table 2.5A

GROUP SERIES: CARD POSITION FREQUENCIES

Card	Direct	Reverse
1	583	538
2	584	557
3	525	586
4	602	535
5	527	548
6	551	559
7	608	576
8	544	564
9	558	521
10	547	586
11	595	553
12	551	546
13	578	578
14	566	557
15	586	561
16	584	535
17	567	558
18	548	602
19	568	571
20	537	576
21	543	554
22	593	624
23	548	544
24	581	567
25	613	553

Table 2.6

GROUP SERIES: DISTRIBUTION OF TOTAL INDIVIDUAL SCORES

Individual Scores	Frequencies Direct	Reverse
30	0	1
31	0	0
32	0	0
33	0	0
34	1	1
35	0	3
36	3	5
37	1	3
38	3	0
39	3	6
40	4	10
41	11	4
42	8	6
43	6	8
44	22	12
45	14	8
46	15	15
47	16	17
48	18	15
49	14	18
50	19	19
51	21	25
52	13	20
53	9	19
54	10	15
55	21	11
56	8	8
57	12	8
58	3	7
59	5	8
60	6	3
61	2	6
62	3	1
63	2	0
64	5	0
65	0	0
66	2	0
67	3	0
68	0	0
69	1	1
70	0	1

Table 3.1

MAIN DT SERIES: SUMMARY OF RESULTS

	Direct Series						Reverse Series			
Subjects	No. of Runs	Score	Av./25	Dev. from n.p.	σ	C.R.	Score	Av./25	Dev. from n.p.	C.R.
Major Subjects										
C.J.	1000	6893	6.89	1893	64.50	29.35	5033	5.03	33	.51
R.S.	500	3013	6.03	513	45.61	11.25	2511	5.02	11	.24
C.B.	700	3852	5.50	352	54.02	6.52	3534	5.05	34	.63
D.M.	500	2584	5.17	84	45.61	1.84	2414	4.83	-86	-1.89
H.D.	500	2530	5.06	30	45.61	.66	2440	4.88	-60	-1.32
Total Major Subjects	3200	18872	5.90	2872	115.40	24.89	15932	4.98	-68	- .59
Minor Subjects										
S.B.	100	624	6.24	124	20.40	6.07	481	4.81	-19	- .93
G.E.	100	622	6.22	122	20.40	5.98	490	4.90	-10	- .49
J.M.	120	710	5.92	110	22.35	4.92	624	5.20	24	1.07
L.L.	140	764	5.46	64	24.15	2.65	676	4.83	-24	- .99
C.L.	130	672	5.17	22	23.26	.94	625	4.81	-25	-1.07
J.R.	140	713	5.09	13	24.15	.54	731	5.22	31	1.28
J.H.	100	500	5.00	0	20.40	.00	486	4.86	-14	- .69
E.C.	150	745	4.97	- 5	24.98	- .20	736	4.91	-14	- .56
Total Minor Subjects	980	5350	5.46	450	63.85	7.04	4849	4.95	-51	- .80
Series Total	4180	24222	5.79	3322	131.89	25.19	20781	4.97	-119	- .90

SUMMARY FOR HIGH-SCORING AND FOR LOW-SCORING SUBJECTS

	Direct Series						Reverse Series			
Subjects	No. of Runs	Score	Av./25	Dev. from n.p.		C.R.	Score	Av../25	Dev. from n.p.	C.R.
High Scorers	2660	16478	6.19	3178	105.21	30.21	13349	5.02	49	.47
Low Scorers	1520	7744	5.09	144	79.53	1.81	7432	4.90	-168	-2.11

Table 3.2

MAIN DT SERIES: DISTRIBUTION OF HIT FREQUENCIES

Subjects		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
C.J.	Direct	2	13	23	53	87	107	152	158	157	108	62	40	20	10	5	2	1
	Reverse	8	24	70	116	215	176	162	106	71	32	11	7	2	0	0	0	0
R.S.	Direct	0	5	18	36	68	77	86	91	61	31	12	14	0	1			
	Reverse	1	10	35	64	101	96	85	52	35	13	5	2	1	0			
C.B.	Direct	0	16	44	55	110	134	117	110	61	31	14	5	1	1	1		
	Reverse	1	13	55	96	135	113	120	86	48	24	4	4	1	0	0		
D.M.	Direct	2	9	26	76	77	94	91	69	29	16	9	2					
	Reverse	2	15	39	80	94	95	72	55	28	12	7	1					
H.D	Direct	2	18	32	56	85	106	82	65	28	23	1	2	0				
	Reverse	4	16	27	78	101	102	72	50	29	11	4	3	3				
Major Total	Direct	6	6	143	276	427	518	528	493	336	209	98	63	21	12	6	2	1
	Reverse	16	78	226	434	646	582	511	349	211	92	31	17	7	0	0	0	0
S.B.	Direct	0	0	1	10	7	17	24	15	13	6	5	2					
	Reverse	1	1	8	18	17	20	16	12	3	2	2	0					
G.E.	Direct	0	1	1	4	11	23	22	12	11	8	5	2					
	Reverse	2	3	6	10	24	16	16	14	6	3	0	0					
J.M.	Direct	0	2	2	13	14	18	27	16	15	7	5	0	1				
	Reverse	0	4	9	10	19	31	16	14	9	6	1	1	0				
L.L.	Direct	0	1	4	20	25	23	22	25	13	3	2	2					
	Reverse	1	2	19	16	22	39	11	11	11	7	0	1					
C.L.	Direct	1	1	7	17	34	22	15	14	8	6	4	0	0	0	1		
	Reverse	1	3	5	22	25	31	23	11	7	1	0	1	0	0	0		
J.R.	Direct	1	4	12	15	23	29	19	21	5	9	1	1					
	Reverse	0	1	8	17	26	30	23	17	11	6	1	0					
E.C.	Direct	1	4	10	23	24	35	20	13	13	4	3	0					
	Reverse	0	2	13	19	31	36	19	18	5	4	2	1					
J.H.	Direct	0	1	8	16	21	15	18	9	6	4	2	0	0				
	Reverse	0	1	11	14	25	8	27	5	2	4	2	0	1				
Minor Total	Direct	3	14	45	118	159	182	167	125	84	47	27	7	1	0	1		
	Reverse	5	17	79	126	189	211	151	102	54	33	8	4	1	0	0		
Series Total	Direct	9	75	188	394	585	701	695	618	420	256	125	70	22	12	7	2	1
	Reverse	21	95	305	560	835	793	662	451	265	125	39	21	8	0	0	0	0

MAIN DT SERIES: DISTRIBUTION SUMMARY FOR HIGH-SCORING AND FOR LOW-SCORING SUBJECTS

High Scorers	Direct	2	38	93	191	322	399	450	427	331	194	105	65	22	12	6	2	1
	Reverse	14	57	202	330	533	491	426	295	183	87	23	15	4				
Low Scorers	Direct	7	37	95	203	263	302	245	191	89	62	20	5	0	0	1		
	Reverse	7	38	103	230	302	302	236	156	82	38	16	6	4	0	0		

Table 3.3A

MAIN DT SERIES: DAILY FLUCTUATION IN SCORING (MAJOR SUBJECTS)

Date	Number of Runs	Direct Average	Reverse Average
Mr. C.J.			
Sept. 30	30	5.60	5.00
Oct. 5	50	6.98	5.56
Oct. 6	50	5.76	5.24
Oct. 7	50	6.76	5.20
Oct. 8	40	6.17	4.95
Oct. 12	80	7.91	5.31
Oct. 13	40	6.95	5.20
Oct. 14	80	7.47	4.97
Oct. 15	30	6.90	4.66
Oct. 18	20	6.55	5.10
Oct. 19	50	6.36	5.26
Oct. 20	10	8.90	4.70
Oct. 21	80	7.78	4.98
Oct. 25	20	5.60	4.10
Oct. 26	10	5.00	4.60
Oct. 28	50	5.74	5.08
Oct. 29	30	7.76	5.73
Nov. 2	40	5.77	4.85
Nov. 3	20	5.65	4.90
Nov. 4	20	6.10	4.65
Nov. 9	30	6.70	4.60
Nov. 10	70	7.54	4.61
Nov. 11	100	7.49	5.03
Mr. R. S.			
Oct. 7	10	5.50	4.90
Oct. 13	30	5.10	4.93
Oct. 15	40	5.43	5.14
Oct. 18	60	6.68	5.22
Oct. 20	60	6.05	5.38
Oct. 25	50	6.80	4.88
Oct. 27	40	6.18	4.83
Nov. 1	80	6.45	4.82
Nov. 3	30	5.23	4.73
Nov. 5	40	5.58	5.15
Nov. 12	20	5.60	4.70
Nov. 19	10	4.30	5.40
Nov. 24	30	6.20	5.10
Mr. C.B.			
Oct. 1	10	4.40	4.80
Oct. 5	70	5.23	5.13
Oct. 7	60	4.93	5.03
Oct. 11	70	4.83	4.83
Oct. 12	60	4.93	5.47
Oct. 19	80	6.49	5.01
Oct. 21	60	6.28	5.30
Oct. 25	40	4.68	5.08
Oct. 28	80	5.79	4.90
Nov. 2	40	5.40	4.98
Nov. 9	60	5.58	4.71
Nov. 16	40	6.05	5.12
Nov. 22	30	5.77	5.10
Mr. H.D.			
Apr. 17	30	5.60	5.33
Apr. 20	20	5.35	5.05
Apr. 21	30	4.23	5.13
Apr. 24	20	5.00	4.75
Apr. 25	130	5.20	4.71
Apr. 26	10	5.90	6.00
Apr. 27	30	5.13	4.96
May 2	30	4.60	5.00
May 4	70	5.01	5.11
May 9	40	4.92	4.78
May 18	30	5.23	4.30
June 1	60	4.93	4.68
Miss D.M.			
Oct. 8	15	6.00	4.40
Oct. 11	15	5.93	3.87
Oct. 13	20	5.35	5.20
Oct. 26	10	5.80	4.50
Oct. 28	30	4.63	5.40
Nov. 1	20	5.20	4.50
Nov. 4	30	4.87	4.13
Nov. 6	20	4.80	5.25
Nov. 11	30	5.13	4.60
Nov. 15	20	5.30	4.85
Nov. 16	6	4.00	4.33
Nov. 20	20	4.85	5.20
Dec. 6	10	4.80	5.40
Feb. 3	10	5.80	4.60
Apr. 5	4	5.00	6.50
Apr. 19	10	6.50	5.60
Apr. 20	30	5.20	4.60
Apr. 22	20	5.65	5.10
Apr. 26	20	5.05	4.95
May 2	10	4.80	4.00
June 1	20	5.20	5.20
June 6	20	5.05	4.40
June 7	10	4.70	4.60
June 16	20	5.80	5.30
June 19	40	4.70	4.85
Aug. 10	40	5.23	4.90

Table 3.3A (cont'd)

MAIN DT SERIES: DAILY FLUCTUATIONS IN SCORING (MINOR SUBJECTS)

Date	Number of Runs	Direct Average	Reverse Average
	Mr. S.B.		
Oct. 29	20	6.59	4.50
Nov. 5	40	6.00	4.98
Nov. 12	40	6.13	4.80
	Mr. G.E.		
Apr. 29	30	6.16	5.00
May 9	20	6.15	4.65
May 23	20	6.30	4.70
May 25	20	6.55	5.00
Oct. 27	10	5.70	5.60
	Miss L.L.		
Nov. 10	30	5.87	4.43
Feb. 23	10	5.80	5.20
Mar. 1	10	5.80	3.70
Mar. 8	60	5.48	5.27
Apr. 27	10	5.10	5.00
Apr. 28	20	4.60	4.40
	Miss J.M.		
Apr. 28	40	5.90	5.17
May 12	30	5.56	5.16
May 23	10	7.80	4.50
May 24	20	6.20	5.65
May 25	20	5.25	5.20

Date	Number of Runs	Direct Average	Reverse Average
	Miss E.C.		
Sept. 28	20	4.45	4.50
Sept. 30	20	5.50	4.90
Oct. 3	10	4.70	5.60
Oct. 5	40	5.10	4.68
Oct. 7	10	4.70	5.50
Oct. 8	10	4.80	6.00
Oct. 12	15	5.20	4.20
Oct. 19	25	4.88	5.08
	Mr. J.H.		
Dec. 2	10	5.00	4.30
Dec. 3	30	4.73	5.22
Dec. 5	30	4.90	4.70
Dec. 6	20	5.40	4.35
Dec. 13	10	5.30	4.80
	Miss C.L.		
Oct. 25	20	6.30	4.50
Oct. 27	40	5.00	5.23
Oct. 28	30	5.23	4.63
Nov. 4	40	4.73	4.68
	Mr. J.R.		
Jan. 24	30	4.40	5.26
Jan. 26	20	5.30	5.40
Feb. 1	30	5.20	4.53
Feb. 6	30	5.53	5.56
Feb. 10	30	5.10	5.40

Table 3.4

MAIN DT SERIES: FREQUENCIES FOR ISOLATED AND FOR CONSECUTIVE HITS

Major Subjects		1	2	3	4	5	6	7	8	9	10	11	12
Jencks	Direct	3571	951	259	75	39	18	4	0	0	0	0	1
	Predicted	3264	627	120	23	4	1	0	0	0	0	0	0
	Reverse	3379	582	123	18	6	2	1	0	0	0	0	0
Simmons	Direct	1687	423	96	29	9	4	1	0	0	0	0	0
	Predicted	1632	314	60	12	2	0	0	0	0	0	0	0
	Reverse	1651	320	57	11	1	0	0	0	0	0	0	0
Bilz	Direct	2323	512	102	30	11	4	0	0	0	0	0	0
	Predicted	2285	439	84	16	3	0	0	0	0	0	0	0
	Reverse	2344	404	95	23	1	0	0	0	0	0	0	0
Martin	Direct	1718	291	69	9	3	2	2	0	0	0	0	0
	Predicted	1632	314	60	12	2	0	0	0	0	0	0	0
	Reverse	1628	301	48	10	0	0	0	0	0	0	0	0
Dixon	Direct	1631	316	66	10	2	2	1	0	0	0	0	0
	Predicted	1632	314	60	12	2	0	0	0	0	0	0	0
	Reverse	1622	287	60	9	3	1	1	0	0	0	0	0
Total Major Subjects	Direct	10930	2493	592	153	64	30	8	0	0	0	0	1
	Predicted	10445	2007	383	74	14	3	1	0	0	0	0	0
	Reverse	10624	1894	383	71	11	3	2	0	0	0	0	0
Minor Subjects		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
S.B.	Direct	368	91	17	3	1	1	0	0	0	0	0	0
	Predicted	326	63	12	2	0	0	0	0	0	0	0	0
	Reverse	310	68	8	0	2	0	0	0	0	0	0	0
G.E.	Direct	356	82	24	5	2	0	0	0	0	0	0	0
	Predicted	326	63	12	2	0	0	0	0	0	0	0	0
	Reverse	333	62	9	1	0	0	0	0	0	0	0	0
J.M.	Direct	430	88	24	4	2	1	0	0	0	0	0	0
	Predicted	392	75	14	3	1	0	0	0	0	0	0	0
	Reverse	458	48	22	1	0	0	0	0	0	0	0	0
L.L.	Direct	472	104	20	6	0	0	0	0	0	0	0	0
	Predicted	457	88	17	3	1	0	0	0	0	0	0	0
	Reverse	411	85	23	4	2	0	0	0	0	0	0	0
C.L.	Direct	432	65	22	7	2	1	0	0	0	0	0	0
	Predicted	424	81	16	3	1	0	0	0	0	0	0	0
	Reverse	414	77	12	4	1	0	0	0	0	0	0	0
J.R.	Direct	456	84	25	2	0	1	0	0	0	0	0	0
	Predicted	457	88	17	3	1	0	0	0	0	0	0	0
	Reverse	495	80	24	1	0	0	0	0	0	0	0	0
J.H.	Direct	338	52	12	3	2	0	0	0	0	0	0	0
	Predicted	326	63	12	2	0	0	0	0	0	0	0	0
	Reverse	339	46	10	5	1	0	0	0	0	0	0	0

Table 3.4 (cont'd)

Minor Subjects		1	2	3	4	5	6	7	8	9	10	11	12
E.C.	Direct	462	103	23	2	0	0	0	0	0	0	0	0
	Predicted	490	94	18	3	0	0	0	0	0	0	0	0
	Reverse	550	59	16	5	0	0	0	0	0	0	0	0
Total Minor Subjects	Direct	3314	669	167	32	9	4	0	0	0	0	0	0
	Predicted	3199	615	118	23	4	1	0	0	0	0	0	0
	Reverse	3311	526	124	21	6	0	0	0	0	0	0	0
Total High Scorers	Direct	9702	2251	542	152	64	28	5	0	0	0	0	0
	Predicted	8682	1668	320	61	12	2	0	0	0	0	0	0
	Reverse	8887	1570	337	58	12	2	1	0	0	0	0	0
Total Low Scorers	Direct	5037	911	217	33	9	6	3	0	0	0	0	0
	Predicted	4961	953	183	35	7	1	0	0	0	0	0	0
	Reverse	5048	850	170	34	5	1	1	0	0	0	0	0
Series Total	Direct	14244	3162	759	185	73	34	8	0	0	0	0	0
	Predicted	13644	2622	503	96	18	4	1	0	0	0	0	0
	Reverse	13935	2420	507	92	17	3	2	0	0	0	0	0

Table 3.5

MAIN DT SERIES: AVERAGE SCORES (PER 25) FOR SUCCESSIVE FIVE CARD GROUPS

Subjects	Direct I	II	III	IV	V	Reverse I	II	III	IV	V
Major Subjects										
R.S.	5.88	6.12	6.24	6.09	5.80	5.09	4.70	5.18	5.24	4.90
C.B.	5.52	5.57	5.56	5.73	5.14	5.08	5.06	5.10	5.07	4.93
H.D.	5.09	4.80	5.27	4.95	5.19	5.02	4.86	4.93	4.95	4.64
D.M.	5.02	5.30	4.91	5.23	5.38	4.85	4.69	4.91	4.78	4.91
Total	5.39	5.46	5.50	5.52	5.35	5.02	4.85	5.04	5.02	4.81
C.J.	7.66	7.31	7.07	6.37	6.07	4.82	5.24	5.07	5.21	4.84
Major Total	6.10	6.04	5.99	5.79	5.58	4.95	4.97	5.05	5.08	4.85
Minor Subjects										
S.B.	6.15	6.55	6.15	6.10	6.25	4.95	4.65	5.15	4.55	4.75
G.E.	6.35	5.65	6.30	7.05	5.75	4.85	5.35	5.05	4.05	5.20
J.M.	5.75	5.93	5.79	6.38	5.75	5.46	5.46	5.63	5.00	4.46
L.L.	5.61	6.00	5.43	5.25	5.00	5.50	4.04	4.89	5.11	4.61
C.L.	4.88	5.38	4.77	5.31	5.50	5.12	4.58	4.58	4.89	4.89
J.R.	4.93	5.29	5.43	5.18	4.64	5.50	5.32	5.14	4.68	5.46
J.H.	5.00	4.65	4.75	4.95	5.65	5.20	4.55	5.00	5.20	5.35
E.C.	5.33	4.87	4.77	5.27	4.60	5.43	5.93	4,60	4.83	4.73
Minor Total	5.46	5.52	5.38	5.63	5.32	5.28	4.85	4.98	4.70	4.92
High Scorers										
Major Subjects (without C.J.)	5.67	5.80	5.84	5.88	5.41	5.08	4.91	5.13	5.14	4.92
Minor Subjects	5.92	6.02	5.87	6.12	5.63	5.23	4.83	5.17	4.73	4.73
Total High Scorers (without C.J.)	5.74	5.86	5.85	5.95	5.47	5.12	4.89	5.14	5.03	4.86
Total High Scorers (with C.J.)	6.46	6.41	6.31	6.10	5.70	5.01	5.02	5.30	5.10	4.85
Low Scorers										
Major Subjects	5.06	5.05	5.09	5.09	5.29	4.92	4.78	4.92	4.87	4.78
Minor Subjects	5.05	5.07	4.95	5.19	5.04	5.33	4.88	4.82	4.68	5.09
Total Low Scorers	5.05	5.06	5.04	5.03	5.20	5.07	4.81	4.88	4.80	4.88
Grand Total (without C.J.)	5.41	5.47	5.46	5.55	5.33	5.10	4.85	5.02	4.90	4.87
Series Total	5.95	5.91	5.85	5.75	5.54	5.03	4.94	5.03	4.98	4.86

Table 3.5A

MAIN DT SERIES:

Hit Frequency with Respect to Card Position (Direct)

Major Subjects

Card	R.S.	C.B.	H.D.	D.M.	C.J.	High Scorers	Low Scorers	Total
1	118	164	109	103	301	583	212	795
2	114	152	103	98	317	583	201	784
3	103	150	93	78	324	577	171	748
4	106	141	102	108	298	545	210	755
5	147	166	102	115	291	604	217	821
6	130	158	96	95	304	592	191	783
7	110	158	99	104	274	542	203	745
8	123	153	96	107	297	573	203	776
9	126	158	94	111	278	562	205	767
10	123	153	95	113	309	585	208	793
11	103	153	99	94	302	558	193	751
12	139	149	111	87	296	584	198	782
13	114	153	113	102	277	544	215	759
14	138	168	101	109	258	564	210	774
15	130	155	103	99	281	566	202	768
16	119	155	90	97	264	538	187	725
17	125	152	103	99	255	532	202	734
18	120	170	107	107	238	528	214	742
19	111	148	93	102	250	509	195	704
20	134	177	102	118	266	577	220	797
21	128	163	96	108	239	530	204	734
22	122	150	97	112	241	513	209	722
23	122	134	122	108	228	484	230	714
24	97	139	110	99	275	511	209	720
25	111	133	94	111	230	474	205	679

Table 3.5A

MAIN DT SERIES:

Hit Frequency with Respect to Card Position (Reverse)

Major Subjects

Card	R.S.	C.B.	H.D.	D.M.	C.J.	High Scorers	Low Scorers	Total
1	111	161	90	81	186	458	171	629
2	104	139	111	98	201	444	209	653
3	105	138	84	102	200	443	186	629
4	91	139	105	108	181	411	213	624
5	98	134	112	96	195	427	208	635
6	99	152	91	84	214	465	175	640
7	80	142	98	108	198	420	206	626
8	94	144	95	93	213	451	188	639
9	96	138	91	106	205	439	197	636
10	101	133	111	78	217	451	189	640
11	105	161	87	106	187	453	193	646
12	91	128	90	95	188	407	185	592
13	114	153	113	102	277	544	215	759
14	110	134	94	96	184	428	190	618
15	98	138	109	92	178	414	101	615
16	100	153	86	92	219	472	178	650
17	106	124	112	87	201	431	199	630
18	99	146	104	102	204	449	206	655
19	115	154	92	101	201	470	193	663
20	104	133	101	96	217	454	197	651
21	90	154	91	99	187	431	190	621
22	101	121	103	115	197	419	218	637
23	106	139	95	82	184	429	177	606
24	101	135	82	114	196	432	196	628
25	92	141	93	81	203	436	174	610

Table 3.5A

MAIN DT SERIES:

Hit Frequency with Respect to Card Position (Direct)

Minor Subjects

Card	S.B.	G.E.	L.L.	J.M.	E.C.	J.H.	C.L.	J.R.	High Scorers	Low Scorers	Total
1	28	23	22	28	34	18	33	26	101	111	212
2	31	29	38	31	26	23	26	32	129	107	236
3	22	19	32	24	35	25	12	30	97	102	199
4	23	32	40	20	30	12	29	23	115	94	209
5	19	24	25	35	35	22	27	27	103	111	214
6	20	25	33	33	26	17	30	38	111	111	222
7	29	24	28	34	31	20	29	28	115	108	223
8	26	18	39	18	27	21	28	23	101	99	200
9	34	24	36	28	25	15	30	22	122	92	214
10	22	22	32	29	37	20	23	37	105	117	222
11	27	23	37	29	30	18	30	27	116	105	221
12	23	33	36	28	36	15	21	27	120	99	219
13	24	22	24	29	27	23	28	26	99	104	203
14	21	23	29	23	26	16	19	44	96	105	201
15	28	25	26	30	24	23	26	28	109	101	210
16	18	31	33	27	28	20	23	23	109	94	203
17	22	27	22	32	24	12	29	25	103	90	193
18	30	34	30	34	33	14	31	33	128	111	239
19	27	25	23	29	40	25	31	28	104	124	228
20	25	24	39	31	33	28	24	36	119	121	240
21	28	15	31	23	24	25	29	28	97	106	203
22	26	27	34	32	31	24	36	35	119	126	245
23	25	30	24	26	31	26	23	19	105	99	204
24	22	21	32	22	33	18	26	28	97	105	202
25	24	22	19	35	19	20	29	20	100	88	188

Table 3.5A

MAIN DT SERIES:

Hit Frequency with Respect to Card Position (Reverse)

Minor Subjects

Card	S.B.	G.E.	L.L.	J.M.	E.C.	J.H.	C.L.	J.R.	High Scorers	Low Scorers	Total
1	24	15	31	27	40	19	22	32	97	113	210
2	22	25	31	28	25	24	28	36	106	113	219
3	22	19	41	23	34	20	28	34	105	116	221
4	15	20	23	26	29	18	30	27	84	104	188
5	16	18	28	27	35	23	25	25	89	108	197
6	24	20	22	27	30	21	24	28	93	103	196
7	15	20	22	24	30	16	26	34	81	106	187
8	22	24	16	32	25	22	19	26	94	92	186
9	17	21	33	26	29	16	23	31	97	99	196
10	15	22	20	22	34	16	27	30	79	107	186
11	19	22	35	22	26	11	25	33	98	95	193
12	20	14	20	34	24	25	22	27	88	98	186
13	24	22	24	29	27	23	28	26	99	104	203
14	23	23	29	31	30	23	22	27	106	102	208
15	17	20	29	19	31	18	22	31	85	102	187
16	17	17	29	26	28	20	31	30	89	109	198
17	19	10	31	25	28	20	19	27	85	94	179
18	18	15	26	25	34	19	27	26	84	106	190
19	21	20	25	24	29	15	23	19	90	86	176
20	16	19	32	20	26	10	27	29	87	92	179
21	17	28	21	18	27	22	23	31	84	103	187
22	23	17	33	24	26	21	31	29	97	107	204
23	21	11	27	23	30	25	27	29	82	111	193
24	13	20	29	18	27	19	29	22	80	97	177
25	21	28	19	24	32	20	17	42	92	111	203

Table 3.5A

MAIN DT SERIES:

Card Position Summary for High- and for Low-Scoring Subjects

Card	High-Scoring Subjects with C.J.	High-Scoring Subjects without C.J.	Low-Scoring Subjects	High-Scoring Subjects with C.J.	High-Scoring Subjects without C.J.	Low-Scoring Subjects
1	684	383	323	555	369	284
2	712	395	308	550	349	322
3	674	350	273	548	348	302
4	660	362	304	495	314	317
5	707	416	328	516	321	316
6	703	399	302	558	344	278
7	657	383	311	501	303	312
8	674	377	302	545	332	280
9	684	406	297	536	331	296
10	690	381	325	530	313	296
11	674	372	298	651	264	288
12	704	408	297	495	307	283
13	643	366	319	643	366	319
14	660	402	315	534	350	292
15	675	394	303	499	321	303
16	647	383	281	561	342	287
17	635	380	292	516	315	293
18	656	418	325	533	329	312
19	613	363	319	560	359	279
20	696	430	341	541	324	289
21	627	388	310	515	328	293
22	632	391	335	516	319	325
23	589	361	329	511	327	288
24	608	333	314	512	316	293
25	574	344	293	528	325	285

Table 4.1

UT SERIES:

Summary of Results for Direct, Reverse, and Matching Series

Subject	Runs	Trials	Hits	Av./25	Dev.	σ	C.R.
Mr. C.J.--direct	1000	25000	7387	7.39	2387	64.50	37.01
Mr. C.J.--reverse	1000	25000	5126	5.13	126	64.50	1.95
Matching Series	1000	25000	4911	4.91	- 89	64.50	- 1.38

Table 4.2

UT SERIES: DISTRIBUTION OF HIT FREQUENCIES

Subject	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Mr. C.J.--direct	0	3	14	22	50	100	159	166	176	154	78	48	21	8	0	0	1
Mr. C.J.--reverse	3	27	63	122	163	209	170	125	67	35	13	2	1	0	0	0	0

Table 4.4

UT SERIES: FREQUENCIES FOR ISOLATED AND FOR CONSECUTIVE HITS

Series	1	2	3	4	5	6	7
Direct	3648	1034	345	82	45	8	5
Reverse	3399	585	122	37	5	3	0
Predicted	3264	627	120	23	4	1	0

Table 4.5

UT SERIES:

Average Scores (per 25) for Successive Five Card Groups

Series	I	II	III	IV	V
Direct	8.85	7.91	7.35	6.67	6.18
Reverse	5.27	5.05	5.35	4.99	4.98

Table 4.3

UT SERIES:

Daily Flucutations in Scoring

Date	Runs	Direct Average	Reverse Average
Jan. 25	50	5.62	5.36
Jan. 26	30	6.47	5.37
Jan. 28	20	6.90	5.15
Jan. 31	20	6.90	5.45
Feb. 8	50	7.04	5.10
Feb. 9	20	7.15	4.80
Feb. 15	20	7.40	6.25
Feb. 19	40	7.60	5.15
Feb. 23	40	8.02	4.92
Feb. 24	40	7.62	5.05
Mar. 1	40	7.58	4.70
Mar. 7	40	7.65	5.48
Mar. 8	40	7.18	4.83
Mar. 29	50	7.78	5.26
Mar. 31	80	7.79	4.99
Apr. 5	60	6.87	4.63
Apr. 7	70	7.44	4.87
Apr. 12	50	7.26	5.14
Apr. 14	50	7.48	5.14
Apr. 19	40	8.13	5.68
Apr. 21	50	7.54	5.50
Apr. 30	40	8.08	4.82
May 3	60	7.66	5.23

Table 4.5A

UT SERIES:

Hit Frequencies with Respect to Card Position

Card	Direct Freq.	Reverse Freq.
1	356	209
2	375	205
3	373	200
4	365	219
5	300	221
6	332	212
7	333	189
8	306	219
9	314	179
10	296	211
11	325	208
12	307	194
13	283	283
14	286	175
15	268	209
16	281	182
17	290	201
18	272	194
19	244	219
20	249	201
21	244	193
22	258	199
23	264	209
24	243	199
25	226	196

Table 5.1

EE SERIES: SUMMARY OF RESULTS

	Direct Series						Reverse Series			
	No. of Runs	Score	Av./25	Dev. from n.p.	σ	C.R.	Score	Av./25	Dev. from n.p.	C.R.
EE_{RL}	400	2414	6.05	414	40.80	10.15	1995	4.98	- 5	- .12
EE_{LR}	140	823	5.88	123	24.13	5.09	663	4.73	-37	-1.53
EE_{25-5}	110	726	6.60	176	21.39	8.22	548	4.98	- 2	- .09
EE_5	60	462	7.70	162	15.80	10.25	315	5.25	+15	+ .95
DT (control)	140	811	5.79	111	24.13	4.60	679	4.85	-21	- .87
Total EE	710	4425	6.23	875	54.36	16.10	3521	4.96	-29	- .53
Total (Series)	850	5236	6.16	986	59.48	16.58	4200	4.94	-50	- .84

Table 5.2

EE SERIES: DISTRIBUTION OF HIT FREQUENCIES

		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
EE_{RL}	Direct	0	6	12	26	41	72	83	67	48	28	9	7	1		
	Reverse	1	9	29	57	73	78	66	47	18	15	6	1	0		
EE_{LR}	Direct	0	2	5	9	20	32	25	11	15	15	4	2			
	Reverse	0	2	12	23	34	22	25	11	7	2	1	1			
EE_{25-5}	Direct	0	0	1	2	13	15	20	29	15	8	3	2	2		
	Reverse	0	2	8	19	17	25	12	16	6	4	0	0	1		
EE_5	Direct	0	0	0	1	3	5	8	9	13	8	10	1	2		
	Reverse	0	2	7	4	9	11	4	14	8	0	1	0	0		
DT (control)	Direct	0	3	3	8	24	29	26	19	14	8	4	1	0	1	
	Reverse	0	6	10	20	27	32	17	12	9	4	2	1	0	0	
Total EE	Direct	0	8	18	38	77	124	136	116	91	59	26	12	5	0	
	Reverse	1	15	56	103	133	136	107	88	39	21	8	2	1	0	
Total (Series)	Direct	0	11	21	46	101	153	162	135	105	67	30	13	5	1	
	Reverse	1	21	66	123	160	168	124	100	48	25	10	3	1	0	

Table 5.3A

EE SERIES:

Daily Fluctuations in Scoring

Date	Number of Runs	Direct Average	Reverse Average
	EE_{RL}		
Sept. 29	40	7.28	5.13
Oct. 7	10	6.00	4.40
Oct. 17	10	4.50	5.20
Dec. 2	10	6.70	5.50
Jan. 6	30	6.43	4.20
Jan. 7	40	6.70	4.65
Jan. 9	40	6.88	5.18
Jan. 11	20	5.25	4.65
Jan. 13	30	5.30	5.00
Jan. 16	20	5.55	5.45
Jan. 18	10	6.20	4.60
Jan. 23	20	5.95	5.10
Jan. 24	20	5.55	5.55
Jan. 25	40	5.52	5.22
Jan. 26	10	4.50	5.10
Jan. 30	30	5.40	4.63
Feb. 2	20	6.15	5.70
	DT (control)		
Sept. 28	20	6.35	3.65
Oct. 7	10	6.00	4.20
Nov. 14	20	4.95	5.00
Nov. 16	20	5.10	4.75
Nov. 17	10	5.70	5.10
Nov. 18	20	5.45	5.15
Nov. 28	10	7.20	6.40
Nov. 30	10	6.90	5.20
Jan. 4	20	5.80	4.95

Date	Number of Runs	Direct Average	Reverse Average
	EE_{LR}		
Feb. 27	20	5.45	4.75
Mar. 1	20	6.35	5.15
Mar. 2	50	5.32	4.68
Mar. 3	20	5.90	4.55
Mar. 4	30	6.77	4.67
	EE_{25-5}		
Mar. 6	10	6.10	4.70
Mar. 30	40	6.43	4.87
Mar. 31	20	7.20	4.75
Apr. 3	40	6.60	5.27
	EE_5		
Apr. 7	10	6.40	4.00
Apr. 11	10	6.90	6.20
Apr. 12	10	7.20	4.00
Apr. 17	10	8.80	6.00
Apr. 19	10	8.10	5.40
Apr. 24	10	8.80	6.00

Table 5.4

EE SERIES:

Frequency for Isolated and for Consecutive Hits

Series		1	2	3	4	5	6	7	8
EE_{RL}	Direct	1361	304	107	27	2	1	0	0
	Predicted	1306	251	48	9	2	0	0	0
	Reverse	1353	235	45	8	1	0	0	0
EE_{LR}	Direct	495	103	20	9	4	1	0	0
	Predicted	457	88	17	3	1	0	0	0
	Reverse	434	83	12	3	3	0	0	0
EE_{25-5}	Direct	359	85	37	6	4	7	0	0
	Predicted	359	68	13	3	0	0	0	0
	Reverse	388	55	13	0	1	1	0	0
EE_5	Direct	249	68	19	5	0	0	0	0
	Predicted	196	38	7	1	0	0	0	0
	Reverse	206	36	7	4	0	0	0	0
DT (control)	Direct	522	95	23	6	0	1	0	0
	Predicted	457	88	17	3	1	0	0	0
	Reverse	452	82	14	2	1	0	0	1
Total EE	Direct	2464	560	183	47	10	9	0	0
	Predicted	2318	445	85	16	3	0	0	0
	Reverse	2381	409	77	15	5	1	0	0
Total (Series)	Direct	2986	655	206	53	10	10	0	0
	Predicted	2774	533	102	20	4	1	0	0
	Reverse	2833	491	91	17	6	1	0	1

Table 5.5A

EE SERIES:

Hit Frequency with Respect to Card Position (Direct)

Card	EE_{RL}	EE_{LR}	EE_{25-5}	EE_5	Total EE	DT
1	140	52	38	21	251	40
2	172	46	42	23	283	34
3	147	44	32	14	237	42
4	135	56	42	17	250	25
5	83	40	31	16	170	32
6	85	33	18	19	155	35
7	83	28	25	17	153	30
8	75	29	19	19	142	26
9	94	28	22	17	161	31
10	85	22	16	12	135	30
11	92	28	20	21	161	33
12	94	28	16	22	160	35
13	75	23	28	17	143	26
14	77	30	21	20	148	34
15	95	33	22	16	166	30
16	83	25	22	15	145	35
17	75	29	23	22	149	30
18	86	30	22	19	157	36
19	103	30	22	12	167	27
20	92	33	24	20	169	28
21	92	30	40	21	183	23
22	87	27	41	22	177	42
23	81	34	52	23	190	35
24	90	34	45	19	188	32
25	93	31	43	18	185	40

Hit Frequency with Respect to Card Position (Reverse)

Card	EE_{RL}	EE_{LR}	EE_{25-5}	EE_5	Total EE	DT
1	87	23	26	10	146	21
2	79	30	15	12	136	30
3	82	25	17	14	138	28
4	84	30	25	14	153	24
5	83	31	19	8	141	31
6	73	38	20	9	140	30
7	92	20	21	7	140	33
8	72	31	17	19	139	19
9	79	21	29	9	138	29
10	83	21	21	12	137	29
11	79	28	21	13	141	22
12	70	27	18	6	121	34
13	75	23	28	17	143	26
14	84	30	26	14	154	23
15	91	29	11	10	141	21
16	81	23	26	17	147	29
17	67	20	30	15	132	25
18	66	25	25	19	135	26
19	82	28	19	9	138	27
20	72	27	22	15	136	34
21	79	25	27	11	142	28
22	86	22	22	6	136	27
23	94	23	26	23	166	27
24	79	22	15	14	130	32
25	76	41	22	12	151	24

Table 5.5

EE SERIES:

Average Scores (per 25) for Successive Five Card Groups

Method	Direct I	Direct II	Direct III	Direct IV	Direct V	Reverse I	Reverse II	Reverse III	Reverse IV	Reverse V
EE_{RL}	8.46	5.28	5.41	5.49	5.54	5.19	4.99	4.99	4.60	5.18
EE_{LR}	8.50	5.00	5.07	5.25	5.57	4.96	4.68	4.89	4.39	4.75
EE_{25-5}	8.41	4.55	4.86	5.14	10.05	4.64	4.91	4.73	5.56	5.09
EE_5	7.58	7.00	8.00	7.33	8.58	4.83	4.67	5.00	6.25	5.50
Total EE	8.39	5.25	5.48	5.54	6.50	5.03	4.89	4.93	4.85	5.11
DT	6.18	5.43	5.64	5.57	6.14	4.79	5.00	4.50	5.04	4.93

Table 6.1

MINOR VARIATIONS: SUMMARY OF RESULTS

	Direct Series						Reverse Series			
Method	No. of Runs	Score	Av./25	Dev. from n.p.	σ	C.R.	Score	Av./25	Dev. from n.p.	C.R.
DT_T	110	899	8.17	349	21.39	16.31	522	4.74	-28	-1.31
DT_D	120	781	6.51	181	22.35	8.10	630	5.25	30	1.34
DT_{TD}	70	511	7.30	161	17.06	9.43	367	5.24	17	1.00
UT_D	90	607	6.74	157	19.35	8.11	446	4.96	- 4	- .21
UT_{50}	30	234	7.80	84	11.17	7.52	152	5.07	2	.18
UT_{75}	9	75	8.33	30	6.12	4.90	40	4.44	- 5	- .82

DT_T and DT_{TD}: Summary of Results for the Nine Pack Check

Method	No. of Runs	Score	Av./25	Dev. from n.p.		C.R.
DT_T	990	4973	5.02	23	64.18	.36
DT_{TD}	630	3149	4.99	- 1	51.20	-.02

Table 6.2

MINOR VARIATIONS: DISTRIBUTION OF HIT FREQUENCIES

Experimental Technique	Series	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Runs	Score
DT_T	Direct	1	0	1	1	4	12	13	9	16	18	18	6	6	1	4	110	899
	Reverse	2	3	10	16	17	23	19	10	8	0	2					110	522
DT_D	Direct	0	1	1	11	18	10	22	14	17	16	4	3	1	0	2	120	781
	Reverse	0	2	7	14	18	25	23	20	5	5	0	0	1			120	630
DT_{TD}	Direct	0	1	1	4	4	6	13	8	12	6	4	7	3	1		70	511
	Reverse	0	2	5	12	7	12	13	7	6	4	2					70	367
UT_D	Direct	0	1	1	6	6	8	16	22	14	7	4	3	2			90	607
	Reverse	1	1	8	9	16	24	15	6	7	1	1	1				90	446
UT_{50}*	Direct	0	0	0	0	3	1	3	5	7	6	2	2	1			30	234
	Reverse	0	1	1	4	6	6	8	1	1	1	0	1				30	152
UT_{75}*	Direct	0	0	0	0	1	1	0	0	2	2	1	2				9	75
	Reverse	0	0	0	1	4	3	1									9	40
Total	Direct	1	3	4	22	36	38	67	58	68	55	33	23	13	2	6	429	3107
	Reverse	3	9	31	56	68	93	79	44	27	11	5	2	1			429	2157

*Recorded frequencies pertain to arbitrary "runs" of 25.

Table 6.3A

MINOR VARIATIONS: DAILY FLUCTUATIONS IN SCORING

Daily Fluctuations in Scoring

Date	Number of Runs	Direct Average	Reverse Average
		DT_T	
Nov. 16	50	8.46	4.80
Nov. 18	60	7.90	4.65
		DT_D	
Nov. 18	30	4.80	5.25
Dec. 1	30	7.83	5.00
Dec. 2	20	7.75	5.30
Jan. 5	40	6.17	5.42
		DT_{TD}	
Nov. 25	70	7.30	5.24
		UT_D	
May 5	30	6.60	4.93
May 10	10	7.50	5.10
May 11	20	6.70	4.75
May 12	30	6.66	5.06
		UT_{50} and UT_{75}	
May 17	59	7.92	4.82

Table 6.4

MINOR VARIATIONS: FREQUENCIES FOR ISOLATED AND FOR CONSECUTIVE HITS

		1	2	3	4	5	6	7
DT_T	Direct	436	118	40	16	5	3	0
	Predicted	359	69	13	3	0	0	0
	Reverse	348	69	8	3	0	0	0
DT_D	Direct	397	111	31	9	4	1	1
	Predicted	392	75	14	3	1	0	0
	Reverse	411	71	25	2	0	0	0
DT_{TD}	Direct	228	72	25	8	5	0	0
	Predicted	228	44	8	2	0	0	0
	Reverse	237	44	11	1	1	0	0
UT_D	Direct	298	77	29	10	3	1	1
	Predicted	293	56	11	2	0	0	0
	Reverse	276	68	6	4	0	0	0
UT_{50}	Direct	98	38	13	4	1	0	0
	Predicted	98	19	4	1	0	0	0
	Reverse	103	17	3	0	0	0	0
UT_{75}	Direct	29	16	5	1	1	0	0
	Predicted	29	6	1	0	0	0	0
	Reverse	30	5	0	0	0	0	0
Total	Direct	1476	432	143	48	19	5	3
	Predicted	1400	269	52	10	2	0	0
	Reverse	1405	274	53	10	1	0	0

Table 6.5A

MINOR VARIATIONS: HIT FREQUENCY WITH RESPECT TO CARD POSITION

Direct

Card	DT_T	DT_D	DT_{TD}	UT_D	UT_{50}		UT_{75}		
1	42	44	22	43	8	7*	1	1*	1**
2	51	52	24	46	8	6	2	0	2
3	55	36	18	33	6	4	3	1	1
4	46	26	24	28	4	4	0	1	0
5	39	33	19	18	5	0	1	0	1
6	38	33	25	20	4	3	2	[illegible]	2
7	36	27	20	18	7	7	2	1	1
8	37	24	21	19	6	2	0	1	1
9	38	21	15	21	6	2	2	0	3
10	32	28	21	16	2	4	2	1	3
11	36	29	21	17	4	7	1	3	0
12	40	27	19	15	1	3	0	0	2
13	42	29	23	22	8	6	1	0	1
14	26	26	18	20	6	5	1	1	1
15	40	29	25	26	5	7	1	1	0
16	34	25	21	19	3	3	1	1	1
17	38	25	23	17	3	7	2	1	0
18	34	35	23	29	7	5	0	1	1
19	26	22	12	26	4	4	0	1	1
20	29	25	22	15	4	4	1	1	0
21	27	30	23	28	6	4	0	1	2
22	35	40	21	28	6	6	0	0	0
23	26	42	22	34	4	3	3	0	1
24	21	39	14	26	6	1	2	1	1
25	31	34	15	23	3	4	1	0	0

Reverse

Card	DT_T	DT_D	DT_{TD}	UT_D	UT_{50}		UT_{75}		
1	13	21	24	12	2	2*	0	0	1**
2	26	26	16	18	4	2	1	0	1
3	17	20	14	22	5	5	0	0	0
4	26	23	11	21	2	4	0	0	0
5	18	23	17	8	3	4	1	0	0
6	11	29	11	25	2	1	0	2	1
7	22	23	21	11	5	1	1	0	1
8	19	26	21	22	7	6	0	3	0
9	26	29	16	19	4	3	0	0	0
10	21	27	9	18	3	5	0	0	3
11	22	27	15	28	3	3	0	1	0
12	20	28	10	23	0	2	0	2	1
13	42	29	23	22	1	1	0	0	2
14	14	27	6	18	4	5	2	0	1
15	13	35	14	15	3	3	1	1	0
16	24	27	15	17	0	4	1	0	0
17	18	26	13	23	2	1	0	0	1
18	21	23	15	18	1	4	1	1	0
19	10	24	16	19	2	3	1	2	0
20	23	30	13	16	3	6	2	0	0
21	22	17	13	17	1	3	0	0	0
22	20	29	16	12	1	2	0	0	1
23	19	22	13	16	2	4	0	0	1
24	31	14	10	13	6	6	0	1	1
25	24	25	15	13	4	2	1	0	0

* Second column records hits for cards 26 to 50.
** Third column records hits for cards 51 to 75.

Table 6.5

MINOR VARIATIONS: AVERAGE SCORES (PER 25) FOR SUCCESSIVE FIVE CARD GROUPS

Method	Direct I	II	III	IV	V	Reverse I	II	III	IV	V
DT_T	10.59	8.23	8.36	7.32	6.36	4.55	4.50	5.05	4.36	5.27
DT_D	7.96	5.54	5.83	5.50	7.71	4.71	5.58	6.08	5.42	4.46
DT_{TD}	7.64	7.29	7.57	7.21	6.79	5.86	5.57	4.86	5.14	4.79
UT_D	9.33	5.22	5.56	5.89	7.72	4.50	5.28	5.89	5.17	3.94
UT_{50}	9.33	7.50	7.67	7.67	6.83	5.83	3.16	5.16	5.16	6.00
UT_{75}	10.55	7.22	8.33	9.44	6.11	4.44	2.22	6.66	4.44	4.44

Table 7.1

SUMMARY FOR ALL EXPERIMENTAL SERIES

	Direct Series						Reverse Series			
Series	No. of Runs	Score	Av./25	Dev. from n.p.	σ	C.R.	Score	Av./25	Dev. from n.p.	C.R.
High Scorers										
Preliminary	1571	10356	6.59	2501	80.85	30.93	7861	5.00	6	.07
Main DT	2660	16478	6.19	3178	105.21	30.21	13349	5.02	49	.47
Special Methods	2279	15730	6.90	4335	97.39	45.41	11483	5.04	88	.90
Total High Scorers	6510	42564	6.54	10014	164.59	60.84	32693	5.02	143	.87
Low Scorers										
Preliminary	1600	8267	5.17	267	81.60	3.27	7948	4.97	- 52	- .64
Main DT	1520	7744	5.09	144	79.53	1.81	7432	4.90	-168	-2.11
Total Individual Subjects	3120	16011	5.13	411	113.95	3.61	15380	4.93	-220	-1.93
Class	2840	14187	5.00	- 13	108.71	- .12	14049	4.95	-151	-1.39
Total Low Scorers	5960	30198	5.07	398	157.49	2.52	29429	4.94	-371	-2.36
Grand Total	12470	72762	5.83	10412	227.81	45.70	62122	4.98	-228	-1.00

Table 6.6

DT_{TD}: HITS DERIVED FROM THE MATCHING OF CALLS AGAINST TARGET AND NON-TARGET PACKS

Pack / Call*	1	2	3	4	5	6	7	8	9	10	Total unintended hits per call	Av./25
1	95	54	44	53	52	60	61	60	60	73	517**	5.22
2	49	82	70	70	50	69	64	56	67	51	546	5.51
3	45	43	85	60	41	54	58	57	54	56	468	4.73
4	57	56	62	93	52	71	47	61	52	51	509	5.14
5	62	62	54	55	83	47	65	67	47	59	508	5.13
6	51	53	66	52	61	101	46	56	53	59	497	5.02
7	59	49	47	55	55	45	83	57	51	62	478	4.83
8	52	42	53	69	54	47	55	84	51	52	475	4.80
9	48	58	55	47	50	53	62	47	98	54	474	4.79
10	58	54	56	55	64	48	58	51	57	88	501	5.06
Total unintended hits per pack (diagonal cells omitted)	481**	471	507	516	479	492	506	512	492	517	4973	
Average per 25	4.86	4.76	5.12	5.21	4.84	4.97	5.11	5.17	4.97	5.22		5.00

* Call 1: Call intended for Pack 1.
** Cell expectancy: 55; column expectancy 495.

Direct Av: 7.30

DT_T: HITS DERIVED FROM THE MATCHING OF CALLS AGAINST TARGET AND NON-TARGET PACKS

Pack / Call*	1	2	3	4	5	6	7	8	9	10	Total unintended hits per call	Av./25
1	40**	32	45	39	30	37	30	39	38	30	320**	5.08
2	32	52	38	40	31	48	37	34	34	31	325	5.16
3	41	40	46	30	41	36	44	38	39	29	338	5.37
4	36	38	27	61	26	44	43	34	29	31	308	4.89
5	27	32	31	33	51	29	39	35	27	37	290	4.60
6	38	25	43	32	43	51	40	29	45	42	337	5.35
7	32	38	34	37	28	34	50	29	29	35	296	4.70
8	36	37	38	36	35	24	36	46	34	39	315	5.00
9	31	36	34	33	34	35	27	44	60	38	312	4.97
10	41	36	27	34	33	35	46	29	24	54	305	4.84
Total unintended hits per pack (diagonal cells omitted)	314**	314	317	314	301	322	342	311	299	312	3146	
Average per 25	4.98	4.98	5.03	4.98	4.77	5.11	5.43	4.93	4.75	4.97		4.99

* Call 1: Call intended for Pack 1.
** Cell expectancy: 35; column (or row) expectancy: 315.

Direct Av: 8.17

Table 7.2

SUMMARY FOR ALL EXPERIMENTAL SERIES: DISTRIBUTION OF HIT FREQUENCIES

		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
High-Scoring Subjects																				
Preliminary	Direct	4	12	42	119	189	204	217	249	200	124	88	53	33	22	9	4	0	1	1
	Reverse	6	36	121	217	279	302	266	165	103	47	18	8	1	1	0	1	0	0	0
Main DT	Direct	2	38	93	191	322	399	450	427	331	194	105	65	22	12	6	2	1		
	Reverse	14	57	202	330	533	491	426	295	183	87	23	15	4	0	0	0	0		
Special Methods	Direct	1	17	39	90	187	291	388	359	349	276	141	84	39	11	6	0	1		
	Reverse	7	57	160	301	391	470	373	269	142	71	28	7	3	0	0	0	0		
Total	Direct	7	67	174	400	692	894	1055	1035	880	594	334	202	94	45	21	6	2	1	1
	Reverse	27	150	483	848	1203	1263	1065	729	428	205	69	30	8	1	0	1	0	0	0
Low-Scoring Subjects																				
Preliminary	Direct	4	36	95	216	264	336	235	209	109	62	26	4	3	0	1				
	Reverse	7	39	124	239	276	307	241	184	114	42	18	7	2	0	0				
Main DT	Direct	7	37	95	203	263	302	245	191	89	62	20	5	0	0	1				
	Reverse	7	38	103	230	302	302	236	156	82	38	16	6	4	0	0				
Total (Individual)	Direct	11	73	190	419	527	638	480	400	198	124	46	9	3	0	2				
	Reverse	14	77	227	469	578	609	477	340	196	80	34	13	6	0	0				
Class	Direct	15	70	190	404	528	572	415	335	168	90	32	14	3	2	2				
	Reverse	17	69	200	414	532	562	421	314	186	80	30	12	3	0	0				
Total	Direct	26	143	380	823	1055	1210	895	735	366	214	78	23	6	2	4				
	Reverse	31	146	427	883	1110	1171	898	654	382	160	64	25	9	0	0				
All Subjects																				
Grand Total	Direct	33	210	554	1223	1753	2104	1950	1770	1246	808	412	225	100	47	25	6	2	1	1
	Reverse	58	296	910	1731	2313	2434	1963	1383	810	365	133	55	17	1	0	1			

Table 7.3A

SUMMARY FOR ALL EXPERIMENTAL SERIES: DAILY FLUCTUATIONS IN SCORING

Nov.	4	60	5.18	4.66
Nov.	5	80	5.78	5.06
Nov.	9	90	6.18	4.68
Nov.	10	100	7.04	4.56
Nov.	11	100	7.49	5.03
Nov.	12	60	5.95	4.76
Nov.	16	90	7.39	4.98
Nov.	18	90	6.89	4.84
Nov.	19	10	4.30	5.40
Nov.	22	30	5.77	5.10
Nov.	23	70	7.30	5.24
Nov.	24	30	6.20	5.10
Dec.	1	30	7.83	5.00
Dec.	2	30	6.83	4.97
Dec.	3	30	4.73	5.43
Dec.	5	30	4.90	4.70
Dec.	5	20	5.40	4.35
Dec.	13	10	5.30	4.80

Jan.	5	40	6.17	5.42
Jan.	25	50	5.62	5.36
Jan.	26	30	6.46	5.36
Jan.	28	20	6.90	5.10
Jan.	31	20	6.90	5.45
Feb.	3	10	5.80	4.60
Feb.	8	50	7.04	5.10
Feb.	9	20	7.15	4.80
Feb.	15	20	7.40	6.25
Feb.	19	40	7.60	5.15
Feb.	23	50	7.58	4.98
Feb.	24	40	7.62	5.05
Mar.	1	50	7.22	5.00
Mar.	7	40	7.65	5.47
Mar.	8	100	6.16	5.09
Mar.	29	50	7.78	5.26
Mar.	31	80	7.78	4.98
Apr.	5	64	6.75	4.75
Apr.	7	70	7.44	4.87
Apr.	12	50	7.26	5.14
Apr.	14	50	7.48	5.14
Apr.	19	530	5.29	4.93
Apr.	20	620	4.93	5.00
Apr.	21	50	7.54	5.50
Apr.	22	800	5.04	4.81
Apr.	26	20	5.05	4.95
Apr.	27	10	5.10	5.00
Apr.	28	60	5.47	4.92
Apr.	29	30	6.16	5.00
Apr.	30	40	8.07	4.82
May	2	10	4.70	4.00
May	3	60	7.66	5.23
May	5	30	6.60	4.93
May	9	20	6.15	4.65
May	10	10	7.50	5.10
May	11	20	6.70	4.75

May	12	60	6.12	5.12
May	23	30	6.80	4.63
May	24	20	6.20	5.65
May	25	40	5.90	5.10
Sept.	28	40	5.40	4.07
Sept.	29	40	7.28	5.13
Sept.	30	20	5.50	4.90
Oct.	3	10	4.70	5.60
Oct.	5	40	5.10	4.68
Oct.	7	20	5.35	4.85
Oct.	8	25	5.52	5.04
Oct.	11	85	5.02	4.65
Oct.	12	15	5.20	4.20
Oct.	13	20	5.25	5.20
Oct.	17	10	4.50	5.40
Oct.	19	25	4.88	5.08
Oct.	25	450	4.79	5.08
Oct.	26	10	5.80	4.50
Oct.	27	550	4.97	4.98
Oct.	28	30	4.63	5.40
Nov.	1	20	5.20	4.50
Nov.	4	30	4.86	4.13
Nov.	6	20	4.80	5.25
Nov.	11	30	5.13	4.60
Nov.	14	20	4.95	5.00
Nov.	15	20	5.30	4.85
Nov.	16	26	4.85	4.65
Nov.	17	10	5.70	5.10
Nov.	18	20	5.45	5.15
Nov.	20	20	4.85	5.20
Nov.	28	10	7.20	6.40
Nov.	30	10	6.90	5.20
Dec.	2	10	6.70	5.50
Dec.	6	10	4.80	5.40

Jan.	10	50	5.60	4.44
Feb.	4	90	5.30	5.30
Feb.	5	20	4.30	4.85
Feb.	6	10	7.30	5.40
Feb.	8	10	5.90	4.60
Feb.	9	30	5.20	4.43
Feb.	10	40	5.52	4.65
Feb.	12	9	6.11	5.44
Feb.	16	10	7.40	4.00
Feb.	17	90	6.28	4.81
Feb.	18	70	5.54	4.94
Feb.	19	66	4.84	4.90
Feb.	20	20	6.00	5.45
Feb.	21	40	4.85	5.22
Feb.	23	94	5.19	4.81
Feb.	25	10	4.50	5.60
Mar.	1	10	5.50	4.80
Mar.	2	50	5.20	5.08
Mar.	3	30	5.63	5.40
Mar.	5	30	5.10	4.86
Mar.	6	30	6.83	4.96

Table 7.3A (cont'd)

Mar.	11	25	6.32	5.84
Mar.	12	36	7.58	5.30
Mar.	14	60	7.58	4.61
Mar.	17	10	6.90	5.30
Mar.	28	20	4.95	6.15
Mar.	29	20	5.25	4.25
Apr.	3	30	7.67	4.87
Apr.	5	30	6.80	4.73
Apr.	6	60	5.13	5.40
Apr.	8	100	4.68	4.18
Apr.	10	25	4.88	4.80
Apr.	12	100	5.84	4.69
Apr.	13	40	5.48	4.85
Apr.	14	20	5.05	4.90
Apr.	15	60	5.21	5.15
Apr.	16	30	5.70	5.07
Apr.	17	30	5.60	5.33
Apr.	19	60	4.80	5.05
Apr.	20	50	4.56	5.20
Apr.	21	60	4.75	4.97
Apr.	22	50	5.16	5.12
Apr.	24	20	5.00	4.75
Apr.	25	130	5.20	4.71
Apr.	26	50	5.90	5.00
Apr.	27	99	5.02	5.08
Apr.	29	90	6.90	4.95
Apr.	30	110	6.17	5.08
May	1	10	5.10	4.80
May	2	30	4.60	5.00
May	3	60	5.60	4.78
May	4	180	6.00	5.24
May	5	80	7.06	5.00
May	6	27	5.44	4.66
May	9	50	4.85	4.90
May	10	50	8.52	5.20
May	11	10	5.40	5.70
May	12	80	5.98	5.10
May	13	30	5.20	4.40
May	17	50	6.90	5.16
May	18	80	5.19	4.89
May	19	120	4.79	5.08
May	20	110	6.43	5.11
May	21	40	7.55	4.75
May	22	30	5.00	5.02
May	25	50	6.00	5.00
May	27	20	6.05	5.10
May	28	30	5.73	5.53
June	1	100	5.53	4.80
June	14	80	8.63	5.05
July	2	40	5.40	5.05
July	3	20	5.95	5.60
July	6	10	5.40	4.00
July	27	30	5.76	4.66
Aug.	24	10	5.40	4.90
Sept.	30	30	5.60	5.00
Oct.	1	10	4.40	4.80
Oct.	5	120	5.96	4.48
Oct.	6	50	5.76	4.24
Oct.	7	120	5.74	5.92
Oct.	8	40	6.17	4.95
Oct.	11	70	4.83	4.83
Oct.	12	140	6.64	5.38
Oct.	13	70	6.15	5.08
Oct.	14	80	7.47	4.97
Oct.	15	70	6.05	4.94
Oct.	18	80	6.65	5.18
Oct.	19	130	6.44	5.15
Oct.	20	70	6.45	5.28
Oct.	21	140	7.14	5.12
Oct.	25	130	5.88	4.76
Oct.	26	10	5.00	4.60
Oct.	27	80	5.58	5.02
Oct.	28	160	5.67	4.91
Oct.	29	50	7.44	5.24
Nov.	1	50	5.40	4.80

Jan.	4	20	5.80	4.95
Jan.	6	30	6.43	4.20
Jan.	7	40	6.70	4.65
Jan.	9	40	6.87	5.17
Jan.	11	20	5.25	4.65
Jan.	13	30	5.20	4.86
Jan.	16	20	5.55	5.45
Jan.	18	10	6.20	4.60
Jan.	23	20	5.95	5.10
Jan.	24	50	4.86	5.38
Jan.	25	40	5.52	5.22
Jan.	26	30	5.03	5.30
Jan.	30	30	5.40	4.63
Feb.	1	30	5.20	4.53
Feb.	2	20	6.15	5.70
Feb.	6	40	5.97	5.52
Feb.	10	30	5.10	5.40
Feb.	27	20	5.45	4.75
Mar.	1	20	6.35	5.15
Mar.	2	50	5.32	4.66
Mar.	3	20	5.90	4.55
Mar.	4	30	6.76	4.63
Mar.	6	10	6.10	4.70
Mar.	30	40	6.42	4.87
Mar.	31	20	7.20	4.75
Apr.	3	40	6.60	5.27
Apr.	7	10	6.40	4.00
Apr.	11	10	6.90	6.20
Apr.	12	10	7.20	4.00
Apr.	17	10	8.80	6.00
Apr.	19	10	8.10	6.40
Apr.	24	10	8.80	6.00
May	17	39	7.92	4.92
June	1	20	5.20	5.20
June	6	20	5.05	4.40
June	7	10	4.70	4.60
June	16	20	5.80	5.30
June	19	40	4.70	4.85
Aug.	10	40	5.25	4.90

Table 7.4

SUMMARY OF ALL EXPERIMENTAL SERIES: FREQUENCIES FOR ISOLATED AND FOR CONSECUTIVE HITS

Low-Scoring Subjects

		1	2	3	4	5	6	7	8	9	10	11	12
Preliminary	Direct	5320	1056	207	42	8	1	0					
	Predicted	5222	1004	193	37	7	1	0					
	Reverse	5130	983	200	47	9	2	1					
Main DT	Direct	5040	911	216	33	9	6	3					
	Predicted	4961	953	183	35	7	1	0					
	Reverse	5048	850	170	34	5	1	1					
Total Individual Subjects	Direct	10360	1967	423	75	17	7	3					
	Predicted	10184	1957	376	72	14	2	1					
	Reverse	10178	1833	370	81	14	3	2					
Class	Direct	9341	1705	331	90	13	3	0					
	Predicted	9270	1781	342	65	13	2	0					
	Reverse	9195	1750	335	65	13	4	0					
Grand Total	Direct	19701	3672	754	165	30	10	3					
	Predicted	19454	3738	717	137	26	5	1					
	Reverse	19373	3583	705	146	27	7	2					

High-Scoring Subjects

		1	2	3	4	5	6	7	8	9	10	11	12
Preliminary	Direct	5584	1372	424	98	48	8	5	4	1	0	0	0
	Predicted	5128	985	189	36	7	2	0	0	0	0	0	0
	Reverse	5167	939	213	36	3	3	0	0	0	0	0	0
Main DT	Direct	9207	2251	542	152	64	28	5	0	0	0	0	1
	Predicted	8682	1668	320	61	12	2	0	0	0	0	0	0
	Reverse	8880	1570	337	58	8	4	2	1	0	0	0	0
All remaining series with Mr. C.J.	Direct	8110	2121	694	183	74	23	8	0	0	0	0	0
	Predicted	7439	1429	274	53	10	2	0	0	0	0	0	0
	Reverse	7712	1367	252	53	11	1	0	1	0	0	0	0
Grand Total	Direct	22901	5744	1660	433	186	59	18	4	1	0	0	1
	Predicted	21249	4083	783	150	29	6	1	0	0	0	0	0
	Reverse	21759	3876	802	147	22	8	2	2	0	0	0	0

Table 7.5

SUMMARY FOR ALL EXPERIMENTAL SERIES:

Average Score (per 25) for Successive Five Card Groups

Subjects	Direct I	II	III	IV	V	Reverse I	II	III	IV	V
Low Scorers	5.16	5.10	5.02	5.17	5.21	4.96	4.84	4.92	4.98	4.85
High Scorers (minus C.J.?)	6.10	6.14	6.20	6.20	6.06	5.10	4.95	5.12	4.98	4.91
Total (minus C.J.)	5.61	5.60	5.58	5.66	5.61	5.03	4.89	5.01	4.98	4.93
Mr. C.J. (minus EE)	8.00	7.18	6.98	6.43	6.35	4.97	5.11	5.24	5.09	4.84
Mr. C.J. EE	8.39	5.25	5.48	5.54	6.50	5.03	4.89	4.93	4.85	5.11
Individual Total	6.50	6.05	6.52	5.88	5.90	5.01	4.96	5.08	5.00	4.92
Class	4.97	4.94	5.06	4.94	5.07	4.87	4.94	4.92	5.00	5.00
Grand Total	6.18	5.80	5.79	5.67	5.71	4.98	4.95	5.04	5.00	4.94

Table 7.5A

SUMMARY FOR ALL EXPERIMENTAL SERIES:

Hit Frequency with Respect to Card Position (Direct)

Card	Low Scorers	High Scorers	Total minus Mr. C.J.	Mr. C.J. (minus EE)	Mr. C.J. EE	Individual Total	Class	Grand Total
1	696	738	1434	950	251	2635	583	3218
2	657	705	1362	985	283	2630	584	3214
3	596	642	1238	978	237	2453	525	2978
4	629	678	1307	920	250	2477	602	3079
5	640	714	1354	821	170	2345	527	2872
6	606	681	1287	873	155	2315	551	2866
7	665	727	1392	816	153	2361	608	2969
8	633	685	1318	825	142	2285	544	2829
9	626	719	1345	827	161	2333	558	2891
10	655	690	1345	835	135	2315	547	2862
11	600	700	1300	844	161	2305	595	2900
12	611	740	1351	844	160	2355	551	2906
13	641	699	1340	810	143	2293	578	2871
14	674	701	1375	762	148	2285	566	2851
15	607	697	1304	802	166	2272	586	2858
16	628	681	1309	772	145	2226	584	2810
17	649	727	1376	772	149	2297	567	2864
18	664	723	1387	746	157	2290	548	2838
19	635	677	1312	708	167	2187	568	2755
20	647	726	1373	743	169	2285	537	2822
21	665	739	1404	702	183	2289	543	2832
22	671	705	1376	766	177	2319	593	2912
23	671	673	1344	755	190	2289	548	2837
24	604	627	1231	756	188	2175	581	2756
25	641	709	1350	715	185	2250	613	2863

Table 7.5A

SUMMARY FOR ALL EXPERIMENTAL SERIES:

Hit Frequency with Respect to Card Position (Reverse)

Card	Low Scorers	High Scorers	Total minus Mr. C.J.	Mr. C.J. (minus EE)	Mr. C.J. EE	Indi-vidual Total	Class	Grand Total
1	588	608	1196	551	146	1893	538	2431
2	635	590	1225	601	136	1962	557	2519
3	606	596	1202	581	138	1921	586	2507
4	643	546	1189	574	153	1916	535	2451
5	626	568	1194	583	141	1918	548	2466
6	619	573	1192	601	140	1933	559	2492
7	625	547	1172	569	140	1881	576	2457
8	585	610	1195	624	139	1958	564	2522
9	594	544	1138	582	138	1858	521	2379
10	594	550	1144	597	137	1878	586	2464
11	616	586	1202	581	141	1924	553	2477
12	599	532	1131	565	121	1817	544	2361
13	641	699	1340	810	143	2293	580	2873
14	595	551	1146	525	154	1825	557	2382
15	618	551	1169	569	141	1879	561	2440
16	632	599	1231	587	147	1965	535	2500
17	613	525	1139	590	132	1860	558	2418
18	653	578	1231	578	135	1944	602	2546
19	599	569	1168	599	138	1905	571	2476
20	612	566	1178	606	136	1920	576	2496
21	620	558	1178	540	142	1860	554	2414
22	631	541	1172	577	136	1885	624	2509
23	536	584	1220	570	166	1956	544	2500
24	623	547	1170	567	130	1867	567	2434
25	577	570	1147	565	151	1863	553	2416

VARIATIONS OF TIME INTERVALS IN PRE-SHUFFLE CARD-CALLING TESTS*

Lois Hutchinson
Parapsychology Laboratory, Duke University

INTRODUCTION

History of the Problem. The hypothesis of precognitive ESP states that future events of a random character can be identified to some degree by means of ESP. Historically, this hypothesis originated outside the psychological laboratory. The literature of the societies for psychical research, of which the report of Saltmarsh[1] is typical, contains many accounts of dreams and waking spontaneous experiences of presumably precognitive character. But the problem of precognition is also one which arises inferentially from the results of laboratory studies of ESP.

As a matter of fact, the problem of the occurrence of precognition may clearly be subjected to experimental test, and the results of some laboratory studies have already been reported. As a first step toward testing the hypothesis, Rhine[2] had subjects make calls for a to-be-determined sequence of events that, so far as could be judged, could not be perceived sensorially, inferred rationally, nor brought about by the subject deliberately. The tests, "pre-shuffle card-calling," were made with ESP cards; the subject was asked to predict the order a pack would have after it had been shuffled by the experimenter. These tests yielded results which deviated significantly from chance expectation, but the interpretation of their bearing upon the precognition

---

* A thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts in the Graduate School of Arts and Sciences of Duke University (1940). The writer is indebted to Dr. J. B. Rhine for his guidance of the research, and to Dr. J. G. Pratt for his many helpful suggestions and painstaking supervision of this study. The assistance of other persons in the experiments is mentioned in the text and is gratefully acknowledged.

1. Saltmarsh, H. F., "Report on Cases of Apparent Precognition," *Proceedings of the S. P. R.*, 1934, Vol. XLII, p. 49.
2. Rhine, J. B., "Experiments Bearing on the Precognition Hypothesis," *Journal of Parapsychology*, March, 1938, Vol. II, p. 38.

hypothesis has since been qualified by the results of later tests, the "ESP shuffle." Rhine, Smith, and Woodruff[3] reported that subjects could apply ESP to the task of shuffling a pack of cards. This suggests that the person shuffling the cards in the pre-shuffle card-calling tests might have used ESP to effect better-than-chance scores, an explanation which would make the precognition hypothesis unnecessary for the interpretation of the results. The ESP shuffle work emphasized the importance of using an impersonal procedure to "randomize" the cards in any adequate test of precognition.

Two other laboratory tests concerning the hypothesis of precognition are those of Carington[4], who asked his subjects to predict the faces that would turn up on the throw of dice, and of Tyrrell[5], who used an electric ESP machine. In Tyrrell's experiment the subject was asked to open one of five boxes which would be mechanically selected to be lighted by an electric bulb following the response of the subject. Although both experimenters state that their results are relevant to the problem of precognition, these results cannot be considered as conclusive. Carington states, "It is clear that these 40 subjects *taken as a group*, afford no justification for supposing that any precognitive faculty is at work, for the total number of successes is exactly equal to expectation." His conclusion that his results are significantly above expectation is made on the basis solely of a selection of individual subjects who scored above and below chance expectation. Such procedure is subject to the counter-hypothesis of selection of data, and the results must be discounted accordingly. Tyrrell's results, while indicating ESP, are not necessarily interpreted as of precognitive character because (1) the adequacy of the random character of the mechanical selector has not for the series in question been fully determined, and (2) the results could be interpreted as the outcome of the subject's clairvoyant knowledge of the mechanical selector's direction of choice.

---

3. Rhine, J. B., Smith, Burke M., and Woodruff, J. L., "Experiments Bearing on the Precognition Hypothesis: II The Role of ESP in the Shuffling of Cards," *Journal of Parapsychology*, June, 1938, Vol. II, p. 119.
4. Carington, Whately, "Preliminary Experiments in Precognitive Guessing," *Journal of the Society for Psychical Research*, June, 1935, Vol. XXIX, p. 86.
5. Smith, Burke M., "The Tyrrell Experiments," *Journal of Parapsychology*, March, 1937, Vol. I, p. 63.

Scope of the Research. In the experiments presented here, the subjects were asked to call decks of ESP cards as they would be arranged at some future time, with random rearrangement intervening. A mechanical shuffler, which has been described in detail by Smith[6], was rotated a fixed number of times to determine the random arrangement of the cards.

Two time intervals are contrasted: (1) calling for checking with the random rearrangement taking place within twenty-four hours, (2) calling for delayed checking, with random rearrangement taking place ten days after the calls are made.

In a first phase of the investigation, to be designated as Series I, the subjects were not told any of their scores until the last results were checked, twenty days after the experiment began. In the second phase, Series II, the subjects were told their scores daily for the ONE-DAY check group, but did not, of course, learn their scores on the TEN-DAY check series until at least ten days later. One aspect of the investigation will be concerned with the relative effects of telling the subject his scores at these two intervals after the time of making the calls.

Statement of Problems. This report will be concerned with three problems: (1) Can calling the order of cards with random rearrangement intervening be demonstrated by normal subjects to an extent greater than chance expectation? (2) If so, how does success at calling the order of cards one day ahead compare with calling the order for ten days ahead? (3) If these two intervals of time affect the scores differently, how is this difference to be explained?

PROCEDURE

General Statement. The experimental procedure involved the use of the text known as "precognitive down through," or PDT, in which the subject predicts the order of cards in ESP decks as they will be after they are shuffled some time after the responses are made.

There were in reality two experiments made with this general procedure. The two experiments will be designated, in the order in which they were performed, as

6. Smith, Burke M., "Note on a Shuffling Machine," *Journal of Parapsychology*, September, 1938, Vol II, p. 231.

Series I and Series II, respectively. As Series II was planned as a separate experiment involving certain conditions different from those of Series I, it will be necessary to describe the two procedures and to consider the results of each part separately. The necessity for this procedure will become evident as the plan of the investigation is explained.

Procedure for Series I. Sixteen subjects, all graduate women students residing in Faculty Apartments, took part in Series I.

In preparation for the experiment, the investigator first wrote out a description of the purpose and plan of the experiment, stating exactly the number of days the experiment would run and estimating that between fifteen and twenty subjects would be used. This prefatory statement was filed with the record librarian of the Parapsychology Laboratory, who then issued to the experimenter the number of record sheets which would be required. These sheets were numbered serially and each one was stamped with the seal of the Laboratory, which was accessible only to the librarian. Record of the serial numbers and of their intended use was kept, and no other similar sheets have been issued with any of the same numbers. The serial numbers therefore served as a convenient means of identifying the data sheets at all times, and the degree to which the numbers are all represented among the "raw data" sheets of the completed experiment is an objective criterion for judging the likelihood of the selection of data.

At the beginning of the series, the experimenter saw each of the subjects personally, gave to each one ten of the record blanks, and told her the purpose of the experiment and the procedure to be followed in making the calls. Written instructions were not read, but the investigator was careful that each subject understood what she was to do and the procedure that was to be followed. Each subject was told that sometime on the evening of February 26, and on each evening thereafter for ten days in all, she was to predict in the ten CALL columns of the appropriate record sheet the order of ten decks of ESP cards as they would be after being shuffled later. The correct order of cards would be recorded in the adjacent CARD column at the time appointed for checking the scores; if the subject preferred, she could take as her task to predict for each space in the CALL column what would later be entered in the CARD space just to the right. The subject was told that the first five runs on each sheet would be filled in and the scores checked within twenty-four hours of the time of making the calls, while the cards for the second five runs would not be shuffled until ten days had elapsed. In order to make all conditions except the varied time internals as similar as possible, the experimenter asked the subjects to fill out the first column on the sheet, then the sixth--which was, of course, the first for the tenth-day check--the second, then the seventh, etc.

The subjects were advised that they would receive for the two conditions similar cash rewards for high scores, according to the following schedule:

$ .05	for	a	score	of	8
.10	"	"	"	"	9
.25	"	"	"	"	10
.50	"	"	"	"	11
1.00	"	"	"	"	12
2.50	"	"	"	"	13
4.00	"	"	"	"	14

They were also informed that they would not be told any of their results until the experiment was completed and checked, that is, at the end of the twentieth day.

The experimenter collected the record sheets from the subjects each morning and turned them over to Mr. E. P. Gibson, who had charge of shuffling the cards. Five separate decks of ESP cards were shuffled on the day following the making of the calls by placing each in the shuffling machine which was rotated exactly three times. The decks were recorded with the cards in the order in which they came from the shuffler, the first deck shuffled as run number 1, the second as number 2, and so on. All the subjects' responses were checked against the same set of five decks. The cards were recorded in the first five CARD columns of each record sheet by Mr. Gibson, using a gelatin duplicator. No sheets which were not stamped at the proper time by the librarian are counted in the totals.

The librarian then gave the stamped sheets for each day to Miss Maxine Slaughter*, who without marking the sheets in any way, counted the number of hits by comparing the entries under the adjacent CALL and CARD columns for the first run, then for the second, and so forth. The experimenter also checked the number of hits, likewise without marking the data sheets in any way, and

* Graduate student in the Department of Psychology, Duke University.

the two records were kept separately. At the conclusion of the experiments, all the records received an independent check by Miss Esther Bond or Miss Maxine Slaughter, and any discrepancies between their scores and those of the experimenter were resolved by carefully consulting the record sheet. Discrepancies were found to have arisen predominantly through overlooking hits, a finding in keeping with the experience of Greenwood[7] in checking and re-checking 20,000 runs. The record sheets of this experiment, still unmarked, are open to further independent check.

At the proper interval for the tenth-day check, the record sheets were again given to the librarian and decks for the second five runs of each sheet were shuffled in the machine as already described and the cards were recorded on each sheet by means of the duplicator. The results were checked and independently re-checked as for the ONE-DAY check series.

All computations entering into this report were made twice. Miss Esther Bond and the experimenter made independent computations, and in case of discrepancies the work was re-checked until complete agreement was obtained.

Procedure for Series II. Twenty-four subjects (9 graduate women students, 5 graduate men students, one undergraduate woman, 2 undergraduate men, one college instructor, one school teacher, one housewife, one mechanical engineer, and 3 secretaries) took part in Series II. The procedure and conditions were the same as in Series I, with the following exceptions:

(1) Instead of making their predictions on ten successive days as previously, the periods of making the calls were divided into two groups of five days each, the groups being separated by an interval of three days' recess. These two parts of the experiment will be designated II-A and II-B, respectively. Series II-A, beginning on the evening of April 9, involved eighteen of the above twenty-four subjects. In Series II-B, which started on April 17, twelve subjects continued from II-A and six additional ones joined the experiment.

(2) Instead of waiting until the last results were checked to hear of their scores, the subjects were told that they would be informed from day to day what

7. Greenwood, J. A., "Analysis of a Large Chance Control Series of ESP Data," *Journal of Parapsychology*, June, 1938, Vol. II, p. 141.

scores they made on the first half of each sheet, that is, the half for the ONE-DAY check. They could not, of course, learn their scores on the TEN-DAY check series until ten days after the calls were made.

(3) The subjects were not specifically told in what order to fill in the CALL columns of their record sheets. Five of the subjects who had taken part in Series I are known to have followed the staggered order used then, but it may be supposed that most of the sheets were filled out in regular order from left to right. That is, the five ONE-DAY check runs were recorded in the first half of each period and the five TEN-DAY check runs in the second half. The possible effect of this order of making the calls will be taken into account in discussing the probable significance of the difference in results between ONE-DAY and TEN-DAY checking.

RESULTS

Results of Series I (Knowledge of scores withheld until end of experiment). There were 1430 runs made in Series I. Of this number, 715 were checked within the same twenty-four hour period and are grouped as ONE-DAY check, while 715 were checked ten days after the sheets were filled out and are grouped as TEN-DAY check. The total deviation for the 715 runs of the ONE-DAY check group is -4; the total deviation for the 715 runs of the TEN-DAY check group is +28. The results of both groups are not significantly different from chance expectation; nor is there any unusual difference between the ONE-DAY and the TEN-DAY checked groups (See Table 1).

Although the results of Series I were inconclusive, one factor of interest emerged. A comparison of the total deviations of each subject for the ONE-DAY checked runs with her total deviation for the TEN-DAY checked runs indicated that 13 pairs of the 16 deviation signs were identical. That is, if a subject had a negative deviation for 50 runs on the ONE-DAY checked series, more often than not her deviation for the 50 runs on the TEN-DAY checked series would also be negative. A binomial evaluation of this result in the experiment gives a probability value of .01, suggesting that the identity of direction of the 13 out of 16 pairs of deviations in a group of unselected subjects may not have been wholly due to chance (see Table 1). In any event, an evaluation based on so few data would be suggestive rather than conclusive.

TABLE 1

Results of Series I showing the subjects used, number of runs of each subject, deviations from chance expectation, and totals.

	ONE-DAY Check Series		TEN-DAY Check Series	
Subjects	Runs	Deviations	Runs	Deviations
C.B.	50	+21	50	- 8
E.B.	15	+ 2	15	+10
H.B.	45	- 8	45	-17
A.F.	50	-14	50	+10
M.H.	50	+ 3	50	- 1
I.M.	50	+ 1	50	+11
N.M.	50	-12	50	- 3
H.R.	50	+ 4	50	+23
J.R.	30	-21	30	-16
M.S.	45	+18	45	+12
E.S.	40	-12	40	- 1
L.T.	40	+ 7	40	+12
E.T.	50	+19	50	+18
G.V.	50	-13	50	-13
H.W.	50	+12	50	+ 5
B.W.	50	-11	50	-14
Total	715	- 4	715	+28

S.D.: 715 runs = 53.48
C.R. = 0.07

S.D.: 715 runs = 53.48
C.R. = 0.52

S.D.: 1430 runs = 75.63
C.R. = 0.42

Results of Series II. In Series II the total number of runs that were checked within the same twenty-four hour period in which the subjects filled out the sheets, equals 835 and has a positive deviation of 180. The average number of hits per run for this group is 5.22.

For the group that was checked ten days after the date the sheets were filled out, there are 835 runs with a negative deviation of 82, and an average number of hits per run of 4.90. (These results are shown in Table 2.)

Series II-A included 850 runs, 425 of which were dhecked one day after, and 425 of which were checked ten days after this date. The series checked after one day

TABLE 2

Results of Series II, showing the subjects used, number of runs of each subject, deviations from chance expectation, and totals.

	ONE-DAY Check Series		TEN-DAY Check Series	
Subjects	Runs	Deviations	Runs	Deviations
A.D.	50	+ 20	50	- 35
B.T.	25	- 18	25	+ 1
D.C.	40	+ 4	40	- 4
D.F.	50	+ 1	50	+ 3
P.G.	25	+ 1	25	- 12
S.M.	25	- 10	25	- 1
A.F.	50	+ 34	50	- 7
M.B.	50	+ 18	50	+ 8
R.P.	25	+ 12	25	- 1
J.G.	50	+ 11	50	+ 5
B.S.	50	+ 17	50	- 3
E.H.	50	+ 10	50	- 18
J.W.	50	- 2	50	+ 11
G.Z.	45	+ 18	45	- 1
E.S.	40	+ 15	40	+ 4
I.M.	45	+ 5	45	- 2
P.S.	25	+ 2	25	+ 7
B.C.	20	+ 3	20	- 2
H.W.	25	+ 16	25	- 19
H.R.	20	+ 6	20	- 3
A.A.F.	20	+ 15	20	- 11
E.T.	25	- 7	25	- 13
M.S.	25	+ 8	25	+ 10
L.P.	5	+ 1	5	+ 1
Total	835	180	835	- 82

has a positive deviation of 99, and the average number of hits per run is 5.23. The series checked after ten days has a negative deviation of 13, and the average number of hits per run is 4.97 (see Table 3).

Series II-B included 820 runs, 410 of which were checked one day after and 410 of which were checked ten days after the date they were filled out. The series checked after one day has a positive deviation of 82, and the average number of hits per run is equal to 5.197. The series checked after ten days has a negative deviation of 66, and the average number of hits per run is 4.83 (see Table 3).

TABLE 3

Series II: Showing runs, deviations, average hits per run for A and B under the two conditions.

	ONE-DAY Check			TEN-DAY Check		
	Runs	Deviation	Ave. Hits per Run	Runs	Deviation	Ave. Hits per Run
Series II-A	425	+99	5.23	425	- 13	4.97
Series II-B	410	+81	5.197	410	- 69	4.83
Totals	835	+180	5.22	835	- 82	4.90

C.R. = 3.11 C.R. = 1.42

C.R. of difference (ONE-DAY and TEN-DAY) = 3.21

The critical ratio for the ONE-DAY checked series of 835 runs is 3.11. The P value for a chi-square treatment is .035. (See frequency distribution, Table 4.) The critical ratio for the TEN-DAY checked series of 835 runs is -1.42. The P value for a chi-square treatment is .35.

In Table 4 it will be seen that a score of 14 occurs; the binomial expectation of this is .000,063. Also there are 13 scores of 11 or larger. The critical ratio of the deviation of these scores from the binomial hypothesis is 3.9. These items are of interest in supporting the deviation-critical ratio evaluation, even though they are obviously selected.

For the sake of completeness, it may be of interest to pool the results of Series I and Series II. The total 3100 runs have a total positive deviation of 122, which is insignificant (C.R. = 1.10). The difference in experimental conditions between the two series and between the ONE-DAY and TEN-DAY sub-divisions warrant the separate analysis given above. If a combination evaluation is desired, the chi-square method is the only one affording due distinction to the separate divisions. Using this method for the four subdivisions, ONE-DAY and TEN-DAY, both for Series I and II, a chi-square of 12.0 is obtained, which with 4 degrees of freedom gives a probability of .017.

A safe and approximate probability of the occurrence of a block of results, such as that of the Series II, ONE-DAY tests which gave a critical ratio of 3.11, may be found simply by multiplying the probability represented by 3.11, which is .00094, by 4 (since there are 4 such subdivisions in the report). The result, .00376, is significant; it represents the probability

TABLE 4

Frequency distribution of run scores for Series II - ONE-DAY Check - by individual subjects and with the total distribution for the experimental series.

Subjects	Frequency of Run Scores															Runs and Dev.
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
A.D.	0	1	3	3	14	9	6	6	2	3	1	2	0	0	0	50: + 20
B.T.	0	1	3	4	7	4	4	0	2	0	0	0	0	0	0	25: − 18
D.C.	0	1	2	6	7	8	5	7	2	2	0	0	0	0	0	40: + 4
D.F.	0	0	4	11	6	9	6	8	5	1	0	0	0	0	0	50: + 1
P.G.	0	2	2	3	2	5	5	3	1	1	1	0	0	0	0	25: + 1
S.M.	0	1	0	8	4	5	3	2	1	1	0	0	0	0	0	25: − 10
A.F.	0	2	2	6	6	10	3	12	4	2	0	2	1	0	0	50: + 34
M.B.	0	0	4	4	8	12	7	8	5	2	0	0	0	0	0	50: + 18
R.P.	0	2	1	1	4	5	4	3	3	1	0	1	0	0	0	25: + 12
J.G.	0	0	2	5	11	11	10	6	4	1	0	0	0	0	0	50: + 11
B.S.	0	1	3	8	7	6	9	9	4	1	2	0	0	0	0	50: + 17
E.H.	0	0	5	9	11	8	5	1	4	3	1	3	0	0	0	50: + 10
J.W.	0	0	4	8	10	11	6	5	4	2	0	0	0	0	0	50: − 2
G.Z.	0	1	1	3	6	14	8	9	2	0	1	0	0	0	0	45: + 18
E.S.	0	0	6	2	5	9	7	5	2	1	3	0	0	0	0	40: + 15
I.M.	0	2	2	2	11	8	12	3	4	1	0	0	0	0	0	45: + 5
P.S.	0	1	2	3	4	6	2	2	4	1	0	0	0	0	0	25: + 2
B.C.	0	0	1	5	2	4	5	1	1	0	0	0	0	1	0	20: + 3
H.W.	0	0	1	5	7	1	3	2	1	3	1	0	0	0	1	25: + 16
H.R.	0	2	0	3	3	3	3	2	1	2	1	0	0	0	0	20: + 6
A.A.F.	0	1	1	1	5	2	4	0	3	0	2	1	0	0	0	20: + 15
E.T.	0	0	3	3	6	4	7	1	0	0	1	0	0	0	0	25: − 7
M.S.	0	0	2	5	5	3	3	1	3	1	1	1	0	0	0	25; + 8
L.E.	0	0	0	0	1	2	2	0	0	0	0	0	0	0	0	5: + 1
Total	0	18	54	108	152	159	129	96	62	29	15	10	1	1	1	835: +180

χ^2 = 17.60 with 9 d.f.; P = .03517

that at least one of the 4 sections of results would have given a positive critical ratio by chance as large as 3.11.

A COMPARISON OF SERIES I WITH SERIES II

Subjects. Sixteen graduate women subjects took part in Series I. Of this number, only four had ever been subjects in any ESP experiment before. For series II-A, eighteen subjects, all new, were used. Series II-B included twelve subjects from Series II-A, five subjects from Series I, and one new subject. The entire group of subjects makes a total of thirty-five. They form a fairly homogeneous group in most respects, since there are no wide deviations in age, occupations, and cultural standing.

Material. The material, standard ESP record sheets and cards, were identical in the two studies, as was the recording and scoring of the data. Subjects were given the record sheet which they filled out and returned to the experimenter, who then took them to the record librarian.

Order of Subjects' Calls. In Series I the subjects were explicitly asked to fill out the CALL column under number 1, then the CALL column under number 6, the CALL column under number 2, then number 7, etc. The experimenter overlooked this point in Series II.

Length of Series. In Series I, the subjects were asked to do ten runs of twenty-five trials for each of ten consecutive days.

In Series II, the subjects were asked to do ten runs of twenty-five trials for each of five consecutive days. After an interval of three days, Series II-B was begun. The subjects were asked to do ten runs of twenty-five trials for each of five consecutive days. Thus, in Series II the ten runs were divided into two groups of five consecutive days each with a three-day interval between, but in Series I the ten runs were done on ten consecutive days.

The Subject's Knowledge of his Scores. The subjects in Series II were told daily of their scores and rewards, but a condition of Series I was that the subjects should not know their scores until the completion of the experiment. This meant that they were told nothing of the experiment except for their initial instructions until the twentieth day after they began filling out the ESP record sheets.

Rewards. The rewards were identical for both studies, except for the effect of nearness. In Series II the subjects knew their scores on the ONE-DAY tests within 24 hours, and accordingly the rewards were less remote than in the TEN-DAY runs and still less so than in Series I.

The Subject-Experimenter Relationship. The subject-experimenter relationship was similar in both studies so far as the instructions and the tasks are concerned, but there was more freedom and the contact was greater in Series II because the conditions permitted discussion of the daily scores, etc.*

Table 5 is a summary of the comparison of Series I and Series II.

TABLE 5

A Summary of the Comparison of Series I and Series II

	ONE-DAY Check: Subjects told daily scores			ONE-DAY Check: Subjects not told scores until end of experiment			TEN-DAY Check: Subjects not told scores until end of experiment		
	No. Runs	Dev.	Ave. Hits Per Run	No. Runs	Dev.	Ave. Hits Per Run	No. Runs	Dev.	Ave. Hits Per Run
Series I				715	- 4	4.99	715	+28	5.04
Series II	835	+180	5.22				835	-82	4.90

C.R. - Series I ONE-DAY Check = -0.07
C.R. - Series I TEN-DAY Check = 0.52
C.R. diff. = 0.42

C.R. - Series II ONE-DAY Check = 3.11
C.R. - Series II TEN-DAY Check = -1.42
C.R. diff. = 3.21

DISCUSSION

It was stated that three problems would be considered in this study. They are: (1) Can pre-shuffle card-calling be demonstrated by normal subjects to an extent greater than chance expectation? (2) If so, how does success at calling the order of cards one day ahead compare with calling them ten days ahead? (3) If these two intervals of time affect the scores differently, how is this difference to be explained? These problems will be considered in the order listed. Finally, the suggestion of conditions probably favorable to the testing of the ESP hypothesis, and a consideration of further experimentation as an outgrowth of this study will be discussed.

* Dr. Rhine reminds the experimenter that she had wanted to do the experiment as in Series II at the start but had been urged to do it as in Series I to insure equal motivation for the ONE-DAY and TEN-DAY tests. Her own interest was obviously greater for the Series II method.

The Demonstration of ESP in Pre-shuffle card-calling Tests. Any consideration of this problem involves an examination of the counter-hypotheses. For discussion, the counter hypotheses have been grouped into four broad categories: (1) the results are accidental, (2) sensory perception was not wholly eliminated, (3) the results are due to errors in the scoring and recording, and (4) the results are due to the incompetence and untrustworthiness of the experimenter and others involved in the recording and scoring. The first of these is concerned entirely with the statistical evaluation of the results; the other three raise questions about the adequacy of the experimental conditions and the competence of the experimenter.

(1) When the commonly accepted measures of probability are applied to the data designated as the ONE-DAY Check, under a chance hypothesis such results will occur but one time in a thousand. The repetition of this experiment with a critical ratio of 3.21 of the difference between the ONE-DAY Check and the TEN-DAY Check series by chance alone is likely to happen only 7 times in 10,000. As already stated, there was no selection of the data. All the record sheets used in the two series are reported in this study with the exceptions noted.[8] All mathematical evaluations were independently checked.

8. In order to make certain that there is no selection of data, every record sheet issued by the librarian to the experimenter for this experiment must be accounted for. In those cases in which the subject failed to fill out his sheet, the blank records were returned to Mr. Gibson. Several of the record sheets were filled out by the subjects but were turned in too late to be stamped by Mr. Gibson. The experimenter has kept a record of all these sheets, even though it was stated that no sheets except those which were stamped on time by Mr. Gibson are considered in the total. The following table shows by subjects, series, date, runs, and deviations the record sheets which were filled out, but not stamped.

Series	Date	Subject	ONE-DAY Check	Dev.	TEN-DAY Check	Dev.
II-A	April 10	L.E.	6 0 3 7 3		5 6 3 3 6	
II-A	April 11	L.E.	7 4 4 4 5	-13	11 3 3 4 1	-1
II-A	April 12	L.E.	7 4 5 0 3		5 9 7 7 2	
II-A	April 13	L.E.	4 8 3 6 4		3 7 3 3 5	
II-A	April 13	G.Z.	5 5 2 8 4	- 1	6 4 5 4 5	-1
II-B	April 21	B.C.	5 11 6 4 2	+ 3	5 4 4 8 7	+3
II-B	April 21	E.S.	3 6 3 2 2	- 8	. . . . .	..

In order to obtain a series of scores to serve as a control on the chance factors inherent in the experiment, with the experimental factor (ESP) eliminated, a cross-check was made by checking the subjects' calls against card records of two days ahead--records they were not intended to match. This was done for the 835 runs of the ONE-DAY part of Series II. An average score of 4.9952 was obtained, and a critical ratio of .07. The chi square of the score frequency distribution gives a probability of .63. The standard deviation of the cross-check distribution is 1.907, which is not significantly different from the theoretical value of 2.0.

From this it may safely be inferred that inadequate shuffling, chance coincidence of card sequences with subject preferences, or habit patterns, and any other factor associated with the randomness of the test material can safely be dismissed as explanatory hypotheses in this instance. This position is strengthened considerably by the many cross-check analyses of similar character made by earlier experimenters, all leading to similar results.[9] Likewise, subjects could not have discovered any possible pattern trends in the cards, had there been any, and have followed them, since they did not see either the cards or the records at any time.

There was no improper selection in the determination of the stopping point of the experiments. The final subseries II-B had been begun with a definite assignment of record sheets, and when this was finished, there was no further time for continuance before writing and submitting the report to the Graduate Office in time for acceptance.

(2) The procedures of the experiment required that the subject fill out his individual record sheet before the card orders against which the calls could be checked were determined. Therefore, it is clearly impossible that the results of this experiment are explainable by the hypothesis of sensory perception of the subjects.

(3) The stamping of the CARDS in columns adjacent to the CALLS by the librarian, and the independent checking and recording of the hits should be sufficient to eliminate the hypothesis of errors in scoring and recording as a possible explanation of the results.

---

9. Greenwood, J.A., and Stuart, C.E., "Reply to Dr. Feller's Critique," see page , this issue of the *Journal of Parapsychology*.

(4) It did not occur to the investigator at the time of planning and conducting the experiment as necessary to assume that any of the persons participating in the recording and checking could be suspected of dishonesty, inasmuch as the record sheets were at different times entirely within the keeping of each of the three persons concerned. Therefore, it is obvious that outright dishonest practices with regard to the data could have been possible. But certain features of the procedure make it highly unlikely that unconscious errors in recording or the loss of data of a "motivational" character could have occurred. The first such feature was the use of numbered record sheets. All but eight of such sheets used by the experimenter in any experiment by her have been returned and acknowledged by the record librarian. Those eight sheets have been accounted for by the subjects to whom they were issued in statements filed with the librarian. A second feature was the procedure of using a hectograph stamping of all the sheets in this experiment by the librarian. A third feature was his use of a shuffling machine rotated a fixed number of times to determine the order of the cards. A fourth feature was the separate recording of the run scores and deviations by Miss Maxine Slaughter and the experimenter. A fifth feature was the independent computation made by Miss Esther Bond and the experimenter.

There will be, no doubt, other objections than the four main hypotheses usually brought against extrasensory perception to the acceptance of the precognition hypothesis. For instance, it may be suggested that the mechanical shuffler does not wholly eliminate the ESP shuffle. However, since the machine was rotated a fixed number of times for each deck (3 times), this counter-hypothesis in its usual form is not applicable. So far as can be judged, if extra-sensory perception did take place, it must of necessity be explainable by hypotheses other than those of clairvoyance and telepathy, which were eliminated by the conditions. The evidence suggests the validity of the hypothesis of precognition, i.e. that significant pre-shuffle card-calling of extra-sensory and non-inferential character has occurred, the calling corresponding to random future events; but no final conclusions can be reached without much more evidence.

If, however, we accept the results at their face value for the moment and assume the hypothesis of precognition to be demonstrated, what inferences as to the nature of this function are we able to make? Has anything been learned regarding the nature of so-called precognition? In view of the fact that the writer considers that the results are of a suggestive rather than a conclusive nature, conclusions to be reached regarding secondary problems must be even more tentative. One finding, at least, relates pre-shuffle card-calling to other modes of ESP: methods successfully used in the demonstration of telepathy and clairvoyance may be used to measure pre-shuffle card-calling. The DT, or calling "down through" a pack of cards method was utilized in this experiment with only the addition of the condition that the key decks against which the calls were checked were not to be arranged nor recorded until some time after the calls were made.

How does the use of different time limits in the same experimental situation affect pre-shuffle card-calling? It has been stated by Rhine[10] that precognition is a logical inference arising from distance experiments in ESP. That is, when distance has been a condition in ESP experiments, there has been no critical effect of space shown in the results. Since time measurements are made in reference to spatial relations, and our conceptions of time and space are inseparable, the probability would seem to be that distance in time would likewise not affect ESP scores one way or the other.

At first glance, the results obtained in Series II seem to show that pre-shuffle card-calling by subjects to an extent greater than chance expectation can be demonstrated as taking place within a twenty-four hour limitation, but will not extend as far in time as ten days. If this were true, it would indeed be a fact of great importance about the nature of precognition; a fact that effectively contradicts our logical inference regarding space and time. However, a closer examination of the results suggests that such an interpretation may be incorrect. Possibly the explanation of the difference might be found in the consideration of the subjects and their reactions to the different testing conditions for the demonstration of ESP ability. An expansion of this argument follows in the next two divisions.

<u>The Contrast of Difference of Time Intervals in Checking</u>. The results of Series I gave no significant difference for the ONE-DAY and TEN-DAY checks. The

10. Rhine, J.B., "The Effect of Distance in ESP Tests," *Journal of Parapsychology*, September, 1937, Vol. 1, p. 184.

ONE-DAY checked series of 715 runs gave a deviation of -4; the TEN-DAY checked series of 715 runs a deviation of +28. In fact, the results of the series as a whole were clearly explainable by chance, comparable to the TEN-DAY check half of Series II.

A striking difference is indicated in the results of Series II. The series of 835 runs designated as ONE-DAY check has a critical ratio of 3.11, and the average number of hits per run is 5.22, while the TEN-DAY check series of an equal number of runs gave a negative critical ratio of 1.4 and the average number of hits per run is 4.90. This difference between the ONE-DAY and TEN-DAY checking intervals yields a critical ratio of +3.21. (See Table 5.)

The Interpretation of the Results Obtained in the Contrasted Time Intervals of Series II. A suggestion was made that the difference between the ONE-DAY and TEN-DAY groups of Series II might be due to the fact that the five calls for the ONE-DAY group were made first in most instances while those of the TEN-DAY group followed. A difference in procedure between Series I and Seriee II makes it possible to use the results of Series I as a control for the effect of the order of making the calls in Series II.

In order to test whether the sequence of calls did affect the results, an investigation of the results according to the order of the subject's calls for Series I was made.

As has been stated, in Series I the subjects made alternate calls for the ONE-DAY and TEN-DAY conditions. A comparison of the first five runs made by the subject and his second five runs was made. The comparison was made by totaling all the runs of all subjects under columns 1-10, and comparing the sum of 1, 2, 3, 6, 7, selected because they were the first half of the runs made, with the sum of 4, 5, 8, 9, 10. The total hits of the first five runs (columns 1, 2, 3, 6, 7) filled out by the subject and disregarding the checking time division equals 3,512. For the second five runs the total of the hits equals 3,662. The difference of 150 (S.D. of the difference is 75.6; C.R. is 1.97) is not significant. It is concluded that there is no appreciable difference shown between a subject's first five runs and his second five, and that in Series II the difference must be due to some other condition.

Any explanation of the contrasted results of the time intervals in Series II may involve as well an explanation of the different results obtained in Series I and in Series II. As was outlined in the treatment of the results, the whole of Series I can be classed as delayed for the subjects in the same sense as the TEN-DAY series of Series II. That is, the subject's knowledge of his scores was delayed in both instances. It is suggested that he had no clear differentiation of the time intervals in Series I, since he knew nothing of his scores until all the data were in. This meant his knowledge of all his scores was delayed for a period of more than twenty days after Series I began. His knowledge of his daily scores in the TEN-DAY series was delayed for a period of ten or more days after Series II began. But in Series II, this lack of knowledge of his daily scores on the TEN-DAY series is in clear contrast with his daily knowledge of scores and rewards for the ONE-DAY check group. The results seem to suggest that the knowledge or lack of knowledge made for a differentiation in his mind of the time intervals in Series II -- a differentiation which was lacking in Series I -- and that the explanation of the difference in results may be traced to this factor.

A "negativistic" attitude of the subject towards tests in which delayed checking is a condition has been previously noted. Rhine[11] says, in the summary of his article "ESP Tests with Enclosed Cards," "It was found that there was a marked correlation between delay in the checking and decline in the scoring." This decline effect appeared in six of the seven test procedures used in the experiment.

The subject, his interests and motives, must be counted as variables in any psychological experimentation. It is believed that in the most successful ESP experiments, the subject should know exactly what he is doing and what his results are at the end of one run or one unit of runs. This is perhaps of particular importance in ESP testing because the subject cannot be said to possess much cognitive clarity about the very nature of his task. He is asked to call symbols spontaneously so that they will match a pattern in someone's mind, or on a slip of paper, or as they will be recorded at some future time on a slip of paper. His introspection while he works or afterwards tells very little about his

11. Rhine, J.B., "ESP Tests with Enclosed Cards," *Journal of Parapsychology*, September, 1938, Vol. II, p. 199.

conscious direction of his task, if indeed there can be said to be a conscious process important in ESP. It is certain that the subject feels the situation to be more real, and as possessing more ego-reference when he is daily told of his scores and encouraged to do well.

The results of Series II clearly support this line of reasoning. The runs which were immediately checked with the subject's knowledge of scores gave a positive deviation (835 runs + 180), an average number of hits per run of 5.22. Those which were checked after ten days and which possessed no immediate interest for the subject gave a negative deviation (835 runs -82), an average number of hits per run of 4.90. Experiments with sealed cards in which the subject's knowledge of his scores was delayed are cited in confirmation of this point.[12]

<u>A Consideration of Other Factors Favorable to the Testing of the ESP Hypothesis</u>. Other factors than the one of the subject's knowledge of his daily scores may have contributed to the difference in the results of Series I and II and are of importance in a consideration of further testing of the precognition hypothesis.

The experimenter believes that the effect of the subject's knowledge or lack of knowledge regarding his daily scores is the probable explanation, but desires to state certain other differences in the conditions of Series I and Series II which may have been factors: namely, (1) the length of the series in consecutive testing days; and (2) the subject-experimenter relationship.

(1) In Series I, the subjects were told the experiment would run ten days, and were given ten record sheets to fill out. Each record sheet represented 250 trials. Those subjects who did 100 runs -- who filled out a record sheet for each of the ten days -- made by the end of the experiment 2500 trials each. Since the experimental days were consecutive, it may be the series was too long for the best results to obtain, although there was no marked difference between the scores for the first day and those for the last in the series.

In Series II-A, the experimenter decided on the grounds of personal convenience and close supervision of the experiment that it would be preferable to reduce the consecutive den-day period to a five-day block. The

12. <u>Ibid</u>., p. 199.

subjects were asked to do ten runs of 25 trials each on record sheets for five consecutive days. The block began April 9 and continued through April 13. The checking of the TEN-DAY group, of course, did not begin until the tenth day from April 9, April 19. A second block, (II-B) was begun after an interval of three days. It ran from April 17 through April 21, with the checking of the TEN-DAY group beginning ten days from April 17 (April 27) and continuing through May 1.

(2) The experimenter feels that there was a marked difference in the kind of contact she had with the subjects during Series I and Series II. In all psychological research, the personal element assumes importance, and in an experimental task which involves two people, the relationship between the two is a factor which must be considered. It might be argued that the conditions of the pre-shuffle card-calling experiment eliminated that nebulous sort of relation we sometimes call "rapport," and that the reward scale was the important motivating force. The experimenter believes that the establishment of a certain mutual interest in the task and a feeling of unification is as important to the securing of high scores in pre-shuffle card-calling as in other branches of ESP research. It is her belief, as well, that a reward scale is merely an added incentive, and perhaps serves as a means of synthesizing and openly expressing the several motives in the situation.

The experimenter had little contact with most of the subjects in Series I beyond the initial instructions. In Series II the experimenter had daily contact with all but two or three of the subjects. It is her belief that the attitudes of the subjects toward the task were more favorable than those of the subjects in Series I.

Consideration for Further Experiments. The suggestion that the subject's interest and motivation arising from his knowledge of his scores and several other favorable factors are responsible for the difference in the results of the contrasted time intervals and make for more favorable testing conditions should be further elaborated experimentally.

Particular factors to be considered in the experiment to be outlined are: (1) the subject's cognitive clarity concerning the experiment, (2) the subject's interest and motivation, (3) the subject-experimenter relationship, and (4) the length of the series.

TENTATIVE CONCLUSIONS

The following tentative conclusions are made:

(1) That in the ONE-DAY check series of Series II there seemed to be a positive demonstration of extrasensory perception, and that the hypothesis of pre-cognition used to describe this demonstration warrants more research.

(2) That the difference in the results of the contrasted time intervals, in Series II, may be explained by the difference in interest and motivation of the subjects towards the two conditions.

(3) That the difference in the results of Series I and Series II may be due to the poorer differentiation in the mind of the subject of the one-day and the ten-day time intervals in the former, and the clearer contrast in the latter series.

(4) That the effect of telling the subject his daily scores probably has a direct relation to (a) the interest and motivation of the subject under the different conditions, and hence to (b) the difference obtained in the results of the contrasted time intervals.

Reno, Nevada.

STATISTICAL ASPECTS OF ESP[1]

Willy K. Feller
Brown University

1. Presumably many a student feels puzzled by the frequent controversies arising out of statistical investigations in various fields. In fact, statistics makes claim to mathematical rigor, and still its practical applications are often disputed or rejected as absurd.[2] I fear it may sometimes produce a feeling that mathematics is, after all, 'une de ces voies rationelles et scientifiques, qui n'est qu'une etroite, courte et sale impasse, au fond de laquelle on se casse le nez inglorieusement'.[3] The fact is that the importance and

---

1. Invited address, delivered at the Duke Mathematical Seminar, on April 24, 1940. The publication of the address was not originally intended, but was suggested both by the Mathematics Department of Duke University and by Professor Rhine. Professor Rhine and other interested persons have insisted that it would be desirable to publish the address in its original form, and I have done my best to conform to this wish: for this, however, I had to rely on incomplete stenographic notes taken during the lecture. The following are the only essential changes:

   (i) The original lecture had been planned as a positive contribution and, accordingly, all was avoided that could be interpreted as criticism of individuals; in particular, the lecture did not contain any reference to or quotation from the current literature, except a tribute to C. Leuba. Subsequent events have induced me to change my attitude, and I have hesitantly, substantiated my considerations by a few examples.

   (ii) The content of s2 was discussed only in private conversations in the Parapsychology Laboratory.

   (iii) Time did not permit completing the last section: the two mathematical examples and the very end were omitted.— Two further incidental departures are marked as such. The new book by Rhine and his collaborators (1) appeared after the present lecture.

2. An outstanding biologist assured me that, in periods of depression, he is accustomed to recover by reading some modern statistical investigations in biology or medicine. They give him, he explained, the comfortable feeling of certainty that Nature would never allow anything to happen that would be probable to a statistician.

3. "Les opinions de M. Jerome Coignard" by A. France. (Oeuvres completes, VIII, Paris, Calmann-Levy ed., 1926, p. 376.)

indisputable successes of mathematical statistics have induced many writers to describe the statistical methods empirically without due attention to the underlying theory. Thus it happens that frequently formulas are given and used without discussion of the fundamental assumptions and far beyond the field of their original scope and their applicability.[4]

The statistical methods involved in ESP-research in particular have been the subject of ardent discussions. It must be admitted that many criticisms were biased or concerned with minor details. On the other hand, ESP-advocates have rejected well-founded and almost obvious criticisms. Today, as far as I can see, nobody denies any longer that the fundamental law of probability, on which all numerical evaluations rest, has been, until quite recently, consistently misinterpreted, and that most of the experiments in ESP were not conducted in accordance with the theoretical requirements. I am referring to the so-called effect of optional stopping, which has been pointed out by several critics. It is a most surprising chapter in the history of science that an ardent wishful thinking could have misled the ESP advocates to deny vigorously an effect which is quite obvious, well known in experience and the subject of a vast literature in probability.[5]

We shall see in ss6-7 that this discovery fundamentally changes the aspect of the empirical material available at present; however, before proceeding to discuss this main point I propose to analyze some more technical details, especially the problem of shuffling (ssE-5).[6]

4. E.g., the tests for goodness of fit, designed to estimate certain parameters of probability distributions, are used frequently in cases where there is absolutely no place for probabilities; we are, in such cases, advised about the numerical values of the odds in favor of such and such assumption, though these words have no sense whatever. For an actual case where such statistical theories confirm with high odds three contradictory assumptions cf. Feller (5).
5. For details and references see below, s6. As far as I am aware, among all experiments supposed to show the existence of ESP the first and only one to take into account the effect of optional stopping is that by Pratt and Woodruff (16). According to these authors (loc. cit., p. 141) "the optional stopping correction was discovered as the present research was nearing completion."
6. The reading of ss3-5 is not necessary for the understanding of the sequel.

Now I wish to make it perfectly clear from the very beginning that the whole interest of the following discussion is concentrated on the statistical methods involved in ESP research. In particular, I am not going to discuss the experimental side, and for the purposes of this discussion it may be assumed that the experiments have been performed in the best possible way.[7] Nor is the aim of this discussion to "explain" anything or to prove that ESP does not exist: This obviously could not be achieved by purely statistical arguments. It may be clearly stated that a few experimental series are recorded with so large deviations from chance that they obviously never could be reasonably "explained" purely statistically.[8] However, results of this type form but an exceedingly small portion of the statistical material. On the whole, the most surprising feature of the material is its extraordinary heterogeneity. Subjects with supposed ESP abilities lose and regain their abilities very rapidly. We hear also of series of guessings consistently below chance. Furthermore, there is a great difference between individual experimenters, which has been the subject of several discussions and investigations: some experimenters are particularly lucky and find some degree of ESP ability on the average in every fifth person tested; in many of the successful experiments subjects guessed at a distance, in some of them even temporally in advance of the actual card-shuffling (implying a kind of precognition[9]). Other investigators, on the contrary, were unable to find any traces of ESP ability.

This in itself certainly would not affect the ESP hypothesis. However, we shall see that the fluctuations of the score-level caused by chance are bound

7. It is of course a pure question of terminology whether one says "the experiments have not been performed according to the requirements of the theory", or "the theory has been misinterpreted." With the first formulation Professor Camp's well-known statement *(J. Parapsychol.*, 1937, I, 305) holds also after the discovery of the so-called optional stopping effect.
8. That is to say, no reasonable allowance for the effects discussed in the sequel would essentially change the aspect of these series. The two record series obtained by Riess (22) (1850 trials with 1349 hits against an expected value of 370) and by Martin and Stribic (14) (50,000 trials with 14,280 hits against expected 10,000) are not even subjected to the effect of optional stopping: in fact, the Riess series was terminated by an illness of the subject with consequent loss of her guessing ability, and in the Martin-Stribic series the number of trials was, according to the experimenters, fixed in advance.
9. Cf. Rhine (20) and Rhine-Smith-Woodruff (21).

sometimes far to exceed the usually accepted limits, and that a convenient use of the method of optional stopping completely changes the situation: to my mind there is not the slightest doubt that this accounts for a part of the divergence in the experiences of different investigators. We shall indeed see that a considerable portion of the so-called significant material can and must be ascribed to chance, and that sometimes conclusions on the nature of ESP were drawn from material including no single significant case. It seems to me that the statistical material available at present affords but one safe conclusion: that ESP ability is, at its best, extremely rare.

To summarize: it is my definite impression that at present the evidence in favor of the ESP hypothesis consists of the few series with extraordinarily high scores: that any defense of and any attack on the hypothesis should be based on this portion of the material alone. I do not think that the ultimate word belongs to statistics: in this respect experiences of the past strongly suggest extreme caution. This extremely cautious attitude of most statisticians is also stated by the English statistician Bartlett: "In card experiments with supposed telepathic or other paranormal subjects, most of us are probably still a little chary about attributing statistically significant deviations to extrasensory perception."[10] And the point is made perfectly clear in the following passage from a letter of R. A. Fisher to H. Rogosin: "Perhaps I may say, with respect to the use of statements of very long odds, that I have before now criticised their cogency on the grounds, not only that the procedure of calculation is often questionable, but that they are much less relevant to the establishment of the facts of nature than would be a demonstration of the reliable reproducibility of the phenomena."[11]

2. The first point which I wish to emphasize is the following: When summing up the evidence in favor of ESP, Rhine combines all published series of which he knows and which had been conducted under a specified set of conditions to exclude sensory cues into a series of 907030 trials with 194605 hits;[12] and then treats this

10. In a lecture before the Royal Statistical Society, cf. (3), p. 14. The remark was not challenged in the subsequent discussion. (The publication came to this country after my address was delivered.)
11. This statement was read by H. Rogosin at the ESP Symposium, cf. (24), p. 267.
12. For the convenience of the reader the figures are here given

series as a continuous experimental series with a critical ratio (C.R.) of 39.90. It must be said that this procedure is unjustifiable, and that no mathematical or statistical reality whatsoever can be attached to this C.R.

The same method is the common practice when an experiment is conducted with several subjects: it is argued that, if no observations are dropped, the trials form a continuous chance series unless ESP intervened. We shall see that, owing to optional stopping, not even this restricted statement is true. The case is, however, incomparably worse when a heterogeneous material scattered throughout the literature is combined into one statistical series: this procedure could be justified were a so-called random sample of the set of all experiments ever conducted.[13] Now the whole material ever published (with a selection only in view of reliability) is certainly any thing but a random sample: would any one maintain that it contains all negative experiments ever performed as it actually contains all positive experiments?

Rhine argues[14] that he has avoided "selection of results" by taking "all the tests of which I had any knowledge". But the strongest possible selection lies in the very fact that the results have been published. It is like an army which is selected from only male individuals and is therefore no suitable basis for statistics of sex-distribution. Indeed, every positive result is, rightly, deemed interesting and published with details; whereas the chance results do not deserve detailed publication, and are seldom published.[15] Nevertheless every

---

(Footnote Continued) according to Rhine's new book (1), table 7, pp. 95-97. In April this book was not published and my reasoning was based on the older figures given by Rhine at the ESP Symposium, cf. (19), p. 255. The same method is used in different places.

13. Experience shows that even then the procedure would be extremely dangerous and that getting a true random sample is much more problematic than it would seem.
14. Loc. cit. in footnote 12.
15. Rhine's table 7, mentioned in footnote 12 contains 34 entries; among these there are but four with negative deviations and six with positive deviations and a C.R. not exceeding 3. However, these entries do not contradict the argument that the material is selected. Indeed, one of the negative C.R. is purely formal since it "was pursued with that as goal" (Rhine, loc. cit., p. 102: the C.R. is there properly treated as positive). All the six experiments with positive but insignificant deviations were

one of us has knowledge of experiments in card guessing, which have been performed at American universities and colleges, and have led only to chance results. Any in the very publications from which Rhine's figures are taken, we find recorded that also other experiments, under changed conditions, were undertaken but without positive results.

Now it should be constantly borne in mind how easy it is to obtain series with an arbitrarily large C.R. by combining perfectly trivial random series. After all, it is easy to obtain any number of series consisting of 1000 trials each with a deviation of 12 hits above chance expectation: the C.R. would be 0.949 and the probability of such a series is about 0.18. But combining 10 such random series into one we get a "significant" C.R. of 3, by combining 100 such series we arrive at a C.R. of 9.487 which indicates a "probability" of about 10^{-20}! Or let us toss a coin 25 times: the probability of obtaining at least 13 "heads" is obviously ½. Suppose now that this experiment is repeated and that, for any reason, only runs of 25 trials with at least 13 heads are counted (=published). Every run represents a deviation of at least ½ above chance expectation (12½ heads) and the probable deviation exceeds 2. In order to get a series of the length of Rhine's table mentioned above[12] let us combine 36280 such runs: the probable deviation above the "chance expectation" (= 36280 × 12.5) exceeds 72560 and we are absolutely certain to obtain a deviation of at least 18140: thus the probable "C.R." would be 152 and at the very worst we would get a C.R. of 38 (against Rhine's C.R. of 39.90). Yet we obtained our series by combining runs of a probability ½ and this without selection after we have agreed to count (publish) only a certain class of runs. And we should have obtained a similar result by concentrating our interest, instead of on single runs, on longer experimental series with a favorable result.

(Footnote continued) conducted for special purposes: to compare different methods or stimuli, examine the precognition hypothesis, covariation, etc. All six are considered as *favorable* for ESP. Two of the remaining three entries with negative deviations (Nos. 20 and 32) are not accessible to me. The last (No. 8) is one of eight series produced by Carpenter and Phalen (4). Seven of them show, in part considerable, significant *positive* deviations, and the experiment as a whole was published to *confirm* the ESP hypothesis (the seven favorable series are dropped by Rhine probably because of a possibility of sensory cues). — Thus the table contains at most (if any) two experiments with chance results, conducted in order to examine the repeatibility of the Duke experiments.

Or to use a simpler language: Suppose we make a statistical study of the proportions of different parts of the human body by collecting all reliable statements and measurements, without selection, which are published in the clinical periodicals: our conclusions would necessarily be queer. And the fault would not be ours: rather it would be that the writers in clinical periodicals concentrate their interest on exceptional cases.

3. One of the most important assumptions underlying the whole mathematical technique of ESP research is that all possible arrangements of the 25 cards in the deck are equally probable. Plausible as it sounds, this assumption is very abstract: for there are 623,360,743,125,120 different arrangements of the five symbols and to verify that they are, on the average, equally frequent, at least 10^{17} experiments would be needed: if the whole of mankind including newborn babies devoted eight hours daily to card experiments, working at the rate of an experiment a minute, it would take 300 years to perform 10^{17} experiments. But even if the experiment confirmed the hypothesis, it would only follow that the usual estimates of the probable fluctuations due to chance hold in series with a length comparable to 10^{17}. To ensure that the estimates hold in practice, and *in every case* that might occur, much more is needed: it would be necessary to prove that the usual method of shuffling, when repeated, *necessarily* produces so-called independent random samples of the set of all arrangements. This means, e.g., that there is no linkage whatsoever between the distribution of cards in a deck before and after shuffling. And for the purposes of ESP it would not even suffice that the possible linkage is statistically negligible on the average over a large number of experiments (perhaps, to a mutual compensation of different effects): it is necessary to know that the usual formula for the standard deviation, etc., holds for *every* subseries.

Now, it is a well-known experience that ordinary card shuffling is far from producing a random distribution. It is well-known that some card-players take advantage of their knowledge of the actual card-distribution in the previous game;[16] and it is also known that the most unexpected card-distributions are recorded to have

16. Of course, this is possible only for very special sorts of interlinkage, and hence not for all shuffling methods. To ensure a perfect game much less is needed than perfect randomness.

actually occurred.[17] This may not sound convincing. But we have also the experience of the repeated attempts to produce empirically random series. The simple methods of dice-throwing, card-shuffling, etc.,proved a failure and were given up.[18, 19]

Now the reader may wonder whether these doubts are not expressions of a too sophisticated mathematical mind: whether any adverse effect could actually be traced in the ESP experiments. This question could, no doubt, be answered exhaustively were it not for the common practice of not publishing complete records of the ESP experiments. As it is, I was able to trace in the literature only one complete record of actual card

17. After all, doubts about the effectiveness of card shuffling have led to Poincare's investigations which, finally, gave the result that infinitely many shufflings would produce a random order.
18. Compare the actual construction of the tables (29) and (11). The latter were constructed with the utmost refinement imaginable. They contain 100 groups of 1000 random numbers each. They were subjected to four different tests of randomness. And it may surprise optimists that "although the complete entity of 100,000 random digits is seen to satisfy these tests, *five of the 100 groups are found to deviate from expectation* when tested as individual groups". It should be noted that precisely this so-called "poor *local* randomness" would be crucial for the ESP experiments.
19. Also the following personal experience may be recorded. The Administration of the Telephone Works of Stockholm ordered, for purposes of random experiments, 100 numbered balls about ½ inch in diameter. The balls were of finest Swedish steel, the numbers almost invisible, and whatever care skilled craftmanship could take to ensure homogeneity was taken. Nevertheless we did not succeed in obtaining random series.

    Recently a friend of mine, an experienced card player, challenged these considerations on the grounds of his personal experience. So we decided to make experiments. Apparently he used one of the less effective methods, but I am told that the shuffling machines perform essentially the same method of shuffling, though probably with them the number of shufflings is larger. Now, he would shuffle the deck 7 to 12 times; the deck was then not cut but simply compared with the original distribution, or its reverse order according to whether the number of shufflings was even or odd. The number of coincidences varied between 10 and 20 instead of the expected number of 5 (we used ESP cards). Even when the deck was cut I was able to guess consistently above chance, although only when I observed from afar the cutting procedure in order to estimate at what place in the deck it was distributions, that published by Riess (23). Moreover, Kennedy's paper (12) contains (opposite p. 216) a photostatic copy of a part of the original record in an experiment of Mac Farland; however, the card distributions only of 10 consecutive runs are given, so that it is only possible to analyze the effect of at most 9 shufflings.

Now the card distribution in the Riess-series seems to be satisfactory, as far as I could make out without a thorough analysis. However, in this experiment it was a rule to perform only two runs on the same day, and the 37 days of experimentation are scattered over a period of several months. It is therefore impossible to ascertain to what degree the randomness is due to excellent shuffling. Suppose, therefore, that in this case the shuffling was perfect: after an analysis of Mac Farland's 10 runs no doubt will be possible that such is not always the case.

Indeed, let us consider the last three runs. If the five symbols are denoted by c (circle), p (cross), s (star), r (square), w (waves) the actual card distributions in the runs number 8, 9, and 10 were

(8) r s w p c s r c r c p p r c c w p w s w s s p r w
(9) w r s w s w p c s s s p c r r c p p r r w c c w p
(10) p r r w c c w p c s s s w p p r c r w r s w s c p

It will be noticed that the cards show a marvelous tendency to cling together. In the 9th and 10th arrangement of the deck the 1, 2, ... places are occupied by the same symbols as the following places in the 8th or 9th arrangement of the deck, respectively,

(8→9): 18 1 19 20 2 3 4 5 6 21 22 23 8 7 9 10 11 12 13 24 25 14 15 16 17

(9→10): 18 19 20 21 22 23 24 25 8 9 10 11 6 7 12 14 13 15 1 2 3 4 5 16 17

It should be borne in mind that the deck was cut: thus after the shuffling proper the first card of the rearranged deck lay below the last one. In other words, the shuffling of the 9th arrangement of the deck left the group of cards numbered 16-25 (i.e., 10 cards) in the same order; in addition, the 5 cards numbered 1-6 clung together, likewise the 4 cards 8-11. In all 17 cards

(Footnote continued) made. This is, of course, not the usual set-up in ESP experiments but proves nevertheless that shuffling does not *necessarily* produce randomness.

have after the shuffling the same successor as in the original deck. A similar remark holds for the shuffling in the 8th run: here 15 cards conserved their successor, and it will be remarked that even the same groups of cards show a tendency to cling together. The linkage between consecutive decks is so strong that it can be observed even when the distribution of the 10th run is compared with that in the 8th run. We get the following reordering:

(8→10) 12 13 24 25 14 15 16 17 5 6 21 22 3 4 23 7 8 9 18 1
19 20 2 10 11,

and it is seen that despite a repetition of the procedure of shuffling and cutting the deck, 13 cards have conserved their successor.

Strictly speaking, it cannot be logically proved that the cards actually did stick together: in a deck of 5 different symbols and five cards of each there are $(5!)^5 = 24,883,200,000$ different arrangements of the cards showing the same arrangements of symbols, and only the latter is known to us. It is, however, apparent that chance alone produces rearrangements of a different kind, and the corresponding odds are easily computed. There is only one reasonable ambiguity: the cards numbered 24 and 25 in the 9th run were actually the same as the cards 6 and 7: it is therefore perfectly possible that the order of the cards was not, as described above, but

(9→10): ... 22 23 6 7 8 9 10 11 24 25 12 ...

This, however, is of no importance, the number of cards with conserved successor being the same. (The largest unchanged group would now consist of 8 cards, the next two of 6 and 5 against 10, 5 and 4 with the first alternative.)

Not all runs are linked to the preceding one in the same net fashion, e.g., it is probable that the deck was changed, or fell to the floor, between the fifth and the sixth run. Between the 6th and the 7th run the group 9-16 was conserved in its original order but for the interchange of the cards numbered 14 and 15. The rearrangement of the third run shows two such interchanges:

(3→4): 21 22 23 24 8 9 10 11 25 13 14 12 15 16 17
18 19 1 20 2 3 4 5 6 7

where 14 cards conserved their successor, etc. After this example it will hardly be denied that linkages of the worst kind actually do occur.[20] Nevertheless, I proceed to further examples.

It has already been pointed out that I was not able to find records of any other card distributions which have actually occurred. However, there are two records from which at least indirect conclusions about the actual card distribution are possible. I was unable to find more such records.[21]

4. L. Warner (31), p. 237, published a complete record of 250 consecutive trials. It is unfortunate for the purposes of the present investigation that the cards used were *not* arranged in decks: from trial to trial "a different deck, newly shuffled, was cut to obtain the next card to be guessed." Obviously this procedure is much more favorable to randomness than the usual shuffling of a deck. The result is, however, not encouraging. Indeed, the order of symbols is expected to be random, each symbol having, in each trial, the probability 1/5. The actual distribution shows deviations from expectation in different aspects. Suffice it to note that the symbol c occurs 71 times against an expected number of 50; this makes a C.R. of 3.32, indicating, according to the usual normal approximation, a probability of 0.00045. This probability should, of course, be multiplied by 5 since there are 5 symbols. Nevertheless,

20. The whole series of which the reproduced 10 runs form a part, was, from another viewpoint, examined by J. L. Kennedy, loc. cit. He also reached the conclusion that the shuffling was poor.

21. The paper by Murphy and Taves (15) may, however, be mentioned as one of special methodological interest. These authors have devoted a careful study to the problem of adequacy of shuffling and the problem of linkage in their experiments (cf. esp. pp. 44 seq. and 62 seq.). Their results have, however, unfortunately, no bearing on our present problem. As a matter of fact, the experiment in question differed essentially from the usual set-up in that several decks were used simultaneously; and the authors find that there was no significant linkage between *different decks*. The comparison of the distribution of a target deck against the succeeding one was a minor point: in this respect only the number of coincidences was counted and found to be within random limits. This, however, is only a very special and unusual type of linkage. For the 10 runs of the MacFarland experiment the number of coincidences is well within random limits: for the last 3 runs they are 3 and 6 respectively against an expected number of 5: yet we say that these three runs are interlinked.

with the ordinary rule, odds of 0.00225 are still considered significant. In addition, there are other deviations: the expected number of occurrences of triples like ccc, etc., would be 2 for each symbol. Actually ccc occurs three times, sss once and the remaining three not at all. The explanation for this poor randomness[22] is probably that the 250 decks, which were used, lay originally in the same order as they are delivered from the factory; they were shuffled and cut, but this procedure can never be relied upon to produce randomness.

Of course, different methods of shuffling differ also with respect to their effectiveness. One of the methods consists in dividing the deck into several groups and reordering them: for this method the linkage can easily be shown to be very considerable. For this and many other methods there is not only the theoretical probability of a linkage; if the shuffler is not familiar with the dangers of shuffling it can easily happen that, in the consecutive removing and reinsertion of cards of the deck, more or less the same cards are affected, or, alternatively, that some group of cards is not affected at all. Both possibilities are rather probable because of the way in which the deck is held in the hand. The better methods of shuffling also often produce only an apparent random order. Consider, for example, the much advertised method in which the pack is divided into two nearly equal parts and the two halves are recombined in an alternating order, subject of course to small random fluctuations. The result may be schematically represented by the following rearrangement of the original order:

1, 14, 2, 15, 3, 16, 4, 17, 5, 18, 6, 19, 7, 20, 8, 21, 9, 22, 10, 23, 11, 24, 12, 25, 13.

This is of course but a scheme, and in reality sometimes two cards might slip out or the two halves might vary in size. It is nevertheless clear that this procedure separates more couples than a random rearrangement would or that one could safely bet to find the cards numbered 23, 24, 25, in this order in the second half of the deck. If we repeat the procedure we might get another rearrangement, e.g., 1, 20, 14, 8, 2, 21, 15, 9, ···; again random fluctuations will affect the order, but it is perfectly clear that the effect cannot be a random order; there

22. I wish to make it perfectly clear that the above consideration does not intend to explain (or, indeed, analyze) Warner's experiment as such. Actually, his subject had no preference for the symbol c (having called c exactly 50 times).

are, e.g., long odds that the cards numbered 24 and 25 will still be, in this order, in the second half of the deck, that the cards 1 and 2 are separated by other cards, etc. The argument can be repeated, although naturally with each new shuffling the effect of random fluctuations increases until it, eventually, predominates. The effectiveness of the method would be increased if the number of shufflings were random; however, there if a natural tendency to use, consciously or unconsciously, always the same number of shufflings.[23]

5. In examining whether the shuffling effect is present in practice, each experimental series must necessarily be considered separately.[24] Indeed, when combining a larger amount of material, coming from different series, the ordering of the runs is arbitrary and the very notion of linkage loses its sense. Furthermore, when the whole material is submitted to random tests such as counting the frequency of pairs of symbols or of gaps between them, the effect of possible poor local randomness disappears owing to the mutual cancellation of the different methods of shuffling, the different initial arrangements, etc. (In some experiments the cards may have a tendency to cling together, as was the case in MacFarland's experiment, in other experiments they may be separated too frequently as was the case with the last method of shuffling discussed above, etc.) In theory, of course, even poor local randomness can be tested on a large and heterogeneous material: but only by a study of the local variation of the standard deviation, which is very laborious and requires a very large amount of material.

The possible effect of poor shuffling on ESP experiments has been frequently discussed with respect to the possible advantage the subject could derive from his knowledge of the card distribution in the previous arrangement. How, as has already been pointed out above, for the purposes of the present considerations, the experimental set-ups are assumed to be the best: it may, e.g., be assumed that the subject gets no information about the actual card distributions, so that no predictions are possible unless ESP exists. However, the

23. Thus, e.g., Woodruff and Pratt (16 , p. 130) explicitly state that they used 5 shufflings for each deck.

24. Originally the following considerations were made during the discussion following the address when my attention was drawn to Greenwood's "chance control series" (6) which was said not to indicate any deviations from randomness. It may be mentioned that this series was intended, and is useful, for other purposes.

theory is based on a formula for the expected limits of the random fluctuations, and no mathematical formula can be based solely on the lack of knowledge on the part of subjects. The main problem is whether the usual estimates for the expected number of hits (i.e., 5 per run of 25 trials) and for the standard deviation hold.

For the first part the answer is affirmative, at least if it is assumed that the deck is properly cut, that is to say, if the deck is cut in such a way as to ensure that all 25 possible resulting arrangements are equally probable. This is true (and mathematically trivial) whatever the original distribution of the deck and whether or not the original distribution is known to the subject; and this remains true even if the deck is not shuffled at all.[25]

It is, however, perfectly clear that, with strict randomness not ensured, the standard deviation may be larger than the value usually assumed. To put it in plain language: The gross average of the number of hits for a large number of experiments with many subjects will not be affected by poor shuffling; but there is the possibility that the mental habits of some subjects will, sometimes, cooperate with the shuffling effects so that, for some subject and some period, the average number of

---

25. However, the assertion is not necessarily true if the deck is not properly cut. With several methods of shuffling the subject may (consciously or unconsciously) derive a slight advantage from the experimenter's habit of always cutting in the inner portion of the deck, that is to say, never to leave the deck as it was (as he is expected to do, on the average, once in 25 runs); in such cases a subject's habit not to call the same symbol at the same place may give him some advantage or disadvantage, as the case may be.

I do not know whether this ever occurs, but the argument in itself is of interest. It must be clearly understood that if the chances of the subject exceed 5 correct guesses per run, then, small as the difference may be, the subject is bound eventually to obtain "significant" deviations from supposed chance-average of 5. This argument was used by Kennedy, who, in certain experiments, discovered recording errors increasing the average score of 5 to 5.087. Assuming, which is plausible or at least possible, that the rate of recording errors is constant throughout the experiments, Kennedy rightly concluded that the recording errors can produce a "significant" deviation in 10,000 runs. Surprisingly, this argument was ridiculed by Stuart (27), p 309 seq.

hits will be sometimes above chance, and sometimes below. Thus the shuffling effect increases both the number of positive and of negative deviations from random expectation, and the average remains unchanged. E.g., if the shuffling tends to reverse the order of cards (as was the case in Mac Farland's experiment) and hence increases the repetition of couples of symbols, this can help a subject inclined to call the same couples; and this effect may be considerably augmented if the subject is told his results (in particular if the subject notices the occurrence of sequences of adjacent hits and tends to repeat such "lucky" combinations of symbols).

It is clear that any kind of linkage between consecutive target decks may (with more or less probability) get into such a correlation with the calls and affect the average number of hits favorably or adversely. Now in practice, if the scoring level is not above chance, then "there is, of course, no point in including these latter data in our efforts to compare ···. If average series are not above chance, there is nothing to say".[26] With this we come again to the question of selection, raised in s2. It may be conceded that in each experiment the recorded totals include all results, so that there is no selection of results in the individual experiments:[27] the fact remains that, generally speaking, only experiments are published which are positive on the whole. And for this reason the practical effect of nonrandomness of the card-distributions is to increase the relative frequency only of series with positive deviations. In theory it increases the probability for large deviations of both signs, but the selection effect of publication accentuates the increase in the positive deviations only and so produces a spurious increase in the average number of hits.

Another serious effect of the linkage between consecutive target-decks lies in the possibility of producing a correlation between the scoring successes in consecutive runs; that is to say, under certain circumstances, it may happen that a successful run increases the chance of success in the next. The full effect of such a correlation will become apparent first in the next paragraph, when we analyze the usual experimental

26. Rhine (18), p. 181.
27. This is true for the experiment by Rhine just mentioned, and I think, for all experiments of the last period. Admittedly, it was not always true.

practice, by which the experimenter may watch the successes, interrupt and start again with the experiment at will and even continue during the pauses the experiments "off record". If there is the slightest correlation between consecutive runs, this procedure clearly amounts to a hidden selection of results.

It might be objected that we are here considering exceptional cases not likely to occur frequently. And so we are. However, it should be borne in mind that we are also concerned with observations which are, on the whole, not frequent. The question is whether the poor randomness may sometimes produce some effect, or in other words, whether the assumptions underlying the usual ESP procedure are plausible or proved.[28]

Concluding these considerations I may mention the practice of some experimenters to check the subject's calls not only against the target deck but also against the reversed order of the latter. I assume that this is done for some special purposes; it should, however, be understood that this method has no bearing on the problem of linkage or chance-correlation between consecutive runs. Indeed, it cannot be expected that a given series of calls should show an exceptional number of coincidences both with the target decks *and* their reversed order. Thus a series with hits above chance expectation must in general be at chance level or less when checked against the reversed target.

However, the whole problem of randomness could easily be avoided by a systematic use of tables of random numbers (11 and 29): the cards could thus be deliberately stacked in a random order unknown to the subject.[29]

6. I proceed now to the question of so-called optional stopping. It is easily explained. If we decide to make, say, 1000 trials, there is a definite probability of obtaining a C.R. of, say, 3. But until quite recently[30] the corresponding formula was, wrongly, applied

28. To forestall futile discussion I wish to emphasize once more that I am not discussing the existence of ESP and not proposing a counter hypothesis; generally speaking, however, I confess that I am unable to see any serious objection to the remark that a counter hypothesis is not applicable to all results.
29. The same suggestion was made independently also in the address of Bartlett mentioned above ([3], p. 14).
30. Cf. footnote 5.

to compute a quite different probability. It was not decided to make 1000 trials but the experiment was continued until a favorable C.R. was obtained: if, in this way, the experiment was terminated, after, say, 1250 trials with a C.R. of 3, the probability of obtaining a C.R. of 3 or more with exactly 1250 trials was computed, instead of the probability of getting it *sometimes* (with reasonable limits). In plain language: the probability of a rainy day at Los Angeles may be negligible for every day in the year: yet, by staying there, you are sure, eventually, to get rain.

The situation being as simple as this, it is surprising that the formula was misinterpreted; still more, that although this criticism was made by psychologists the validity of the argument was again and again vigorously denied.[31] Thus it came about that a simple mathematical fact was finally established by experiments rather than by arguments: In [13] Leuba described a series of pure chance matchings, between two decks, conducted in accordance with the usual experimental set-up. He made a total of 13,410 trials with 87 "subjects" and found *seven* among them who did have, at one time or another, 200 consecutive trials with a C.R. of 2.5 or more. One subject got 1000 trials with an average of 6.3 per 25, giving a C.R. of 4.1 (according to the usual interpretation the probability for this would be .00002). One subject even averaged 7.4 hits per 25 trials during 250 trials and his average score over the whole experiment (725 trials) was 6.3 (C.R. of 3.5). In the same issue of the Journal of Parapsychology Greenwood [7] reported on similar experiments, though conducted under somewhat more favorable conditions. After these results the effect of optional stopping was no longer denied.[32]

31. Today this statement is, in turn, denied, and it is instructive to quote one instance: (Greville [9], p. 251):

"The statement is occasionally made that by continuing long enough one is certain eventually to obtain a fairly high average total score. This is clearly at variance with the recognized laws of probability, but in order to meet the objection more specifically I have worked out a formula which gives, with a high degree of accuracy, the probability that a given score will ever be reached in a series of infinite length ···. For example, the probability in an *infinite series* of calls of *ever* securing, at *any* time after the first 10,000 calls, a total average score of 5½ correct guesses per 25 can be shown to be less than one in a million."

In reality this probability is 1 (certainty) and one is sure to get, infinitely often, an arbitrarily large C.R.

32. In the ESP literature the theoretical discovery of the effect

Now the next question concerns the actual limits of the effect of optional stopping. It is certainly dangerous to belittle it with fallacious arguments. Thus Stuart ([28], p. 199) argues: "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 ...." Now, the notion of "gambling system" is a matter of definition: according to the definition usual in the theory of probability no gambling system is possible no matter whether the capital or the number of plays be finite or infinite. On the other hand, there are long odds that a systematic use of the advantage of optional stopping will work satisfactorily. Indeed, this is an everyday experience, and many a "refutation" of the theory of probability rests upon it.[33]

The effect of optional stopping is certainly not capable of "explaining" record series like the one produced by Riess, or indeed any series in which the subject

(footnote continued) of optional stopping is ascribed to Greenwood and Greville's [8]. As a matter of fact it is difficult to say when it was discovered since, in reality, we are only concerned with the correct interpretation of a formula which never before had been consistently misinterpreted. That in an infinite series any C.R. is bound to occur infinitely often is an immediate consequence of the very law in probability on which the whole ESP theory is based. A long series of mathematical investigations is devoted to refinements of this statement. Thus it is known since 1914 that in a series of n trials the order of magnitude of the greatest observed C.R. is expected to be $\sqrt{\log \log n}$, and this quantity is steadily increasing. This gives an idea of what can be achieved in practice by stopping at favorable points. — Generally speaking, we are here concerned with a trivial special case of deep and interesting problems to which a considerable part of the modern theory of probability is devoted.

33. Well known is the fact that many persons have succeeded in visiting Monte Carlo year after year and never losing. I myself have a friend who did it systematically, firmly believing in his ability of prediction. I never succeeded in convincing him that his experience was a common one. Indeed, if we suppose that he started gambling with a capital of $1000, he has obviously a considerable chance to gain $200 before ever losing the original capital. Stopping at this point, he has reasonable odds in favor of repeating the same experience say, 10 times. There must be many persons with this experience. If we collect the data about them and combine into a single continuous series we again may get an arbitrarily large C.R. and refute the chance-hypothesis. No impractical long series are needed.

scored *consistently* above chance expectation.[34] It is even probable that the *average* effect of optional stopping (i.e., when the average is taken over the set of all thinkable experiments) would not be of decisive influence.[35] We are, however, concerned with a selected material, and the only question of practical interest can be: what is the probable effect of optional stopping at its best? We shall see that it fundamentally changes the aspect of a large portion of the experimental work produced until now.

In [10] Greville gives a summary of his theoretical investigations on the frequencies of occurrence of different C.R.'s caused by the optional stopping. He reaches the conclusion (p. 90) that "these figures would indicate the adoption as a criterion of significance of a C.R. of 5½ as roughly equivalent to the 2½ which is customary under ordinary circumstances." And this estimate leaves out of account the effect of linkage which at its worst may considerably increase the outer limits. Greville's estimate is the only one I know of which is reached by mathematical deductions[36] and there it can certainly not be lowered if one wishes to keep on the safe side.

Now a glance at the literature shows how fundamentally Greville's estimate changes the whole aspect. It should be compared with the rule of the "Handbook for Testing Extra-Sensory Perception" by Stuart and Pratt (p. 59): "If the C.R. is between 2.5 and 3.0, you can be relatively certain that the results did not happen by chance. If it is between 3.0 and 4.0, you can be very sure."[37] It must be understood that, according to

34. Cf. footnote 28.
35. Therefore no reliability can be attached to the numerical values of the effect as it occurred in the experiments of Leuba and Greenwood referred to above: they may be below chance and, at least, they cannot be assumed to be essentially above chance; hence they provide no limits.
36. Greville's paper is not yet published (at least it was not at the time when Rhine's new book appeared). I am unable to follow Greenwood's far more optimistic "definite but tentative recommendation" ([7], pp. 225 and 229), which is reached on his empirical material. Cf. also footnote 35.
37. In his table 7 referred to above ([1], pp. 95-97), Rhine sums up the main experiments "which acceptably exclude sensory cues." The table contains 34 entries with a total of 907,030 trials; of those only 14 entries with a total of 250,305 trials show a C.R. of at least 5½. Among those one single entry accounts for

Greville, a C.R. of 5½ (i.e., 2½ in the usual interpretation), though commonly defined as "significant" must occur fairly often. Indeed, according to the normal approximation the probability of a C.R. exceeding 2.5 would be 0.0062. A better approximation was computed by Stuart and Greenwood ([26], p. 303) according to whom the probability in question is, roughly[38], about 0.008, which means that because of optional stopping a C.R. of 5.5 may occur once in 125 experiments.[39] And, it must be insisted, this figure relates to the set of all *started* experiments: in practice, naturally, the experiments which turn out to have an unfavorable start are soon dropped, since they are unlikely to yield interesting results. In view of the time and interest devoted to ESP at many American colleges it is certainly not surprising to hear that some experiments resulted in a C.R. of 5.5.[40,41]

(Footnote continued) 96,700 trials. Not even all of these 14 entries could make us feel "very sure" according to Greville's estimate combined with the Handbook's rule.

38. Strictly speaking, this probability depends on the length of the experimental series. For a series of 625 trials Stuart and Greville find that the probability is 0.0081 or 0.0093, respectively, according as the binomial or the matching hypothesis is accepted. For 2500 trials the corresponding figures are .0075 and .0082.

39. The figure given is, of course, according to Greville, a bound rather than an estimate. On the other hand, his estimate leaves out of account at least the effect of linkage.

40. When discussing the various proposed recommendations as to which C.R. would be considered "significant" it is, unfortunately, not always understood that, for special experimental designs the estimate is the more conservative the *lower* the chosen limit is. Such is the case when some items are tested as to the expedience of their being subjected to further experimentation, and when those items are conserved which obtain a significant C.R.

41. It should be noticed that there is also an effect of optional beginning parallel to the effect of optional stopping. As a typical example the report [2] may be quoted. The writer states that "···when beginning the experiments ··· I kept no permanent records, merely recording runs whenever and wherever I found it convenient···. As the results began to surprise me, I began to use the record pad and made the observations at my house ···." With this favorable beginning the writer achieved a series of 1000 trials with a C.R. of 3.44, which will no longer surprise. The next 1000 had a C.R. of 2.66, the third of .86 and the writer called "3500 additional cards with varying scores, at one time *steadily below expectation*···." The writer then divides

However, even while accepting far lower limits of significance than a C.R. of 5.5 would be, we still can find reports of experiments reaching conclusions on ESP, and not including a single case of significance -- that is to say, based on purely chance material. Before quoting a couple of examples I have to mention the common practice of making experiments with several subjects, combining the obtained series into one large series and computing the corresponding C.R. as if this series were a continuous experimental series. It was already pointed out (s2) that if you stop each subject at a point slightly above chance (and you are practically sure that you can achieve this at will) you will eventually combine many series with absolutely insignificant but positive C.R. into one series with a highly "significant" C.R. We shall see that this is exactly what happens in practice.[42]

Now the same method is applied even in experiments with a single subject. When the scores drop the experimenter may[43] "'quit scoring for a time'. He may stop the testing for a minute, an hour, or a day ··· or he may suggest continuing the same procedure off 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."

(Footnote continued) the last mentioned 3500 calls "into three groups (1) Morning calls when in good shape, (2) Evening calls, after work, (3) Morning calls when sleep had been lacking or I awoke with a sick headache". The last two groups show negative C.R., the first 2.99. Such a division is, of course, practically always possible and by no means significant.

A related source of error was repeatedly criticized (cf. Warner [30], p. 92), and is now, I assume, completely abandoned. It consists in making so-called preliminary tests and continuing only with successful subjects but including the preliminary results in the actual test made thereafter. Mathematically this changes the absolute probability of obtaining a specified C.R. into the conditional probability of obtaining it when it is known that the first few runs have been successful. The numerical difference is considerable.

42. Against this obvious argument it has been objected that the combined series would not show a scoring level *consistently* above chance. This is equally true as beside the point if we are not informed about the total number of hits at each trial in the series: as we are not.

43. Quoted from Stuart ([28], p. 200), where Rhine's method is defended against Kellogg.

Because of optional stopping this method is clearly unjustifiable since it means the addition of many small random errors (the point will be elucidated below by a simple mathematical example). However, in this case the method is obviously far more dangerous than it was in the previous example, because of the possible effect of linkage. If there is any correlation between the number of hits in consecutive runs, this method is certainly most efficient in selecting positive results.

To see how all these methods work in practice to create spurious significance let us, as a typical example, consider Shulman's "Study of Card Guessing in Psychotic Subjects" [25]. The patients were grouped into 14 clinical classifications. The corresponding C.R. are

0.53, 0.25, -0.37, -0.58 -0.08, 0.09, -1.37,
0.51, -0.14, 3.388, -1.937, 1.34, -1.62, -0.56

which certainly yield an example of a random distribution. One group, however, the manic-depressive, have a "significant" C.R. of 3.388. Now we do not even need Greville's rule that a C.R. of 5.5 should be considered as roughly equivalent to 2.5 and usual conditions to understand that the C.R. of 3.388 is not very surprising. It should also be noticed that we are here concerned with the probability that among 14 independent groups one at least should attain a C.R. of 3.388, and thus the rule of "significance" loses its foundation. The worst is that this C.R. of 3.388 is purely spurious. Indeed, the group consists of 12 patients and they attained the following C.R.

-0.49, 1.50, 0.90, 1.50, -0.39, 1.73,
0.79, 1.19, 2.01, 1.45, 1.64, 1.56.

This series does not contain anything extraordinary: it is a typical example of stopping each individual series at a moment slightly above chance and getting in this way a spurious C.R. for the whole series. The largest C.R., 2.01, was obtained in a series of 500 trials. Even without correction for optional stopping, the probability of getting, in 500 trials, a C.R. of 2.01 or more exceeds .03: the probability that one among 12 subjects gets such a C.R. exceeds 1/3 -- without any corrections. Similar remarks hold for the other investigations, e.g., for the investigation on children by Louisa E. Rhine [17] which does not include a single case of any significance.

7. I may end by an example to show in simple but exact terms, what happens when several small series are combined into one. The example is not chosen so as to give any idea of the possible effect of the optional stopping: on the contrary, it simply neglects the effect of optional stopping altogether. The example is given because it is the simplest one I was able to find, and seems nevertheless to show clearly that *in principle* the whole mathematical problem is changed when the technique described above is used.

Let us consider a card guessing experiment, with independent trials, the probability of a hit being, in each trial, 1/5 (thus only the restriction to complete runs of 25 is removed for sake of simplicity). The guessing is done by the same subject. We adopt in advance the rule that the experiment shall be stopped when the subject has made a specified number, N, of hits. Thus there is no place for optional stopping -- the point of stopping may, but need not, be particularly favorable.

Let, in an actual experiment, n be the number of trials before stopping. In order to compute the probability of this event "N hits in n trials," we have, according to the common routine, to compute the quantity

$$X = \frac{N - \frac{1}{5}n}{\sqrt{\frac{4}{25}n}} = \frac{5N - n}{2\sqrt{n}}$$

We propose to show that this is *not* the true C.R., in other words, that the desired probability cannot be obtained by looking up the probability associated with X according to the normal law.

It is convenient for the following to restate the problem in order to see its connection with our main problem. We may split up the whole series of n trials into N subseries, the first ending with the first hit, the next with the second hit, etc. With our convention when to stop it obviously no longer matters whether the guessing is done by one or by N subjects. The whole experiment consists of N tests: each of them consisting in letting a subject guess until he scored his first hit. Thus a "test" may consist of a single trial (probability for this = 1/5), of two trials (probability 4/5 . 1/5, etc. In other words, the length of a single test is a random variable: it is easily verified that its mean value is 5, and its standard deviation $2\sqrt{5}$. Hence the total number of trials in N tests is again a random variable with the mean value 5N and the standard deviation $2\sqrt{5N}$.

In this formulation the experimental set-up is strictly the classical one, and to calculate the probability of a certain pair (n,N) we have to compute the true C.R.

$$C = \frac{n - 5n}{2\sqrt{5N}}.$$

It is seen that

$$X = -C\sqrt{\frac{5N}{n}},$$

so that X is not the true C.R. of the problem. To get a numerical illustration choose N = 180 and suppose that the whole experiment is terminated after n = 6000 trials. Then the true C.R. is C = -5, whereas the usual procedure would lead to an estimate of X = 6.124. Similarly, if n = 720, the usual method leads to a spurious C.R. of X = 3.354 instead of the correct value -C = 7. For larger N the relative difference between -C and X decreases and in the limit it vanishes. Thus, in this case the numerical difference is not essential, but we have also agreed to choose an unfavorable rule of stopping and have *abandoned* optional stopping in order better to illustrate the theoretical point.

A more radical example of a similar character would consist in simply tossing a coin with the understanding that the experiment has to end when the number of "heads" exceeds the number of "tails" by N_0. The experiment can be restated as a making of N tests, each consisting in tossing a coin until the number of "heads" exceeds the number of "tails" by one. (This again means only splitting the whole series into N subseries.) Suppose the whole experiment takes n trials. To judge the probability of this event one would, according to the usual routine, proceed to calculate the "C.R.", namely,

$$X = \frac{N - \frac{n}{2}}{\sqrt{\frac{1}{4}n}} = \frac{2N - n}{\sqrt{n}}$$

Yet, in reality, the number n of trials in our experiment is a random variable with *infinite mean value*: It is known that the probability of a given n, i.e.. of a certain X, can by no means be judged from the normal probabilities.[44] Hence, though the experiment can easily be conducted in practice, the "C.R." computed in the usual way is completely meaningless.

To advocates of ESP these arguments may seem a reaction of too sophisticated a mathematical mind and the requirement to keep on the safe side in all estimates may sometimes appear hard. Yet it is only the consequence of the decision to submit the ESP hypothesis to strictly statistical proof. If there is ESP, it seems natural to introduce optional stopping, etc., in order to allow for periods of fatigue and the like. Yet this is possible only at the price of introducing upper estimates for the probabilities in question. On the other hand, one should bear in mind that the ESP ability must necessarily, though with diminished effect, become statistically significant also if the right of optional stopping be given up altogether.[45]

Or else there remains the classical way which according to so experienced a statistician as *R. A. Fisher* still seems more "relevant to the establishment of facts of nature":[46] to find the most favorable conditions for ESP and to improve them until a reliable reproducibility of the phenomena under favorable conditions is definitely proved.

There is also a compromise between both methods, which has been frequently recommended: to substitute a deck of blank cards, or a deck consisting of identical symbols, for the usual one so as to change the frequency of occurrence of different symbols in a way unknown to the subject. This was declined by the authorities on ESP ([1], p. 17) on the ground that "this is to assume that ESP is a wholly conscious process and one that cannot be tricked". Nevertheless, it may be worth an experiment. For according to the same authorities, belief in ESP runs through all ages: but all ages tested it by trying to trick it. So, according to Shakespeare, when Joan La Pucelle came to see the Dauphin, he tried to trick her:

"Go, call her in. -- But first, to try her skill,

44. I omit the proof which is not quite simple. Despite the infinite mean value the law of large numbers holds (the early naive discussions about the so-called paradox of St. Petersburg have been completely cleared up.)
45. Cf. footnote 8.
46. Cf. the quotation above and of s1.

Reignier, stand thou as Dauphin in my place:
Question her proudly; let thy looks be stern:
By this means shall we sound what skill she
hath."

But she cannot be tricked:

"Reignier, is't thou who thinkest to beguile me?
Where is the Dauphin? Come, come from behind;
I know thee well, though never seen before."

REFERENCES

1. Pratt, J. G., Rhine, J. B., Smith, Burke M., Stuart, Charles E., and Greenwood, J. A. *Extra-Sensory Perception after Sixty Years.* New York, H. Holt and Co., 1940. (This book was not published at the time of the Address.)

----

2. A scientist tests his own ESP ability. *J. Parapsychol.*, 1938, II, 65-70. With editorial comment.
3. Bartlett, M. S. The present position of mathematical statistics. *J. Royal Stat. Soc.*, 1940, 103, 1-19 and discussion 20-29.
4. Carpenter, C. R., and Phalen, H. R. An experiment in card guessing. *J. Parapsychol.*, 1937, I, 31-43.
5. Feller, W. On the logistic law of growth and its empirical verifications in biology. *Acta Biotheor.*, 1940, 5, 51-66.
6. Greenwood, J. A. Analysis of a large chance control series of ESP data. *J. Parapsychol.*, 1938, II, 138-146.
7. Greenwood, J. A. An empirical investigation of some sampling problems. *J. Parapsychol.*, 1938, II, 222-230.
8. Greenwood, J. A., and Greville, T. N. E. On the probability of attaining a given standard deviation ratio in an infinite series of trials. *Ann. Math. Stat.*, 1939, 10, 297-298.
9. Greville on ESP and mathematics. The ESP Symposium at the A. P. A. *J. Parapsychol.*, 1938, II, 248-252.
10. Greville, T. N. E. A summary of mathematical advances bearing on ESP research. *J. Parapsychol.*, 1939, III, 86-92.
11. Kendall, M. G., and Smith, B. Babington. *Tables of Random Sampling Numbers.* (Tracts for Computers, No. 24.) Cambridge University Press, 1939.

12. Kennedy, J. L. A critical review of "Discrimination shown between experimenters by subjects" by J. D. MacFarland. *J. Parapsychol.*, 1939, III, 213-225.
13. Leuba, Cl. An experiment to test the role of chance in ESP research. *J. Parapsychol.*, 1938, II, 217-221.
14. Martin, D. R., and Stribic, F. P. Studies in extra-sensory perception, I and II. *J. Parapsychol.*, 1938, II, 23-30 and 287-295.
15. Murphy, G., and Taves, E. Covariance methods in the comparison of extra-sensory tasks. *J. Parapsychol.*, 1939, III, 38-78.
16. Pratt, J. G., and Woodruff, J. L. Size of stimulus symbols in extra-sensory perception. *J. Parapsychol.*, 1939, III, 121-158.
17. Rhine, Louisa E. Some stimulus variations in extra-sensory perception with child subjects. *J. Parapsychol.*, 1937, I, 102-113.
18. Rhine, J. B. The effect of distance in ESP tests. *J. Parapsychol.*, 1937, I, 172-184.
19. Rhine on exclusion of sensory cues. The ESP Symposium at the A. P. A. *J. Parapsychol.*, 1938, II, 254-259.
20. Rhine, J. B. Experiments bearing on the precognition hypothesis. *J. Parapsychol.*, 1938, II, 38-54.
21. Rhine, J. B., Smith, Burke M., and Woodruff, J. L. Experiments bearing on the precognition hypothesis: II. The role of ESP in the shuffling of cards. *J. Parapsychol.*, 1938, II, 119-131.
22. Riess, B. F. A case of high scores in card guessing at a distance. *J. Parapsychol.*, 1937, I, 260-263.
23. Riess, B. F. Further data from a case of high scores in card-guessing. *J. Parapsychol.*, 1939, III, 79-84.
24. Rogosin, H. in the discussion at the ESP Symposium at the A. P. A. (Quotation from a letter of R. A. Fisher.) *J. Parapsychol.*, 1938, II, 266-267.
25. Shulman, R. A study of card-guessing in psychotic subjects. *J. Parapsychol.*, 1938, II, 95-106.
26. Stuart, C. E., and Greenwood, J. A. A review of criticisms of the mathematical evaluation of ESP data. *J. Parapsychol.*, 1937, I, 295-304.
27. Stuart, C. E., A review of recent criticisms of ESP research. *J. Parapsychol.*, 1938, II, 308-321.
28. Stuart, C. E., A review of recent criticisms of ESP research, II. *J. Parapsychol.*, 1939, III, 194-205.

29. Tippett, L. H. C. Tables of Random Sampling Numbers. (Tracts for Computers, No. 15.) Cambridge University Press, 1927.
30. Warner, L. The role of luck in ESP data. *J. Parapsychol.*, 1937, I, 84-92.
31. Warner, L. A test case. *J. Parapsychol.*, 1937, I, 234-238.

A REVIEW OF DR. FELLER'S CRITIQUE

J. A. Greenwood and C. E. Stuart

In the preceding article (5), Dr. Feller has raised three questions about ESP research:

(1) Is the shuffling of the cards in ESP tests adequate to provide a random target series?

(2) Does evaluation of published results, as a whole or in groups, constitute, because of supposedly unpublished chance findings, an improper selection of data for conclusions regarding ESP?

(3) Has the ESP hypothesis been improperly favored by stopping experimental series at a suitable point?

These are not new questions. Their familiarity in the literature of the ESP research is affirmed by the fact that they have been discussed by several of the major critics within the last six years. (20: pp. 414-415). The fact that they have been so discussed makes both the proper statement of the problems and the evidence bearing upon them readily accessible.

We shall examine each of the issues to determine: (a) What ESP research results are exempt from the bearing of the question; and (b) What effect may be expected in research not exempted.

Adequacy of Shuffling

Statement of the Question. Is the hypothesis of inadequate shuffling of the cards in ESP tests adequate to account for the significant results reported? Inadequate shuffling allows persistence of unbroken sequences in successive card orders. Under the hypothesis, these may coincide unduly with habitual sequences of the subject's calls. Or the subject may be supposed to make his calls in patterns deliberately aimed at duplicating sequences previously observed to recur among the cards.

It will be shown here that (a) the major work bearing upon ESP is exempt from any significant bearing of the hypothesis of inadequate shuffling; (b) the hypothesis is inadequate to account for even that minor section of the research to which it is relevant; and (c) the appearance of applicability of this hypothesis, as expressed by Dr. Feller, is contributed by spurious features of his demonstrations.

Work That is Exempt. Much of the more important ESP research is clearly exempt, by virtue of the experimental conditions and the nature of the results, from explanation by the hypothesis of inadequate shuffling. Sufficient ESP work is conceded by Dr. Feller himself to be exempt to satisfy the needs of the ESP hypothesis. He agrees that: "It may be clearly stated that a few experimental series are recorded with so large deviations from chance that they obviously never could be reasonably 'explained' purely statistically. That is to say, no reasonable allowance for the effects discussed in the sequel would essentially change the aspect of those series. The two record series by Riess .... and by Martin and Stribic are not even subjected to the effect of optional stopping ...." and "....I wish to make it perfectly clear that the above consideration [lack of randomness] does not intend to explain .... Warner's experiment as such."

There are other comparable series for which the grounds of exemption are equally good. The more obvious of these are listed in Section B of Table I, with the grounds for exemption noted in each case.

Table I

Work Exempted from Inadequate Shuffling Hypothesis

A. Reports Agreed to be Exempted			
Reported by	Runs	Av. per Run	C.R.
Riess (23)	74	18.24	53.6
Martin and Stribic, I (13)	1000	6.89	29.9
Martin and Stribic, II (14)	1000	7.38	37.7
Warner (28)	10	9.30	6.8
B. Other Reports Obviously Exempted by Conditions or Controls*			
Rhine (Pearce-Pratt Series) (19)	74	7.53	10.9
Rhine (PT) (19; 20: p. 416)	26	6.81	4.6
Murphy and Taves (15)	2000	5.04	1.0
Pratt and Woodruff (16)	2400	5.20	5.0
Soal (26)	Displacement Study		
Carington (4)	Analysis of Drawings		

* The Pearce-Pratt series and the Pure Telepathy series are both high-scoring experiments in which cross-checks show no effective

It is worth noting that every one of the series discussed in Chapter VI of *Extra-Sensory Perception After Sixty Years* as crucially supporting the ESP hypothesis falls into Table I, two of them with Dr. Feller's agreement. It should be kept clear, therefore, that the discussion of this question from this point on is a largely academic one, so far as the status of the ESP research is concerned.

The Question Answered. Although it is clear from the above table that the *a priori* likelihood of inadequate shuffling, as explaining any important research report, is small, the hypothesis itself may be dealt with in terms of its three aspects: (1) the theoretical expectation; (2) authoritative judgment; (3) and the evidence from empirical statistical checks.

(1) That mathematically perfect randomness is theoretically impossible using card shuffling as a basis can be proved to be true. This proof, however, simply affirms the fact that no mathematical formulation can be rigorously paralleled in any set of actual observations. This completely skeptical attitude has always been applicable but has never been useful in applied mathematics. The closeness of approximation of a real situation to a mathematical hypothesis is customarily a matter of belief based upon trial and observation of the hypothesis in question as applied. If the results are uniformly observed to be the same as if the hypothesis were

---

(Footnote continued) degree of linkage. The Pratt and Woodruff series is a low-scoring experiment, but the precautions observed affirm the validity of the cross-check demonstration that no effective degree of card-linkage recurred.

Of the Murphy and Taves work, Dr. Feller says: "These authors have devoted a careful study to the problem of adequacy of shuffling and the problem of linkage in their experiments..... Their results have, however,.... no bearing on our present problem." Literally interpreted, this comment is contradictory since "a careful study" could scarcely yield results of no bearing. However, the problem of linkage is shown to have no reasonable bearing upon their covariation results.

Soal's experiments utilized card orders specially randomized from tables of numbers. Carington's drawings were randomized by a special technique of selection, the validity of which has never been challenged.

The number of items in this table could be greatly increased by discussion of less obvious cases, but such elaboration is unnecessary since we are concerned here only with the demonstration that *at least one* significant research report may be excepted from the inadequate shuffling hypothesis.

actually satisfied, then we are content to use the theoretical measures as a means of evaluating the practical results. Our confidence in so doing increases with the number of verifications which have accumulated, without any exception of extra-chance character.

If poor shuffling is postulated, then there will be some kind of linkage observable between successive card orders in an experiment. This poor shuffling may result from mechanical features of the cards which foster retention of certain orders and avoidance of others. If subject preferences accidently retain orders similar to those retained in the cards, the effect of this linkage will result in the variance of correspondences between calls and cards being different from the variance of random expectation. Let us denote this as the hypothesis of *objective linkage*.

Another theoretical consideration is that shuffling techniques may foster retention of certain *changes* in order. The subject, if allowed to see the cards or the card records, might observe carefully the changes in previous card orders and thereby infer information about succeeding card orders. The effect of this, of course, would be to increase the number of correct correspondences. Let us denote this hypothesis as that of *subjective linkage*.

(2) That shuffling is adequate to produce a practical degree of randomness of certain types has been the considered conclusion of outstanding authorities in the mathematics of probability who have considered this problem. Emile Borel (2: p. 33) and Paul Levy (10: pp. 48-49) have independently concurred that a reasonable number of shuffles of elementary types produce effective individual card randomness. That is, the probability that any specified card shall be in any specified place in the deck is $1/n$, where n is the number of cards in the deck. This view would rule out the effective possibility of subjective linkage* and permit the problem to be wholly concerned with the hypothesis of objective linkage.

(3) In order to meet directly the question of objective linkage it is necessary to know the empirical variance of ESP studies. The variance we need must be

* Subjective linkage is, of course, not possible in those experiments in which the subject did not have any knowledge of the card order or of past records (e.g., Riess (23), Warner (28), Rhine enclosed card series (17)).

based on some aspect of the data which excludes possible changes in variance introduced by the hypothetical ESP factor itself. This variance is directly available from cross-check studies. If such patterns persist and card orders persist, then a cross-check involving the same population of cards and calls as the ESP test but with calls checked against card orders other than those intended will give the empirical variance we want. If objective linkage is effectively present, this variance should be significantly different from expectation upon the hypothesis of randomness.

Of the six crucial ESP series, the five which produced significant C.R.'s were cross-checked. None of these cross-checks gave a significant average, and none gave a variance significantly different from binomial expectation of 4.00. The results of these cross-check studies are listed in Table II.

TABLE II

Cross Check Data of Crucial Series (20: p. 43)

Series	Method of Cross-Check	Runs	ESP Av.	Check Av.	Check Variance	p
Rhine (P.T.)	Calls vs. 3rd card run preceding	26	6.80	4.88	3.64	.53
Rhine (Pearce-Pratt)	Calls vs. 3rd. card run following	74	7.53	5.20	4.81	.10
Pratt and Woodruff	Calls vs. 3rd. card run following	2400	5.20	5.02	4.04	.37
Warner	Calls vs. 3rd. card run following	10	9.30	5.60	4.04	.34
Riess	Calls vs. 4th card run following	74	18.24	4.78	3.44	.77

The last two columns of Table II give the evidence relevant to the linkage question. Under "Check Variance" are the observed variances of the cross-checks. To find whether these were greater than chance expectation, each was substituted in the formula

$$\chi^2 = \frac{Ns^2}{\sigma^2}$$

wherein N is the number of runs, s^2 is the observed variance, and σ^2 is the theoretical variance of 4.00

expected on a binomial hypothesis. The last column gives the p-values of the χ^2 values so computed. The figures denote the probability that a random sample from a population with variance 4.00 would give an observed variance greater than the one tabulated. If p were less than .01, the observed variance would be significantly greater than chance expectation. None of the p-values noted are significant.

Although the variances appear quite different and are drawn from somewhat heterogeneous samples, a test of their homogeneity* yields a p-value slightly greater than .40, showing that the differences among the variances are attributable to chance variation.

Since the variances observed are negligibly different from those expected upon the hypothesis of random distribution, it may be concluded that any lack of randomness in the shuffled cards was not effective in any way upon the results.

A relevant, but somewhat less crucial study which supports the results just stated has been made of the total chance check data of Table II of *Extra-Sensory Perception After Sixty Years*. The C.R. of each of the 24 chance checks was found. The hypothesis was tested that each of them was a random sample from a normal universe with a mean of zero and a variance of one. This was done in two ways: forming a chi square by squaring the critical ratios, and summing with 24 degrees of freedom; and by using the approximate standard deviation of a standard deviation. *Both methods yielded probabilities greater than .01.* If the variance applicable to these individual chance check series were appreciably greater than 4.00, then computing the individual C.R.'s using that variance should have resulted in a dispersion of the C.R.'s about zero significantly greater than the expected value of one.

<u>Dr. Feller's Position</u>. Why, then, it is pertinent to ask, do the conclusions of the foregoing discussion vary so greatly from those implied by Dr. Feller? The main point of departure appears to be Dr. Feller's major attention to supposed illustrations of card linkage so superficially plausible that their actual bearing was never tested. Unfortunately, these illustrations were themselves misleading, as may be seen in Dr. Feller's discussion of them.

---

* Rider (22: p. 102)

Dr. Feller examines the symbol orders of a record sheet from MacFarland's data (8). The 9th and 10th runs of that record were as follows:

(9)	wrsws	wpcss	spcrr	cpprr	wccwp
(10)	prrwc	cwpcs	sswpp	rcrwr	swscp

Considering the 9th order as numbered serially from 1 to 25, he points out that it is possible to arrange those numbers to match the proper symbol in the 10th order in the following way:

(9-10) 18,19,20,21,22 23,24,25,8,9 10,11,6,7,12
14,13,15,1,2 3,4,5,16,17

Of this Dr. Feller says: "It should be borne in mind that the deck was cut: thus after the shuffling proper the first card of the rearranged deck lay below the last one. In other words, the shuffling of the 9th arrangement of the deck left the group of cards numbered 16-25 (i.e., 10 cards) in the same order; in addition, the 5 cards numbered 1-5 clung together; likewise, the 4 cards 8-11."

In these remarks an inaccuracy of only one word occurs; namely, "cards" is used in place of the proper word "symbols." But this apparently trivial error actually introduces a major assumption: *that the order of numbers listed shows the true relation, card for card, between the 9th and the 10th target decks.*

Actually there are 24,883,200,000 possible orders that the *cards* might have taken to produce the order of *symbols* observed. The order of *numbers* Dr. Feller lists is a deliberately selected one of these possibilities. It is that order which would require the least reordering of cards to produce the observed symbol order; or, in other words, that arrangement in which the most numbers can be written consecutively in their customary order. But the *a priori* probability that this will be the true order is 1/24,883,200,000. And even if the cards were poorly shuffled, it would still be extremely improbable that the shuffling would produce *the least possible* disturbance of the cards. It is, therefore, virtually certain that Dr. Feller's analyses in no case presented the actual rearrangement of the cards.*

* In fact the orders are, of course, deliberately selected to be those apparently most favorable to the inadequate shuffling hypothesis.

But if the number orders presented are false insofar as they represent the actual card orders, these deductions from them regarding shuffling or card linkage are completely erroneous no matter how convincing the deductions appear. Such statements as ".... the same groups of cards show a tendency to cling together...." and ".... it is probable that the deck was changed, or fell to the floor, between the fifth and sixth runs...." describe nothing that happened in the experiment.

The lack of relevance of these deductive statements can be readily illustrated. We construct the following successive ESP deck orders from Tippett's table of random numbers (27). (Twelve pairs of runs were observed.) In this case there is no shuffling or cutting, and there is no possible relation between the *card* orders.

Order 1	prprs	rcrcc	wwwrp	pcswc	wspss
Order 2	wspws	ppspr	prsrc	crwwc	rscwc

Although we know that there is no relation between the cards in these series let us assume that Order 2 is the result of Order 1 after shuffling and, following Dr. Feller's analysis, give Order 2 numbers, assuming, as he did, that the highest possible numbers of cards remain together.

(1-2) 21,22,23,13,24 15,16,25,1,2 3,4,5,6,7 9,8,11,12,10 14,18,17,19,20

On Dr. Feller's assumption that these numbers represent the *true* relation between the cards, we can reconstruct easily the supposed act of shuffling which produced the second order. It is evident that the deck was finally cut between the seventeenth and eighteenth cards; the order before the cut being, therefore:

1,2,3,4,5 6,7,9,8,11 12,10,14,18,17 19,20,21,22,23 13,24,15,16,25

It is evident, further, that the shuffling displaced only four cards, the 10th, the 13th, and the 15th and 16th as a pair. Two other pairs (8 and 9; 17 and 18) were permuted without displacement. Since the shuffling *disturbed only eight cards in the middle of the deck,* we can deduce with confidence the actual shuffling act, which appears to have consisted of the single act of lifting the middle section of the deck loosely and sliding the lifted cards back into the deck with two or three motions.

We readily recognize these deductions as ridiculous when we know that there was no shuffling and cutting of cards at all. But if these orders had occurred among ESP cards randomized by shuffling, there would be no such protection against the persuasive force of these deductions. We could easily, like Dr. Feller, agree that "after this example it can hardly be denied that linkages of the worst sort do occur."

Dr. Feller was aware of the faults in his illustrations, and he notes correctly that: "Strictly speaking, it cannot be logically proved that the cards actually did stick together." But how, then, does an example show "that linkages of the worst sort do occur?" This contradiction is never resolved. It is evident, therefore, that the approach does not provide a basis for rational judgment regarding the first of his three major questions.*

Resumé. In answer to the question: "Is the shuffling of the cards in ESP tests adequate to provide a random target series?" it has been possible to show that shuffling is adequate within the degree of randomness required. (a) The lack of randomness postulated in this que tion is not adequate to explain significant results in numerous crucial ESP experiments. (b) The probable effect of lack of randomness, even if assumed to be present in some instances, is shown to be negligible. (c) The variation of these conclusions from those implied by Dr. Feller evidently follows from his major dependence on spurious demonstrations of card linkage.

Completeness of Publication of Results

Statement of Problem. Is there a sampling error due to biased selection through supposed publication of only positive ESP results? If the results of all experiments are not published, may not those which do achieve publication be the favorable ones with positive conclusions? As a sample of "all ESP results," they would accordingly constitute a biased sample. Any attempt to average or pool published results may, then, according to these hypotheses, lead to erroneous conclusions because of improper selection.

* It should be emphasized that we do not here even consider the problem of whether a significant degree or effect of linkage actually occurs in the MacFarland series (11). The statistical methods for studying linkage in the data are straightforward and well-known. Dr. Feller's analysis is not one of them

We will show that (a) the major work crucially bearing upon the ESP hypothesis is exempt from the hypothesis of improper sampling in publication, and, indeed, that all individual research reports are exempt; (b) pooled results on which the question might bear have been sampled properly; and (c) that the criticism of Dr. Feller could only have arisen from erroneous suppositions regarding the pooling done.

Work That Is Exempt. There has been no question raised as to the right of the individual investigator to report and evaluate his own work as a unit, nor is the right of the investigator to report his work in its experimental subdivisions questioned. Hence, any adequately reported individual research is exempt from the hypothesis of improper selection by incomplete publication. In fact, to list the individual research items properly exempt from criticism on this score would be to present almost the entire experimental literature on ESP. Application of this hypothesis is possible, then, only when results of different published experimental series are pooled or averaged.

The Question Answered. Two kinds of pooling are possible and useful: (1) Pooling such as occurred in Rhine's original ESP monograph (18) and in Soal's recent total report (26). In each case, there is assembled the total result of all the several experiments of an individual investigator, work which may properly be conceived as constituting one large experimental project. It is assumed that he publishes all observations belonging to the project in question.

(2) The second kind of pooling is that which occurs when results of various experiments having a common condition are summarized. This kind of pooling occurs in the tables of *Extra-Sensory Perception After Sixty Years*. These summaries have never been offered, however, as crucially requisite for proving the ESP hypothesis. (For example, the crucial evidence in the book mentioned, given in Chapter VI, is all in the form of individual research series.) The tables are obviously directed to answer a secondary question: What is the effect of combining all the available results of tests of a given type and condition, favorable and unfavorable? They permit, for example, the consideration of the effect of "negative series" cancelling deviations from positive work. A critical ratio from such a pooled table permits intelligent judgment upon these and other secondary questions. This kind of pooling is directly within the bearing of the above hypothesis.

But even though (i) this type of pooling is not essential to the evaluation of ESP data and (ii) though it is not used except as an incidental means of keeping the total data in view, it is clearly a legitimate procedure and may be defended in two ways: (1) It is a simple matter to determine the outer bound of work which could be supposed to exist unpublished, assuming it to have a chance average, without cancelling the significance of the deviations noted in a given table or pool. In fact, Table XII of *Extra-Sensory Perception After Sixty Years* gives the result of such a calculation. The amount of unpublished chance work necessary to reduce Table VII (the one attacked by Dr. Feller) would require 170,000,000 trials. This is almost 200 times the amount of work reported in that table. It would be incontestably ridiculous to suppose that so much ESP research had ever been done, much less that so much that is worthy of scientific consideration could have been conducted without some knowledge of it reaching those interested in publication of ESP research.

(2) A second answer to this question is that in the volume mentioned, Tables VIII-XI constitute the admittedly most important summaries of ESP research. These summaries are made on the basis of the superior methods and precautions used. Since many of these have been developed in the Parapsychology Laboratory and since several of these precautions are of quite recent development, it is extremely unlikely that any completely independent work belonging in those tables could be wholly unknown. For example, all of the research summarized in Table VIII was known in the Parapsychology Laboratory before it was published and no completed work meeting the conditions and known to the Parapsychology Laboratory remained unpublished. There is, thus, every moral certainty that Tables VIII-XI actually include all work done under the experimental conditions required for inclusion. These tables constitute, therefore, permissible samples of ESP research and their significantly positive results cannot be attributed to selection of only positive cases for publication.

Dr. Feller's Position. Again, the difference between these conclusions and those suggested by Dr. Feller should be accounted for if possible. We believe this can be done by pointing to the natural but erroneous supposition of the critic that all experimenters with positive results hasten to publish and all with negative or chance results hesitate to do so. The erroneous character of this supposition is indicated in *Extra-Sensory Perception*

After Sixty Years, p. 75, where it was noted that of known and deliberately unpublished reports, thirteen had results favorable to the ESP hypothesis and seven had chance results.

In the second place, Dr. Feller implies that the case of ESP is necessarily based upon pooled data. This is, as stated above, a mistaken assumption; in ESP research--as in all other research--the basic unit of evidence is the original research report.

Resumé. In considering the question of sampling error due to biased selection in publication, the following points are noted: (a) All the crucial ESP evidence is exempt from the bearing of this hypothesis. (b) The hypothesis is shown to deal only with a secondary question: the right to summarize arithmetically the reports published. It is then shown that the summaries attacked are duly safeguarded against the source of error implied in the hypothesis. (c) Dr. Feller's different conclusion appears to depend upon two erroneous suppositions: that only positive results are published and that the case of ESP is based upon summarized data.

Optional Stopping Effects

Statement of the Question. Has the ESP hypothesis been improperly favored by stopping experimental series at a suitable point? The customary statistical formulas for significance assume that the stopping point in an experiment is independent of the statistical characteristic being measured. In ESP tests we know that chance fluctuations alone may occasionally produce runs of scoring above or below the expected chance average. The experimenter, by watching for these chance variations might, it is supposed, under the optional stopping hypothesis, produce consistently positive or consistently negative averages simply by stopping his experiment at a favorable point.

It will be shown here that (a) the major part of the work crucially bearing upon the ESP hypothesis is exempt from this hypothesis; (b) statistical methods have been devised to correct for this hypothesis and in practice these corrections have been trivial; and (c) the apparent bearing of this hypothesis has been exaggerated by Dr. Feller and the correct criteria mis-applied.

Work That is Exempt. There are several grounds for exempting work from the adverse bearing of the optional stopping hypothesis. Not all of these need be invoked and discussed here. And it is not necessary to list all the exempt series, especially those requiring lengthy description of conditions, since the following instances of exception are ample to maintain the case for ESP.

Dr. Feller's own concession that "the two record series by Riess .... and by Martin and Stribic are not even subjected to the effect of optional stopping...." permits exception of reports by

1. Riess (23)
2. Martin and Stribic, I (13)
3. Martin and Stribic, II (14)

Cases in which scoring rate was so high that the data remained positively significant at all points after the first few runs do not permit effective option. This requirement excludes reports by

4. Rhine (Pearce-Pratt Series) (19)
5. Woodruff and George (29)
6. MacFarland and George (12)

One experiment fully designed to limit option and to permit statistical correction for the limited option was the report by

7. Pratt and Woodruff (16)

Other series depending for their significance upon post-terminal analyses, and in which current knowledge of the statistical item was unavailable, include those by

8. Carington (4)
9. Soal (26)

Since, thus, there are major crucial experiments favoring the ESP hypothesis whose significance is not explainable by the optional stopping hypothesis, it is evident that optional stopping is not a necessary condition for significance in ESP research.

The Question Answered. Since the case for ESP is not involved in it, the actual technical issues of optional stopping may be viewed in a better perspective. There are two points to be cleared up at the outset: first, even among the ESP series not specifically exempted above, there is none in which optional stopping is known to have been used as a matter of policy; that is, in which the experimenter has continued the experiment to the point where it reached a significant C.R. and then stopped the experiment at that point. *Ample evidence of this lies in the wide distribution of C.R's, very few of them being marginal* (see, for example, Table XXIX, *Extra-Sensory Perception After Sixty Years).* The question, therefore, must deal with the possible effect of stopping in spite of the fact that the attainment of a given C.R. probably seldom, if ever, actually determined the stopping point.

Second, the optional stopping hypothesis is relevant (if at all) only in regard to the point at which an experimental series, *taken as a whole,* is terminated; that is, the end point of the statistical data upon which the experimental conclusions are based. Thus, for example, if an experiment includes work with ten subdivisions and the experimenter bases his conclusions upon the total work rather than upon the individual subseries (e.g., work of individual subjects), then an optional stopping hypothesis may be applied only to the total results. The limits of individual subject performances are not stopping points from the viewpoint of the work as a whole and are accordingly irrelevant to the hypothesis.*

Fortunately there are means of answering the question of the effect of optional stopping in ESP tests to an extent at least sufficient for present needs. There are two statistical techniques developed by which the maximum possible effect may be computed. Both of these are aimed at revising the probability obtained upon a simple sampling hypothesis to permit the assumption that option was used to the extent allowed by the methods. They require only a slight limitation in experimental procedure.

The first method, suggested by Greville (7), necessitates the preassignment of the length of the projected series to be between two rather wide-apart limits. Greville gives three such possible categories of choice and with his formula could compute any desired number of them. The experimenter is enabled to select the largest C.R. obtained in the prescribed interval. The Greville

* This may be easily seen if we consider a chance series to be made up by a number of throws of dice. Let us assume that we have a stock of ten pairs of true dice. For the chance series, it would be irrelevant if, for a time, one pair, and later another, were used to provide the data. A change of subject is the same as change of dice; in a chance series there could be no effect from option in such changes.

formula then gives an *upper bound* to the probability that by chance a C.R. as large or larger than the observed one could have been found in the prescribed interval of trials. This upper bound is *not an approximation to the true probability*. And since the departure from the true probability is a function of the number of trials in the prescribed interval, it would be an advantage to be able to decrease the size of the interval in the experiment.

The practical difficulty of the restriction of interval led to another method suggested by Greenwood. The difficulty could be partly met by grouping the trials into as large blocks as convenient and treating the blocks as single trials. If the size of the blocks were determined at the outset, as well as an upper bound to the number of blocks, option would be possible, but so limited as to permit straightforward treatment. The value nP, wherein n is the upper bound to the number of blocks and P the probability obtained as of simple sampling, is an obvious upper bound to the true probability. In addition, if a little more restriction can be conveniently agreed upon by the experimenter (such as the second block being half as large as the first) a more accurate approximation to the true probability can be found.*

These methods are of recent development; so their application has been possible only in the most recent ESP research. The second method discussed was applied to the Pratt-Woodruff series (16). The probability of obtaining the observed result was upon a simple sampling hypothesis, 3×10^{-7}. The probability corrected for optional stopping was 5×10^{-6}. It was thus possible to establish the fact that, *assuming the full effect of optional stopping in that experiment, the significance of the result was not altered by the assumption.*

This application (though not its applicability) is unique in the field of ESP research. We know of no other case in any field using the critical ratio method in which an experiment has been designed and evaluated to rule out the optional stopping hypothesis.

Dr. Feller's Position. Dr. Feller's discussion of optional stopping is misleading because of his

* The mathematical treatment of this method will be dealt with in full in an article devoted to general discussion of optional stopping and optional selection methods, soon to be presented for publication.

improper assumption that the *probable effect* of such stopping is known, whereas merely an *upper bound* is at present obtainable. This is evident in his treatment of Greville's formula.

Feller quotes Greville's statement that with optional stopping a C.R. of 5.5 would be "roughly equivalent" to the customary 2.5, but leaves to a footnote the content of Greville's very important qualification in his succeeding sentence: "It should be emphasized, however, that the probabilities given are *merely upper bounds, not approximations.*" (7: p. 90) (Italics ours.)

Thus Greville explicitly emphasizes the fact that his development dealt only with the establishment of a *sufficient* condition for elimination of optional stopping effects. Dr. Feller indicates that he is aware of this qualification; yet his discussion and application of the Greville criterion are based entirely upon the assumption that it is a *necessary* condition for significance. His inference that a C.R. of 5.5 has a chance probability of .008 is a specific elaboration of this mistaken assumption.

The Greville formula does not give a chance probability value to any C.R. either above or below the criterion, and Greville himself has indicated that it does not presuppose a chance explanation of C.R. values less than 5.5. All of Dr. Feller's uses of the Greville criterion make the latter improper assumption.

A further instance of improper use of the optional stopping hypothesis occurs in Dr. Feller's attacks upon the validity of the Shulman (25) and the L. E. Rhine (21) experiments because of the probability of optionally stopping the work of the individual subjects. In neither report did the experimenters base their conclusions upon the assumption that the work of individual subjects was individually significant. Correction for the effect of optional stopping in these experiments must, therefore, be applied to the results which are claimed as significant; that is, the total pooled data of the subjects with L. E. Rhine (and the pooled data for each subdivision in Shulman's report). From the standpoint of optional stopping, since the series did not stop with individual subjects, the correction would be the same as if one subject had done all the work.

<u>Resumé</u>. In regard to the question: Has the ESP hypothesis been improperly favored by stopping experimental

series at a suitable point? we have noted: (a) That many important ESP series are exempt from this hypothesis. The possibility of optional stopping effect is therefore not a necessary condition for significant results in ESP research. (b) Statistical methods have been devised to correct for this hypothesis when the assumptions of simple sampling have been used. In practice these corrections have been trivial in their effect upon ESP research conclusions. It is reasonable to conclude, therefore, that the probable effect of optional stopping in research series from which it is not excluded is not sufficient to explain the significant results reported. (c) That Dr. Feller's discussion of the hypothesis has exaggerated the probable effect because of false assumptions leading to misapplication of the corrective criterion.

Minor Criticisms

In the foregoing we have been concerned with the three major questions raised by Dr. Feller's article. There remain a number of minor criticisms either specifically offered or clearly implied by his remarks. With the major issues resolved, these minor points are in many ways trivial, but attention to them is justified for the sake of completeness.

(1) Dr. Feller's treatment of the question of biased sampling implies that ESP investigators exclude from report relevant data because the results are not above chance.

This hypothesis is specifically dealt with in *Extra-Sensory Perception After Sixty Years* (pp. 118-121). The resumé here may therefore be brief and fragmentary: (i) Certain work (e.g., that of Pratt and Woodruff) was safeguarded against this possibility by the device of serially numbered record sheets: (ii) other series (e.g., that of Pearce-Pratt), by duplicate records independently preserved; (iii) still others (as that of Warner) had two experimenters present until the entire series (relatively short) was over and the total known; (iv) and finally, one such as the Riess series with scores so high that the loss of data could not have accounted for it in any case. Many authors explicitly state that all the test results are included in the report. There is no good reason known for not accepting an author's statement on such matters.

(2) Dr. Feller insists that many events which are generally thought of as unusual are really not unusual when the broad field of probabilities is considered. Insofar as this point of view constitutes a general warning against too ready acceptance of an unusual event as significant of some unusual process, there can be no objection to emphasis upon this aspect of probability theory. It is necessary, however, to differentiate carefully the three implications of this view. The first implication---that probability theory never provides basis for certainty of conclusion in research--is self-evident, as we have noted previously. The second--that judgment of the degree to which an event is unusual should be based upon intelligently conceived probability theory--is sound scientific procedure. The third implication--that since any probability, no matter how small, may be regarded as possible in the long run, it should not, therefore, be considered meaningful at all--is an unsound confusion of these first two implications.

(3) The statement that "sometimes conclusions on the nature of ESP were drawn from material including no single significant case," implies that such a conclusion is improper. *If so, this criticism would apply to all statistical analysis.* ESP conclusions are always and necessarily drawn from material that may be subdivided into units, no one of which is significant. A single call with a one-fifth probability can never be singly significant. Individual runs are seldom significant, even in highly significant experiments. Statistical methods are devised to derive conclusions from grouped data, a necessity which would not arise if each individual datum were adequate to establish the conclusion.

(4) Dr. Feller says that "ESP advocates have rejected well-founded and almost obvious criticisms ... (specifically) the so-called effect of optional stopping, which has been pointed out by several critics."

Lacking a specific reference by Dr. Feller, we can only say that we know of no ground for this statement. ESP advocates have, quite properly, rejected the optional stopping hypothesis as an acceptable counter-hypothesis for crucial ESP research, as we have done here. On the other hand, the question of the effect of optional stopping upon the critical ratios in ESP tests has been a specific research problem and the Parapsychology Laboratory has openly fostered both discussion and research on this problem for several years. (3; 6; 9)

(5) Dr. Feller quotes M.S. Bartlett and R. A. Fisher as authorities expressing the need of caution in arriving at conclusions about ESP from statistical demonstrations. Review of the context (1) of the Bartlett remark reveals that he was emphasizing the difficulty of ascribing an observed significance to any real factor or process in an experiment. The Fisher quotation is similarly the affirmation of a general evaluative view in statistics--that consistent and repeated demonstrations of significance with reasonably small probabilities yield information of a more useful kind than a few observations of significance with extremely small probabilities. It should be noted that in both cases the writers are emphasizing dependence of statistical *interpretation* upon the controls observed in experiments. *In no sense are they challenging either the validity of the statistical methods or the experimental results in ESP research.* Fisher's approval of the soundness of the statistical methods was specifically given in 1935 (18B: p. 44).

Of course the ESP research may revert to a common sense evaluation by comparison of score averages and cross-check averages, and thus avoid the issue of statistical interpretation entirely.

A fair test of the fruitfulness of a critical review is the amount of change necessary in contemporary methods of research to avoid the faults indicated. From our study of Dr. Feller's critique, it appears that no changes in method, either experimental or statistical, are required.

Summary

The three major issues raised in Dr. Feller's discussion concern the effect upon the conclusions of ESP research of (1) inadequate shuffling; (2) biased sampling from incomplete publication; and (3) optional stopping.

We have shown that there is a considerable amount of experimental work manifestly exempt from explanation by these hypotheses.

We have gone further to show that when each hypothesis is considered in terms of the experimental evidence on ESP, the probable effect in no case constitutes sufficient condition for the research results observed.

It is our judgment that the chief fault of Dr. Feller's argument has arisen from his large preoccupation with plausible illustration (not the actual effect) of the hypotheses. This gave the misleading impression that the issues concerned the *possibility* of inadequate shuffling, biased sampling, and optional stopping. The actual issues, the *probable real* effect of the hypotheses within ESP research, he left almost completely undiscussed.

References

1. Bartlett, M. S. The present position of mathematical statistics. *J. Royal Stat. Soc.*, 1940, 103, 1-19.
2. Borel, É. Theorie Mathématique du Bridge. Paris: Gauthier-Villars, 1940.
3. Bugelski, R., and Bugelski, S. A further attempt to test the role of chance in ESP experiments. *J. Parapsychol.*, 1940, 4, 142-148.
4. Carington, W. W. Experiments on the paranormal cognition of drawings. *J. Parapsychol.*, 1940, 4, 1-117.
5. Feller, W. Statistical aspects of ESP. *J. Parapsychol.*, 4, (This Number).
6. Greenwood, J. A. An empirical investigation of some sampling problems. *J. Parapsychol.*, 1838, 2, 222-230.
7. Greville, T. N. E. A summary of mathematical advances bearing on ESP research. *J. Parapsychol.*, 1939, 3, 86-92.
8. Kennedy, J. L. A critical review of "Discrimination shown between experimenters by subjects" by J. D. MacFarland, *J. Parapsychol.*, 1939, 3, 213-225.
9. Lemmon, V. W. The role of selection in ESP data. *J. Parapsychol.*, 1939, 3, 104-106.
10. Levy, P. Calcul des Probabilités. Paris: Gauthier-Villars, 1938.
11. MacFarland, J. D. Discrimination shown between experimenters by subjects. *J. Parapsychol.*, 1938, 2, 160-170.
12. MacFarland, J.D., and George, R.W. Extra-sensory perception of normal and distorted symbols. *J. Parapsychol.*, 1937, 1, 93-101.
13. Martin, D.R., and Stribic, F.R. Studies in extrasensory perception: I. An analysis of 25,000 trials. *J. Parapsychol.*, 1938, 2, 23-30.
14. Martin, D.R., and Stribic, F.P. Studies in extrasensory perception: II. An analysis of a second series of 25,000 trials. *J. Parapsychol.*, 1938, 2, 287-295.
15. Murphy, G., and Taves, E. Covariance methods in the comparison of extra-sensory tasks. *J. Parapsychol.*, 1939, 3, 38-78.

16. Pratt, J. G., and Woodruff, J. L. Size of stimulus symbols in extra-sensory perception. *J. Parapsychol.*, 1939, 3, 121-158.
17. Rhine, J. B. ESP tests with enclosed cards. *J. Parapsychol.*, 1938, 2, 199-216.
18. Rhine, J. B. Extra-Sensory Perception. (A) Boston: Bruce Humphries, 1934; (B) London: Faber and Faber, 1935.
19. Rhine, J. B. Some basic experiments in extra-sensory perception. *J. Parapsychol.*, 1937, 1, 70-80.
20. Rhine, J. B., Pratt, J. G., Stuart, C. E., Smith, B. M., and Greenwood, J. A. Extra-Sensory Perception After Sixty Years. New York: Holt, 1940.
21. Rhine, L. E. Some stimulus variations in extra-sensory perception with child subjects. *J. Parapsychol.*, 1937, 1, 102-113.
22. Rider, P. R. An Introduction to Modern Statistical Methods. New York: John Wiley and Sons, 1939.
23. Riess, B. F. A case of high scores in card guessing at a distance. *J. Parapsychol.*, 1937, 1, 260-263.
24. Riess, B. F. Further data from a case of high scores in card-guessing. *J. Parapsychol.*, 1939, 3, 79-84.
25. Shulman, R. A study of card-guessing in psychotic subjects. *J. Parapsychol.*, 1938, 2, 95-106.
26. Soal, S. G. Fresh light on card-guessing--some new effects. *Proc. Soc. Psych. Res.*, 1940, 46, 152-198.
27. Tippett, L.H.C. Random Sampling Numbers. (Tracts for Computers, No. 15). London: Cambridge Univ. Press, 1927.
28. Warner, L. A test case. *J. Parapsychol.*, 1937, 1, 234-238.
29. Woodruff, J. L., and George, R. W. Experiments in extra-sensory perception. *J. Parapsychol.*, 1937, 1, 18-30.

J.A. Greenwood
Dept. of Mathematics
Duke University

C.E. Stuart
Parapsychology Laboratory
Duke University

A PERCEPTION RATIO STATISTIC FOR ESP TESTS

A. A. Foster

Experimenters reporting results of ESP tests have used gross results, average hits per run, and critical ratios as illustrating extra-chance scoring. They have not, however, attempted to state quantitatively a measure of the extra-chance factor significant scoring purportedly reveals.

A simple quantification of the perception involved in an ESP test is the number of symbols actually identified by the subject. These cannot be directly counted, but the number may be approximated by inferring from the total score how many symbols had to be identified in order to attain that score. Let the total number of symbols or trials be N and the number of kinds of symbol and their mode of presentation be such that the expectancy of successfully naming a symbol by chance alone is p; i.e., the expectancy is pN if chance alone is operative. Now if X symbols are actually perceived, these X would be correctly identified and thus would constitute "hits." There remain N - X symbols which have not been perceived, and out of this number the chance expectation is p(N - X) hits. If H denotes the total number of hits, X the number of percepts, and C the number of hits actually occasioned by chance, we have

$$H = X + C \qquad (1)$$

The most probable value of C is p(N - X). The exact value, however, will be p' (N - X), wherein p' differs from p by some small chance deviant.

Equation (1) may be restated

$$H = X + p'(N - X) \qquad (2)$$

in which H and N can be observed. The equation solved for x becomes

$$X = \frac{H - p'N}{1 - p'} \qquad (3)$$

From the Bernouilli Theorem it is deducible that for a sufficiently large N, provided N - X is also sufficiently large, the error of substituting p for p' is negligible with relative certainty. With the theoretically known p substituted for p', equation (3) gives an estimate of x or the number of "percepts" that have occurred.

It is desirable to express the result as a ratio in order to facilitate comparisons between experiments. Mathematically it is simplest to use

$$\text{Perception Ratio} = \frac{X}{N}, \text{ or}$$

$$\text{P.R.} = \frac{H - pN}{(1 - p)N} \tag{4}$$

It is often convenient to state the result in terms of the percentage of the symbols which have been perceived. This suggests as a useful term the "ESP Quotient":

$$\text{ESP Q} = \frac{100\ (H - pN)}{(1 - p)N} \tag{5}$$

The above formula may be derived by a somewhat different line of reasoning, as follows: Suppose H hits to have been made. The excess over expectancy is given by the relation H - pN. It is reasonable to suppose that, when N and N - X are large, at least that many percepts have been made. However, if H - pN percepts have been made in such places as to add to the probable chance score, it is reasonable to suppose that percepts have occurred in similar ratio among those cases already selected by chance to be hits. To obtain a measure of percepts we should therefore increase the excess of hits over expectancy in the ratio of N to N - pN which gives

$$X = H - pN \times \frac{N}{N - pN} \text{ or}$$

$$X = \frac{H - pN}{1 - p} \text{ as before.}$$

For the particular case of ESP cards, the value of p is 1/5;

$$X = \frac{H - \frac{1}{5}N}{\frac{4}{5}} \tag{3a}$$

and equation (4) becomes

$$\text{P.R.} = \frac{H - \frac{1}{5}N}{\frac{4}{5}N} \tag{4a}$$

if h be used to denote average hits per run of twenty-five, we may write

$$P.R. = \frac{h - 5}{20} \tag{4b}$$

and equation (5) becomes

$$ESP\ Q = 5(h - 5) \tag{5a}$$

At this point it is imperative to consider the accuracy and applicability of the foregoing formulas within ESP test results. The doctrine of economy of hypotheses would bar their use in results which did not produce a deviation from chance expectation to a significant degree. That is, there must be in the data relative certainty of the presence of an extra-chance factor. On the other hand, however, if the measure is applied only to significant series, it is not likely to be a true measure. The latter difficulty arises when, within a group of experiments, some series give significant results and others give non-significant results. If enough of the series are significant, it is proper to postulate an extra-chance factor producing the deviations. But it is also reasonable to expect that of the expected chance deviations within the material the significant group would include somewhat more of the positive cases than the non-significant group. For example, if the extra-chance factors were small and constant throughout the material, then the deviation produced might be added to chance positive deviations to make the total significant, but might be simply reduced to fall within the chance range if added to a chance negative deviation. For this particular problem the selection of only significant series for consideration is not a statistically permissible sample.

It must therefore be made explicit that use of the ESP Q should depend upon the validity of two assumptions: (1) that ESP has occurred within the data under examination, and (2) that from the standpoint of a chance hypothesis this data are a statistically permissible sample. Under these circumstances the ESP Q is, while not an exact measure, the <u>best</u> approximation to the quantitative effect of a postulated ESP process.

A further question is that of reliability of the ESP Q. The statistical constant needed is the standard deviation. Before presenting any such measure in this case it is important to stress the fact that ordinarily two rather different variates are collectively described by a single S.D. For example, it might be stated that the mean weight of a certain animal was 20 gms.; S.D., 5. This S.D. is almost wholly a measure of the range of size for the creature concerned since the actual weight at any one measurement could be determined within what are comparatively very narrow limits. On the other hand, when the atomic weight of an element is quoted at 1.00; S.D., .0013, it is tacitly assumed that the atomic weight is a constant and that the different values found for it represent experimental error only. Or to cite an intermediate case, a psychologist might, after employing several tests, state the I.Q. of a certain group to be 105; S.D., 12. Here the S.D. would express both the inherent variance within the group and the uncertainty attending the use of the tests employed.

Any S.D. associated with an ESP Q would be most comparable to the last instance cited above with the addition of certain complicating factors peculiar to the type of test conducted. Since a major use of S.D. is to afford a measure of significance of difference in experimental results by noting the ratio

$$\frac{M - M_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

it would be necessary to be sure that the means compared were actually comparable so that their corresponding S.D.'s are likewise comparable. For example, suppose we have

Group	Technique	ESP Q	S. D.
A	x	Q_1	σ_1
A	y	Q_2	σ_2
B	x	Q_3	σ_3
B	y	Q_4	σ_4

Q_1Q_2, Q_1Q_3, Q_2Q_4, and Q_3Q_4 are comparable; whereas Q_1Q_4 and Q_2Q_3 are not unless control groups be built up. Several different features might be illustrated by the use of an ESP Q and its S.D. Suppose, for instance, an experimenter had had A subjects make B runs of twenty-five cards each. It would be possible to state:

1. That the ESP Q of a given subject for that experiment was so and so with a S.D. based on the root mean square deviation of the ESP Q values indicated by the successive h value which that subject actually achieved.

2. The ESP Q of the group might be given with a S.D. compounded of the sigmas of the individual subjects.

3. Another investigator might group the ultimate findings above with those of other experiments and arrive at a grand mean ESP Q with its associated S.D. computed from the σ values of the constituent quotients. Obviously, their categorical limits would have to be strictly reckoned with in any comparison of values.

"Low-aim" scores in which the subject attempts to avoid "hits" may be evaluated similarly to the customary method. For low-aim, the p value for ESP cards becomes 1 - 1/5 and formula (4b) becomes

$$P.R. = \frac{5 - h}{5} \qquad (4c)$$

This formula is identical with one previously and independently derived by Prof. R. H. Thouless. (Vide "Dr. Rhine's Recent Experiment on Telepathy and Clairvoyance and a Reconsideration of J. E. Coover's Conclusion on Telepathy," Proc. Soc. for Psy. Research, Vol. XLIII, 1935, pp. 24, 32, 37.)

Use of the ESP Q may be exemplified in the following problem. Before the use of ESP cards, frequently used bases for ESP tests were the common numbers from zero to nine. The question arises whether a chance probability of 1/10 in the test material is better than that of 1/5. It was noted that from all results reported of tests utilizing the 1/10 chance basis a total of 34,078 trials gave 3,985 hits. All tests using 1/5 chance basis amounted to 2,758,354 trials with 604,403 hits.[1] Substitution in formula (5) gives ESP Q's of 1.9 and 2.4 respectively. Although the reliability of the values is unknown, the difference suggests the superiority, in terms of the number of items "perceived," of a 1/5 probability in the test material.

1. "Extra-Sensory Perception after Sixty Years," p. 271.

IS ESP DIAMETRIC?

By A. A. Foster

When the ESP test is like that of Blind Matching, wherein two unknown cards are matched in an attempt to produce correct matchings, the perception may consist of the identification of the two cards followed by an inference that they belong together, or the perception may involve the direct experience of "similarity" without necessarily including identification of elements adjudged similar. The former possibility I shall call the "circumferential" hypothesis; the latter possibility, the "diametric" hypothesis.

Let the probability of success in a given operation be p_0 on a basis of pure chance. Let the probability of non-success be q_0. In N trials the expectation is $p_0N = H_0$. Suppose actual performance results in H_1 successes or hits. Assuming that the ratio H/N is reasonably stable for large N, let us denote the value which this approaches as p_1.

Let us assume that p_1 which is of the nature of an empirical probability (such as a death rate) is the result of joint operation of two factors. One of these factors we call chance and denote its rate of production of successes by p_0. The other factor we might call ESP; and denote its rate of incidence by p_2. Let us denote the chance event associated with p_0 by E_0 and the ESP event associated with p_2 by E_2; let E_1 denote an event regardless of its association. It is possible to conceive situations where an an occurrence of an E_1 is occasioned by an occurrence of

(a) E_0 or E_2; then $p_1 = p_0 + p_2$

(b) E_0 and/or E_2; then $p_1 = p_0 + p_2 - p_0p_2$

(c) E_0 and E_2; then $p_1 = p_0p_2$

(d) E_0 and not E_2; then $p_1 = p_0q_2$

(e) E_2 and not E_0; then $p_1 = q_0p_2$

In general $p_1 = f(p_0,p_2)$ wherein, as previously defined, p_0 is the pure chance probability of incidence, and p_2 is the rate of incidence of a perturbing factor, possibly ESP.

If we consider the case of simple matching of a deck of ESP cards to five visible key cards (the Open Matching Procedure), we have the relation

$$p_1 = p_0 + p_2 - p_0 p_2 \qquad (1)$$

since perception and/or chance may occasion a hit. But, as noted above, H_1/T as an observed rate may be considered a measure of p_1. Therefore for ESP cards we have

$$\frac{H_1}{T} = (1/5 + p_2 - 1/5p_2) \qquad (2)$$

whence, for a run of 25

$$\frac{H_1}{T} = \frac{h - 5}{20} \qquad (3)$$

where h is the number of hits or successes in the run. (This is, of course, the ESP Q which was defined and derived by other means in a previous article.)

We will now consider the somewhat more complicated procedure of endeavoring to match cards from a deck to five key cards which are themselves known sensorially (the Blind Matching procedure). If ESP were to follow what we ordinarily call logical steps, the process would be as follows: An attempt would be made to envisage the key cards and then an effort made to envisage the deck cards and place them opposite the assumed corresponding key cards. The probability of a chance success in the first operation we will denote by p_{oK} and that for the second operation we will denote by p_{oD}, and for the complete operation p_{oBM}. For ESP cards we have the chance expectation

$$p_{oK} = p_{oD} = p_{oBM} = 1/5$$

Let the rates of incidence of the corresponding perturbing factors be p_{2K}, p_{2D} and p_{2BM} and let the corresponding rate of incidence be p_{1K}, p_{1D}, and p_{1BM}. Let the corresponding items for open matching be $p_{oOM} = 1/5$, p'_{OM} and p_{1OM}. <u>A priori</u> one would expect p_{2OM}, p_{2K}, and p_{2D} to be of the same order of magnitude. Now if either p_{2K} or p_{2D} be zero, p_{1BM} will tend towards 1/5 as a limit no matter how high the other of the above factors might be. Further, a perfect score over a long series could only be obtained if both p_{2K} and p_{2D} approximated very closely to unity.

In practice the values of p_{2K} and p_{2D}, or even p_{1K} and p_{1D} are not calculable from the records. However, if p_{2K} and p_{2D} exist at all, their product is p_{2BM}. This follows from the fact that if we postulate ESP to take place in steps, p_{2BM} is dependent upon coincidences of p'_{2K} and p_{2D}.

p_{2K} and p_{2D} are of the same order as p_{2OM}. Therefore, if ESP takes place by logical steps, it is to be expected that p_{2BM} is of the order of $(p_{2OM})^2$.

Actual experimental results indicate that p_{2BM} is much closer to p_{2OM} than to $(p_{2OM})^2$. Rhine and associates found that in total comparable OM and BM series, the ESP Q's were 7.2 and 2.6 respectively.[1] A circumferential hypothesis would predict the BM results to be of the order of ESP Q = 0.5. These suggest that the ESP function proceeds <u>diametrically</u> to its end quite independently of the ordinary circumferential steps of logic. Or, in other words, at levels of action ordinarily associated with ESP experiments, little if anything is positively cognized; the perception is rather, as far as we know, one of correct or incorrect response.

To facilitate further research into the question of diametric or circumferential action of ESP, the following types of experiment are proposed:

1. A group of subjects should try:

(a) OM

(b) BM

(c) AOM, or Arbitrary Open Matching, in which a success would consist, not in matching like to like but in associating the cards according to an arbitrarily predetermined but unrevealed code. For example, the pairs whose association would be termed a hit might be, for example, (OL), (Λ≈), (≈+), (LΛ), and (+O).

(d) ABM, or Arbitrary Blind Matching, which would differ from AOM only in that the key cards are face down.

Two sub-conditions of this experiment suggest themselves. In one case, let the arbitrary code remain fixed throughout and in the other let it change after every run. An ESP machine would greatly facilitate this action. The particular point in introducing procedures (c) and (d) is that on a circumferential hypothesis, the ESP Q to be expected from them is of the order $(p'_{OM})^3$ which would usually be very small indeed.

1. Extra-Sensory Perception after Sixty Years; p. 318.

2. A group of subjects should try:

(a) calling whether a die of six faces showed an odd or an even number.

(b) calling odd or even according to the major number of such faces presented by five dice rolled at once. (We must employ the number of odd and even faces out of a total odd number rather than the sum of their face indications, since the probabilities for odd and even sums are different.)

If ESP is diametric, both case (a) and (b) are simple dichotomous decisions between odd and even. However, if ESP be circumferential, the task of envisaging five dice would seem at least five times as difficult as envisaging one, and the score attained should reflect this difficulty.

A CRITICISM OF DR. PRATT'S USE OF CHAPMAN'S "STATISTICS OF THE METHOD OF CORRECT MATCHINGS" IN THE EVALUATION OF ESP IN DRAWINGS[1]

Douglas G. Ellson
Stanford University

In 1937, J. G. Pratt published an analysis of the data obtained in a series of experiments on extrasensory perception by the late Dr. C. Hilton Rice (*2*). A part of these data consisted of two sets of 64 drawings obtained in informal ESP experiments. A number of "agents" had been used, working over a period of weeks. The agent, seated alone in a room, made an original sketch (O) of some simple object or design. At the same time, Dr. Rice in another room and without knowledge of the O attempted to sketch a reproduction (R) of the original. If similarities between members of the 64 pairs of drawings obtained in this way were more frequent than could be expected by chance, the operation of ESP would be indicated.[2]

After discussing earlier methods of evaluating the presence of ESP in investigations using drawings, Pratt says:

> "There is a more direct method of evaluating such work. This is a procedure called the method of correct matchings, which has an established statistical basis. The method is one which has been widely used in psychological work, but this is its first application in a parapsychological study." (*2*, pp. 251-252)

It is the writer's opinion that Pratt's interpretation of the statistic which Chapman (*1*) presents for use with the method of correct matchings is incorrect.

---

1. This paper has been accepted and approved by the Stanford Committee on Psychical Research as Communication No. 9 from the Psychical Research Laboratory.
2. This assumes that sources of communication (other than ESP) between the two persons taking part in the experiment are excluded. Pratt indicates that this latter condition is not met satisfactorily in the Rice experiment, and consequently does not interpret the extra-chance similarity which he obtains by the use of Chapman's method as evidence for ESP.

First, let us examine Chapman's contribution (*1*). It presents a statistical basis for evaluating the results of experiments in which a group of judges attempt to match two sets of paired items. An illustration will make clear the application and interpretation of the statistic:

> Four judges are given 10 samples of handwriting and 10 character sketches, each sketch describing one of the persons whose handwriting is represented. The judges attempt to match each sample of writing with the corresponding description of the writer, and succeed in matching correctly 4, 4, 5, and 7 pairs respectively. What is the expectation that they would do this by chance, i.e., that 4 judges would average 5 correct matchings?

Chapman presents a table (*1*, Table VII, pp. 296-297) which he describes as follows: It "gives the probability (p) of obtaining a mean number of correct matchings as great as s or greater from n independent trials (e.g. from n judges or S's) when the number of things to be matched (t) in each series is four or more." Thus we have:

p = the probability of obtaining a mean number of correct matchings as great as s or greater

s = the mean number of correct matchings

n = the number of judges

t = the number of things in each series (must be 4 or more for the table to be applicable)

In the above example, $n = 4$, $s = 5$, and $t = 10$. With these values, we find from Chapman's tables that $p < .001$, that is, the probability of s being as great as five is less than 1 in 1000.

We may conclude from this small value of p that the judging was not done on a chance basis, i.e., that presumably the judges were able to recognize some relationships between corresponding members of pairs and to match handwriting samples with the correct character sketches.

It is this statistic which Pratt applies to the Rice data. His treatment is best seen from a direct quotation:

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

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

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

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

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

(3) The 6 judges who acted in the second matching condition and 3 additional ones were asked to judge in 32 matching units consisting of 2 O's and 2 R's each. The 4 drawings making up each group had been obtained together in two tests. The object of having them matched by the judges was to see whether an R tended to be more similar to the O of the same test than to one other O chosen at random. In each of the 32 matching units, a judge had to be right for both pairs or wrong for both, with a chance probability of 1/2 for either result. The results may be considered,

8. In forming these sub-groups, the R's were taken by groups of 4 as they appeared in the shuffled set of slips, and the corresponding O's were selected from the other shuffled set of slips and presented in the order in which they were found. For example, the first matching unit was presented as follows:

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

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

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

therefore, on the basis of 32 trials for each judge with an average chance expectation of 16 successes. The actual number of units in which correct judgments were given were 21, 20, 19, 18, 24, 21, 21, 23, and 22. This gives a deviation of 45 beyond mean chance expectancy, which would on the average happen less than once in a million times.[9]

The three applications here made of the method of correct matchings (64 with 64, 4 with 4, and 2 with 2) all give highly significant results."

Pratt then interprets the probability values obtained by these procedures from Chapman's tables as indicated by the following:

"Thus far we have been concerned only with finding and applying a suitable method for determining whether there is anything more than a chance similarity between agent's and subject's drawings in the 64 tests. *The method applied shows that there is a significant degree of resemblance,* and this fact raises the second question of how to account for the results. Was the more-than-chance relation shown in these drawings due to ESP?" (2, p. 254, italics ours).

It would appear that two criticisms may be made of the treatment indicated in the quotations given above: (a) the interpretation of Chapman's tables in terms of resemblances between the two series of drawings is not justified, and (b) Pratt uses an inflated *n* in the calculation of his critical ratio. Each of these criticisms will be discussed in detail below.

(a) The interpretation of Chapman's tables in terms of resemblances between the two sets of drawings is not justified.

Unfortunately, it is not possible to quote passages directly from Chapman to demonstrate this point. It is evident that he did not foresee the misinterpretation which Pratt makes. However, it can be seen that Chapman's statement that his table "gives the probability (p) of obtaining a mean number of correct matchings as great as *s* or greater" is very different from Pratt's statement that "the method applied shows that there is a significant degree of resemblance." A statement concerning reliability of a judgment is interpreted as a statement of the probability of occurrence

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

of the events judged. Perhaps the clearest way of demonstrating the difference is by the use of an illustration in which the probabilities relevant to the events to be judged are known independently.

Let us suppose that two subjects each make five sketches, and that both subjects were told that only the five ESP card symbols were to be sketched. These subjects draw their sketches in the following order:

Trial	Subject I	Subject II
1*	Circle	Circle
2	Plus	Waves
3*	Star	Star
4*	Square	Square
5	Waves	Plus

Starred trials are "hits," i.e., the drawings are similar. The probability of three hits (pairs of similar drawings) being sketched on the same trials may be computed directly by means of conventional probability formulas. Three hits in five trials would be expected approximately once in 11 times.

If the sketches made by the two subjects are now given to five judges who were told to match drawings in one series with similar drawings in the other, we should expect all five judges to return the drawings in the following pairs (not necessarily in this order):

Circle.....Circle*
Star.......Star*
Plus.......Plus
Square.....Square*
Waves......Waves

Each of the judges would have matched correctly the starred pairs. The average number of correct matchings for the five judges would thus be 3.00. Looking in Chapman's table under $n = 5$ and $s = 3.00$ (t being greater than 4), we find $p < .001$. From this we should conclude that the judges were not matching on the basis of chance alone, since if they had done so, they would have averaged 3.00 or more correct matchings less than once in 1000 times. We might extend this conclusion to say that the judges could recognize similarities between drawings in the two series. We could not, however, interpret this as Pratt has done in his parallel case, and conclude that *the number of similar drawings made on the same trial by the two subjects*[3] would be

3. It is assumed from context that this is what Pratt means by the term "degree of resemblance" between the two series.

expected less than once in 1,000 times, since we have already shown that this would occur *once in 11 times*.

This probability (1 in 11) represents the chance expectation that Subject II should sketch three or more of the five drawings on the same trials as Subject I, *or* that *one* judge should make three or more correct matchings. Obviously the chances are far less for two or more judges to make three correct matchings (i.e., to average 3.00). Pratt's interpretation assumes the probabilities in the two cases to be the same.

(b) Pratt uses an inflated n in the calculation of his critical ratio.

At the end of his description of his third application of Chapman's statistic (*2*, p. 253, see quote above) Pratt states that a (total) deviation of 45 correct matchings obtained by 9 judges would happen by chance less than once in a million times. This probability figure is obtained from a critical ratio of 5.3, which is given, together with the figures on which it is based, in his footnote 9. The critical ratio is computed by means of the formula:

$$C.R. = dev./\sqrt{npq}$$

where dev. = 45, n (trials) = 288, p = $\frac{1}{2}$, and q = $\frac{1}{2}$.

Two of the figures used in this computation may be questioned. The deviation (45) is the total number of correct matchings made by all 9 judges in excess of chance expectation. The n is obtained by multiplying the number of matching units (32) by the number of judges (9), which gives a total of 288 trials.

The use of an inflated n will be obvious at once to the majority of readers. However, for those who are not familiar with the effect of this statistical error, a demonstration may be made by submitting our hypothetical example to the same treatment as that used by Pratt.

Two subjects coincidentally made three pairs of similar drawings in five trials, the chance expectation of this occurrence being 1 in 11. When the two sets of 5 drawings are submitted to 9 judges, they would presumably each make 3 correct matchings, a total deviation of 18 beyond the chance expectation of 9 correct drawings. If the number of trials (n) is taken as the the number of items in each series (5) multiplied by the number of judges (9), then n = 45. Substituting

d = 18, n = 45, p = 1/5, and q = 4/5 in the formula, we obtain a critical ratio of 6.71, i.e., the probabilities of obtaining a deviation of 18 with an n of 45 are less than 1 in 1,000,000. Thus by multiplying the number of items in the series by the number of persons matching the two series, and by taking the *total* number of correct matchings it is possible to obtain almost infinitesimal chance probability values in a situation where the actual probability is 1 in 11.

When the correct n for these data (5) is substituted for the inflated n and the deviation in the sample (2.0) substituted for the total deviation, we obtain a critical ratio of 2.23, which permits a much closer approximation of our known probability of 1 in 11. Similarly, if the correct n for Pratt's data (32) is used and the mean deviation obtained by the judges (5.0) is assumed to be the actual deviation in the sample, his critical ratio of 5.3 becomes 1.77.[4] The probability that the deviation would be obtained by chance is raised from less than 1 in 1,000,000 to approximately 1 in 26.

To summarize: We have pointed out that Chapman's statistics of the method of correct matchings give the probabilities of obtaining a given mean number of correct matchings by a group of judges, and that Pratt's interpretation, which assumes this figure to be the probability of obtaining a given degree of resemblance between paired items in the two series, is in error. The fact that Pratt proposes the use of an inflated n was also demonstrated.

REFERENCES

1. Chapman, D.W. The statistics of the method of correct matchings. *Amer. J. Psychol.*, 1934, 46, 287-298.
2. Pratt, J.G. The work of Dr. C. Hilton Rice in extrasensory perception. *J. Parapsychol.*, 1937, 1, 239-259.

Note: Since this paper was written, Pratt's article has received further notice. It is referred to on pages 65 and 84 of *Extra-Sensory Perception after Sixty Years*, by Rhine, Pratt, Stuart, Smith, and Greenwood. These references do not recognize the errors pointed out in the present paper, but in Appendix 10 (pp. 383-384 of

4. This critical ratio of 1.77 would be lowered somewhat more were the variability of the judges' scores taken into account.

the same volume) the correct application of Chapman's statistic is presented (presumably by Greenwood). In addition, a paper by Greenwood (A caution in the use of the method of correct matchings. *Am. J. Psychol.*, 1940, 53, 614-615) points out the possibility of misinterpreting Chapman's method and of making the error which Pratt made. However, reference to Pratt's paper is omitted.

COMMENT ON DR. ELLSON'S CRITICISM

by

J. G. Pratt

I appreciate the courtesy of Dr. Ellson and the editors of the *Journal of Parapsychology* in permitting me to publish simultaneously a brief statement about the foregoing article. This is not done to debate any of the fundamental issues Dr. Ellson has raised, but rather in the interest of correcting some unwarranted impressions which his discussion seems to me to create.

Dr. Ellson says that I misinterpreted the Chapman statistic. The fact is that my interpretation of the method was the very one that Chapman intended and that the "error" Dr. Ellson imputes to me alone is almost as widespread in the psychological literature as the use of the method of correct matchings itself. As Dr. Ellson's note suggests, credit for discovering the ambiguity likely to arise in the use of this method goes to Dr. J. A. Greenwood. The authors of *Extra-Sensory Perception after Sixty Years* were at that time well aware of the inapplicability of the Chapman statistic. If I interpret Dr. Ellson correctly, he wishes his readers to believe that we attempted to conceal the error in my paper. (This impression is confirmed by his specific emphasis of this charge in a review of our book in the *Psychological Bulletin* for December, 1940.) By way of reply I quote from page 65 of our book the first of the two references upon which he bases this contention: "*Attempts to Evaluate Non-Quantitative Material*. . . . Pratt evaluated the results of Rice by means of the method of correct matchings, the statistical basis for which has been supplied by Chapman (as mentioned in Chapter II). The present status of the use of non-quantitative material as the basis of experiment seems to offer no fully satisfactory method of evaluating the results (but see p. 383)." Page 383 criticizes the Chapman method and sets the conditions for its proper use! I do not see how Dr. Ellson could have failed to get the connection.

LETTERS TO THE EDITORS

Editorial Note: The editors again express their profound appreciation of the assistance given by the Board of Review, which continues to help in many ways in maintaining high standards of scientific reporting. In a few cases, unfortunately, the cost of resetting material cannot be borne by the Journal's limited income, but in all essential points the Board's guidance has been gratefully accepted.

G. M.

B. F. R.

December 16, 1940

Board of Review, Journal of Parapsychology

Memorandum to the Editors.

Subject: Review of Studies in Extra-Sensory Perception, III. A Review of All University of Colorado Experiments by Dorothy R. Martin and Frances P. Stribic.

This manuscript has been reviewed by the Board and the following members have submitted opinions which are included in this report: E. R. Hilgard, R. R. Willoughby, I. Lorge, J. L. Kennedy, L. Long, L. Dick and S. B. Sells.

This paper was well received by the committee. Dr. Willoughby stated that there is a need for more of this kind of careful reporting. Dr. Long and Miss Dick felt that the authors should be complimented on the quality as well as on the quantity of the experiments.

Dr. Hilgard stated that "because of the obviously unsatisfactory conditions in the preliminary series, it is unfortunate to present grand totals in which better conditions are mixed indiscriminately with poorer ones, as in Table U.1."

Dr. Lorge believed that the paper was written in good reportorial style and recommended its publication. He made the following detailed suggestions: that on page IV the expression in No. 5, "happy confluence...etc" be changed; in the introduction he would delete the line "Much of this.....to offer," and take out the word "nevertheless" two sentences away; On page 2 he would write "adequate supply of the special cards"; on page 4 he suggests that the 55555 check should be explained; On page 11 the sentences immediately following the words general psychology should read, "group was composed of individual volunteers"; with reference to the table he recommends that the table heads be set up with No. of Runs, Total Hits, Average Hits per Run of 25. The last suggestion might be consistently adopted throughout the text.

It is recommended that the above suggestions would improve the reporting of the experiments and should be incorporated in the published report.

Saul B. Sells,
Chairman.
Department of Education,
Brooklyn College.

December 16, 1940

Board of Review, Journal of Parapsychology

Memorandum to Dr. Gardner Murphy, Editor

Subject: Review of Variations of Time Intervals in Pre-Shuffle Card-Calling Tests by Lois Hutchinson.

This manuscript has been reviewed by the Board and the following members have submitted opinions concerning it: E. R. Hilgard, R. R. Willoughby, I. Lorge, J. L. Kennedy, L. Long, L. Dick and S. B. Sells.

The Board feels that certain revisions indicated in the discussion below should be incorporated in the paper prior to publication. It is suggested that these changes will make the report clearer and easier to follow.

Dr. Lorge has suggested that the report was clearly presented throughout, except for the conclusions. He feels that the paper would be strengthened if series 2 and series 1 were defined in the conclusions. He also states that it is important to recognize that the empirical distribution of hits is not normal binomial. Perhaps some more work should be done with the whole problem of individual favoritism in runs or groups of runs.

Dr. Long has made the same criticism concerning the distribution of hits.

Dr. Willoughby does not understand where the gelatin duplicator or hectograph comes in since the card orders were recorded apparently in private by a person who could not help seeing the calls already made, which were recorded beside the blanks where he was to make his record. This should be explained.

Dr. Hilgard states that "the results are cautiously handled. The interpretation of probabilities still permits of some choice. For example, the significant deviations of the one-day checked scores (the only ones for which significance is claimed) would occur only once in 1600 times as judged by the critical ratio of 3.11, but would occur once in 30 times in terms of the Chi-square treatment, as the probability values are presented, on page 10 of the manuscript." This point should be clarified in the final copy.

[The duplicator saves time. It also helps slightly in guarding against error, but is not the chief method described for this purpose, The hectograph helps to prevent post hoc "selection" of record sheets.

It is agreed that the critical ratio method and the chi-square method give different p-values. This is the general situation in experimental psychology, and one of the reasons why a p-value as such is usually somewhat arbitrary.--Ed.]

Saul B. Sells,
Chairman.
Department of Education,
Brooklyn College.

The Journal of Parapsychology

Contents of the three preceding issues

[illegible] 1939

Editorial Comment

[illegible] for an ESP [illegible]

[illegible]

A Machine for Research in Extra-Sensory Perception

[illegible]

Report of [illegible] Experiments on Thought-Transference

J. E. Coover

[illegible] Experimental Testing of "Extra-Sensory Perception" [illegible] Card-Sorting

[illegible] and Others

[illegible] Methods in the Comparison of Extra-Sensory [illegible]

Gardner Murphy and Ernest Taves

[illegible] Data from a [illegible] Score in Card-Guessing

[illegible]

Symposium on ESP [illegible] the Meeting of the Southern Society for Philosophy and Psychology

DECEMBER, 1939

Editorial Comment

Size of Stimulus Symbols in Extra-Sensory Perception

J. G. Pratt and J. L. Woodruff

A Method for ESP [illegible]

J. H. [illegible]

A Covariance [illegible]

J. A. Greenwood

A Theory of Extra-Sensory Perception

C. L. [illegible]

A Review of Recent Criticisms [illegible], II

C. E. Stuart

[illegible] Nature of Extra-Sensory Perception

J. L. Kennedy

JUNE, 1940

Editorial Comment

Experiments on the Paranormal Cognition of Drawings

W. Whately Carington

Some Observations on the Experiments with Drawings

W. Whately Carington

[illegible] of Kennedy's Study of the [illegible] Data

Charles E. Stuart

A Further Attempt to Test the Role of [illegible] in ESP Experiments

[illegible] and S. [illegible]

[illegible] the [illegible] Square-Root

[illegible]

[illegible] to Parapsychology [illegible]