The Journal of Parapsychology

A SCIENTIFIC QUARTERLY DEALING WITH EXTRA-SENSORY PERCEPTION AND RELATED TOPICS

Volume 7 DECEMBER, 1943 Number 4

Contents

Editorial: ESP, PK, and the Survival Hypothesis 223

A Large Series of PK Tests 228
EDMOND P. GIBSON, LOTTIE H. GIBSON, and J. B. RHINE

Position Effects in the Large Gibson Series 238
BETTY M. HUMPHREY and J. B. RHINE

The PK Effect: The McDougall One-Die Series 252
J. B. RHINE and BETTY M. HUMPHREY

ESP Tests under Various Physiological Conditions 264
J. L. WOODRUFF

STATISTICAL ARTICLES

Frequency Distributions of ESP Scores for Certain Selected Call-Patterns 272
T. N. E. GREVILLE

On the Evaluation of Differences of Success Ratios from Binomial Populations 277
LT. J. A. GREENWOOD, U.S.N.R.

Glossary 281

DUKE UNIVERSITY PRESS

DURHAM, N. C.

The Journal *of* Parapsychology

Editors
J. B. RHINE and J. G. PRATT, *Editors*
C. E. STUART, *Associate Editor*
LT. J. A. GREENWOOD, U.S.N.R.
Statistical Editor
DOROTHY H. POPE
Managing Editor

This journal is published on the fifteenth day of March, June, September, and December by Duke University Press, Durham, North Carolina.

Contributions submitted for publication and all editorial communications should be addressed to the Managing Editor, Dorothy H. Pope, College Station, Durham, North Carolina. Correspondence with the editors is advised before submitting articles other than reports of experimentation. Since it does not forward manuscripts by registered mail, Duke University Press cannot guarantee that they will not be lost in transit, and contributors are urged to keep copies of their papers. All contributions should be typewritten, double spaced. References should be given in the form adopted by THE JOURNAL.

Reprints may be ordered when the proof is returned. THE JOURNAL will bear one half the cost of reprints up to two hundred copies. Correspondence concerning subscriptions, change of address, back numbers, and other business communications should be addressed to the Parapsychology Laboratory, College Station, Durham, N. C.

The subscription price is $4.00 a year; single current numbers $1.00. The rate for back volumes is $5.00; for single numbers $1.25. Missing numbers will be supplied free when lost in the mails if written notice is given within one month of the date of issue. All remittances should be made payable to the Parapsychology Laboratory.

Entered as second class matter at the post office at Durham, N. C.

DUKE UNIVERSITY PRESS

The Journal *of*
Parapsychology

Volume 7 DECEMBER, 1943 Number 4

EDITORIAL

ESP, PK, AND THE SURVIVAL HYPOTHESIS

WHAT IS the bearing, if any, of the ESP and PK researches on the hypothesis of survival, that is, on the view that something of human personality survives the destruction of the body? Is the survival hypothesis any nearer to being established today as a consequence of many years of investigation of ESP and PK? Have the findings of these subjects rendered it any more plausible and acceptable; or, on the other hand, has the work on these unusual capacities only served to furnish easier alternative explanations for the evidence hitherto regarded as supporting the case for survival?

These questions will, we fully realize, evoke large differences of interpretation among those who are familiar with the work in ESP and PK; for some of those who have been engaged in such work have given little or no thought to the possibility of a relation between their findings and a survival hypothesis, whereas others have regarded their explorations in ESP or PK testing as an introduction inevitably leading up to the survival question. But whatever the reaction may be, the remarks to follow here are intended only to open the discussion and to invite reflection that may help in guiding future investigations. If there is serious disagreement, it is to be hoped that a final appeal can be made to experiment.

We shall begin with the *adverse* bearing which the experimental findings of ESP and PK have on the status of the survival hypothesis. In the first place these findings are going to make it a great deal harder to prove the survival hypothesis, if proof is possible. It

will be harder because the researches have considerably strengthened some of the counter-explanations with which the survival hypothesis has to compete. It is fundamental logic to say that so long as the evidence bearing upon the survival question can be rationally accounted for by principles already established, verification is impossible.

A shift in standards of judgment, therefore, will be necessary henceforth in considering the evidence for survival. The range of ESP is now known not be limited spatially and is even found to transcend time restrictions as well. Accordingly, the evidential value of obscure knowledge conveyed in mediumistic utterances is much reduced. Now that we know that it is possible for certain persons to apprehend events, however distant in space and whatever the barriers that may intervene, we have to recognize the *possibility* that the medium is such a subject when we are amazed at the information conveyed in her material. That is, it is not to be supposed that the very fact that the knowledge is extrasensorially obtained is evidence of its origin in the minds of deceased personalities.

And now the PK research has invaded the "physical phenomena" branch of evidence and altered its bearing on the survival hypothesis. Whether or not the physical occurrences associated with mediumship or the more spontaneous phenomena of poltergeist or hauntings be taken seriously as evidence of survival, all such data will need re-examination in the light of the PK experiments; for obviously if PK is a fairly common ability, it might well be applicable to any paranormal physical effect that is accurately observed and reported. To be sure, many of those who have been concerned with the study of the survival hypothesis in the past have been in some degree attentive to these alternative explanations, but in recent years our knowledge dealing with the nature and prevalence of ESP and PK has developed considerably, and these capacities have now achieved a much more definite experimental status. The time has come for a clear-cut appraisal of the state of the evidence which has from time to time been described as supporting the survival hypothesis with a view to determining its significance in the new perspective of recent ESP and PK research.

But the researches in ESP and PK have *favorable* bearings, too, on the survival hypothesis. Indeed, to those who are acquainted with

the ESP and PK investigations the hypothesis will seem more reasonable as a consequence. This effect is achieved by the general shift in the rational conception of man's place in nature which is required by the findings. It is nothing less than a reversal of the trend of scientific education as it has developed since the eighteenth century.

The general drift of scientific thinking concerning the nature of man has been mechanistic. Science has naturally been studying man from the easier, more objective approach of the physical sciences. The more familiar yardsticks of these sciences have been applied first as a matter of course and have fostered the impression that the thought-world of man is but an incidental evolutionary by-product of cerebral organization. Studies of the localization of mental functions in the brain and, in general, the close relation between mental processes and nervous action have only served to stamp in this concept of mind as a mere epiphenomenon of brain physics.

The upshot of this line of study is that some variety of physical interpretation of personality reflecting the characteristic physical theory of the day is invariably accepted. Whether it be old-fashioned materialism on the atomic plan or a more sophisticated analogy of dynamics with field theory as its background, the interpretation is basically of the same general type: the ultimate identification of mental process with physical law. If this were scientifically established fact, we should have to bow to it and accept it whether we liked to or not. But we see in the successive examples of physical analogies the familiar pattern of untested speculation which is all too easily confused with critical scientific thinking.

The ESP and PK experiments introduce a radically different type of evidence which opposes this trend toward a wholly physical interpretation of man. For ESP, as has long been established, does not fit into a physical mode of description. It is not subject to decline with the square of the distance. It is independent of angles and barriers, and even of the dimension of time. And the PK work, as far as it has gone, confirms this evidence. It shows that the mental system has a determinative influence which produces registrable effects without any conceivable physical intermediation. These effects reveal the stamp of intelligent purpose and do so in a way that is nonetheless physically a cause-effect phenomenon. A type of law-

fulness peculiar to mind and contrary to physics is increasingly evident in the ESP and PK researches.

Without these researches and with only the facts of the biological sciences to go on, it is hard to see how any kind of immortality would be possible. The brain-dominating or cerebrocentric view of personality would not allow it. The brain is too primarily and too completely the center of man. But if the psyche is a force and a factor in its own right, with laws and ways peculiarly nonphysical, the survival hypothesis has at least a logical chance. If the mind is different from the physical brain system, it *could* have a different destiny, *could* perhaps be independent, separable, unique. This degree of simple possibility must not, of course, be mistaken for probability; but mere logical possibility is itself very important. In fact, without it few scientists will undertake an investigation. "Science knows no impossibles," it is true, but practically the work of scientists, like that of everyone else, is very much guided by impressions of the possibility and impossibility of all that is undertaken. It should be easier, therefore, as a consequence of the ESP and PK work, for an inquiry to be launched on the survival hypothesis if the methods can be found.

There are other more specific bearings of the ESP and PK experiments on the survival hypothesis. If on any hypothesis of survival there is to be any awareness of things in the universe, perception of any kind (and what would a surviving mind be without awareness?) such perception would necessarily be extrasensory. It has been important, then, to discover whether such extrasensory perception can and does occur, and how general the capacity is! While it *could* be that some order of ESP might function in an incorporeal state of existence and not appear in the normal lifetime (and *vice versa*), it is certainly most reassuring to the investigator of the survival hypothesis to know in advance that this capacity is demonstrated to be a fairly common one, perhaps possessed by all persons, even though too easily inhibited to show up in the tests now in use.

Similarly, if a "spirit" personality of any shape or form is to acquaint the world of corporeal men of his existence, he has to start some kind of causal chain of events that begins with the nonorganic (presumably nonphysical) state in which he exists and ends up in some such concrete communication as a voice (sound waves) or as writing. Whether or not there is an intermediary or medium, this

reduces to the same thing—in brief, to psychophysical or psychokinetic causation. So we have to have PK ability in the explanation of any kind of discoverable survival or it would not make sense. And to be worth-while investigating, it has to make sense, at least to those who gamble their time and efforts on the research.

We have to know, then, that man possesses ESP and PK capacities in order to make any tentative conception of an existence beyond the transitions of bodily death a reasonable one. Without them, such survival could not occur and be discovered. The evidence of these abilities is not at all in itself positive evidence of such survival and could not in the nature of the case be so construed. But again, to establish that such survival is *possible* in certain important respects is to render an otherwise extremely absurd notion much more reasonable and therefore more likely to be investigated by men of rational habits.

"Immortality" and "survival" are scarcely more than nebulous concepts, and they are constantly changing with the growth of knowledge. Accordingly, it is hard to pin down the problem to any specific criteria with which we can deal in such a discussion. We are all doubtless more or less tradition-bound in our thinking of the survival hypothesis; but if we can single out any very general criteria of survival, they are those with respect to change in time and space. We are mortal, we say, if we cease to exist as individual, determining entities in the space-time world. Contrarily, to be immortal we need to escape the destructive changes of space and time, to exist and act in a nonspatial, nontemporal order.

Is it not then provocative, to say the least, to discover certain capacities of mind that appear to operate across the boundaries of space and time within which our sensorial, bodily system has to live and move? Here, surely, if ever, "hope sees a star" and the urge toward an inquiry into the question of survival receives valuable impetus and encouragement. Here too, of course, the warnings against over-confidence and a reminder of the difficulties are in order. But along with our reflections on the possibilities and the difficulties of the problem, let us keep the sobering thought before us that we are all, in the psychical sciences, still plainly ignorant of the fundamental realities of mental life. This will help us to avoid the dogmatic denials which, like dogmatic beliefs, discourage that exploration which alone can give a reliable answer.

A LARGE SERIES OF PK TESTS

By Edmond P. Gibson, Lottie H. Gibson,
and J. B. Rhine

ABSTRACT: In these tests Mr. and Mrs. Gibson, who were the principal subjects, threw six dice at a time from a cup, nearly 150,000 single die-throws in their main experiment alone. At one time or another, they threw for each of the six faces of the die as target face. The results were above chance expectancy for every face except one and that one was only very slightly below. Hence, the results cannot be explained as due to any defect or loading in the dice. These results are the most impressive reported to date. They represent odds against chance of more than a trillion to one. This large series provides an exceptional body of data for analysis.

Mr. Gibson, formerly of the Parapsychology Laboratory, is familiar to readers of this JOURNAL as a contributor both to the PK and to the ESP researches.—Ed.

In an earlier report, we dealt with experiments on the PK effect which had been conducted by Dr. C. E. Stuart in collaboration with one of us, E.P.G., in September, 1935. In that article (1) we described tests in which the subjects tried to influence the fall of dice by direct volitional action without mechanical intermediation. The statistical results of that report, as well as those of the various other publications on the PK research (see list of references at end of this paper), all appear to show that such influence is possible.

In the work here reported the attempt was made to stress the *length* of the series rather than some unique or varied departure from earlier work and to obtain a really large set of data, on the principle that the larger the better, other things being equal. This effort has produced the largest series of single die-throws yet reported in the PK literature, one amounting to well over 200,000, when all conditions of the experiment are included. The value of mere size of a series is considerable; it consists in more than weight of numbers alone. Analyses may be carried out on large series which would be relatively unreliable and unprofitable on smaller batches of data.

Because of the complexity of the task of reporting, we have divided the work into two parts. The present paper gives the general results with the common methods of evaluation that have been applied to ESP and PK data up to the time of writing. In a paper to follow (in this issue) there is an account of the special analyses of these data that have been carried out to date. But we regard this collection of data as especially suitable for continued analytical study and it is anticipated that still further developments will be forthcoming in due time. It is no exaggeration to say that the analyses as described in this report, lengthy and laborious though they have been, cover only preliminary questions. The work of both papers together, then, is only introductory, so extensive is the field of problems opened up for research.

The Experimental Program

The research was carried out in Grand Rapids, Michigan, during the period extending from the fall of 1935 over into 1937. It resulted from the visit of Dr. Stuart in early September, 1935, which was referred to in the earlier report of the Grand Rapids work; and although there was a degree of collaboration with the Parapsychology Laboratory at Duke, the testing was done entirely by E.P.G. and his wife, L.H.G., in whose home the work was conducted. L.H.G. was the subject in more than two-thirds of the tests, and in more than two-thirds of the remainder, E.P.G. was the subject. Twelve other subjects, mainly friends of the Gibson family, took part as subjects, contributing the remainder of the records. The tests were carried out in the evening hours or on week-ends when E.P.G. was at leisure from his professional work.

The dice were common 9/16-inch cubes made of transparent plastic material with white spots. Three happened to be greenish and three red in color. No effort was made in advance to determine their fine measurements, since only experimental controls on the trueness of the dice can be regarded as adequate for the purposes of these tests. The dice had been used, however, in the earlier research and had, as explained in that report, proved adequately true under similar test conditions.

Since in the earlier Grand Rapids experiments in PK the more successful tests had been with the casting of six dice at a time instead

of two, six was made the standard number per throw for the entire series. The dice were shaken and thrown from a cylindrical cardboard container or cup of about a quart capacity. They were thrown either onto a blanketed table-top or a rug on the floor, commonly against a baffle or backstop set up to aid in the randomization of the throwing. They were read and recorded by the observer, who called the score aloud. Then the subject replaced them in the cup, shook them, and threw again. The recording in all tests was so done as to allow the subject to observe the scores put down. The subject's interest in the scoring afforded a double check and made the chances of error relatively small.

Standard record forms were used throughout, and the score record of each throw was entered at once when the reading of the dice was made. The throw scores were recorded in short columns of six entries, and each column was totaled. These columns were grouped into sets of 10 columns with 6 entries in each, making 60 throws of 6 dice, or 15 runs to the set. There were commonly two of these sets across the page, left to right, and a maximum of eight on the page, though usually it was only four. Hence, there were anywhere from one to eight sets on a page. Sometimes one of these would be done at a sitting, sometimes two, three, or even four, depending largely upon the subject's interest, which in turn often depended upon his success. In a continuous series such as this, however, the stopping and starting are of no statistical concern as all data are included. Obviously, if a series is due to chance, the stopping and starting points will not alter the case.

The subjects from the beginning had favorable attitudes; that is, they were either convinced by the previous work that PK is a real effect in nature or they were willing to give it a good tryout on the strength of their friendly interest in the work E.P.G. was doing. The two main subjects, L.H.G. and E.P.G., had by this time become quite convinced that PK is a genuine phenomenon, though E.P.G. possessed less confidence in his own capacity than L.H.G. did in hers.

Although L.H.G. and E.P.G. sometimes cast the dice without an observer present, primarily in order to observe the effect of working alone, none of the results of these unwitnessed tests are included in the main series on which this report is based. These

unwitnessed test results are, however, of value in the discussion; and they are especially valuable in the secondary analyses conducted in order to get at the nature of the PK process. Since the analyses are made for effects not anticipated by the subject at the time of the experiment, we cannot regard the circumstance of witnessing of importance in weighing the analytical findings. Accordingly, the unwitnessed work will be included here but given separate evaluation.

There are three main groupings of the witnessed test results: first, the L.H.G. Series, consisting of 104,184 die-throws—these represent 17,364 throws of six dice or, in terms of the standard run of 24 die-throws, 4,341 runs; second, the E.P.G. Series with 885 runs (of 24); and third, the Miscellaneous Series with 807 runs. These make a total of 6,033 runs. In all of the L.H.G. Series, E.P.G. was the observer, and in the E.P.G. Series the relation was reversed. E.P.G. was the observer in all of the Miscellaneous Series; thus he was a witness in all the witnessed experiments reported in this paper.

Results

The Question of Chance

The total of 6,033 witnessed runs reported here gave an average score of only 4.18 where an average of 4.00 is expected by chance. But over so long a series, this deviation of the average score of .18 represents a positive deviation of 1,057 hits. With the SD of ±141.73 for 6,033 runs, this gives a CR of 7.46, one of the outstanding CR's of all the published (and unpublished) PK reports to date. There is, of course, no doubt that we have something besides chance to deal with in these data, for the odds against such results are trillions to one.

Since the later analyses will utilize the unwitnessed test data, it is appropriate to mention here that the total experimental results, pooling witnessed and unwitnessed tests, give 9,123 runs, an average score of 4.23, and a CR of 11.83.

It will be seen from the results of the witnessed tests summarized in Table 1 that L.H.G. contributed 4,341 runs at an average score of 4.17, and a CR of 6.14. This is obviously the principal contribution; and this large block of data should, even at this relatively low average score, provide good material, not only for the exclusion

of the chance hypothesis but also for the analyses needed in studying the workings of the PK effect.

Table 1

RESULTS OF WITNESSED AND UNWITNESSED PK TESTS WITH SIX DICE PER THROW

A. WITNESSED TESTS

Subject	Throws	Runs	Average per Run*	Total Deviation	SD	CR
L.H.G.	17,364	4,341	4.17	+ 738	±120.24	6.14
E.P.G.	3,540	885	4.15	+ 130	± 54.29	2.39
Misc.	3,228	807	4.23	+ 189	± 51.84	3.65
Total	24,132	6,033	4.18	+1,057	±141.73	7.46

B. UNWITNESSED TESTS

Subject	Throws	Runs	Average per Run*	Total Deviation	SD	CR
L.H.G.	9,120	2,280	4.44	+ 996	± 87.13	11.43
E.P.G.	3,240	810	4.01	+ 8	± 51.93	.15
Total	12,360	3,090	4.33	+1,004	±101.44	9.90
Grand Total	36,492	9,123	4.23	+2,061	±174.29	11.83

*Mean chance expectation is 4.00 hits per run of 24 die-throws for a specified face.

E.P.G., with 885 runs at an average almost the same as that of L.H.G., gives a CR of 2.39 ($P = .01$) which is just significant, while the Miscellaneous Series gives a higher average and a significant CR of 3.65 with its 807 runs. These uncommonly large blocks of data, even though much smaller than L.H.G.'s, are worthy of special analysis for further relationships, but the primary concern at this stage is to determine whether the experiment furnishes evidence of the PK effect. To do this requires the exclusion of a number of relevant counter-hypotheses, and we therefore turn our attention next to the results which bear on them. This is done on the assumption that the chance hypothesis is already ruled out of further consideration.

The Problem of the Trueness of the Dice

The question of whether faulty dice could have produced all the significant results obtained can be answered by data from a number

of sources. First, let us submit the simplest case against this hypothesis. It can best be stated after inspecting the results, given in Table 2, showing the average scores for each subject and for each face of the die.

Table 2

COMPARISON OF SCORE AVERAGES MADE WITH VARIOUS FACES AS TARGETS

Subject	1-Face		2-Face		3-Face		4-Face		5-Face		6-Face	
	Runs	Av. Score	Runs	Av. Score	Runs	Av. Score	Runs	Av. Score	Runs	Av. Score	Runs	Av. Score
L.H.G.	913.5	4.00	328.5	4.05	448.5	4.00	313.5	4.12	1,408.5	4.25	928.5	4.36
E.P.G.	90.0	3.77	135.0	3.83	165.0	4.31	15.0	4.47	90.0	4.24	390.0	4.24
Misc....	30.0	3.63	70.5	4.30	30.0	4.17	75.0	3.97	133.5	4.15	468.0	4.33
Total.	1,033.5	3.97	534.0	4.03	643.5	4.09	403.5	4.10	1,632.0	4.24	1,786.5	4.33

Plainly the hypothesis of distorted dice fails to explain this table. For if it did, the higher averages of 4.33, 4.24, and 4.10 produced by the six-, five-, and four-faces (which give most of the significance to the results) would have to be compensated for by *negative* deviations given by the other three faces taken together. *The dice could not favor one side or corner without equally disfavoring another.* But the one-, two-, and three-faces actually add up to a positive deviation also! It matters not that it is a small one (+34), and the number of runs is not important here. There is simply *no negative deviation on any face adequate to support the biased-dice hypothesis.* The only face giving a negative deviation is the one-face, and it is only .03 below expectation.

The gradual increase of score average from the one-face to the six-face might be due either to preference for the higher faces (which is common) or to bias in the dice (possibly to inequalities introduced in marking the spots). We need not decide here between the two possibilities, for even if the dice were biased to favor sixes and disfavor ones, the case against chance, and hence for PK, is as clear as before; for when we pool the hits for ones and sixes, thus evening up any effect of bias, we still get a positive deviation of 550, which has a CR of 5.68. This is very far above the requirement for significance. We may repeat again, therefore, that the hypothesis of biased dice is not the explanation.

These data, then, refute the hypothesis of defective dice too completely to need further discussion.

This is not surprising, however, for as we have said, the six dice which were used had already been tested by Stuart and Gibson and were found adequately true for use with the six-face as a target. (In the Stuart and Gibson check, the 60 control runs gave a deviation of −3 for the six-face.) Since the tests for this same target in the present report are in themselves significant (CR=7.60), this Stuart and Gibson check bears as definitely against the faulty-dice hypothesis here as it did in the previous report.

Piling control upon control, beyond all normal requirements, we might go on to add that one of the strongest checks on the hypothesis of defective dice in this research lies in the work of E.P.G. in which he used all faces as targets as a matter of routine and threw an equal number of times for each face in turn. As shown in Table 3, the 90 runs by E.P.G. from these around-the-die tests give a total deviation of +43 and a CR of 2.48. This is significant and is even higher than the CR which E.P.G. obtained for his entire 885 runs. L.H.G. did not do as well in this section of her work, yet scored positively (average score 4.09), and the pooled total of both subjects gives an average score of 4.16 and a CR of 1.91. The average of 4.16 in this portion of the research in which no defect of the dice could have affected the score, all faces having equal representation, is so close to that of the average of these two subjects for their total work (4.15 and 4.17) that it leaves little ground for the hypothesis of defective dice.

Table 3

RESULTS OF TESTS BASED ON EQUAL NUMBER OF THROWS FOR EACH FACE OF THE DIE

Subject	Throws	Runs	Average Score	Total Deviation	SD	CR
L.H.G.....	1,584	396	4.09	+34	±36.31	0.94
E.P.G.....	360	90	4.48	+43	±17.32	2.48
Total.....	1,944	486	4.16	+77	±40.24	1.91

Other Counter-Hypotheses

The statement of the results has presented the findings of this research as far as they bear upon the question of chance and that of

the trueness of the dice. On these issues we feel confident there will be no serious question left. The results do not fit in with either hypothesis.

The experimental method itself was intended to provide adequate precautions against the introduction of skilled throwing. The use of the cup, the shaking of the dice, and the throwing of the dice upon a blanketed surface and against a baffle seem to allow no reasonable opportunity for the use of motor skill, conscious or unconscious.

There are further questions which it has become customary in parapsychological reports to consider. Among these are the possibility of recording errors and loss or selection of records. These, however, are essentially counter-hypotheses which concern the individual competence and trustworthiness of the experimenter. The common method of meeting such questions in scientific research is that of confirmatory repetition of the experiment in other investigations. If the series of published researches has not already passed the point of supplying a sufficient number of confirmations of the conclusion that PK occurs, the reader who is still unsatisfied on that point is simply invited to await the continuation of the train of reports scheduled to follow.

Another way of meeting all the questions that are listed above as counter-hypotheses to the PK effect is to consider the analyses for position effects reported in the paper to follow immediately in this issue. The analytical findings in that paper represent an independent research in themselves though they are based upon the data which have just been reported.

The Effects of Witnessing

The results obtained by E.P.G. in his 885 runs of witnessed tests averaged 4.15, whereas in his 810 unwitnessed runs he averaged only 4.01. The witnessed work gives a significant CR of 2.39. The unwitnessed approximates the mean chance expectation of 4.00. This is of special interest because a similar effect will appear in a later report of E.P.G.'s work under other circumstances. It seems very reasonable to suggest that the stimulus to perform at his best requires the presence of an interested observer. In the witnessed work, his observer was his wife, L.H.G.

Quite a contrary effect is found in the case of L.H.G. herself.

In her 3,861 runs of witnessed work, she averaged 4.16. (This consists only of that witnessed work which was unmixed with unwitnessed tests on the same page.) In the 1,695 comparable unwitnessed runs, she averaged 4.41. This gives a deviation of the average score for the unwitnessed that is more than twice that of the witnessed (.41 to .16). Now a similar relationship is found in the "mixed" pages, that is, in those pages in which part of the record was for witnessed and part for unwitnessed runs. In this section of her work, there were 480 witnessed runs which gave a score average of 4.26 and 585 unwitnessed runs which averaged 4.52. Here the deviation of the average score for the unwitnessed is twice that of the witnessed (.52 to .26).

It would be in conformity with general observation to expect the subject to do better scoring alone than in the presence of an observer. We cannot, from this experiment, conclude with any degree of finality that E.P.G. requires an observer and that L.H.G. is to some extent handicapped by one, but we may say that that is what is suggested by the results. As other relevant findings accumulate, it may be possible to generalize further.

Concluding Remarks

A large series of tests has been made involving the throwing of six dice at a time from a cup and this series has yielded a very large total positive deviation. The odds against getting so large a deviation by chance alone are trillions to one. This series gives one of the most impressive CR's in the history of the PK research.

Not only is the investigation unexplainable by the chance hypothesis, but it is likewise not subject to the explanation in terms of faulty dice. The simple fact that all faces of the dice but one gave positive deviations, and that one gave only a small negative deviation, is of conclusive bearing upon this important alternative to the PK hypothesis.

The chief value of the investigation lies in its size, together with the uniformity of its procedure, and the relative homogeneity of its conditions and records which permit extensive analyses—analyses which may contribute to the understanding of the working of the PK effect. The paper to follow represents the first results of such analyses.

References

1. Gibson, Edmond P., and Rhine, J. B. The PK effect: III. Some introductory series. *J. Parapsychol.*, 1943, **7**, 118-134.
2. Hilton, Homer, Jr., Baer, George, and Rhine, J. B. A comparison of three sizes of dice in PK tests. *J. Parapsychol.*, 1943, **7**, 172-190.
3. Hilton, Homer, Jr., and Rhine, J. B. A second comparison of three sizes of dice in PK tests. *J. Parapsychol.*, 1943, **7**, 191-206.
4. Reeves, Margaret P., and Rhine, J. B. The PK effect: II. A study in declines. *J. Parapsychol.*, 1943, **7**, 76-93.
5. Rhine, J. B. Dice thrown by cup and machine in PK tests. *J. Parapsychol.*, 1943, **7**, 207-217.
6. Rhine, Louisa E., and Rhine, J. B. The psychokinetic effect: I. The first experiment. *J. Parapsychol.*, 1943, **7**, 20-43.

POSITION EFFECTS IN THE LARGE GIBSON SERIES

By Betty M. Humphrey and J. B. Rhine

ABSTRACT: The records of the Gibson series reported in the preceding article were studied with a view to discovering whether they would show a variation in scoring on the record page similar to that already reported in a number of preceding PK articles. Two units of the record were used as the basis of this study—the set, which consisted of ten columns of six entries each, and the page, which consisted commonly of four to six such sets. The set was found to have a very definite left-right or horizontal decline, most of the hits above chance occurring in the first column of the set. In the other direction, the vertical, it was found that the first of the six throw-scores in the column was very high as was also the last, whereas the four middle throws were relatively low. This gave a terminal salience effect, that is, a standing-out of the ends, which produced a U-curve. Dividing the set and the entire page into four parts brought out quarter-set and quarter-page distributions that are of interest in the light of similar effects already reported. These patterns seem lawful and psychologically meaningful. The differences between the parts of the set and between the parts of the page are in most cases significant so that the patterns cannot be ascribed to chance. The particular value of these declines at present lies in the fact that they constitute a very superior order of evidence against faulty dice and the other more common issues raised to counter the PK hypothesis.—Ed.

Introduction

The experimental work upon which the present account is based was described in the paper immediately preceding this one in this issue (2). The procedure and conditions were given and the general test results presented in that paper. These test results were found to be highly significant and were therefore not reasonably attributable to chance. They were obtained from experiments in which six dice were thrown at a time from a cup, and it seemed very improbable that any type of skilled throwing could have produced the significant CR. Likewise, the results were so distributed over the different target faces of the die as to show that no defect or distortion in the dice could reasonably explain the data. Therefore, taking the report as it stands, the conclusion is that the PK hypothesis is the most likely explanation of the results.

But this phase of the investigation, adequate as it is for meeting the principal counter-hypotheses, represents only a portion of the study to which the experimental data have been subjected. Already the reader of these PK reports is familiar with the analyses which have been carried out in several instances to determine the hit distribution in the experimental records with reference to position of the trial. The evidence obtained from the decline analyses that have already been reported has constituted a substantial part of the evidence for the PK effect and has been particularly effective in discriminating between that effect and various other possible explanations that have to be considered.

The experimental work of this report represents an exceptional opportunity for such analyses, principally because of the homogeneity of the records and of the size of the series. If we include the unwitnessed runs—as we may properly do since analyses such as these could not have been anticipated by the subjects when they took part in the experiments seven or eight years ago—we have a total of 9,123 runs for analysis, with 2,061 hits above expectation. More than two-thirds of these runs were made by a single subject, L.H.G., and thus we have a considerable degree of homogeneity which is very advantageous in the interpretation of the analyses.

We are interested here in analyses of hit distributions of the data grouped in three different ways: according to set, page, and temporal periods of the experiment. The set has already been described in the preceding paper as a section of the record page. In this experiment it consisted of a block of scores, six to the column and ten columns to the set, each score representing the number of hits obtained for the specified target face from a throw of six dice. The set was the standard unit of recording for this research; accordingly, it is our main interest to note the effect of the position of the throw in the set upon the distribution of hits. The horizontal and vertical distributions in the set will be presented separately since the relative distribution of hits from left to right in the ten columns of the set (horizontal) and the distribution of hits in the six-score column (vertical) are of interest in themselves. Finally, in as far as these two distributions merge in the form of a more general one, there will be a further treatment, based on the set, called a quarter distribution.

In general, the analyses for position effects have hitherto been based upon the record page (or, in some instances, the half-page where a double quota of runs was recorded on the page). In this particular research, however, not all of the records were kept with due regard for formal page structure. The analyses for position effects were of course not anticipated. In almost half of the runs of the series, the set records were so irregularly distributed on the record sheet as to prohibit any general partition of the page into equal parts. However, in the 5,025 runs of L.H.G.'s work, witnessed and unwitnessed, with uniform conditions on the page (that is, without any change of test condition during the tests recorded on that page), an analysis of hit distribution could be made. It is thus possible to present the hit distribution of the page for a substantial section of the series and also to compare it with the distribution of the set for this same section of the data.

Finally, we shall present the chronological distribution of hits for the same section of the series on which the page distribution has been worked out. In this chronological distribution, the work of the subject is subdivided into three temporal periods and the average rate of scoring determined. The size of the subdivisions is as nearly equal as is possible without the breaking-up of a session. The reason for the selection of the same group of runs for this chronological distribution as that used in the page distribution analysis is to afford a proper basis of comparison. The sample is sufficient to represent the main subject, L.H.G.

The test data have been filed with the Parapsychology Laboratory, Duke University, since 1937. The analyses, carried out under the direction of B.M.H., were done in 1942 and 1943. They have involved the independent checking of the records of the entire experiment four times over, and in each instance the results were checked against the independently checked totals of the original records as presented in the preceding paper. In addition, all computations have been completely double-checked. The possibility of an appreciable amount of error is therefore greatly reduced. If, in the judgment of critical readers, however, the importance of the findings warrants a further recheck, the records are open to qualified persons who may be interested in making such analyses.

Distribution within the Set

Horizontal Distribution

Horizontal decline has been an almost invariable accompaniment of the analyses of PK test results. In previous analyses, however, there was, in most cases, a longer record column to deal with, one which ranged from 12 to 24 entries. In this study the record columns, as has been stated, contained only six entries. Since there were 10 such columns in each set, the long axis of the set was from left to right. Two sets were located end to end across the page. The results were computed at the end of the set by totaling each column and then adding up the 10 column totals.

The results are very striking in the way in which the first column of the set stands out above the rest. The results for the whole experiment are shown in Table 1 in terms of deviation or number of hits above expectation for each column of the set. The first column shows a total of 738 hits above expectation, whereas the next highest one, which is the third column, gives 204 hits above. The difference between the first column and *each other column in the set* is a highly significant one; in fact, every one gives a CR above 6.00. For example, there is a high CR of the difference of 7.87 between the first and second columns.

Table 1

Horizontal Distribution of Hits in the Set

Subdivision of Results	Total Runs	Total Dev.	Deviation by Column in the Set									
			1	2	3	4	5	6	7	8	9	10
L.H.G. Witnessed	3,861	+ 614	+198	+ 77	+ 58	+ 8	+ 46*	+ 89*	+ 59*	+29*	+ 0*	+ 50*
L.H.G. Unwitnessed	1,695	+ 691	+317	+ 63	+ 42	+34	− 4	+ 55	+ 52	+22	+ 36	+ 74
L.H.G. Mixed†	1,065	+ 429	+161	+ 6	+ 33	+35	+ 44	+ 28	+ 32	+ 2	+ 47	+ 41
L.H.G. Total	6,621	+1,734	+676	+146	+133	+77	+ 86*	+172*	+143*	+53*	+ 83*	+165*
E.P.G.	1,695	+ 138	+ 36	+ 5	+ 55	− 1	+ 21	+ 6	− 21	+ 5	+ 20	+ 12
Misc.	765‡	+ 166	+ 26	− 25	+ 16	+19	+ 39	+ 11	+ 40	+10	+ 19	+ 11
Grand Total	9,081‡	+2,038	+738	+126	+204	+95	+146*	+189*	+162*	+68*	+122*	+188*

$CR_{diff.\ (1-2)} = 7.87$§

$CR_{diff.\ (1-3)} = 6.86$§

$CR_{diff.\ (1-4)} = 8.26$§

*These columns have 1.5 runs fewer in each instance than do the first four columns of the set.
†This section includes pages containing both witnessed and unwitnessed records.
‡A total of 42 runs (dev. of +23) has to be omitted because they were not recorded as sets.
§By the arc sine transformation method, these CR's of the difference are:
$CR_{diff.\ (1-2)} = 7.54$; $CR_{diff.\ (1-3)} = 6.55$; $CR_{diff.\ (1-4)} = 7.93$.

An inspection of the table will show that it is the work of L.H.G. primarily that produces this remarkable salience of the first column and in her case it is a consistent effect since the three large subdivisions of her data given in the table all show this characteristic distribution. On the other hand, the work of E.P.G. and that of the Miscellaneous Section show a marked drop from the first to the second column—not because the first column is exceptionally outstanding, but because the second is low in both cases. See Figure 1 for a graphical representation of the horizontal distribution of the set.

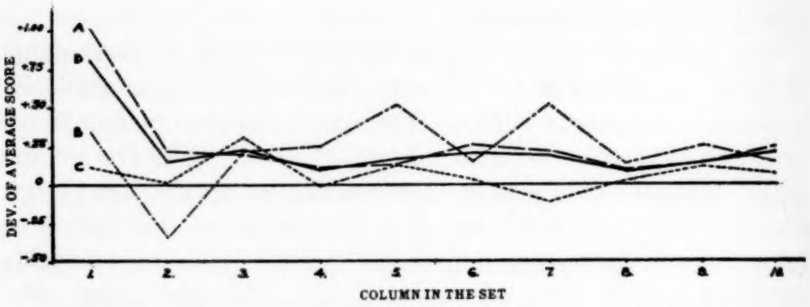


FIG. 1.—Deviation of average score for horizontal distribution of columns in the set. A. L.H.G.'s total; B. Miscellaneous total; C. E.P.G.'s total; D. Grand total.

Vertical Distribution

It may be recalled that in the first report by Gibson and Rhine (3), the first and last places in the six-entry column gave very much higher positive deviations for the entire work than did the four middle places in the column. The figures are as follows: +41, +12, −2, +16, −2, +34. These total deviations for the six places in the column average 37.5 hits above expectation for the two end positions as against 6.0 for the interior positions. It is an interesting fact that this earlier work, too, was produced mainly by L.H.G.

When a similar distribution study was made on the work dealt with in this paper, it was found that L.H.G. continued in very much the same general distribution pattern which she had followed earlier. The results are shown below in Table 2 and graphically represented in Figure 2.

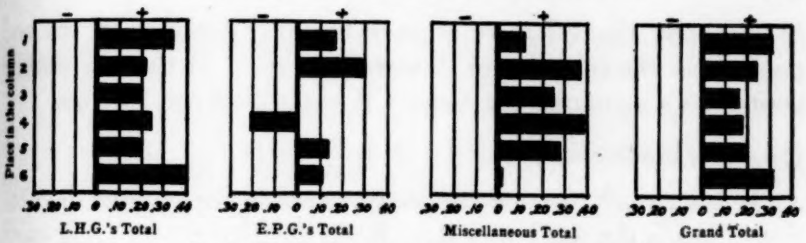


FIG. 2.—Deviation of average score for vertical distribution of hits in the column.

Table 2

VERTICAL DISTRIBUTION OF HITS IN THE SET IN TERMS OF TOTAL DEVIATION FOR POSITION IN THE COLUMN

Position in Column	L.H.G.'s Total (6,621 runs)	E.P.G.'s Total (1,695 runs)	Misc. Total (807 runs)	Grand Total (9,123 runs)
1.....	+ 374	+ 47	+ 17	+ 438
2.....	+ 235	+ 83	+ 48	+ 366
3.....	+ 199	0	+ 34	+ 233
4.....	+ 260	− 59	+ 51	+ 252
5.....	+ 223	+ 38	+ 38	+ 299
6.....	+ 443	+ 29	+ 1	+ 473
Total.....	+1,734	+138	+189	+2,061

$CR_{diff.}\ [(1+6)-(3+4)]=2.99^{*}$

*By the arc sine transformation method, this CR is 2.90.

Under L.H.G.'s Total column, it will be seen that the end positions stand well out from any of the four interior places. From comparison with the columns for E.P.G. and the Miscellaneous Series, it is clear that L.H.G. alone is responsible for this distribution. So marked is this effect that the pooling of the work as a whole for the grand total of 9,123 runs shows a still very pronounced salience of the terminal positions. It might be pointed out that though this terminal effect is the main feature in the vertical distribution, there is a *slight* vertical decline from the upper to the lower half of the column. The interesting question arises here whether the features of this vertical distribution—marked terminal salience and little decline—are due to the short length of the column since the longer columns in earlier reports have not shown them.

The statistical evaluation of this vertical distribution cannot be of any special importance, but it may be of interest to some readers

to know that the difference between the total deviation of the end positions in the column and the total deviations of the two middle positions is a significant one with a CR of the difference of 2.99.

Quarter Distribution of the Set

It has already become customary to obtain the quarter distribution of hits in the set. This is done by dividing the set into four parts by median horizontal and vertical lines and counting the hits in each quarter thus formed. In earlier investigations, this quarter distribution of the set has been found to be quite illuminating, particularly in the Gibson and Rhine paper and in that by Rhine on the comparison of cup and mechanical throwing (7), in both of which the set was the distinctive unit in the recording. There is a marked pattern identifiable with the work of L.H.G. in this quarter distribution just as there is in the vertical distribution. In the results as summarized in Table 3 below, the character of the distribution given by the grand total is principally due to L.H.G.'s data. The first quarter gives the highest deviation of score average as is usual; but the fourth quarter, while considerably lower than the first, is still not the lowest, and this constitutes the exceptional feature. As Section A of Table 3 shows, L.H.G.'s work in its two main subdivisions, witnessed and unwitnessed, displays this peculiarity of increased rate of scoring in the fourth quarter. In Section B of the table, the QD (quarter distribution) of L.H.G.'s total work is presented along with those of E.P.G.'s total and the Miscellaneous total.

The differences in deviation between the quarters for the grand total are so marked that several significant CR's are found. It is of considerable interest to note that there is a significant drop from the upper left to the lower right, referred to as the diagonal decline. This difference has a CR of 2.38 (P=.01) and the decline is in the direction expected on the basis of previous studies. The difference between the entire left half and the entire right, the horizontal decline, gives a CR of the difference of 3.49. The difference between the total of the first and fourth quarters and the total of the second and third gives a very high CR of 5.43. A small correction of the probability for this CR is necessary for the degree of selection. But these are not especially important figures for our purposes since there

Table 3

QUARTER DISTRIBUTION OF HITS IN THE SET IN TERMS OF DEVIATION OF AVERAGE SCORE

A. QD's for L.H.G.'s Results.

L.H.G.'s Witnessed Runs = 3,840*	
+.32	+.02
+.08	+.21

L.H.G.'s Unwitnessed Runs = 1,695	
+.59	+.01
+.50	+.53

B. QD's for Subjects' Totals and the Grand Total.

L.H.G.'s Total Runs = 6,600* Dev. = +1,725	
+.45	+.03
+.23	+.33

E.P.G.'s Total Runs = 1,695 Dev. = + 138	
+.28	+.02
.00	+.02

Misc. Total Runs = 765† Dev. = +166	
+.15	+.27
+.25	+.20

Grand Total Runs = 9,060‡ Dev. = +2,029	
1 +.392	3 +.050
2 +.189	4 +.264

$CR_{diff.}\ (1-4) = 2.38$§
$CR_{diff.}\ (L-R) = 3.49$
$CR_{diff.}[(1+4)-(2+3)] = 5.43$

*Twenty-one runs with a deviation of +9 are omitted here as not suitable for QD analysis.
†Forty-two runs with a deviation of +23 are omitted here as not suitable for QD analysis.
‡A total of 63 runs with a deviation of +32 is omitted here as not suitable for QD analysis.
§By the arc sine transformation method, this CR = 2.33.

is surely no need in this discussion for amplifying further the fact that there are significant internal differences. The important thing is to keep in mind the trend of differences as they relate to similar observations in other work. For that reason the much less significant CR of the difference between the first and fourth quarters stands out above the other less predictable but larger CR's of the difference.

DISTRIBUTION ON THE PAGE

Some of the previous reports on the PK work have been supplemented by quarter-page distribution studies. In these QD analyses, the page was divided into quarters and the distribution of hits obtained for each quarter. As explained above, this could be done for the present investigation only on the total of 5,025 runs. The other record pages were either made up of mixed conditions or were

so recorded that no mental structuring of the page would have been possible on the part of the subject. For example, if on one page of a subseries there were two sets recorded horizontally across the page and on the next page there were two sets recorded one below the other, it would not be reasonable to expect to obtain consistent and predictable results from merging these two record sheets. The record sheets used for the page distribution analysis were those from subdivisions in which the recording was consistently done on pages filled out in a standard manner. In one such block of results, L.H.G.'s Witnessed Series, four sets were recorded to the page, two across the page horizontally and two just below them. For the quarter distribution of this page, then, it was simply a matter of drawing a vertical line at the mid-point of the page and a horizontal line between the upper and lower pairs of sets. One set thus fell into each quarter. For L.H.G.'s unwitnessed runs, six sets were done on each page and in order to make a quarter-page distribution it was necessary to draw the horizontal line at the mid-point of the second row of sets, thus putting a set and a half in each quarter.

These quarter declines for the page are shown below on the right in Table 4, and there it can be seen at a glance that a somewhat different distribution is found from the quarter decline of the set shown above in Table 3. The upper left quarter is still the highest in level of scoring. The fourth quarter is very near the lowest but is not quite so low as the upper right, though it is lower than the lower left. Again there is a diagonal decline from the first to the fourth quarter, giving a CR of the difference of 1.75, but this is not significant. It should be noted, however, that this page distribution represents a much smaller collection of data than the QD of the set in Table 3. Proportionately the difference between the deviations of the average scores in quarters one and four is about the same. For a more strict comparison, the QD of the set is given (on the left), using the same group of records as was used for the QD of the page (on the right).

The QD for the page is sufficiently close to the typical QD to be of considerable interest in the larger view of these studies, for it is in line with a general trend, found in the PK research as a whole, which is the subject of a report to appear later.

Table 4

QD's of the Page and Set for L.H.G.'s Work Alone in Terms of Deviation of the Average Score

QD of the Set Runs = 5,265* Dev. = +1,261	
+.408	+.033
+.199	+.318

$CR_{diff.\ (1-4)} = 1.26$

QD of the Page Runs = 5,025† Dev. = +1,179	
+.318	+.182
+.235	+.190

$CR_{diff.\ (1-4)} = 1.75$

*Twenty-one runs with a deviation of +9 were omitted here as not suitable for QD analysis.
†Two hundred sixty-one runs with a deviation of +91 were omitted here as not suitable for QD analysis.

Chronological Distribution

L.H.G.'s work for each condition, witnessed and unwitnessed, was divided into three subdivisions on the basis of the period of time in which it was conducted, except that no session was split up. The session falling on the dividing line was thrown into the subdivision in which it would cause the smallest departure from the evenly divided number of runs. Thus we obtained from the analysis the results given in Table 5. There it may be seen that a decline is evident again though it is not an extremely marked one. There are no significant differences; the range in average score from 4.30 to 4.20 is not at all striking. However, in the general collection of these findings this item may contribute something to the total picture. The same records were used as were analyzed for the QD of the page, namely, 5,286 runs. This was done to facilitate comparison between the effects obtained.

Table 5

Chronological Distribution of Hits in the Work of L.H.G.

Period	Runs	Deviation	Av. Score
1st.....	1,800	+535	4.30
2nd.....	1,836	+405	4.22
3rd.....	1,650	+330	4.20

Discussion

Altogether, the analysis of the work of this report has shown a fairly consistent picture of decline: a decline to the right in the

horizontal distribution, a decline from the first to the second and third places in the vertical distribution (though with an incline on the last two trials of the column), a decline diagonally from first to fourth quarter in the QD of the set and of the page, and finally, a decline in the temporal sequence of the series.

It appears to us to be a sufficiently burdensome task upon the reader merely to familiarize himself with these main outlines of distribution without here pausing to undertake an extensive general survey of these findings against the background of similar ones from other reports. The latter is a task, rather, for a survey devoted primarily to comparison and review of the position effects. It will be readily recognized that the results of these analyses, incomplete though they necessarily are, add a great deal to the weight of evidence concerning the position effects and that it will be easier to evaluate them the longer they are allowed to accumulate.

The remarkable drop after the first column in the horizontal distribution is the statistically outstanding effect. The strong upturn in score average at the lower end of the vertical column is perhaps the next most noteworthy phenomenon. The internal consistency of these effects in the work of L.H.G. and the fact that they are both peculiar to her work add psychological significance to them both. Here, as a matter of fact, are two peculiar and significant effects of L.H.G.'s test performance of which she knew nothing until years later when the analyses were completed, yet which in some way are tied up with her particular personality. The other effects noted in the analyses—for example, the chronological decline and the diagonal decline from the first to the fourth quarter in the quarter distributions of the set and page—are not so striking nor are they peculiar to L.H.G. These declines are the normal thing as far as generalization is possible at this time from the various distribution analyses of other reports.

But the two items first mentioned, the sharp drop in the horizontal distribution and the standing-out of the upper and lower ends in the vertical distribution call for explanation in terms of the subject's individuality. L.H.G.'s QD of the set is likewise a personal and peculiar one. That we find declines in these data and even that we find diagonal declines in a QD is not apparently a matter of individual peculiarity, although there are a few subjects whose results have failed to show them. Neither is the finding of a

horizontal decline unusual, but the sharp drop after the first short column seems more distinctly associated with L.H.G. than anything else that has been encountered. Therefore, it is legitimate to raise the question, even if it cannot be answered at this time, as to what characteristics of L.H.G. could have produced these effects. What possible circumstances might have favored them? Would they have been the same with a different structure of the set? Would there have been a difference with different numbers of dice per throw? Was there a difference between the sets on the page for instance? At once it is clear that more analyses are needed. It may be that the assembling of further facts will bring out more understanding of what has happened, and this is, therefore, our only course to pursue.

But without further interpretations which are in turn dependent upon more knowledge, there is an immediate advantage to be gained from the distribution analyses herein reported. First of all, they lift the whole discussion of counter-hypotheses to a level on which decision is clear-cut and final. These differences which appear in declines and in other position effects, with their highly significant CR's in some cases, can leave no possible doubt of the inapplicability of the faulty-dice hypothesis. This explanation, like chance, is refuted beyond the need for further debate. Furthermore, even if the use of a cup and six dice at a time allowed any question about the possibility of manipulation of the dice, we cannot suppose that the horizontal distribution described above could have been the result of skillful dice-throwing. The hypotheses of lost records and recording errors recede into insignificance in the face of the CR's of the difference arising from the distribution analyses. None of these counter-explanations could account for the significant internal effects. Rather, such effects are the functions of mental peculiarities, traits of the mind that are not yet well understood perhaps but which are sufficiently stable and real to persist through long series of test performances. Similar effects of position are already well recognized in the realm of ESP psychology; indeed, they represent a subject of interest in a number of laboratories at the present time. Moreover, position effects that may be more or less related to those described in the PK analyses—such as memory curves and learning gradients—are known in general psychology.

The emphasis has been upon declines in this discussion of position effects. The presence, however, of the terminal salience effect found in the vertical distribution, particularly in L.H.G.'s work, reminds us not only that comparable effects were reported in certain earlier PK papers, such as that of Gibson and Rhine (3) and of Reeves and Rhine (6), but also that a similar effect may be predicted to appear frequently in future analyses. To the student of ESP literature, terminal salience is already a familiar phenomenon (4, 5, 8, 9). It has a fair background of research of its own.[1] At this point we would be overzealous to speculate unduly upon this parallel, but it cannot be pointed out too often that such similarities between the ESP and the PK analyses are producing a growing need of finding out by deliberate experimental attack what the actual relationship may be which produces these similar effects in ESP and PK. Logically it seems to us they must be related and that perhaps these experimental parallels may be the clues of what that relationship is.

Concluding Remarks

We have analyzed the total records of the large Gibson series, witnessed and unwitnessed. These 9,123 runs afford the most extensive material for analysis yet made available for the study of position effects in hit distributions.

The most significant of the various distribution effects encountered in the analyses is that of the horizontal decline within the set. Significant differences found between the first and other columns in the set give CR's ranging from 6.00 to 8.00. The effect was due entirely to one subject, L.H.G., but appears consistently in the three main subdivisions of her work.

The vertical distribution of hits in the column represents another outstanding result of the analysis. This distribution shows strong terminal salience in the distribution of the column for the entire series. The standing-out of the beginning and end trials was characteristic also of the first Gibson report (3). Again the effect is due to L.H.G. who was the principal subject also in the earlier Gibson study.

[1] It is especially interesting to compare the curve of L.H.G.'s hit distribution for the column in this dice work with her U-curve in the ESP card-calling tests (1: p. 275).

The quarter distribution of the set gives a diagonal decline with a significant difference between the upper left and lower right quarter of the set. The quarter-page analysis is limited to only 5,025 runs and likewise has a diagonal decline. The CR of the difference, however, though suggestive is not significant.

The chronological distribution shows a decline with the temporal sequence of the tests though the differences are not significant. The declines having to do with the structure of the record page—that is, the horizontal, vertical, and quarter declines—are all produced by the work of L.H.G. as the primary contributor. This suggests that these distribution effects are to some extent tied up with traits of the individual subject.

The chief significance of the distribution effects at this stage of the PK research lies in their value as evidence for that effect. The significant internal differences are not attributable to faulty dice, skill in throwing, recording errors, loss of records, or, in fact, any of the common run of counter-hypotheses. The larger significance of the hit distribution is left for the time being for a fuller survey of the evidence of such effects from the various PK researches.

References

1. Gibson, Edmond P. A study of comparative performance in several ESP procedures. *J. Parapsychol.*, 1937, **1**, 264-275.
2. Gibson, Edmond P., Gibson, Lottie H., and Rhine, J. B. A large series of PK tests. *J. Parapsychol.*, [this number].
3. Gibson, Edmond P., and Rhine, J. B. The PK effect: III. Some introductory series. *J. Parapsychol.*, 1943, **7**, 118-134.
4. Humphrey, Betty M., and Rhine, J. B. A confirmatory study of salience in precognition tests. *J. Parapsychol.*, 1942, **6**, 190-219.
5. Pratt, J. G., Humphrey, Betty M., and Rhine, J. B. A check on salience relations in ESP data. *J. Parapsychol.*, **6**, 44-51.
6. Reeves, Margaret P., and Rhine, J. B. The PK effect: II. A study in declines. *J. Parapsychol.*, 1943, **7**, 76-93.
7. Rhine, J. B. Dice thrown by cup and machine in PK tests. *J. Parapsychol.*, 1943, **7**, 207-217.
8. ———. Evidence of precognition in the covariation of salience ratios. *J. Parapsychol.*, 1942, **6**, 111-143.
9. ———. Terminal salience in ESP performance. *J. Parapsychol.*, 1941, **5**, 183-244.

THE PK EFFECT: THE McDOUGALL ONE-DIE SERIES

By J. B. RHINE and BETTY M. HUMPHREY

ABSTRACT: In this PK experiment Professor William McDougall, with the help of eight others as subjects, attempted to influence a single die so that it would fall with a specified face turned up. There were 269 runs, each run consisting of 24 throws of the die, in the course of which all six faces were used as the target. By the more common method of evaluation, the odds against these results occurring by chance are only of the order of 20 to 1. But by the equally valid chi-square method, the odds are about 400 to 1. Moreover, the patterning of hits on the record page showed a distribution which was quite significant. For example, the odds against finding such a difference in rate of scoring as that which occurred between the upper left and lower right quarters of the page are over 1,000 to 1. The unique feature of the McDougall Series lies in the fact that the distribution of hits on the record page shows the exact reversal of what has come to be regarded as the typical page distribution. The conditions that might have induced this reversal are described. These extrachance distributions and their exceptional consistency remove all reasonable likelihood of the McDougall Series being due to chance, faulty dice, skilled throwing, or other common hypotheses.—Ed.

IN THE FALL OF 1935, the late Professor William McDougall, F.R.S., then Head of the Department of Psychology at Duke University, filed with the Parapsychology Laboratory two record books filled with experimental data. They contained the records of an experiment involving the testing of the PK hypothesis which he made during the preceding five months while he was visiting his family in England. Besides himself, he had eight other subjects, most of them his relatives.

In the test, the subject attempted to influence a single die so that when thrown it would fall with a specified face turned up. All six faces were represented as target face but not equally so, since each subject was permitted to make his own choice of the face for which the run or column of 24 entries would be made. An ordinary die of common commercial quality was used. It was thrown onto a cloth-covered table from a vase-shaped cup with a narrowed neck.

From three to six members of the family circle were always

present. They sat around the table during the testing, Professor McDougall himself, however, doing the recording except when he was performing as subject.

The record books were of the "blue book" type commonly used in college examinations. Ten runs were recorded to the page, 269 runs in all. Each page was given over to a particular subject, but the runs were alternated from subject to subject among those present. The records had all been checked for successes when the books were turned over to J.B.R., and they were then filed without further analysis or evaluation until, in the spring of 1943, they were rechecked and analyzed at the Parapsychology Laboratory under the direction of B.M.H.

Professor McDougall died in 1938 before any attempt was made either to appraise the significance of his work or to prepare a report of it for publication. All that can be said under the circumstances is that these tests led him to a continued interest in the PK problem so that he conducted other experiments at a later date. These later ones were carried out under quite different circumstances and will be the subject of further analysis and publication.

Results

The 269 runs, of 24 throws each, gave a total deviation of + 49 and an average score of 4.18 as against the expectation of 4.00. This, with a standard deviation of ±29.93, gives a CR of 1.64 which is not a significant result. The distributions of deviation over the different faces of the die along with the totals just mentioned are shown in Table 1. There it will be seen that the six-face was the favorite but that the twos and threes gave higher average scores.

If the die was not distorted so as to affect the results appreciably, the CR's in the table are of interest. One is significant, and two others are suggestive. Combined by the chi-square method, they give a total chi-square of 20.37, which with six degrees of freedom is significant, having a corresponding P of .0025. We have, then, to accept the view either that the die is faulty or else that it was influenced in the throwing to an extrachance degree. This situation calls, then, for a further step if we are to explain the results of Table 1. Both the chance hypothesis and that of a faulty die have to be eliminated *together* if the results are to have any significance for the PK hypothesis.

Table 1

THE MCDOUGALL ONE-DIE SERIES, RESULTS BY FACES

Face	Runs	Average Score	Total Deviation	SD	CR	χ^2
1.....	21	4.14	+ 3	± 8.36	.36	.13
2.....	65	4.72	+47	±14.71	3.20	10.24
3.....	35	4.28	+10	±10.80	.93	.86
4.....	32	3.34	−21	±10.33	2.03	4.12
5.....	6	2.50	− 9	± 4.47	2.01	4.04
6.....	110	4.17	+19	±19.14	.99	.98
Totals....	269	4.18	+49	±29.93	1.64	$\Sigma\chi^2 = 20.37$ d.f. = 6 P = .0025

Vertical and Horizontal Distribution

When the McDougall data was first studied for the distribution of hits in the run and in the sequence of runs on the page of the record book, it was noted that instead of the vertical decline commonly found in PK records, there was a very marked *reversal* of distribution, namely, a considerable incline from − 12 to + 46 from the top half to the bottom half of the run. Not only was this true of the work as a whole but also of each of the two books: - 11 to + 36 in the first, and − 1 to +10 in the second. (In this and in the other distribution studies, 39 runs with a deviation of + 15 were omitted because the page records were broken up by entries made on different days. This might reasonably have interfered with the structuring effect of the page in the subject's performance. Because this break in the structuring would necessarily exclude these records from the horizontal and QD analyses, it seemed best for comparative purposes to exclude them from the vertical analysis also. Their inclusion in the vertical distribution, however, would show an incline from − 5 to + 54, with a CR of the difference of 1.97. The interpretation of the results is not in any way affected by this disposition of these records.)

Such a reversal was of immediate interest and it justified an exhaustive examination of the data for other position effects. The next distribution to be examined was that in the horizontal direction, the distribution across the page from left to right. Here a similar reversal was found: instead of the usual horizontal decline, the scoring inclined even more than it had in the vertical direction,

rising from − 19 on the left half of the page to + 53 on the right. Here again the effect was divided between the two books (− 20 to + 45 in Book I, and + 1 to + 8 in Book II). This consistency of scoring in the opposite direction from that expected was quite impressive after the long line of declines that had, up to that point, been found in other series of test data. The analytical results mentioned are given in Table 2.

Table 2

VERTICAL AND HORIZONTAL DISTRIBUTION OF HITS ON THE PAGE

Book	Vertical				Horizontal			
	Top		Bottom		Left		Right	
	Runs	Dev.	Runs	Dev.	Runs	Dev.	Runs	Dev.
I.....	90	−11	90	+36	90	−20	90	+45
II.....	25	− 1	25	+10	25	+ 1	25	+ 8
Total...	115	−12	115	+46	115	−19	115	+53

$CR_{T-B}=2.10$ $CR_{L-R}=2.60$

Up to this point the analysis shows two effects which are the reverse of those commonly found in the scoring in PK experiments, and these two exceptions were divided between the two record books. This, in itself, constitutes a very interesting result. Statistically, the fact that the trend is not in the direction expected from past findings is more than offset by the consistency shown. More than that, the differences are impressive. The CR of the difference between the total top and bottom halves of the vertical distribution is 2.10, and that between the left and right halves is 2.60. Combined as independent measures they give a chi-square of 11.17 which, with two degrees of freedom, is equivalent to a P of .004.

There remains, however, an important question to be cleared up before accepting the results of the horizontal analysis. There was a variation of target face from column to column on the page. This was so varied that the records are regarded as providing in effect an equalization of the target face variation for the two halves (left-right) of the page. The distribution of target faces on the page for the full 230 runs analyzed for position effects, is given below in Table 3, showing that this distribution of targets could not

account for the large left-right incline. We can estimate how much could at most be fairly attributed to such a source by means of the table of deviation of the average scores for each target face as shown in the table. By consulting the rate of deviation and the frequency of the target faces, it may be estimated that less than one-sixth of the horizontal difference of the series is attributable to the small irregularity of the target-face distribution.

Table 3

TARGET FACE DISTRIBUTION AND RATE OF DEVIATION

Target Face	Frequency of Use of Target Faces in the Columns of the Page										Total Runs	Dev.
	1st col.	2nd col.	3rd col.	4th col.	5th col.	6th col.	7th col.	8th col.	9th col.	10th col.		
1.....	0	1	2	2	2	1	1	2	1	1	13	− 8
2.....	5	7	6	5	6	7	7	7	7	7	64	+48
3.....	4	3	3	3	4	3	3	3	3	3	32	+10
4.....	5	3	3	3	3	3	3	3	3	3	32	−21
5.....	0	2	0	1	0	1	0	0	0	0	4	− 9
6.....	9	7	9	9	8	8	9	8	9	9	85	+14
Total....	23	23	23	23	23	23	23	23	23	23	230	+34

Quarter-Page Distribution

From these findings it was obvious that a QD (quarter-page distribution of hits) would show a reversal of the usual pattern. Again in the QD analysis the two books were treated separately, and the results of the analysis are given below in Table 4. The deviations of the quarters of the record page for Book I, Book II, and the total are shown separately. The striking similarity of proportion in the QD's of the two books is evident. The first book starts out with a much more striking negative deviation than the second; but the latter, even though only slightly negative in its first quarter, is still essentially similar in that its upper left quarter is the lowest of the four. These are by no means chance distributions; rather, they are strongly significant. Even Book I alone gives a significant difference between the first and fourth quarters. The total obtained by pooling the records of the two books is likewise very significant. The CR of the difference between the first and fourth quarters of the total results

(3.32) is so large that it would not lose its significance if corrected for the unexpected reversal of direction.[1]

Table 4

QUARTER-PAGE DISTRIBUTION

	Book I (180 runs) Left	Right		Book II (50 runs) Left	Right		Total (230 runs) Left	Right	
Top.....	−32	+21	−11	− 2	+ 1	− 1	−34	+22	−12
Bottom..	+12	+24	+36	+ 3	+ 7	+10	+15	+31	+46
	−20	+45		+ 1	+ 8		−19	+53	
	$CR_{diff.\ 1-4}$=3.23			$CR_{diff.\ 1-4}$= .99			$CR_{diff.\ 1-4}$=3.32		

The degree to which this QD pattern is derived from the various subjects is also of interest. There were three main subjects, A.A.M., W.M., and K.M., with 40, 50, and 100 runs respectively, and the other 40 of the 230 runs suitable for QD analysis were done by three of the other five subjects. The QD's for the three main subjects and for the miscellaneous group are given below as Table 5.

Table 5

QD's FOR THREE MAIN SUBJECTS AND MISCELLANEOUS GROUP

A.A.M. (40 runs)		W.M. (50 runs)		K.M. (100 runs)		Misc. (40 runs)		Total (230 runs)	
−1	+1	−7	+5	−10	+ 4	−16	+12	−34	+22
−1	+3	+1	+5	+13	+13	+ 2	+10	+15	+31

There is a complete consistency in the rise from a negative beginning (upper left quarter) to a large positive fourth quarter (lower right). Whatever the determining conditions or traits, the "reversed decline" or incline produced in the McDougall Series is hardly an accident; nor is it the work of any one individual. We shall presently discuss its significance.

[1] The additional precaution has been taken of applying a supplementary statistical test, the "Student's" *t*-test, to this difference. It gives a fully confirmatory result: $t = 3.70$, with a corresponding (doubled) probability of .0002. Further evaluated by the arc sine transformation method proposed by Dr. Greenwood in his article in this issue, the CR is 3.50 and the corresponding (doubled) P is .0004.

To complete the report on the study of position effects, there are two small items to be discussed. The first three runs on the page for the entire series were found to have the following deviations, each deviation representing 27 runs: + 11, − 23, + 7. It is interesting to note this very striking drop on the second run (which is even more marked if the records of the first book are taken alone: + 3, − 28, + 5, with 18 runs represented by each deviation). It would appear that in the second run there is some curious and statistically significant effect ($CR_{diff.\ 1-2} = 2.54$) which may offer some clue to the explanation of the reversal of trend in the position effects in the McDougall data. This decline is the more noteworthy when contrasted with the incline which was found for the general horizontal distribution of the page as a whole. A study of the data on target faces in Table 3 does not reveal an explanation in terms of sequence of targets.

The other item of interest is the chronological distribution. By this is meant the course of variation in rate of scoring over the series as a whole with the work of each subject analyzed separately. There are not enough data available as yet on the entire PK research to allow of a generalization regarding this type of distribution; but declines found in such distribution analyses are of general interest because of their frequent mention in the ESP literature. The results of the examination of the McDougall records on this point are given in Table 6. The data for each subject are divided according to temporal sequence in the experiment as nearly as possible into three sections without breaking up any given session.

Table 6

CHRONOLOGICAL DISTRIBUTION

Section	Runs	Deviation	Av. Score
1st.....	130	+31	4.24
2nd.....	81	+17	4.21
3rd.....	58	+ 1	4.02

The drop in scoring that occurred during the last period may be regarded as in line with the more general evidence of chronological distributions in ESP and PK, although this evidence is not sufficiently striking at this stage to support a prediction. The chrono-

logical distribution here appears to be quite independent of the page distribution effects described above. There are no grounds at this stage of our knowledge for anticipating them to be otherwise.

Discussion

It is now possible in the light of the analyses of position effects to discuss, with some possibility of reaching definite conclusions regarding them, the main hypotheses that counter the PK explanation of the results. The chance hypothesis is of necessity the first consideration. In a number of instances there have been found significant differences in the analyses reviewed above. These are not all independent, but we need not be concerned about any other difference than that shown by the total results in Table 4 with a CR of the difference of 3.32 between the first and fourth quarters. This CR and the internal consistency of the QD's as described above remove any doubt of the extrachance character of the McDougall Series.

The same significant QD presents a sufficient case against the hypothesis that the results may have been produced by physical imbalance in the die used. Such imbalance would not, of course, be a fluctuating one, and it could not be supposed to have given the patterns of distribution which were found. A similar argument holds against the view that the subjects might have consciously or unconsciously used some degree of skill in throwing the die so as to favor the face for which it was thrown. It will, of course, be recalled that a small-necked, vase-shaped dice cup was used throughout the series, and any manipulation would appear out of the question. But in addition, it cannot reasonably be supposed that the subjects persistently, over five months of experimentation, began the record page with scoring well below chance as a result of skill in throwing the die and in the course of the page recurrently reversed the objective so as to bring up the record total for the last quarter to a point about as far *above* expectation as the initial quarter had been *below*. (See the Total section of Table 4.) It would take as much skill (statistically speaking it would take five times as much) to produce the negative deviation of − 34 as it would to produce the positive deviation of + 31. It would have to be supposed that in the first quarter the subject did not want to score and then only changed his mind when he came to the second quarter; but this is

ridiculous because there are five runs in the left half of the page and it would mean that the subject would have to change his attitude at the mid-point of each of these runs. Furthermore, the subjects actually did their best scoring during the first 130 runs out of the 269 that were done. As Table 6 has shown, these first 130 runs averaged 4.24 and there was a drop almost to mean chance expectation in the final chronological section of 58 runs. This does not at all fit in with the hypothesis of the acquisition of skill in the course of practice. There is, in fact, nothing to support that hypothesis. We need only remember that the picture given by the quarter distribution is the picture for every page in the entire series. It is the composite of all the sessions so that the incline shown is not such an incline as could conceivably result from acquired skill. It is something which has to be accounted for within the conditions of a single session as a consequence of an effect of page structure that is much more subtle than manual skill in the handling of the dice.

The main alternatives to the PK hypothesis have been considered. There are of course others. But from the distribution analyses the same argument might be developed against the hypothesis that errors in recording produced the deviation. Why should errors take the form found in the distribution of the QD? How could the error hypothesis explain a negative deviation in the first quarter and a positive deviation in the fourth? The argument against lost records or selection of data is simple: the record books were bound. They are, like all published and unpublished records, available at the Parapsychology Laboratory for examination by qualified students of this field. The optional-stopping issue is not applicable to the QD type of evidence. Thus it seems unnecessary to give lengthy attention to such counter-hypotheses as these (in themselves unusual questions to raise in an academic research discussion).

If, then, as seems reasonable, the McDougall research may be interpreted as going beyond the hypotheses of chance, faulty dice, skilled throwing, as well as various minor questions, it will be worthwhile to give it serious consideration for the bearing it has and the contribution it will make to the knowledge we have of conditions affecting PK. Most important by far is the fact that its QD reverses all precedents in the direction of the distribution. It is so consistently an incline series, both in horizontal and in vertical directions,

that it indicates the presence of a different condition or determinant than we have dealt with in the earlier reports of QD analyses. Any attempt to explain one such instance of reversal would be necessarily hypothetical, perhaps bordering on the speculative, but there is a need for summarizing all the relevant circumstances in the interest of guiding further research.

One of the variants represented in this series is the fact that the subjects *worked as a group.* More than that, they took turns in the throwing, each doing a run at a turn. This made a single column of 24 entries. No other work has been reported that was done in a similar way. We cannot yet determine whether these conditions were responsible for the reversal of the QD. It seems possible that the tedium, not only of throwing a single die from a vase 24 times, but of waiting for three to five others to do the same, might induce subconscious strain that could adversely influence the delicate PK process—even make it miss the target at first. But as scores came in and competitive interest took its place, this initial impatience perhaps gave way to real effort to score high. To some extent this occurred in the individual run, but over the page the effect was cumulative.

There is another set of conditions that merits consideration as possibly relevant to this reversal of QD. The PK test could hardly have been introduced to the McDougall group without some degree of intellectual conflict. All the subjects were old enough and sufficiently sophisticated to require some rational explanation of the objective of the tests and some justification for taking the hypothesis seriously enough even to make a game out of testing it. Professor McDougall's age and eminence and the respect which doubtless all members of the group felt for him were enough to secure outward participation in the experiment, but it seems likely that nothing could have kept all those concerned from feeling at least slightly foolish in thinking they could will the die they were throwing so as to influence it in any appreciable manner. This is necessarily conjectural but quite reasonable. If it be a correct picture of the initial state of mind, it is not difficult to suppose further that a state of mild conflict existed in the subject as he began to throw the die, especially with several others looking on. He was, on the one hand, trying to get into the spirit of the test and actually give his attention to the

die; on the other, he felt a bit silly over it and thought that anyhow it was probably all harmless nonsense. Then, as the first run advanced, he found himself increasingly absorbed in the objective operation—the game aspect—of the test. As the check-up of the first round of scores was completed, the figures induced rational thought which engendered intellectual difficulties. Thus the second run got the biggest depressing effect of the page. But this initial negative effect grew less thereafter with mere repetition as each run was completed, until by the end of the page and session the subject was fairly well involved in the game and accordingly integrated in his concentration of attention on the moving object. As he went down the column, this getting away from initial inhibiting attitudes favored a vertical incline; and with each run on the page the subject's earlier handicapping attitude petered out as he became cumulatively absorbed in the throwing.

But it is less important to hit upon the precise motivation concerned here than it is to observe and interpret correctly the subtle changes that came in as the test advanced over the page. In this respect, the reversal of the QD and the production of the significant diagonal incline are not much out of line with the typical QD. That is, whatever the attitude at the start, *it changed* on the way down the page and across the page. The introduction of a consistent and significant horizontal and vertical distribution change, regardless of direction, is the effect of paramount interest at this stage. The direction of the change and the initial attitude are important, but in a secondary way.

As the PK reports are issued, readers will note that a comparison of the rate of success with different numbers of dice per throw is being made possible. The fact that the McDougall Series was *entirely* based upon the throwing of a single die is therefore of special importance; that is, the important thing to point out here is that there was no throwing of any larger number at a time and consequently no comparison within the experimental series. In the three reports in which the throwing of a single die has been mentioned (**1, 2, 3**), only chance results were given. (The number of trials was small in these instances, for the one-die throwing was a minor phase of the experiment.) It is reasonable to suggest that had there been a comparison of different numbers of dice per throw

in the McDougall Series, the one-die test might have suffered by contrast and that it is the comparison and not the absolute number per throw that counts. Not the number itself, but the relation, is the thing.

Concluding Remarks

At any rate, this series shows that the PK effect can be demonstrated by throwing one die at a time if no competing situation is provided which offers a more interesting rate of progress. On this point alone it is believed the McDougall research makes a sufficient contribution—that is, in its demonstration of the PK hypothesis in tests with the throwing of a single die. It has proved, too, that group testing of PK may be carried out with success, at least under certain conditions. Its unique QD adds a great deal to the growing collection of results of the analyses of position effects. It is doubtless this feature of the work, especially the reversal of the usual QD trend, that will represent its most valuable contribution. To this it is safe to say there will be much future reference.

References

1. Hilton, Homer, Jr., Baer, George, and Rhine, J. B. A comparison of three sizes of dice in PK tests. *J. Parapsychol.*, 1943, **7**, 172-190.
2. Hilton, Homer, Jr., and Rhine, J. B. A second comparison of three sizes of dice in PK tests. *J. Parapsychol.*, 1943, **7**, 191-206.
3. Rhine, J. B. Dice thrown by cup and machine in PK tests. *J. Parapsychol.*, 1943, **7**, 207-217.

ESP TESTS UNDER VARIOUS PHYSIOLOGICAL CONDITIONS

By J. L. Woodruff[1]

ABSTRACT: ESP tests were given to subjects participating in experiments designed primarily to test the effects of low temperature, oxygen deprivation, and certain drugs upon the motor tasks essential to aviation. The ESP tests which were administered prior to the physiological tests were used as the control series; those administered after the physiological tests comprised the experimental series. Seven combinations of drugs, temperature, and oxygen were used in the experiment, and one control combination of normal temperature, normal oxygen, and a sugar capsule in place of the drug.

Of the total of eight combinations, the only one which gave significant results for the ESP tests was that of normal temperature and oxygen with the substitute for the drug. Under these normal conditions, however, the ESP control series gave quite opposite effects from the experimental series under the same conditions. Scoring above chance expectation occurred before the aviation tests, whereas after the two or three hours in the test chamber, the subject scored very low, giving a significant negative deviation from chance expectation. The suggested explanation is that boredom in the aviation test session induced a score-depressing attitude in the subject by the time the ESP tests of the experimental series were given. There is a significant CR of the difference of 3.68 between the results of the tests given before and those given after the aviation test session. It is apparent that chance is not the explanation.

The author, Dr. Woodruff, is now serving with the armed forces of the United States.—Ed.

The Problem

This report is concerned with experimental work on ESP carried out as an adjunct to confidential research work in Aviation Medicine. Because the material is strictly confidential, it is impossible for the writer to be specific concerning the nature of all the variables employed in the experiment. For example, certain drugs were used,

[1] The work reported here was carried out at the Psychology Laboratory of the College of the City of New York. I wish to acknowledge with thanks the permission granted by Dr. Gardner Murphy, Chairman of the Department of Psychology, and by Dr. J. E. Barmack of that department to carry out the research in extrasensory perception. The writer, at the time, held the Harvard University Hodgson Fellowship in Psychical Research and the present paper is based upon a portion of the report submitted to that committee.

but it cannot be stated specifically what they were. One of the conditions consisted of oxygen deprivation and another of lowered temperature, but the extent of these changes has to be left unspecified. As it happens, however, the confidential nature of the investigation need be no serious limitation to the consideration of the findings, as far as these are concerned with the problems of ESP. This is due to the fact that tests associated with the particular conditions just mentioned gave only chance results. The significant experimental findings are connected with conditions which can be described without restriction.

In view of the outcome of the tests in which drugs were used, there is no need here to recall for the reader the various experimental findings in earlier investigations in which drugs were used. It is sufficient perhaps to mention only that such work as that of Brugmans (1) on the effect of alcohol, and of Rhine (4, pp. 45, 78-79) and of Clark and Sharp (2) on the effects of sodium amytal and caffeine on the subjects' performance in ESP tests has, to say the least, justified continued interest and further experimentation on the possible effect of drugs on the ESP subject.

The condition which emerges from this investigation as the one primarily influential in the production of the results is boredom. Boredom and novelty are related, and it may be well to keep in mind the findings of Pratt and Woodruff (3) that when one type of ESP stimulus was used for a period of time, the scoring rate declined, only to rise again with the introduction of each variation in the type of stimulus. Rhine pointed to the decline of scoring rate of subjects working over a long period of time with the same materials and conditions (4, pp. 159-60). He suggested that the monotony of the procedure lowered the motivation of the subject.

The ESP research which is reported here was an outgrowth of an investigation directed primarily to the testing of the effects of low oxygen content and low temperature on the performance of a simple motor task. The task was roughly similar to that performed by the pilot of an airplane. Drugs were then introduced with a view to counteracting the deleterious effects of the oxygen deprivation, the low temperature, and the performance of a very monotonous task. The increase in the number of errors in the performance of

the motor task was taken as measuring the deleterious effects mentioned.

The difference in scoring rate in the guessing of five decks of ESP cards immediately before the long experimental session began and of five more decks at its conclusion was to be taken as measuring the effect of the factors named on ESP performance. Naturally, it was also assumed that if no significant differences between control and experimental series were found or if no significant deviations from the expected number of correct responses were found, the experimental results obtained would fail to contribute evidence on the ESP hypothesis.

In the ESP section of the experiment, the test consisted of the subject's attempting to identify ESP cards from the standard ESP deck—five decks prior to the motor task and five at the conclusion of the physiological test period. The first five decks were intended as a control series and the last five as the experimental series for ESP tests.

Experimental Conditions

The principal apparatus was an oxygen chamber in which the oxygen content of the air could be varied by the introduction of nitrogen. The motor task was performed on a pursuitmeter designed by Dr. J. E. Barmack of the Psychology Department of the College of the City of New York. The task involved was performed continuously by the subjects for a period of several hours with only brief pauses at fifteen-minute intervals.

In the ESP tests, the following procedure was used: A set of five ESP decks was given to each subject as soon as he entered the experimental room. Each deck was enclosed in a cardboard box and the five boxes were glued in a row side by side on a cardboard strip. Appropriate record sheets were provided so that the subject wrote his calls for the cards in box *1* in column *1* of the record sheet. Likewise, the calls for the cards in box *2* were placed in the second column of the record sheet. Thus there were five columns of calls corresponding to the five boxes of cards. After the subject had made his calls, the sheet on which they were written and the strip of cardboard bearing the boxes of cards were taken from him. It will be seen that this test is the one familiar in the ESP liter-

ature as the DT or "down through" procedure since the deck of cards was called without their being removed one by one as the calling proceeded.

The subject was then put through the preliminaries of the aviation procedure—a brief physical examination, administration of the drug, etc.—after which he was introduced into the oxygen chamber for a period of several hours. During this period, he performed the prescribed motor task under one of the eight combinations of conditions of temperature and oxygen called for by the design of the main experiment. As already mentioned, the precise nature of these conditions cannot be given. The sequence of these combinations was arbitrarily varied so that no combination could be inferred in advance by the subject. At the end of this long period, while still in the oxygen chamber under the conditions indicated, the subject was asked to try to identify another set of five boxes of ESP cards. Throughout the entire experiment, he was in full view of the experimenter through a glass window.

The ESP test given before the entrance into the oxygen chamber was the control test for the experimental variables of the chamber and the drugs. It is designated *Control* in the tables of results and the one given at the end of the period in the oxygen chamber is designated *Experimental*. Lactose was employed in certain combinations as a control on the drugs. The other drugs will be designated as either *x* or *y*. The following precautions were taken in the administration of the ESP tests: The cards were enclosed in cardboard boxes at all times while in the subject's sight. A number of sets of cards were used and they were shuffled very thoroughly between experimental sessions. The same cards were not used for the control sessions and the experimental sessions. During the administration of the ESP tests, subjects were observed by some member of the research staff. This prevented any tampering with the boxes.

After each subject had completed his set of calls, the cards and calls were taken from him, and the sheets bearing the calls were placed in one of the cabinets available. The actual order of the cards in the deck was written down on another set of sheets and the decks were shuffled vigorously, but with no specific number of shuffles. They were then returned to the boxes for use in a subsequent experiment. The list of cards as recorded by a member of the

research staff (usually myself) was then compared with the call record made by the subject. Hits were indicated by a small mark in an appropriate column next to the call column. Since the call record and the card record were on different sheets, there is no likelihood that the call record was an influence in the writing of the card record. The records are available for additional rechecking.

It will be noted in Table 1 that the number of runs is not the same for the control conditions and the experimental conditions in all cases. This is to be accounted for by the fact that occasionally the apparatus broke down in the middle of an experimental session. When this occurred, no ESP test was given under experimental conditions. In such cases the control tests were nevertheless included in the results for the sake of completeness. There were eight experimental combinations of conditions for which experimental tests were to be given. Some subjects dropped out before completion of the series, but their results were included as far as they went in order to avoid improper selection of data.

The 14 subjects who participated were for the most part N.Y.A. students at City College. The experimenter had the impression that most of them regarded the ESP test as somewhat of a nuisance, particularly that part which was given at the end of the long and tiresome experimental period.

Results

There were 1,265 runs (31,625 trials) completed in the ESP experiment. As the tables to follow indicate, it was only under the normal set of conditions that significant results were obtained. These results in terms of number of runs and total deviation from expectation are given in the first column of Table 1. The 85 runs of the control group—that is, the tests which were given at the beginning of the session in the experimental chamber—gave a total of 466 hits which is 41 above mean chance expectation. The experimental tests given at the end of the session have only 300 hits for the total of 70 runs, whereas 350 are expected. Thus a *negative* deviation of 50 is given which is even further below the mean than the control series deviation is above.

It is noteworthy that no other combination of conditions gave even a highly suggestive deviation, positive or negative; nor is there

any marked trend evident in the results shown in the table. The hypothesis of chance must therefore be regarded as the reasonable explanation for all the results shown except those of the first column.

Table 1

RESULTS FOR ALL CONDITIONS IN TERMS OF DEVIATION FROM EXPECTED NUMBER OF HITS

	Lactose Normal Temp. Normal Oxygen	Lactose Low Temp. Normal Oxygen	Lactose Low Temp. Low Oxygen	Lactose Normal Temp. Low Oxygen	Drug *x* Low Temp. Low Oxygen	Drug *x* Normal Temp. Low Oxygen	Drug *y* Low Temp. Low Oxygen	Drug *y* Normal Temp. Low Oxygen
Control*								
Runs....	85	75	80	85	85	80	75	85
Dev.....	+41	−23	−11	+ 5	−11	+ 7	+16	− 6
Exper.								
Runs....	70	75	80	80	75	80	75	80
Dev.....	−50	+13	−20	+ 1	+ 5	−10	− 9	−15

*The eight combinations of special conditions were effective only on the *experimental* runs. The control runs were made before the conditions were introduced.

In view of the fact that attention centers mainly on those results of the tests made under the normal conditions, it will be of interest to give further details of the evaluation of that section of the experimental data. Of course it is important to recognize that the evaluation is being made on a selective basis in the sense that one out of the total of eight sets of conditions is being evaluated apart from the total. The results, however, as Table 2 presents them, are such that due correction may be made in as far as such correction is regarded as statistically appropriate. The control series is seen to give a CR

Table 2

RESULTS FOR THE LACTOSE, NORMAL TEMPERATURE, AND NORMAL OXYGEN CONDITION

	Runs	Deviation	Av. Score	SD	CR	$CR_{diff.}$
Control.....	85	+41	5.48	±18.44	2.22	3.68
Experimental...	70	−50	4.29	±16.73	2.99	

of 2.22, which is highly suggestive. The experimental series has a CR of 2.99 and is significant. The CR of the difference is 3.68. This CR may well be corrected for the selection mentioned by multiplying the probability which it represents by eight. The probability would still be quite significant. It would be significant also if it were to be further corrected for the use of an alternative method

of evaluation: the CR of the deviation instead of the CR of the difference. That is, if the probability representing a CR of 3.68 were to be corrected on the assumption that a selection of one out of 16 possible evaluations were represented, it would still be a definitely significant value. The question could still be carried further, but this would seem to cover any debatable point.

It may be of interest to note here the estimated probabilities for the principal CR's represented. The CR of 2.99 has a corresponding probability of .0014 which should be doubled to allow for the evaluation of the deviation in either direction from the mean. The probability of a CR of 3.68 is less than .0001.

Discussion

The significant results obtained in this study were obtained under conditions in which the only relevant variable seems to be that of increasing boredom. Perhaps the most interesting aspect of the results is the fact that the experimental session in lactose, normal temperature, and normal oxygen gave *significantly low* scoring. The question arises as to why the results on this experimental session should be so much lower than on any other experimental session. This set of conditions was the only one in which boredom alone prevailed; that is, in which the other variables were absent. In such a circumstance, the boredom factor, which must be taken into account on every experimental session, might well be at a maximum. That is, if a real drug were given, or if the temperature or oxygen content were varied, such variations might well be diverting and might minimize the boredom effect. Thus even if those changes added to the discomfort of the subject, they would, at the same time, partially counteract a feeling of boredom. It is suggested, then, that the supposedly more intense boredom under the experimental condition may have caused the most marked negative deviation; subjects who are intensely bored will not score at an expected level, but considerably below that level. Of the 14 subjects who participated under these conditions, 12 had negative deviations and the other two had no deviations. *Thus no subject had a positive deviation under these conditions.* This alone is indeed an impressive circumstance.

Concluding Remarks

All of the seven combinations of special physiological conditions, including drugs, temperature, and oxygen variation, produced only insignificant results in the ESP tests when the results of the 14 subjects were pooled. On the other hand, the normal conditions of temperature and oxygen with a sugar substitute for the drug gave scores sufficiently above expectation in the initial or control tests to be worthy of some interest. And the results of the experimental series of ESP tests for these conditions were distinctly significant though their significance is based upon a negative rather than a positive deviation. These two series, the control and the experimental, have a very significant difference. Correction even for the fact that these results represent only a part of the total ESP work does not alter the fact that the work is significantly different from expectation on the theory of chance.

The best interpretation of the significant negative deviation of the experimental series referred to appears to be that it is due to an effect of extreme boredom resulting from the confinement for several hours in the experimental chamber. The absence of the physiological variables present in the other seven combinations of the conditions is interpreted as contributing to the boredom in the control series. It is, however, recognized that this suggestion is hypothetical and that further work would be required to render it a conclusive explanation.

It seems safe in the end to view the results as representing evidence of the occurrence of ESP and to recognize that some factors in the varying conditions were responsible for the differences in the results, whether or not the suggestion of the effect of boredom be the correct one.

References

1. Brugmans, H. J. Some experiments in telepathy performed in the Psychological Institute of the University of Groningen. *Compte-Rendu du Premier Congrès International des Recherches Psychiques*, 1921.
2. Clark, C. C., and Sharp, Vernon. The effect of sodium amytal and caffeine upon ESP test performance. [Unpublished manuscript.]
3. Pratt, J. G., and Woodruff, J. L. Size of stimulus symbols in extra-sensory perception. *J. Parapsychol.*, 1939, **3**, 121-158.
4. Rhine, J. B. Extra-Sensory Perception. Boston: Bruce Humphries, 1934.

FREQUENCY DISTRIBUTIONS OF ESP SCORES FOR CERTAIN SELECTED CALL-PATTERNS

By T. N. E. Greville

ABSTRACT: This note gives the complete frequency distributions obtained in matching a standard ESP deck against six selected call-series patterns. Comparison of these distributions shows that, within the range of variance encountered in practice, the frequency distribution varies in a consistent and regular manner with variation in the call-pattern. Specifically, the frequencies of scores 4 to 7, inclusive, increase, and the frequencies of all other scores decrease, with decreasing variance.

Dr. Greville, a mathematician of probability, is in the service of the U. S. government in Washington, D. C. On several occasions he has made contributions to the mathematics of the ESP research.—Ed.

In the usual ESP experiment, the subject makes a series of 25 calls, choosing for each call one of the five standard symbols; and the series of calls is then matched against the order of the cards in a shuffled deck. The number of correspondences or agreements between the call series and the shuffled deck (or "target" series) is the score for the particular run. The chance or non-chance character of a particular set of data is judged by comparing the scores actually obtained with the theoretical frequency distribution of scores based on chance alone. In the earliest work with ESP cards it was assumed that a subject's calls were independent of one another: that he was equally likely to name any one of the five symbols on a given call, irrespective of the pattern of previous calls. This assumption leads to the binomial distribution, in which the frequency distribution of scores is given by the successive terms of the binomial expansion $(4/5 + 1/5)^{25}$. This provides a simple and convenient mathematical formulation of the problem, but it is doubtful whether such a condition is ever fully realized. It would mean, for example, that a subject would be as likely to call one suit 25 times as any other particular pattern of calls.

A more precise evaluation of a particular run requires the computation of the frequency distribution associated with the call-pattern

actually observed. This frequency distribution is obtained by matching the observed arrangement of the call series with all possible arrangements of the target deck, and considering the number of correspondences or "hits" which would result from each matching. Since the target deck contains five of each suit, the frequency distribution obtained will depend only on the particular partition of 25 which characterizes the call series (e. g., 6-6-5-5-3). It will not be affected by merely changing the order of the symbols making up the call series, or by interchanging certain symbols while retaining the same numerical division of 25.

At an early stage in the controversy over the mathematics of ESP, the question was raised as to the effect, on the frequency distribution of scores, of preferences on the part of the subject for certain symbols, or of other departures from true randomness in the call series (see, for example, **6**, pp. 204-205). It was shown that the average number of hits remains at five (so long as the target deck contains five cards of each suit), but that the variance changes with the distribution of the call deck. Greenwood (2) gave formulae for the first four moments of the frequency distribution of hits in matching two decks of arbitrary composition, and showed that in the standard ESP situation the maximum variance ($4\frac{1}{6}$, as compared with 4 for the binomial distribution) is given by a call series containing five of each symbol. Hence, by using this value in all cases one secures a conservative evaluation of the critical ratio. On the other hand, Greenwood has also shown (1) that the average variance obtained from empirical data closely approximates that of the binomial. This finding justifies the use of the binomial assumption for ordinary purposes—a procedure which facilitates the solution of many complicated problems.

Nevertheless, it is desirable to know in complete detail how the frequencies of the various scores are affected by changes in the composition of the call series and a frequency distribution is not, of course, completely characterized by four moments. It is not feasible, or necessary, to compute the distribution for all the possible partitions of 25. Comparison of the frequency distributions for a few selected cases gives a very clear picture of the effect of variation in the call pattern. Table 1 gives six such distributions, calculated by a method previously described by the author (3). The six cases selected are:

(i) The 5-5-5-5-5 case, which is the one of greatest theoretical interest. It produces the maximum variance and corresponds to the actual conditions of certain experimental conditions such as the ESP shuffle and a technique described by Martin (**5**, p. 187).

(ii) The 7-6-5-4-3 pattern, which is the one found by Greenwood (**1**: p. 141) to occur most frequently in the empirical data studied by him.

(iii) The 6-6-5-5-3 pattern, which was found by Greenwood to be second in frequency of occurrence.

(iv) The 8-6-5-4-2 pattern, which is of interest because it produces the same variance as the binomial distribution.

(v) The 12-6-4-2-1 pattern, which produces the smallest variance of all those occurring in Greenwood's data.

(vi) As an extreme case, the 25-0-0-0-0 pattern, which gives a zero variance.

The frequencies for the first two cases were previously given in the book *Extra-Sensory Perception After Sixty Years* (p. 28), and those for the 5 x 5 case appeared also in a previous article by the author (**3**, p. 59).

For convenience, the six distributions are arranged in Table 1 in the order of decreasing variance. This differs from the order given above only in placing case (iii) before case (ii). If the extreme 25-0-0-0-0 case is not considered, the remaining columns show a consistent increase with decreasing variance in the frequencies of scores 4 to 7, inclusive; and a consistent decrease with decreasing variance in the frequencies of all other scores. Thus, the effect of varying the call-pattern is particularly simple and consistent over the range of variance observed by Greenwood in the actual data studied by him. Speaking academically, there must of course be a point at which the frequencies of scores 4, 6, and 7 cease to increase and start to approach zero. However, rough calculation indicates that this will occur far outside the range likely to be encountered in actual experimentation.

Table 2 shows the relative frequencies for the same six distributions. For comparison, the relative frequencies of the binomial distribution are also given. It will be noted that the relative frequencies of the binomial and those for the 8-6-5-4-2 call-pattern, which gives the same variance, never differ by more than one unit in the fourth decimal place.

Table 1

FREQUENCY DISTRIBUTIONS OBTAINED IN MATCHING CERTAIN SELECTED CALL-PATTERNS AGAINST A STANDARD ESP DECK

Pattern....	5-5-5-5-5	6-6-5-5-3	7-6-5-4-3	8-6-5-4-2	12-6-4-2-1	25-0-0-0-0
Variance...	4.1667	4.1167	4.0833	4.0000	3.5333	0.0000
Score			Frequencies			
0.....	2,671,644,472,544	2,577,670,089,912	2,509,330,196,142	2,354,856,492,576	1,480,088,199,744	0
1.....	15,865,811,944,500	15,528,566,535,102	15,290,449,805,862	14,726,280,336,096	11,386,416,954,528	0
2.....	45,731,782,136,000	45,270,450,712,800	44,959,125,271,788	44,173,830,305,472	39,407,433,923,088	0
3.....	85,133,419,148,000	84,982,983,832,134	84,902,784,854,706	84,636,797,800,464	83,111,759,050,320	0
4.....	114,899,955,351,500	115,319,526,021,000	115,627,769,033,616	116,341,344,089,376	120,995,244,160,560	0
5.....	119,649,702,080,045	120,385,073,021,868	120,886,029,287,352	122,149,592,016,144	130,027,208,783,472	623,360,743,125,120
6.....	99,898,095,752,500	100,468,688,333,868	100,836,658,021,302	101,811,728,480,784	107,436,482,833,680	0
7.....	68,610,815,305,250	68,770,933,033,140	68,857,292,427,840	69,114,991,852,032	70,055,972,322,336	0
8.....	39,465,495,670,000	39,305,995,752,528	39,186,278,015,724	38,894,208,678,528	36,655,943,884,560	0
9.....	19,259,638,091,625	18,999,164,736,900	18,822,597,375,204	18,368,944,854,624	15,540,360,675,840	0
10.....	8,049,523,797,520	7,837,874,595,216	7,699,903,009,344	7,342,102,246,752	5,358,325,884,912	0
11.....	2,900,903,187,000	2,777,312,705,076	2,699,094,568,716	2,496,866,093,904	1,499,341,635,792	0
12.....	905,725,731,000	848,844,579,960	813,829,852,836	724,339,268,016	337,343,230,224	0
13.....	245,747,376,750	224,281,785,096	211,441,374,648	179,258,236,752	59,887,341,360	0
14.....	58,035,729,000	51,254,965,440	47,323,693,368	37,743,395,088	8,101,813,104	0
15.....	11,932,686,260	10,118,020,392	9,102,772,224	6,721,117,536	783,965,952	0
16.....	2,133,510,000	1,719,702,792	1,497,546,414	1,001,896,896	47,135,088	0
17.....	331,120,250	250,292,430	209,033,622	122,926,944	1,330,560	0
18.....	44,350,000	30,915,360	24,447,276	12,082,560	0	0
19.....	5,141,000	3,203,190	2,347,506	905,856	0	0
20.....	497,500	271,440	179,172	47,376	0	0
21.....	44,125	18,540	10,080	1,344	0	0
22.....	2,500	900	378	0	0	0
23.....	250	36	0	0	0	0
24.....	0	0	0	0	0	0
25.....	1	0	0	0	0	0
Total.....	623,360,743,125,120	623,360,743,125,120	623,360,743,125,120	623,360,743,125,120	623,360,743,125,120	623,360,743,125,120

Table 2

RELATIVE FREQUENCIES OBTAINED IN MATCHING CERTAIN SELECTED CALL-PATTERNS AGAINST A STANDARD ESP DECK

Pattern..... Score	5-5-5-5-5	6-6-5-5-3	7-6-5-4-3	8-6-5-4-2	12-6-4-2-1	25-0-0-0-0	Binomial
0.....	.0042859	.0041351	.0040255	.0037777	.0023744	.0000000	.0037779
1.....	.0254521	.0249110	.0245291	.0236240	.0182662	.0000000	.0236118
2.....	.0733633	.0726232	.0721238	.0708640	.0632177	.0000000	.0708355
3.....	.1365717	.1363303	.1362017	.1357750	.1333285	.0000000	.1357680
4.....	.1843234	.1849965	.1854909	.1866357	.1941015	.0000000	.1866811
5.....	.1919430	.1931226	.1939263	.1959533	.2085906	1.0000000	.1960151
6.....	.1602573	.1611726	.1617629	.1633271	.1723504	.0000000	.1633459
7.....	.1100660	.1103228	.1104614	.1108748	.1123843	.0000000	.1108419
8.....	.0633108	.0630550	.0628629	.0623944	.0588037	.0000000	.0623486
9.....	.0308965	.0304786	.0301954	.0294676	.0249300	.0000000	.0294424
10.....	.0129131	.0125736	.0123522	.0117783	.0085959	.0000000	.0117769
11.....	.0046537	.0044554	.0043299	.0040055	.0024053	.0000000	.0040149
12.....	.0014530	.0013617	.0013056	.0011620	.0005412	.0000000	.0011710
13.....	.0003942	.0003598	.0003392	.0002876	.0000961	.0000000	.0002928
14.....	.0000931	.0000822	.0000759	.0000605	.0000130	.0000000	.0000627
15.....	.0000191	.0000162	.0000146	.0000108	.0000013	.0000000	.0000115
16.....	.0000034	.0000028	.0000024	.0000016	.0000001	.0000000	.0000018
17.....	.0000005	.0000004	.0000003	.0000002	.0000000	.0000000	.0000002
18.....	.0000001	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000
19.....	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000

REFERENCES

1. BATTIN, I. L. On the problem of multiple matching. *Annals of Math. Statistics*, 1942, **13**, 294-305.
2. GREENWOOD, J. A. Analysis of a large chance control series of ESP data. *J. Parapsychol.*, 1938, **2**, 138-146.
3. ———. The first four moments of a general matching problem. *Annals of Eugenics*, 1940, **10**, 290-292.
4. GREVILLE, T. N. E. Exact probabilities for the matching hypothesis. *J. Parapsychol.*, 1938, **2**, 55-59.
5. ———. The frequency distribution of a general matching problem. *Annals of Math. Statistics*, 1941, **12**, 350-354.
6. MARTIN, D. R. Chance and extra-chance results in card matching. *J. Parapsychol.*, 1937, **1**, 185-190.
7. WILLOUGHBY, R. R. A critique of Rhine's *Extra-Sensory Perception. J. Abn. and Soc. Psychol.*, 1935, **30**, 199-207.

ON THE EVALUATION OF DIFFERENCES OF SUCCESS RATIOS FROM BINOMIAL POPULATIONS

By Lt. J. A. Greenwood, U.S.N.R.

The distribution of the success ratio in a set of N trials with constant probability p of success on a single trial is not far from normal for moderate values of p and for large N. That is, one can get a fair idea of the actual probability of a difference of two observed success ratios $P_1 - P_2$ using the familiar $\sigma^2_{P_1-P_2} = \frac{p_1 q_1}{N_1} + \frac{p_2 q_2}{N_2}$ and entering the normal tables. Here we let p_1, p_2 designate the population parameters, N_1, N_2 the respective numbers of trials, and $q_1 = 1 - p_1$, $q_2 = 1 - p_2$. However, when p_1 and p_2 are unknown, recourse must be had to other methods since the variance of $P_1 - P_2$ is then unknown.

The method to be suggested involves the inverse sine transformation on the binomial variates. The fact is that the sampled binomial variates can be transformed so that the new variates (asymptotically with N) become normal and have a variance equal to ¼ (1: p. 146). The great power of this method is that the variance is stabilized and made asymptotically independent of the binomial population parameter p.

Let

$$T' = \sqrt{N} \text{ arc sin } \sqrt{P}, \text{ in radians, and} \tag{1}$$

$$t' = \sqrt{N} \text{ arc sin } \sqrt{p}, \text{ in radians.}$$

Then $T' - t'$ is approximately normally distributed for large N with mean zero and variance equal to ¼.

Place subscripts 1 and 2 on the symbols of (1) to designate variates, parameters, and numbers of trials of two binomial populations.

In order to obtain the significance of the difference between P_1 and P_2 we let

$$T = \frac{T'}{\sqrt{N}} = \text{arc sin } \sqrt{P},$$

(2)

$$t = \frac{t'}{\sqrt{N}} = \text{arc sin } \sqrt{p}.$$

Then $T' - t$ is nearly normal with mean zero and variance $\frac{1}{4}N$, and $T_1 - T_2$ is approximately normally distributed with mean $t_1 - t_2$ and variance $\frac{1}{4}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)$.

Suppose the hypothesis to be tested is that $t_1 = t_2$. This implies that $T_1 - T_2$ is nearly normal with mean zero and variance $\frac{1}{4}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)$. One finds the critical ratio from

$$(3) \quad \text{CR} = \frac{2(T_1 - T_2)}{\left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{\frac{1}{2}}} \quad *$$

and enters the normal tables for the probability. An example follows.

In series 1, four dice were thrown for a total of 1,440 single die-throws for sixes, giving 296 successes. In series 2, the same dice were thrown for a total of 1,920 single die-throws for sixes, giving 310 successes. Do the results differ significantly? The computation is shown schematically below.

Series	N	P	$\sqrt{P}$	T (radians)
1.....	1,440	.206	.454	.472
2.....	1,920	.161	.401	.413

From (3)

$$\text{CR} = \frac{2(.472 - .413)}{\left(\frac{1}{1{,}440} + \frac{1}{1{,}920}\right)^{\frac{1}{2}}} = 3.38.$$

The difference is significant at the .0008 probability level.

* If T is found in degrees instead of radians, that means that the scale of the variate T has been multiplied by $180/\pi = 57.3$ and so the standard deviation or denominator in (3) must also be multiplied by 57.3. Then

$$\text{CR} = \frac{2(T_1 - T_2) \quad \text{in degrees}}{57.3\left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{\frac{1}{2}}}$$

The actual conditions of the illustration given were intended to be typical of those of the PK experiments with dice published in the *Journal*. As it is well known that dice are generally biased, it is not safe to assume that there is a probability of 1/6 of throwing a given face on one throw with a single die. However, we have surmounted that difficulty with the inverse sine transformation but are yet confronted with the following one. Since several dice were thrown at one time, the results must be assumed to have come from a combination of *different* binomial populations with unknown parameters, p. In general the basic population cannot therefore be said to be binomial and doubt is cast upon the validity of the results thus obtained as if it were binomial. I shall now show that in such a situation the computed CR is not greater than the true one (is generally less), and therefore with regard to that discrepancy the results err on the side of conservatism.

Suppose r dice are thrown and let the respective parameters (probabilities of success) be $p_1, p_2, \ldots, p_r$. Let $P_1, P_2, \ldots, P_r$ be corresponding statistics (observed success ratios or estimates of the parameters). The distributions of $P_1, P_2, \ldots, P_r$ are asymptotically normal and so is that of $P_0 = \frac{1}{r}(P_1 + P_2 + \ldots + P_r)$. In reference 2, p. 365, it is shown that if the actual situation of throwing the r unequal dice is replaced by the hypothetical throw of r identical dice of parameter $p_0 = \frac{1}{r}(p_1 + p_2 + \ldots + p_r)$, then the variance of the number of successes of the throw of the uniform dice is greater than that for the single throw of the non-uniform dice. Note that the two situations have the same expected value or mean.

In the inverse sine method of evaluation, we have implicitly substituted the hypothetical situation for the actual and have reduced the hypothetical binomial to a special normal form (asymptotically) in order to obtain a probability. But note that this is merely another means of obtaining the probability inherent in the hypothetical binomial result P_0 itself. Since the latter probability is associated with a hypothetical variance which exceeds the actual variance, the CR computed by the inverse sine method is less (not greater, if we admit that the dice might all have the same value of the parameter) than is warranted by the actual situation. This is of course all on

the assumption that the number of trials is sufficiently large that the asymptotic character of the variates has asserted itself in the results.

References

1. Curtiss, J. H. On transformations used in the analysis of variance. *The Annals of Mathematical Statistics*, 1943, **14**, No. 2, 107-122.
2. Yule, G. U., and Kendall, M. G. An Introduction to the Theory of Statistics. London: G. Griffin and Co., Ltd., 1937. Pp. xiii + 570.

GLOSSARY

In order to avoid constant redefining of commonly recurring terms in papers appearing in this journal, the following definitions are submitted for convenient reference. In case of any discrepancy between glossary and usage in the text of an article, the latter should be followed. Words defined elsewhere in the glossary are italicized in the text of the definitions.

AGENT: In tests for *telepathy*, the person whose mental states are to be apprehended by the *percipient*. In *GESP* tests, the person who looks at the *stimulus object*.

AVERAGE SCORE: Average number of *hits* per *run*.

BM (BLIND MATCHING): The technique in which the *subject matches* a *deck* of *ESP cards* to five *key cards* which are laid out face-down before him in an unknown order. Unless otherwise stated, the order is also unknown to the experimenter.

BT (BEFORE TOUCHING): The technique in which the top card of the face-down *deck* is *called* and, after being called, is laid aside for *checking* at the end of the *run*. Each card in the *deck* is treated in the same way.

CALL v.: To attempt to identify a *target* or *stimulus object* (or mental state of an *agent* in *telepathy*).

CALL n.: The *response* described above; also the resulting selection.

CHANCE: The complex of undefined causal factors irrelevant to the purpose at hand.

CHANCE EXPECTATION = MEAN CHANCE EXPECTATION: The most likely *score* if only *chance* obtains.

CHANCE AVERAGE: *Mean chance expectation* in terms of *average* per *run*.

CHECK: To determine a *score* after the completion of a *run* by comparing the order of the *subject's calls* with the order of cards in the deck.

CHI-SQUARE: A sum of quantities each of which is a *deviation* squared divided by an expected value. Also a sum of the squares of *CR's*. (Occasionally the square of a simple CR may be used as chi-square.)

CLAIRVOYANCE: *Extra-sensory perception* of objective events as distinguished from *telepathic* perception (of the mental or subjective events of another person).

COVARIATION: Correlation evaluated in terms of theoretical means and *standard deviations*.

CR (CRITICAL RATIO): A measure to determine whether or not the observed *deviation* is *significantly* greater than the expected random fluctuation about the *average*. The CR is obtained by dividing the observed *deviation* by the *standard deviation*. (The *probability* of a given CR may be obtained by consulting tables of the probability integral, such as Pearson's.)

CR OF THE DIFFERENCE: The observed difference between the *score averages* of two samples of data divided by the *standard deviation of the difference*. (Where the samples to be compared are of equal number of runs, the difference between total *hits* may be divided by the *SD* of the total number of *runs* of both samples.)

DECK: Twenty-five *ESP cards*, five of each suit.

DEVIATION: The amount an observed number of *hits* or an *average score* varies from the *mean chance expectation* or *chance average*. A *deviation* may be total (for a series of *runs*) or average (per *run*).

DIE-THROW: The throwing or mechanical release of a single die regardless of the number thrown at the same time.

DT (DOWN THROUGH): The technique in which the cards are *called* down through the *deck* before any are removed or *checked*.

EMPIRICAL CONTROL: An experiment which wholly or partially follows the main experiment with the exception that the conditions are designed to exclude the possibility of *ESP*.

ESP (EXTRA-SENSORY PERCEPTION): Response to an external event (perception) not presented to any known sense.

ESP CARDS: Cards, each bearing one of the following five symbols: star, circle, three parallel wavy lines (called "waves"), square, plus.

ESP SYMBOLS: See plate opposite page 1, this journal, Vol. 1, March 1937.

ESP TESTS: A considerable number of techniques come under this heading which are conveniently represented by initials, the principal ones being: *BT, DT, PT, GESP, BM, OM, STM*.

EXPECTATION; see CHANCE.

EXTRA-CHANCE: Not due to *chance* alone.

FREE MATERIAL: *Stimulus objects* that are not limited to a known number of categories.

GESP (GENERAL EXTRA-SENSORY PERCEPTION): A technique designed to test the occurrence of *extra-sensory perception*, permitting either *telepathy* or *clairvoyance* or both to operate.

HIT: The correct correspondence of a *subject's call* or response with a *stimulus card* or *object*.

HIT FREQUENCY DISTRIBUTION: The grouping of the total *hits* in a *series* of *runs* with respect to their original position in the *run*.

KEY CARD: One of the five cards (where there are five suits) against which the cards of the test *deck* (i.e., *target cards*) in the *matching* tests (*OM, BM, STM,* etc.) are *matched.*

MATCHING: A form of *calling* in which a *target card* is placed opposite the *key card* which the *subject* selects to identify it. Also, in the evaluation of *free material,* the act of a judge in identifying a given *response* with a *stimulus object.*

MEAN CHANCE EXPECTATION; see CHANCE.

OM (OPEN MATCHING): The technique in which a *subject matches* a *deck* of *ESP cards* to five *key cards* which are face-up before him.

P (PROBABILITY): A mathematical estimate of the expected relative frequency of a given event if chance alone were operative.

PARAPSYCHOLOGY: A division of psychology dealing with the "extra-normal"—those psychical effects which appear not to fall within the scope of what is at present normal and recognized law.

PERCIPIENT: The person who makes the *calls* in a test situation.

PK (PSYCHOKINESIS): The direct influence exerted on a physical system by a *subject* without any known intermediate energy or instrumentation.

RESPONSE: The act of the *subject* in attempting to identify the *stimulus object.*

RSR (RUN SALIENCE RATIO): A measure of *salience* within the *run.*

RUN: A succession of *trials,* usually the *calling* of a *deck* of 25 *ESP cards* or symbols. In *PK tests,* 24 single *die-throws* regardless of the number of dice thrown at the same time.

SALIENCE: The relation of rate of success in the end *segments* of the *run* to that of the middle *segments;* also the relation of the rate of success in the end *trials* of the *segment* to that of the middle *trials.*

TERMINAL SALIENCE: A higher rate of *deviation* in the end *segments* of the *run* (or in the end *trials* of the *segment*) than in the middle *segments* (or *trials*).

MIDDLE SALIENCE: A higher rate of *deviation* in the middle *segments* of the *run* (or in the middle *trials* of the *segment*) than in the end *segments* (or *trials*).

SCORE: The number of hits made in one *run.*

TOTAL SCORE: *Score* of any number of *runs.*

AVERAGE SCORE: *Total score* divided by number of *runs.*

SCREEN: An opaque barrier used between the *subject* and the card or *agent.* The main types of screens are illustrated in this journal on their first introduction in print.

SD (STANDARD DEVIATION): The theoretical root mean square of the *deviations*. It is obtained from the formula $\sqrt{npq}$, in which n is the number of single *trials*, p the *probability* of success per *trial*, and q the *probability* of failure. (For *ESP cards*, $SD = 2\sqrt{\text{no. of } runs}$.)

SD OF THE DIFFERENCE: For both *ESP cards* and *PK tests* using dice, the *SD* of the difference is equal to $\sigma_s \sqrt{1/R_1 + 1/R_2}$ where σ_s is the *SD* of a single *run* and R_1 and R_2 are the number of *runs* in the respective samples compared. This gives the *SD* of the difference for *run score averages*.

SEGMENT: One of the five consecutive sets of five *calls* in a *run* of 25 *trials*. The first five *calls* would constitute the first *segment;* the second five, the second, etc.

SERIES: Several *runs* that are grouped in accordance with a stated principle.

SIGNIFICANCE: A numerical result is significant when it equals or surpasses some criterion of degree of chance improbability. Common criteria are: a probability value of .01 or less, or a *deviation* in the expected direction such that the *critical ratio* is 2.5 or greater.

SR (SALIENCE RATIO): A measure of the relation of the rate of success in the end *segments* of the *run* (or in the end *trials* of the *segment*) and that of the middle *segments* (or *trials*). (For details of the manner of obtaining SR's, see Vol. 5, pp. 193-195.)

SSR (SEGMENTAL SALIENCE RATIO): A measure of *salience* within the *segments* of the *run*.

STIMULUS OBJECT: The *ESP card* or drawing or other object, some identifying characteristic of which is to be apprehended by the *subject*.

STM (SCREENED TOUCH MATCHING): The technique in which the *subject* makes his *call* by pointing to one of five positions or exposed *symbols* under a special *screen*. The experimenter places the *target card* so designated in the position pointed to. The *screen* blocks all vision by the *subject* of the *cards* and their manipulation by the experimenter.

SUBJECT: The person who is experimented upon. Most commonly the *percipient* in *ESP*, though also the *agent* in *telepathy*.

TARGET: In *ESP tests*, the *stimulus object*. In *PK tests*, the faces of the die (or combination of faces) which the *subject* attempts to bring up in the act of throwing.

TARGET CARD: The *card* which the *percipient* is attempting to perceive (i.e., to identify or otherwise indicate a knowledge of).

TARGET DECK: The *deck* of cards the order of which the *subject* is attempting to identify.

TELEPATHY: *Extra-sensory perception* of the mental activities of another person. It does not include the *clairvoyant* perception of objective events.

TRIAL: A single attempt to identify a *stimulus object*.

THE PARAPSYCHOLOGY LABORATORY OF DUKE UNIVERSITY

The Parapsychology Laboratory is a branch of the Department of Psychology of Duke University. Its beginnings date from 1930 when its research program was launched by members of that department under the sponsorship of Professor William McDougall and with the favorable interest of the President of the University—at that time, Dr. William Preston Few. The first report of the work done was issued in book form in 1934 under the title *Extra-Sensory Perception;* in 1937, the *Journal of Parapsychology* was established; and in 1940, five members of the Laboratory staff collaborated to publish *Extra-Sensory Perception After Sixty Years* which summarized not only the work of the Laboratory but similar research, past and present, in other places as well.

Problems of Research. The main topic of investigation has been extra-sensory perception in its various phases—telepathy, clairvoyance, and precognition. Other work has been planned, and to some extent undertaken; but the main objective has been to arrive first at a fair understanding of extra-sensory perception itself before turning definitely to other problems of parapsychology. However, an experimental staff is being trained and methods developed for a broad and inclusive program which awaits only the time and the necessary financial assistance for its execution.

Financial Support. The Laboratory is supported jointly by Duke University and by a small group of donors who have thus far desired to remain anonymous. The Prince Memorial Fund, established in 1934, provided two fellowships in Parapsychology at Duke over a five-year period. In 1935, the William McDougall Research Fund was created at the University with a special committee, appointed by the major contributor and approved by President Few, to administer it. This is called the Parapsychology Research Committee and is self-perpetuating. In order to insure that the purposes of the contributors will always be respected, the Committee has been given the legal authority to constitute itself a board of trust, independent of the University should the need ever arise.

Thus far, the Laboratory has had no permanent endowment and has been maintained by current contributions. This situation has resulted in a lack of long-term security and has worked definitely against the planning of large undertakings. It has also occasioned the serious loss of valuable, trained personnel.

In the judgment of this Committee, it is highly important that, either by will or gift, grants be made of sufficient size to establish an endowment that would guarantee not only the permanence but the stabilization of the research program of the Parapsychology Laboratory.

(Signed) ALICE M. BALDWIN, *Dean, Woman's College, Duke University (Chairman)*
HORNELL HART, *Professor of Sociology, Duke University*
O. K. MERRITT, *President, Renfro Hosiery Mills*
J. B. RHINE, *Professor of Psychology, Duke University*
The Parapsychology Research Committee, Duke University