The Journal of Parapsychology

A SCIENTIFIC QUARTERLY DEALING WITH EXTRASENSORY PERCEPTION, THE PSYCHOKINETIC EFFECT, AND RELATED TOPICS

Volume 8 JUNE, 1944 Number 2

Contents

DUKE UNIVERSITY PRESS
DURHAM, N. C.

The Journal *of* Parapsychology

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 8 JUNE, 1944 Number 2

EDITORIAL

TEN YEARS' RETROSPECT OF ESP RESEARCH

AT THE TIME of writing it is almost exactly ten years since the monograph, *Extra-Sensory Perception*, a report of the first three years of the parapsychological investigation at Duke, appeared in publication. For a number of reasons the ten years that have followed may be marked off in the history of parapsychology as a distinctive decade, one that displayed features and trends that are of interest to the student who is looking for factual guidance in the future.

One conspicuous feature in the bibliography of ESP is the spread of research activity in the decade that followed the publication of the monograph mentioned. A greater number of important investigations of ESP occurred in this ten-year period than in all the previous history of the subject, a history that covers at least fifty years if we take the founding of the Society for Psychical Research in London in 1882 as a starting point. It is not only in reports of research that the expansion in this field is noteworthy; critical and popular discussions of the research show comparable growth. Most significant of all of the evidences of expansion is the extent to which the investigation of ESP has invaded the collegiate laboratory of psychology and involved the participation of psychologists themselves. Also, in the psychical research societies, to which the work of the half-century of parapsychology prior to

1934 is largely due, an equally marked resurgence of experimental activity in the province of ESP took place.

Much more important, however, than the mere extent of inquiry is the question: What did the decade to which we are looking back yield in the way of discoveries of value concerning ESP? For one thing, the question of the occurrence of ESP was brought for the first time to sharp focus for the psychological profession. The subject of ESP, although a controversial one, became a topic for serious classroom and laboratory consideration, became the subject of numerous graduate and undergraduate theses in this country and abroad. In the researches conducted, experimental discrimination between telepathy and clairvoyance was taken care of for the first time (although some parapsychologists feel that there are still further distinctions to be made). ESP capacity was found among a wide range of subjects tested—male and female, children and adults, blind and seeing, normal and abnormal. The laws of motivation were found to apply to success in ESP performance much as they do in other delicate abilities. The effect of rewards, interest, novelty, competition, and the like was found to be favorable, and that of boredom, distraction, frustration to be adverse. The curves of performance—that is, the patterning of hits in the test—likewise seemed to be in line with those of other cognitive processes such as memory, sensory perception, and learning. ESP increasingly became identified with normal recognized processes.

But with regard to the physical world, the ESP process showed distinct peculiarities in every respect. In contrast to its relation to *sensory* perception, physical law seemed to play no part in the process of *extrasensory* perception. Whereas visual stimuli are less intense the farther away from the eye, ESP success did not fall off with the removal of the stimulus to greater distances. Other physical conditions, too, such as the angle at which the object was placed or the interposition of barriers, had no influence on the results. This lack of effect of physical barriers had been anticipated in earlier findings, but in the past decade a considerable accumulation of tests converging upon the question of the physical basis of ESP brought the matter to a decision. ESP was found to be unaffected by the spatial conditions imposed.

The experimentation on the question of ESP of the future, or

precognition, first appeared in the decade under review. During that time this research went through a series of ups and downs, but with a degree of progress which by the close of the decade may be said to have fairly well established that most dubious of hypotheses—dubious, that is, from the *a priori* point of view of the general science of yesterday.

Another "first" in parapsychology history belongs in the last ten years. That is the research in PK or the psychokinetic effect. The PK research is more closely related to the ESP investigations than at first appears. It was no mere coincidence that the PK experiments were undertaken in 1934 at the time when the first flush of victory over the problems of the relationship between ESP and the space-time system had been felt. Now it seems increasingly clear that we may speak, not of distinct ESP *and* PK processes, but of the "ESP-PK-process complex."

* * * * *

Even if we take only these more salient achievements of the decade and leave out the long list of worthy accomplishments not mentioned here, we have enough to permit us to raise the questions which are the main issue of this editorial: To what is this ten-year record of parapsychology research primarily due? What can we see behind the developments reviewed that will most help us in the program of research ahead of us?

We submit a suggestion: That the greatly increased research activity of the past decade is primarily attributable, directly or indirectly, to a *method*. Naturally, no single word can adequately represent all that is involved; but we think that more than any other single factor, the introduction of an easy method of adequate testing for ESP capacity was responsible for the promptness with which the academic and lay world took up the subject of extrasensory perception. There were several favorable features involved in the method. The ESP tests could be easily taken by almost anyone, and they were fun; they were much like a game. Again, almost any intelligent and careful person could administer them. They required no search for rare and special individuals. They were easily varied so as to throw light on the nature of the phenomenon. The test bristled with challenging possibilities for amateur or pro-

fessional exploration, and it supplied a procedure which nearly everyone could use and understand.

Students familiar with the history of parapsychology will recognize how little real innovation was involved in the new test procedures over those used in earlier researches. But the little there was changed work into play, a task into a game, and the technicality of evaluation into simple scoring.

The interpretation we are giving seems much more plausible if we set it off against the background of the history of the other sciences. Outstanding developments and periods of expansion in the investigation of many fields have followed upon the introduction of a new method, a method which has often centered in a newly invented instrument. We know what a leap forward astronomy took when Galileo turned the telescope heavenward and what new worlds of biological research followed the introduction of the microscope. Less tangible instruments such as calculus and statistical method have proved equally influential in the introduction of new eras in science.

* * * * *

What concerns us most in this discussion is the significance which this point regarding the role of method may hold for the future. In looking back, we see that the ESP test method not only carried us forward more rapidly toward the solution of the problems attacked at the outset, but it also brought us triumphantly up to the investigation of problems which we had not contemplated at the beginning as being readily subject to experimental solution. For example, the success with the ESP tests introduced, in a quite natural and logical sequence, the old issue of precognition which, in spite of its age, had never been brought to systematic experimental test before 1933. It was not that people did not know how to test for precognition; our point is that the introduction of a fortunate method for ESP testing started a development of research activity which took precognition in its methodological stride. The same, to a modified extent, is true of the PK effect. Under one heading or another (telekinesis, for example), this problem of paranormal psychophysical effects had often been dealt with. Never before, however, had it been dealt with under such control and with such exhaustive inquiry as in the PK researches based on dice throwing

that have gone on during the last ten years. The fact that the PK tests developed as a close adjunct of the ESP tests *gave them the same methodological features.*

Naturally we hope to apply this happy device of inventing a strategic *method* in future research in parapsychology. There are many great problems before us today—more, as a matter of fact, than we were conscious of ten years ago. Of all these numerous problems, that which, entirely on the basis of its scientific and social significance, looms largest and most crucial to most students is the question of whether personality, for all its demonstrated *extraphysical* characteristics, is also in certain respects *extrabiological;* that is, possessed of functions that are not completely dependent upon the living organism. Whether we call it the immortality problem, the survival hypothesis, or something else, it reduces to the question of *whether personality is entirely limited to life.* We are, of course, aware that men have been concerned for thousands of years with the question of the relation between life and mind, but the various efforts made to explain this relation are admittedly inconclusive so far as the critical student of science is concerned.

The question then is this: Are we in this case again merely awaiting a new and more penetrative method? Has it been lack of an adequate procedure or approach that has left the issue a necessarily suspended one? It may help us very much if we answer this question of method before we attempt to plan our inquiries for the years that lie ahead.

* * * * *

But new insights and approaches are not likely to come while a field is under strain and controversy. At the outset of the investigation of any highly controversial question the explorer is thrown very much on the defensive—so much so that he concentrates almost wholly on developing methods that will safeguard him against error. He is likely to be mentally swamped by the critical demands of his undertaking. He hardly—at least, rarely—dares think freely on the constructive exploratory side of his task. But it is on this more positive venturesome aspect that the need for free pioneering is greatest. *We must be free to think, without fear or inhibitions, of the means by which we may get at the truth we are seeking.* We

have all the more need for independence from pressure groups, either friendly or unfriendly, as well as from all factors of time and of economic and professional insecurity.

Freedom and security and an ideal research situation will not of themselves produce the new insight and procedure needed to close the gap of knowledge on the relation between life and mind; but we can assert with confidence that this result will be much more likely to occur with them, and it can hardly be expected without them. No true ESP explorer worthy of the name, however, will sit idly by awaiting the coming of an ideal time and place for the job he is eager to do. He is going to go on in the hope that, with broadly conceived interest and cooperative effort, the conditions will be laid down in time to make his efforts yield their maximum fruit. Then the decade that has passed will be only a prelude to a truly significant understanding of human personality and its place in the world of reality.

THE PK EFFECT: MECHANICAL THROWING OF THREE DICE

By EDMOND P. GIBSON, LOTTIE H. GIBSON, and J. B. RHINE

ABSTRACT: In the PK tests which are described in this paper, three dice were thrown mechanically for 35,784 throws, with approximately equal numbers of throws for each face of the die. Results were significantly above the scoring expected to occur by chance. The number of throws of the different faces was equalized in order to prevent any imperfections in the dice from producing the high scoring. Thus the PK hypothesis is again confirmed.

The results were analyzed for position effects; that is, for the influence which the position of the trial on the record page has on the scoring rate. It was found that scoring declined with progress through the run. It also declined horizontally on the record page. There was a very marked decline from the upper left-hand quarter of the record page to the lower right. The same types of declines occurred in the smaller units of the record page which are called sets, each consisting of three runs. Here the decline was more marked. There was also a chronological decline from the first and second to the third temporal subdivision of the experiment. These distribution differences represent in themselves an independent study confirmatory not only of the PK effect but also of the position effects of which they are typical examples.—Ed.

IN OUR previous report (1) we described experiments in which the PK phenomenon was demonstrated by the throwing of six dice at a time. We felt justified in the conclusion that no alternative hypothesis was applicable as an explanation of these results, reinforced as the conclusions were by the analyses for position effects (3). At the close of the earlier work we became interested in discovering what effect, if any, the mechanical handling of the dice would have upon the PK process; the research presented here deals with the experiment which was carried out with that objective. The principal modification of method, then, was the substitution of an electrically driven rotating cage for the cup used to throw the dice in the earlier experiment. The use of three dice at a throw instead of six represented a minor but somewhat unique change. This was the first time this number per throw had been used for the PK work.

It represents a compromise between the need to keep the number of dice within the limits of what can be read easily as the machine rotates and the desire to use as many dice as possible at a throw, a practice which the interests of both subject and experimenter seem to warrant in order to speed things up.

The present report falls naturally into three sections: the first describes the experiment and its results as evaluated by the general procedures in use; the second section presents the results of an analysis for position effects; and the third provides a discussion of the results of both the experiment and the analysis.

I. The Experiment

Procedure

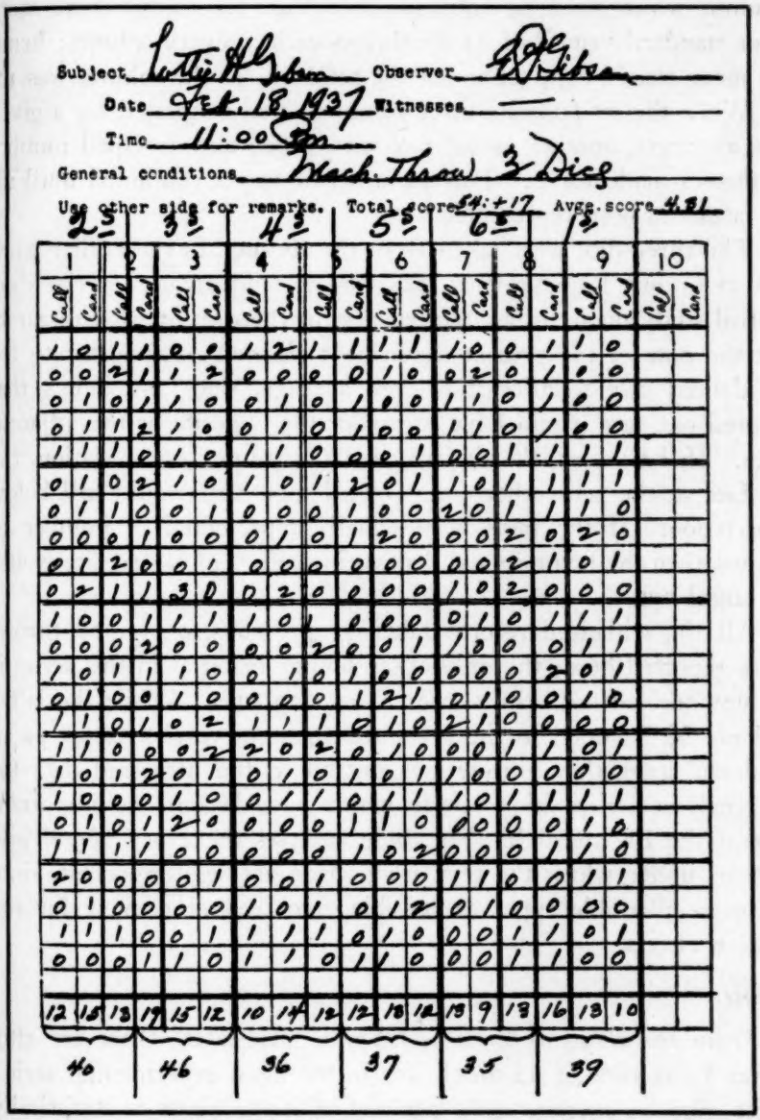
The entire experimental work was done in the early spring months of 1937, from February to May. L.H.G., wife of E.P.G., was the main subject, and E.P.G. himself was a subject for a large part of the series. A group of 13 other subjects took part in another section of the tests. Two of these 13 minor subjects were the children of E.P.G., aged eight and ten years; the rest were adult friends in the community.

E.P.G. was in actual charge of the experiments, and the work was conducted at his home in Grand Rapids, Michigan, in leisure hours from full-time professional duties. All witnessed work of this report was done with E.P.G. as observer and recorder except when he himself was the subject. In that case, L.H.G. recorded for him.

The procedure was simple. A rectangular, electrically driven cage was used to throw the dice. Two sides of the cage were of wire mesh and the other two sides and the ends were of wood. The cage is shown in Fig. 1. Three white, half-inch plastic dice of common variety were in the cage. The subject was asked to try to influence the dice by direct mental activity ("will power") so as to cause them to fall with a designated face turned up. This target face was chosen by the subject, specified in advance for a given set of three columns, and so indicated on the record sheet. When the subject was ready, the switch was turned on, and the cage started rotating. It was controlled by rheostat and was run at a rate to suit the subject, approximately one half-revolution every four-to-six seconds.

FIG. 1.—Motor-driven cage used for throwing dice.

At each half-revolution, when the dice dropped to the lower end of the cage, the observer called aloud the number of hits, and recorded it. The range of possible scores for each throw was, of course, from zero to three. The subject had full vision of the cage and


Subject Betty H. Gibson Observer R. G. Gibson
Date Feb. 18, 1937 Witnesses —
Time 11:00 pm
General conditions Mech. Throw 3 Dice
Use other side for remarks. Total score 54: +17 Avgs. score 4.91

1		2		3		4		5		6		7		8		9		10	
Call	Card	Call	Card	Call	Card	Call	Card	Call	Card	Call	Card	Call	Card	Call	Card	Call	Card	Call	Card
1	0	1	1	0	0	1	2	1	1	1	1	1	0	0	1	1	0		
0	0	2	1	1	2	0	0	0	0	1	0	0	2	0	1	0	0		
0	1	0	1	1	0	0	1	0	1	0	0	1	1	0	1	0	0		
0	1	1	2	1	0	0	1	0	1	0	0	1	1	0	1	0	0		
1	1	1	0	1	0	1	1	0	0	0	1	0	0	1	1	1	1		
0	1	0	2	1	0	1	0	1	2	0	1	1	0	1	1	1	1		
0	1	1	0	0	1	0	0	0	1	0	0	2	0	1	1	1	0		
0	0	0	1	0	0	0	1	0	1	2	0	0	0	2	0	2	0		
0	1	2	0	0	1	0	0	0	0	0	0	0	0	2	1	0	1		
0	2	1	1	3	0	0	2	0	0	0	0	1	0	2	0	0	0		
1	0	0	1	1	1	1	0	1	0	0	0	0	1	0	1	1	1		
0	0	0	2	0	0	0	0	2	0	0	1	1	0	0	0	1	1		
1	0	1	1	1	0	0	1	0	1	0	0	0	0	0	2	0	1		
0	1	0	0	1	0	0	0	0	1	2	1	0	1	0	0	1	0		
1	1	0	1	0	2	1	1	1	0	1	0	2	1	0	0	0	0		
1	1	0	0	0	2	2	0	2	0	1	0	0	0	1	1	0	1		
1	0	1	2	0	1	0	1	0	1	0	1	1	0	0	0	0	0		
1	0	0	0	1	0	0	1	1	0	1	1	1	0	1	1	1	1		
0	1	0	1	2	0	0	0	0	1	1	0	0	0	0	0	1	0		
1	1	1	1	0	0	0	0	0	1	0	2	0	0	0	1	1	0		
2	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	1	1		
0	1	0	0	0	0	1	0	1	0	1	2	0	1	0	0	0	0		
1	1	1	0	0	1	1	1	1	0	1	0	0	0	1	1	0	1		
0	0	0	1	1	0	1	1	0	1	1	0	0	0	1	1	0	0		
12	15	13	17	15	12	10	14	12	12	10	12	13	9	13	16	13	10		
40			46			36			37			35			39				

FIG. 2.—Sample record sheet showing the structure of a typical page and set. One hit for every two entries would be expected from chance.

could check on the accuracy of the readings. If two dice came to rest on top of each other, such falls were rejected and the trial was repeated. This happened infrequently, however.

The test results were recorded for the throws of three dice in columns which held 24 entries. (See Fig. 2.) Thus there were three standard "runs" of 24 die-throws each in every column; hence the mean chance expectation for the total hits of the column was 12.

When the set (usually three columns) was completed for a given face as target, another target face was chosen and an equal number of throws made for it. This variation of target continued until all six faces had been represented.

The three dice were taken from the six used in the earlier work (2) at Grand Rapids; thus they had the advantage of the dice-control study made to test the trueness of the dice for that research. But the plan of the present experiment allowed no opportunity for unbalanced dice to affect the score average, since any defect that favored one face would have to hinder the opposite or the adjacent faces. If, let us say, the six-face gave a large positive deviation, the one-face or the four other faces (or all together) would fall below expectation. If the faces were equally represented in number of throws, then the favoring and disfavoring effects of faulty dice would all cancel out.

All the witnessed records taken—1,491 runs, or 11,928 throws —are reported here; the only other similar material in this series is the unwitnessed test results, comprising 423 runs. While the latter will not be included in the main statement of results except as an incidental item, they are quite as suitable as the witnessed data for the analyses for position effects which have become an important part of the PK research. For these analyses the experimental precautions under which the tests were made are less important, since the more plausible types of possible error cannot account for the position effects.

Results

From the Point of View of Chance. As stated, there are altogether 1,491 runs of 24 die-throws in the main experimental series. These give an average score per run of 4.13, which is statistically significant. The total deviation is +191, and the standard deviation,

±70.46. The CR is 2.71, a significant value, with a probability of .0035. It is made even more reassuring by the fact that so large a series is involved. (There is a total of 35,784 single die-throws.)

If the 423 runs of the unwitnessed work, mostly by E.P.G. himself, are pooled with these results, the effect is not very different, for the CR then is 2.74. Taken by itself the unwitnessed work is insignificant. It may be recalled that in our previous report (1) *the same effect resulted when E.P.G. worked alone:* he dropped to a score average close to expectation, although he scored better with L.H.G. recording.

Table 1

TOTAL PK TESTS WITH THE THREE-DICE, CAGE PROCEDURE

Conditions	Runs	Average	Total Deviation	SD	CR
Witnessed tests.....	1,491	4.13	+191	±70.46	2.71
Unwitnessed tests.....	423	4.07	+ 28	±37.54	.75
Total.....	1,914	4.11	+219	±79.84	2.74

Since the total results, with or without the unwitnessed section, are significantly different from chance expectation, the chance hypothesis has to be rejected as inapplicable to these data. No improper sampling or selection is involved. The report includes all the test results, and the optional stopping hypothesis (that the exercise of choice in terminating the experimental series could have produced this set of results in a chance series) is ruled out on the ground that all of E.P.G.'s PK experiments (1, 2) from the beginning have been extrachance.

Consideration of the Hypothesis of Faulty Dice. The experiment was planned to control fully the question of adequate trueness of the dice. In fact, the extrachance results are *ipso facto* evidence that lies beyond the reach of the hypothesis of faulty dice. There was a slight inequality in the numbers of throws for the various faces of the die, as may be seen in the totals of Table 2. The number of standard runs, the average score per run for each of the six faces, and the total of all are also given.

Table 2

AVERAGE RUN SCORES FOR ALL TARGET FACES

Subject	One-Face		Two-Face		Three-Face		Four-Face		Five-Face		Six-Face		Total	
	Runs	Average	Runs	Average	Runs	Average	Runs	Average	Runs	Average	Runs	Average	Runs	Average
L.H.G.....	105	3.92	105	3.91	105	4.51	105	4.20	105	4.34	105	4.42	630	4.22
E.P.G.....	72	4.15	72	4.43	72	3.83	72	3.79	72	4.06	72	4.08	432	4.06
Miscellaneous..	66	3.91	66	3.76	66	3.94	57	3.88	84	4.19	90	4.50	429	4.07
Total.....	243	3.99	243	4.02	243	4.16	234	4.00	261	4.21	267	4.36	1,491	4.13

The table shows that the small inequality in the numbers of runs occurred in the series with miscellaneous subjects; it was due to interruption of the session in six instances before the rounds were entirely made. However, the exclusion of these fragments or incomplete rounds, which total 87 runs (see Table 3) and have a positive deviation of 20 ± 17.02, does not reduce the CR of the remaining 1,404 runs below the criterion of significance. The CR of the complete rounds is 2.50. Accordingly, the results meet both the statistical and the experimental requirements for coping with the hypothesis of faulty dice.

Table 3

FULL AND INCOMPLETE ROUNDS

Condition	Runs	Total Deviation	SD	CR
Complete rounds:.....	1,404	+171	±68.38	2.50
Incomplete rounds.....	87	+ 20	±17.02	1.18
Total:.....	1,491	+191	±70.46	2.71

It is of interest to note, in supplementing the case against dice defects as an explanation, that in Table 2 the averages for the individual target faces (like the extrachance deviation of the entire series) definitely do not fit in with the view that the dice were at fault; for while the average of the whole research was highest for the six-face, *L.H.G., the main subject, did best on the three-face.* Curiously, then, the other subjects went *below expectation* on the

three-face. Moreover, E.P.G. favored the two-face above all; and here, too, the other subjects went below chance. While the miscellaneous subjects scored highest on the six-face, it was only second best for L.H.G. and third best for E.P.G. All this reduces to the suggestion that "preferences," and other motivational factors, not physical characteristics, are the main determinants, though we cannot say they are the only ones.

Also, while we are referring to Table 2, it is worth while to note that only one face has a negative deviation in the totals, the one-face; and it is negative only to the extent of .01, having 3.99 for an average score. It is manifestly impossible for dice to be "crooked" to the extent of favoring one face—say, the five- or six-face in this work—without reducing one or more of the other faces below mean chance expectation; that is, unless some additional factor like PK comes in.

II. Analysis for Position Effects

The second phase of the work reported in this paper deals with the analyses conducted in the Parapsychology Laboratory at Duke University in 1942 and 1943 to discover whether the position of the trial on the record page or other units of the experiment affected the rate of scoring. Studies of the distribution of hits were made on various subdivisions of the experimental records. The results of these analyses will be presented below under the headings appropriate to the types of distribution studied. This work of analysis still continues and will be the subject of further reports.

Horizontal Distribution

The first study of the distribution of hits was based on the record page as a unit. (See Fig. 2 for illustration of the record page.) We shall report first the results obtained by comparing the left half of the record page with the right half in terms of total deviation and average score. The results of this comparison are given below in Table 4. It will be seen that the left half of the page gave a total positive deviation of 161, and the right half, one of 58. The averages are 4.16 and 4.06 hits per run, respectively. From the details given in the table it will be seen that this horizontal decline is contributed by the work of L.H.G. and the miscellaneous subjects. E.P.G., both in the witnessed and unwitnessed series, reversed the

Table 4

HORIZONTAL DISTRIBUTION OF HITS ON LEFT AND RIGHT HALVES OF PAGE

Conditions	Subject	Left Half			Right Half		
		Runs	Deviation	Average Score	Runs	Deviation	Average Score
Witnessed.....	L.H.G.	315	+109	4.34	315	+29	4.09
	E.P.G.	216	− 1	3.99	216	+26	4.12
	Misc.	258	+ 47	4.18	171	−19	3.78
Unwitnessed::::	E.P.G.	162	− 6	3.96	162	+12	4.07
	Misc.	63	+ 12	4.19	36	+10	4.28
Total.....	.....	1,014	+161	4.16	900	+58	4.06

direction of distribution, as he did in the earlier experiment (1).

The distribution of target faces for the sets of runs across the record page is sufficiently varied that the score decline from left to right could not be explained in terms of predominance of favored target faces on the left.

For that matter, a still more striking decline from left to right is shown within the three-run sets themselves. This horizontal decline in the set is presented separately in Table 5. The pooled total deviations for the three runs of the set are +153, +89, and −23, respectively, the difference between the first and third runs being statistically significant (CR=2.65). The main point to which we invite attention at the moment is the fact that here there is no difference in target face in the three runs of the set. Obviously, in

Table 5

DECLINE OF SCORING IN THE RUNS OF THE SET

Subject	Witnessed								
	First Run			Second Run			Third Run		
	Runs	Deviation	Aver. Score	Runs	Deviation	Aver. Score	Runs	Deviation	Aver. Score
L.H.G....	216	+ 88	4.41	216	+40	4.19	198	+10	4.05
E.P.G.::::	144	+ 47	4.33	144	−10	3.93	144	−12	3.92
Misc.....	150	+ 9	4.06	150	+17	4.11	129	+ 2	4.02
Total.....	510	+144	4.28	510	+47	4.09	471	0	4.00

$CR_{d(1-3)} = 2.40$

TABLE 5 (continued)

UNWITNESSED

Subject	First Run			Second Run			Third Run		
	Runs	Deviation	Aver. Score	Runs	Deviation	Aver. Score	Runs	Deviation	Aver. Score
E.P.G.	108	+ 12	4.11	108	+27	4.25	108	−33	3.69
Misc.	33	− 3	3.91	33	+15	4.45	33	+10	4.30
Total	141	+ 9	4.06	141	+42	4.30	141	−23	3.84
BOTH POOLED									
Grand Total	651	+153	4.24	651	+89	4.14	612	−23	3.96

$CR_{d(1-3)} = 2.65$

comparing the decline in the set with that for the page, it is clear that the set is more of a functional or structural unit for the subject than is the page as a whole, insofar as this horizontal decline is a measure. E.P.G.'s distribution within the set was more or less in conformity with L.H.G.'s, instead of in the opposite direction as was the case with the distribution on the page.

Vertical Distribution

The distribution of hits in the column, or the vertical distribution of hits on the page, is equally as interesting as the horizontal distribution. The results of the vertical distribution study are shown below in Table 6. Here the same data as were represented in the two

Table 6

COMPARISON OF SCORING ON UPPER AND LOWER HALF-PAGES: COLUMN DECLINE

Conditions	Subject	Upper Half			Lower Half		
		Runs	Deviation	Average Score	Runs	Deviation	Average Score
Witnessed	L.H.G.	315	+120	4.38	315	+18	4.06
	E.P.G.	216	+ 25	4.12	216	0	4.00
	Misc.	214.5	+ 15	4.07	214.5	+13	4.06
Unwitnessed	E.P.G.	162	0	4.00	162	+ 6	4.04
	Misc.	49.5	+ 16	4.32	49.5	+ 6	4.12
Total	.....	957	+176	4.18	957	+43	4.04

tables preceding are redistributed so as to show the comparison of the scoring in the upper and lower halves of the page. Incidentally, the vertical distributions for the page and for the set are the same, since the set runs the full vertical length of the page.

In the table just given, the upper half shows a deviation of +176, and the lower half, of +43. This is a four-to-one relation. It may be interesting to note how this vertical decline is distributed in terms of quarter-columns instead of half-columns. With the column divided into four segments, the total deviation is distributed as follows, going down the column: +97, +79, +55, −12. It is noteworthy that the largest drop is from the third to the fourth segment.

Quarter Distribution (QD) of the Page and Set

In view of the fact that both a horizontal and a vertical decline were shown in the distribution study of the record page, we knew that we should find a marked decline from the upper left quarter to the lower right quarter of the page. When each record page was divided into four parts and the distribution of hits in each was determined, the pooled records of this series gave the following deviations for the four quarters of the page.

QD of the Page

507 +112 4.221 1	450 +64 4.142 3
507 +49 4.097 2	450 − 6 3.987 4

$CR_{d(1-4)} = 2.12$

This is what has come to be regarded as a typical quarter distribution; that is, the upper left corner is highest and the lower right is lowest of the four. The customary method of evaluation is to find the difference between the first and fourth quarters (upper left and lower right). In this case there is a marked diagonal decline in these two quarters with a CR of the difference of 2.12. This is

not significant, but it is highly suggestive of an extrachance factor operating in the distribution of the hits.

In order to obtain the quarter distribution of the set, it was necessary to omit the middle column, leaving a block of data for the set analysis which consisted of two columns of 24 entries each. The hits of the upper half of the first column made up the first quarter; the lower half of the first column, the second quarter; the upper half of the second column, the third quarter; and the lower half of the second column, the fourth quarter. The results obtained from this quarter distribution of the set are somewhat more striking than those of the QD of the page. For example, the average scores of

QD of the Set

328.5 +113 4.344 1	328.5 − 6 3.982 3
328.5 +62 4.189 2	328.5 −25 3.924 4

$CR_{d(1-4)} = 3.15$

the first and fourth quarters have a greater difference than those of the QD of the page. The CR of the 1 − 4 difference of the set is 3.15, which is quite significant. The QD of the page and that of the set, while not entirely independent, are largely so. There is a substantial degree of confirmation, then, in the QD of the set; and in any case, the QD either of the page or of the set may be taken as offering support of the strongest character for the general establishment of the PK hypothesis.

Chronological Distribution

Under the heading of chronological distribution we report the distribution of hits over the series as a whole, giving separate treatment to each subject's data. The work of a given subject was divided into three temporal subdivisions, all as nearly uniform as possible. One restriction was that no given experimental session

should be subdivided. The session on the borderline was to go to that subdivision which, with the session included, would show the smallest departure from equal distribution in number of runs. With the data for each subject thus distributed, the subdivisions were pooled for the group as a whole into three general chronological units. These are shown in Table 7 below. Obviously the subdivisions were quite unequal, but there was no more consistent way of arriving at the objective of the inquiry—namely, to discover whether the subjects who participated in these tests showed any appreciable change in scoring rate as they participated in the experiment.

Table 7

CHRONOLOGICAL DISTRIBUTION

Temporal Subdivision	Runs	Total Deviation	Average Score
First	960	+146	4.152
Second	522	+ 87	4.167
Third	432	− 14	3.968

From the table it will be seen that there is a decline in score average from the second to the third subdivision, but not from the first to the second. The differences are not significant.

III. DISCUSSION

The General Case for the PK Effect

The approximate equalization of the number of throws for the target faces of the die offers a satisfactory control on the hypothesis of faulty dice. The use of the dice-throwing machine eliminates the hypothesis of skilled throwing. Other issues that might be raised, such as recording errors, loss of records, and optional stopping, are successfully met by the significant CR of the 1 − 4 difference obtained from the QD of the set. Other evidence might be adduced to bear upon such counterhypotheses, but most readers, especially those familiar with the cumulative case for the PK hypothesis, would find it superfluous.

Position Effects and Their Nature

It is obvious that the outstanding feature of this investigation is the existence of a typical QD of the page and a very significant QD

of the set. Continued analysis of position effects in this material is yielding more evidence of the same trend of distribution in smaller subdivisions of the record material. The impression we get is that the working of the PK process over the series as a whole is very lawfully patterned by the order or "structure" of the test situation; for example, the physical make-up of the record sheet.

This structuring effect seems to be an inhibitory one which depresses the scores as the subject goes along. In the QD of the page, the first quarter gives the highest scores; in the QD of the set, the deviation of the average score for the first quarter is half again as large as that of the first quarter of the page (.34 as compared to .22). Similarly, for L.H.G.'s work alone (the comparable figures are not yet available for the other subjects on the point in question) we find that if we take the upper left *eighth* of the set instead of the quarter, we get a still higher rate of scoring, very much the best eighth of the set. The trials in that portion constitute 54 runs with a total positive deviation of 56, giving an average score of 5.04, or 1.04 above expectation. This is nearly five times the deviation of the score average which L.H.G. shows in her work as a whole. In those first six trials in the first column of the set, L.H.G. exerts her very best effect on the dice. The rest of the set shows little evidence of PK in comparison. The first eighth taken alone is highly significant; the other seven-eighths together are insignificant. The last (lower right) eighth of the set for L.H.G.'s work gives a *negative* deviation of 8. This is the only negative eighth in the set.

But to say that the effect of the set and page structure is inhibitory and that the real uninhibited PK function operates only very briefly at the beginning of the set is not to explain how this inhibitory influence is exerted. The effect of the terminal *versus* middle positions upon success in scoring in ESP runs has already been discussed by one of us (J.B.R.) in connection with the phenomenon of terminal salience.[1] It seems reasonable to suggest speculatively a similar line of thought concerning the PK process: that PK, like ESP, operates with great difficulty because of the fact that normally the much more widely used and more relied upon processes of the mind usurp the control and direction of cognitive activity. The

[1] Dr. Schmeidler's paper in this number relates this effect to the primacy and finality effects in learning and memory studies.

farther the subject proceeds into a column, a set, or a page, the more the established habits and associations which the PK process must evade mount up, get in the way, and entangle the easily inhibited PK effects. If this be true, success in improving the rate of scoring can best be obtained by devising test situations which offer greater spontaneity and less organization or structure.

Special Features

There is a point of some novelty in the use of the three dice at a throw. No good basis is available for a close comparison of these results with other numbers of dice per throw because, of course, there are differences in condition other than the number of dice used. But L.H.G. herself did somewhat better with six dice per throw in the earlier work reported. Also her results were better with six dice than with two in the first PK work in which she was a subject (2). Was the drop in the present series attributable to the use of three dice or to the use of the machine? We may point to the fact that wherever the machine or semimechanical devices have been used in the past, the results have been better than with comparable series in which the dice were cup-thrown. However, we cannot do more than to suggest that probably L.H.G. would have done better had she been able to use six dice in the machine; but this was not a practicable method.

It has already been mentioned that E.P.G. showed a horizontal incline in scoring rate on the page as a whole. He did this not only in the witnessed but also in the unwitnessed work, and he showed a similar effect in his test results previously reported (1). This atypical case suggests that the position effects are, to some extent at least, individual responses due to personal traits. If so, the fact should be discoverable through appropriately planned research.

Summary

A lengthy series of tests was conducted with the throwing of three dice in a mechanically rotated cage. The results were significant, and the hypothesis of chance was dismissed. The use of the cage and the throwing of approximately equal numbers of runs for all the faces of the die offer excellent controls on the question of whether the manner of throwing or the physical imperfections of the

dice can account for the results obtained. All of the common questions as to how the results may have been produced are considered, and none is found to be satisfactory except the PK hypothesis.

Not only is the PK hypothesis effectively confirmed by this experimental series, but also further evidence of its lawful operation is manifested in the typical quarter-distribution effects that are shown. The QD of the page and the QD of the set are both typical, the latter being significant. These results in themselves offer a further and a distinctly superior confirmation of the PK hypothesis.

In the interpretation of the quarter distribution, it is suggested that the effect of the structure of the page and of the set is an inhibitory one, brought about by the formation of interfering associations of habit that the subject undergoes as he progresses in the set or the page.

References

1. Gibson, Edmond P., Gibson, L. H., and Rhine, J. B. A large series of PK tests. *J. Parapsychol.*, 1943, 7, 228-38.
2. Gibson, Edmond P., and Rhine, J. B. The PK effect: III. Some introductory series. *J. Parapsychol.*, 1943, 7, 118-34.
3. Humphrey, Betty M., and Rhine, J. B. Position effects in the large Gibson Series. *J. Parapsychol.*, 1943, 7, 238-51.

POSITION EFFECTS AS PSYCHOLOGICAL PHENOMENA

By GERTRUDE RAFFEL SCHMEIDLER[1]
Harvard Psychological Clinic

ABSTRACT: Dr. Schmeidler tells here what she found when she examined the records of her earlier ESP experiments for position effects; that is, for the effect of the position of the trial in the run or other unit upon the rate of success. There were two groups of subjects, the A group, consisting of subjects who were open-minded on the question of the ESP hypothesis, and the B group, consisting of subjects who regarded ESP as impossible. Dr. Schmeidler found position effects in evidence in four different structural units of the experiment: the five-trial segment, the 25-trial run, the consecutive runs which made up the 10- to 50-run session, and the three series of the experiment as a whole. The position effects are described separately for the two groups of subjects and for the results of both groups pooled.

The original ESP tests had been conducted to discover the effect of attitude upon success in performance. The open-minded subjects had done significantly better than those who denied the possibility of ESP. The main findings in this study of position effects represent a standing-out of high scoring at the beginning of the record unit, and at the end when the subject knew when the end was coming. Dr. Schmeidler points to the similarity between these position effects and the primacy and finality effects encountered in the study of learning and memory.—Ed.

TO A PSYCHOLOGIST who has only recently begun experimentation with extrasensory phenomena, it is natural to differentiate between findings that are peculiar to this field and those that seem to follow familiar laws. This paper deals with an example of the latter, analyzing position effects in terms of certain of the psychological concepts of learning and motivation. It is undertaken in the hope that analogies between the data of psychology and of parapsychology may be fruitful, just as in academic psychology the analogy between the laws of perception and of memory has been productive of better understanding of both.

[1] This research was made possible by Dr. Gardner Murphy, through moneys available to him from the Hodgson Fund of Harvard University.

This paper will describe as briefly as possible the relevant psychological explanations of position effects as they occur in such diverse activities as learning nonsense syllables and remembering childhood experiences, then present the position effects found in an experiment in clairvoyance, and finally attempt to formulate a general law which will include these data of the clairvoyance experiment within the larger framework of psychological events.

A Psychological Description of Position Effects

When the psychology of learning was first being explored, textbook descriptions were composed of material drawn from a relatively scant number of experiments, and from common observation. At that time it was taken for granted that material which was learned first could be remembered especially well and a law of novelty was suggested to describe the fact. A corresponding advantage was shown by material learned last, and this was described by a law of recency.

But as more research was done and thinking became more sophisticated, the picture changed. The law of newness faded into oblivion, to be replaced for a time by a law of vividness. Novelty would be a special case under vividness, for first experiences are usually more interesting and more impressive than are repetitions of those experiences.

Then even the law of vividness grew unfashionable. Laboratory psychologists prefer material which can be quantified; and more and more frequently, in setting up a learning experiment, they used tasks composed of equivalent and interchangeable units, such as right-angle turns in a maze or three-letter nonsense syllables. With such material, vividness became a less important concept, and other more specialized accounts of intraserial position effects began to emerge, couched in terms of the excitatory or inhibitory processes aroused by each stimulus.

In the typical modern experiment in rote learning, the subject—whether rat or human—is introduced to the material to be learned and is allowed to become so familiar with it that it is no longer of great interest to him. He goes over his task again and again until he is presumably rather bored and is gratified each time he reaches the end. He learns the last units before any of the others. The

first units are sometimes more readily learned than the middle group, and sometimes not. When primacy and finality effects occur, they are commonly attributed to a greater initial interest on the subject's part or greater enthusiasm as the subject comes near his goal, to the possibility of rehearsal of the first unit or units during later learning, and to a complex pattern of proactive and retroactive interference effects among the individual units. The last factor is usually considered intrinsic in the learning process; all others are frequently regarded as irrelevant and, insofar as possible, are "controlled" so that they can be ruled out of the picture.

Thus a psychological description of position effects might be summarized in some such fashion as this: In a well-controlled learning experiment, there is typically a slight advantage of the first, and a great advantage of the last units of a task over the middle units. These primacy and finality effects are probably consequent on the nature of the learning process. But in earlier badly controlled studies of learning and memory, there were typically a strong primacy effect and a weaker finality effect. Although these results were usually pronounced, they tended to appear and disappear in an erratic way. They were due very largely to changes in the subject's attitude, for he was usually more interested when he began the task than at any other time; and he was most gratified when he completed it. These are valid psychological events; but they are found in so many nonlearning situations that they could probably best be described as phenomena of motivation.

Our discussion of position effects in the following pages will not be concerned with intraserial inhibition. We shall, however, endeavor to show that the varying interest and enthusiasm which produced such strong but changeable position patterns in the earlier experimentation in learning are operative in ESP; and we shall try to account for the salience patterns found so frequently in paranormal investigation in exactly the same terms as were used for those position effects.

Position Effects in a Clairvoyance Experiment

Only a brief account will be given here of the conditions under which the data were obtained, since they have been described in detail elsewhere (**14, 15**).

Within the general framework of an attempt to study unconscious motivation in ESP, subjects were asked at the beginning of the experiment about their attitude toward telepathy and clairvoyance. Those who professed complete disbelief were tested under routine laboratory conditions; all the others, whether they were convinced of the validity of parapsychology or merely showed slight interest in it, were tested under conditions that were somewhat more flexible and agreeable (but just as fully controlled). It was predicted that the former group would score at or below chance and that the latter group would score above chance. This prediction was fulfilled. Although at the conscious level all subjects were trying to make correct calls, those who neither believed that ESP occurred nor felt much interest in the problem scored at chance; and there was some indication that those who had identified themselves with the argument that the ESP hypothesis was false and who were eager to disprove it made scores which averaged below chance.

Table 1

DISTRIBUTION OF DEVIATION BY TRIALS IN THE SEGMENT AND SEGMENTS IN THE RUN

	Trials in Segment					Segments in Run				
	1	2	3	4	5	1	2	3	4	5
	Section A. Subjects Who Admitted Possibility of ESP									
Deviation	− 3	+41	+ 7	+39	+36	+40	+18	+29	+19	+14
	Section B. Subjects Who Denied Possibility of ESP									
Deviation	− 8	−36	− 8	−22	+40	+10	−26	−24	−24	+30
	Section C. All Subjects Pooled									
Deviation	−11	+ 5	− 1	+17	+76	+50	− 8	+ 5	− 5	+44

Position Effects within the Segment

The DT method was used. The subjects wrote their calls on record sheets where a column of 25 boxes was assigned for each run. The column was divided into groups of five segments. On some of the sheets there was a double line under each segment; on the

others the segments were separated from each other by a blank space of half an inch. The instructions made no mention of the division of the run into segments, nor was the subject advised to guess in a rhythm of blocks of five. It is therefore the more remarkable that a striking finality effect is present in the data of the more negative group and in the data of the experiment as a whole when the five positions within each segment are compared (Table 1). This finality effect is absent when the data of the open-minded subjects are considered separately.

Position Effects within the Run

The picture of run distribution of hits shows a primary effect for each of the groups of subjects and also a finality effect for the subjects who rejected the ESP hypothesis. (See Table 1 and Fig. 1.) Both effects are apparent when all the data are pooled.

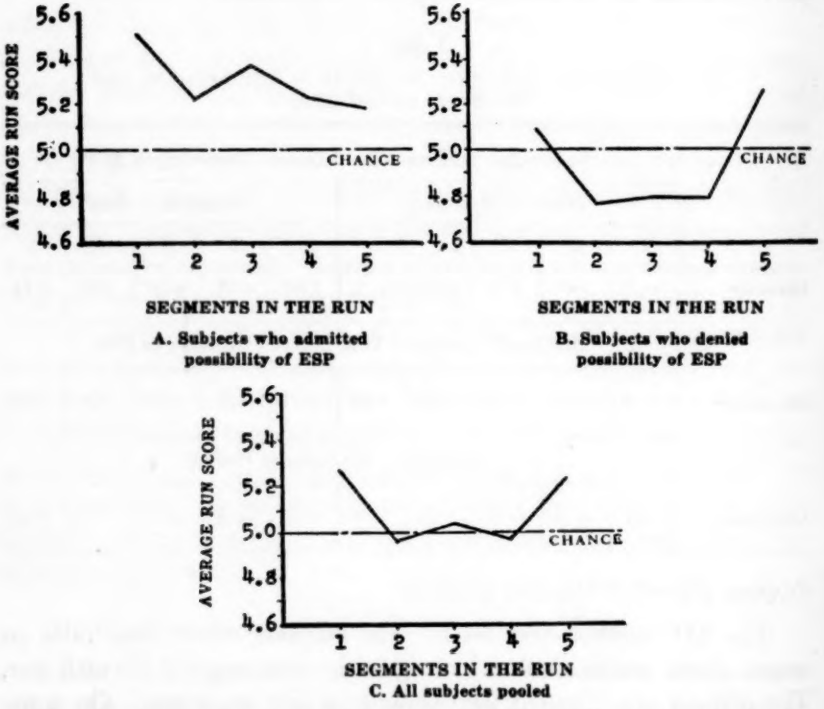


FIG. 1.—Average run scores of successive segments within the run for the two groups of subjects and for their pooled total.

It is interesting to compare the pooled results with the corresponding data of a precognition experiment with unselected subjects reported by Humphrey and Rhine (2). Since precognition is rejected by so many, even among those who admit the possibility of ESP, it is probably safe to assume that some of the subjects in the Humphrey and Rhine experiment felt the same skepticism as did our own group of unbelievers. We might therefore predict that if attitude toward the task is an important determinant of the score, there will be a similarity between the results in the two cases.

Such a similarity occurs, for in both experiments the average run scores are almost at chance, but the salience pattern is striking. (See Fig. 2.) The scores for the first and last segments are above

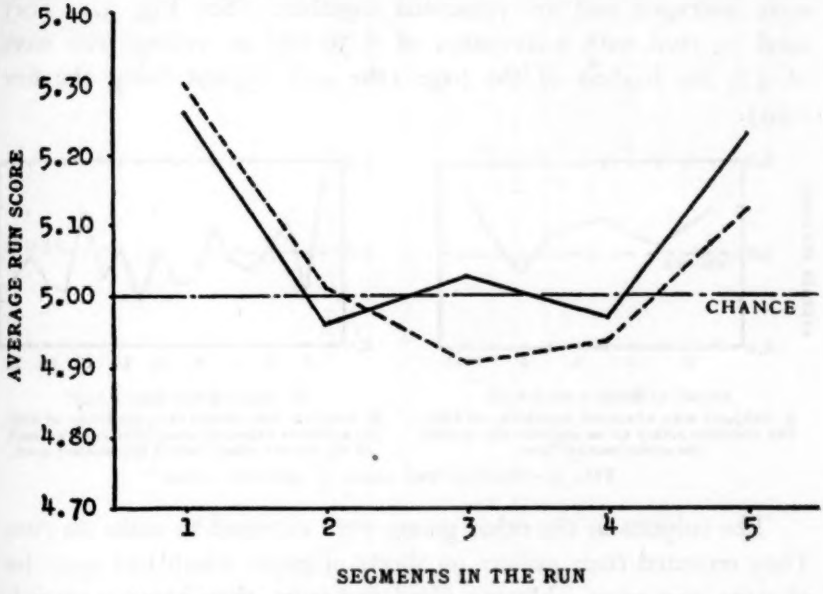


FIG. 2.—Comparison of position effects within the run in the Humphrey and Rhine precognition series (dotted line) and in the pooled results reported in the present paper (solid line).

chance; those for the middle segments are at chance or below. Perhaps in this U-shaped curve of position effects, with an average at chance, we have a characteristic scoring pattern for a group of subjects, some of whom have an inner conviction that their task is impossible.

Position Effects within the Session

One of the differences in procedure between the two groups used here must be explained more carefully since it has influenced the pattern of position effects showing the relation between success and time sequence in the test session. (See Table 2 and Fig. 3.)

The subjects who believed it possible to have paranormal sensitivity never made more than ten runs in a session. After each run we compared their guesses with the target and chatted about the experiment. They were allowed to stop when they pleased; thus the last run came in different positions for the various subjects. *The subject in this group always knew which run was to be the last.* To give an accurate picture of the finality effect, all the last runs were averaged and are presented together. (See Fig. 3.) They total 51 runs with a deviation of + 36 and an average run score of 5.7, the highest of the page (the next highest being the first runs).

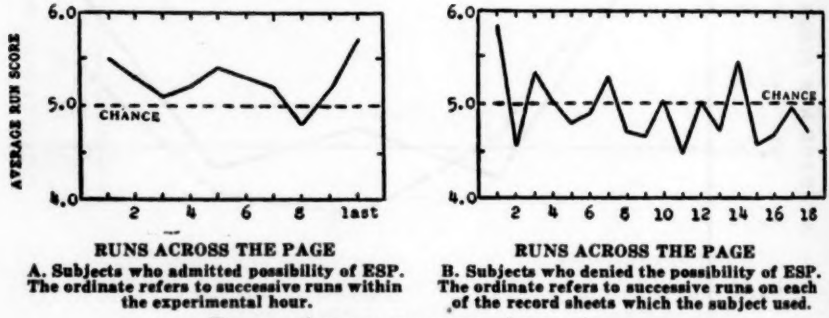


FIG. 3.—Average run scores of successive runs.

The subjects in the other group were expected to make 50 runs. They recorded their guesses on sheets of paper which had space for 18 runs on a page. Having filled one page, they began a second; after filling the second, they began a third; before they had finished the third page, the experiment was completed. Their comments showed that they were surprised when the end of the experiment came; either they had forgotten the original instructions, calling for 50 runs, or else had not made the calculations necessary to show which column would hold the last run. Data for the successive runs on a page are given in Table 2. The scores of the three pages of

Table 2

THE SCORES OF SUCCESSIVE RUNS WITHIN THE EXPERIMENTAL HOUR

Subjects who admitted the possibility of paranormal perception				Subjects who denied the possibility of paranormal perception						
Position of run within the hour	Number of runs in this position	Total dev.	Average run score	Position of run within the hour	Number of runs in this position	Total dev.	Average run score	Position of run on page	Total dev.	Average run score
				Page 1						
1	51	+ 26	5.5	1	12	+10	5.8	1	+30	5.8
2	49	+ 13	5.3	2	12	0	5.0	2	−15	4.6
3	47	+ 4	5.1	3	12	+13	6.1	3	+12	5.3
4	47	+ 10	5.2	4	12	− 9	4.2	4	+ 1	5.0
5	40	+ 17	5.4	5	12	+ 4	5.3	5	− 7	4.8
6	32	+ 11	5.3	6	12	− 5	4.6	6	− 4	4.9
7	28	+ 5	5.2	7	12	+12	6.0	7	+10	5.3
8	24	− 5	4.8	8	12	+ 1	5.1	8	−11	4.7
9	20	+ 3	5.2	9	12	− 2	4.8	9	−12	4.6
Last	51	+ 36	5.7	10	12	− 9	4.2	10	+ 1	5.0
				11	12	− 5	4.6	11	−17	4.5
Total....	389	+120	5.3	12	12	− 1	4.9	12	0	5.0
				13	12	− 3	4.8	13	− 9	4.7
				14	12	− 1	4.9	14	+14	5.4
				15	12	− 6	4.5	15	−11	4.5
				16	12	−11	4.1	16	− 8	4.7
				17	12	− 2	4.8	17	− 1	5.0
				18	12	− 2	4.8	18	− 7	4.7
				Page 2						
				19	12	+ 8	5.7			
				20	12	−10	4.2			
				21	12	− 4	4.7			
				22	12	+ 9	5.8			
				23	12	− 5	4.6			
				24	12	0	5.0			
				25	12	− 3	4.8			
				26	12	−11	4.1			
				27	12	+ 6	5.5			
				28	12	+ 5	5.4			
				29	12	− 9	4.2			
				30	12	+ 2	5.2			
				31	12	− 5	4.6			
				32	12	+10	5.8			
				33	12	− 5	4.6			
				34	12	+ 3	5.2			
				35	12	+ 1	5.1			
				36	12	− 5	4.6			
				Page 3						
				37	12	+12	6.0			
				38	11	− 5	4.5			
				39	11	+ 3	5.3			
				40	11	+ 1	5.1			
				41	10	− 6	4.4			
				42	10	+ 1	5.1			
				43	10	+ 1	5.1			
				44	10	− 1	4.9			
				45	10	−16	3.4			
				46	10	+ 5	5.5			
				47	10	− 3	4.7			
				48	10	− 1	4.9			
				49	9	− 1	4.9			
				50	8	+ 5	5.6			
				Total..	574	−34	4.9			

the less favorably disposed group are listed separately and are also pooled to show how the page structure itself may have influenced scoring. These pooled data are shown on the right. Figure 3 shows the curves of scoring for the pooled distribution on the page for the two groups of subjects. Both the curves of Figure 3 show the familiar pattern of decline, for the scores on the first run are high and the general tendency of the curves is downward. The finality effect is obvious for the subjects who knew which run was to be the last.

Series Decline

A fourth example of a position effect in the experiment appears from a comparison of the three series of which the experiment was composed. These three series consist of 329, 302, and 332 runs, respectively, as already described in previous reports (**14, 15**). None was significant taken alone, but the three together gave a significant CR of the difference between the *favorably* and the *unfavorably* disposed groups (P = .001).

Comparison of these three series for position effects shows that the subjects for whom scores above chance were predicted had a slightly higher average run score in the first series than in the others. (See Fig. 4.) The other group did not appreciably alter its scoring level from series to series. Thus there was a small decline effect for the experiment taken as a whole. Similar declines in scoring rate in the course of the experiment have been so common in the history of ESP research as to represent the predominant trend.

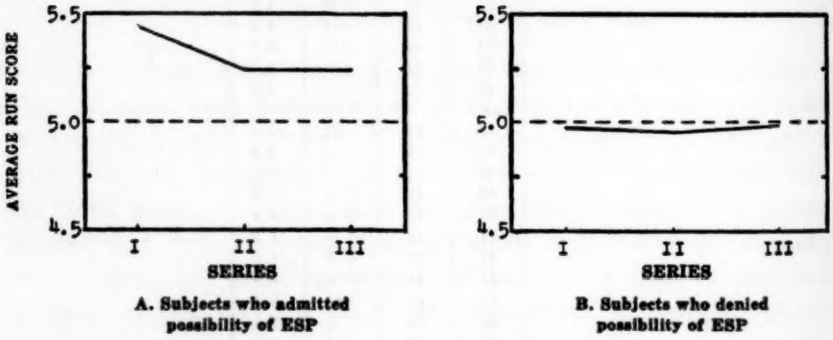


FIG. 4.—The average run scores of successive series, each of which was a subunit of the experiment.

Position Effects as Related to the Subject's Attitude

All these data seem to point toward position patterns that parallel those found in the earlier experiments in learning. Either primacy or finality effects appeared wherever we looked for them: in the segment, the run, the experimental session, and the experiment as a whole. It might have been difficult to predict correctly for each case whether the initial or the final position would show the greater advantage—but this too was a characteristic of the early memory work.

We can feel more confident in pointing out this parallel because the experiment is far from standing alone as a study of position patterns in ESP. In the last few years analysis in terms of the salience of the end trials of the segment and the end segments of the run has become routine in the ESP research and it has repeatedly yielded positive results. (For example, see **2, 11.**)

In broad outline the findings on the salience effect in this research confirm earlier conclusions on the topic. It is our hope that a clue to the explanation of such effects will be found in the hypothesis that they are indices of shifts of attitude, mood, or motivation on the part of the subject. This has often been assumed, as in Rhine and Humphrey's ingenious analysis of the data of a PK experiment conducted by McDougall (**12**). His results show position effects but they are the reverse of the usual ones; and an explanation is offered in terms of the unusual atmosphere presumably induced by his conditions. The authors invoke the subjects' respect for the experimenter, their intellectual conflict about PK, and their cumulative competitive interest in an experiment formally organized to simulate a game.

Similar explanations can be made of the position effects reported in the previous section. As the subjects approached their task in good faith, they may well have started each new run with a good resolution—which petered out after the first few guesses. The hopeful subjects may have tried to continue at the same high level through the entire run; while the pessimistic ones, more aware of boredom, may have prodded themselves again as the end of the run came in sight. This would account for the primacy effect having occurred in both groups, while the finality effect was present only for those subjects who did not accept the ESP hypothesis.

The initial spurts, indicated in Figure 3, may be explained if we postulate what is not unlikely: that all the subjects felt interest or enthusiasm at the beginning of the experiment, and that the group tested under routine conditions had a resurgence of the same lift when they could start fresh on a clean sheet of paper. If the last column had been designated, there would have been a clearer test of a finality effect; under the experimental conditions, none appeared for this group. The subjects tested under pleasanter conditions always seemed responsive when the experimenter made some such comment as, "This is the last run. Let's see what you can do on it!" and in their case a finality effect was marked.

The advantage of the first series over the two later ones probably follows the same pattern. As the experiment continued, the experimenter's approach to the subjects became perceptibly more routine; and since the situation is a social one, this probably had some slight effect on the subjects' approach to their own task. This would account for the higher scores in the first series of the experiment than in the two others.

As was shown in the first section of this paper, similar patterns of advantage in certain serial positions are commonplace in learning experiments. A rough parallel can be drawn between the explanations that we have just offered and those that the earlier psychologists, such as William James, gave for the "conditions of association." Many of our points could be generalized under a law of novelty or James's law of vividness; and probably all of the others under a second law which he proposed for associative readiness, that of emotional congruity (3, p. 577). Both of these laws could also be applied to many of the differential effects other than salience that are found in ESP experiments: a new set of targets will be more vivid to the subject than a familiar set, for example; and the introduction of new targets is followed by an increase in score (9). When the subject has just been given candy (14) or kind words (8) or the enlivening effects of new targets (9), pleasing the experimenter by making scores above chance has more emotional congruity for him. He is in the mood, and his scores rise.

William James used these reasons to explain why certain associations are made rather than others. It is striking and probably more than coincidental that they apply equally well to the pattern of suc-

cessful guesses in ESP. It implies that the two situations have more in common than the contrast between learning and paranormal perception might lead one to suppose; and we suggest that the obvious factor common to both is the subject, whose interest and emotional patterns help in both cases to determine his response.

In addition to the analogy between position effects in serial learning and in ESP, another and more curious analogy might be drawn between the two fields. In both, we sometimes find a subject's responses so incorrect that we are led to postulate inhibition or blocking. Rhine, for example, describes a subject whose scores were remarkably high. But one afternoon he wanted to leave the laboratory early and on this occasion his scores fell far below chance (**10**, pp. 62-63). There are similar cases in experiments which I am now conducting where scores spectacularly below chance seem associated with resistance, poor adjustment, or guilt on the part of the subject; and there was some informal indication of the same effect in the experiment whose data are given above. Such low scores are strikingly analogous to the lapses of memory and slips of the tongue that have so convincingly been shown to be indicators of unconscious resistance.

To round out the picture, let us compare for a moment the normal learning behavior that we continually see about us with spontaneous cases of paranormal perception. Murphy suggests that the need for contact on the part of agent or percipient is an important factor in spontaneous cases of telepathy (**5**, **6**, **7**). And we all take for granted the parallel case: that a well-adjusted person, by and large, will remember rather well the things that are most useful for him to remember.

We have seen that the similarities between extrasensory perception and memory are many. But a remarkable aspect of the facts which have been adduced is that they are concerned with learning phenomena which are now seldom studied *in the field of learning*. Amnesia, slips of the tongue, and emotional congruity are of primary interest to workers in psychological dynamics, students of motivation and emotion. Miller even goes so far as to say: "All theories of forgetting conceived by case study and incubated in the clinic state that, when recall is not possible, it is because the memories do not dovetail with some drive, need, sentiment or instinct of the organ-

ism" (4, p. 227); that is, that memory lapses are due to emotional aberration rather than to inadequate learning or retention.

Concluding Remarks

We conclude, therefore, that where memory and ESP resemble each other—and these resemblances are frequent and striking—the similarity is probably due to a factor common to both situations: the motivation of the subject, whose needs as well as his capabilities systematically influence his response. It is the author's belief that salience is an index of the presence of varying emotions and attitudes, and that the pattern of high and low scores can profitably be studied by the methods used in psychological studies of motivation.

References

1. Greenwood, J. A. The statistics of salience ratios. *J. Parapsychol.*, 1941, **5**, 245-49.
2. Humphrey, Betty M., and Rhine, J. B. A confirmatory study of salience in precognition tests. *J. Parapsychol.*, 1942, **6**, 190-219.
3. James, W. *The Principles of Psychology*. New York: Henry Holt, 1890. Vol. I.
4. Miller, J. G. *Unconsciousness*. New York: J. Wiley and Sons, 1942.
5. Murphy, G. Spontaneous telepathy and the problem of survival. *J. Parapsychol.*, 1943, **7**, 50-60.
6. ———. Psychical phenomena and human needs. *J. Amer. Soc. psych. Res.*, 1943 **37**, 163-91.
7. ———. Removal of impediments to the paranormal. *J. Amer. Soc. psych. Res.*, 1944, **38**, 2-23.
8. Pratt, J. G., and Price, Margaret. The experimenter-subject relationship in tests for ESP. *J. Parapsychol.*, 1938, **2**, 84-94.
9. Pratt, J. G., and Woodruff, J. L. Size of stimulus symbols in extra-sensory perception. *J. Parapsychol.*, 1939, **3**, 121-58.
10. Rhine, J. B. *Extra-Sensory Perception*. Boston: Bruce Humphries, 1935.
11. ———. Terminal salience in ESP performance. *J. Parapsychol.*, 1941, **5**, 183-244.
12. Rhine, J. B., and Humphrey, Betty M. The PK effect: the McDougall One-Die Series. *J. Parapsychol.*, 1943, **7**, 252-63.

13. ———. The PK effect: special evidence from hit patterns. I. Quarter distributions of the page. *J. Parapsychol.*, 1944, **8**, 18-60.

14. SCHMEIDLER, G. R. Predicting good and bad scores in a clairvoyance experiment: a preliminary report. *J. Amer. Soc. psych. Res.*, 1943, **37**, 103-10.

15. ———. Predicting good and bad scores in a clairvoyance experiment: a final report. *J. Amer. Soc. psych. Res.*, 1943, **37**, 210-21.

THE EVALUATION OF SALIENCE IN DR. SCHMEIDLER'S ESP DATA

By BETTY M. HUMPHREY and J. B. RHINE

THE ARTICLE by Dr. Schmeidler in this issue challenges the interest of parapsychologists who have been making a special study of salience effects in ESP data. It seemed to us worth while, therefore, to apply to the data of her Table 1 the standard methods of analysis for the measurement of salience which have been developed in the last few years. We submit the results of the analysis and suggest that a finding has emerged that may have some bearing upon the interpretation of salience effects in general.

For convenience in presenting the salience evaluation, the data obtained from the hit distributions in the segment and the run have been assembled in the table below. In Section A are given the deviations for the five positions in the segment and those for the five segments of the run for the group of favorably inclined subjects. In Section B comparable deviations are given for the other group. Section C gives corresponding data for the two groups together. The chi-squares (CR^2) are given for the deviations for each position represented.

From these chi-squares the salience ratios, or SR's (see Glossary) are computed. The procedure for computation has already been published (2). The estimation of the salience effect is in terms of SSR's, or salience ratios for the segment, and RSR's, or salience ratios for the run. These ratios are given in the table underneath the data from which they are computed.

SALIENCE DATA AND THEIR EVALUATION

	Section A. Subjects Who Admitted Possibility of ESP									
	Trials in Segment					Segments in Run				
	1	2	3	4	5	1	2	3	4	5
Deviation	− 3	+41	+ 7	+39	+36	+40	+18	+29	+19	+14
χ^2	.17	2.32	.40	2.21	2.04	5.15	1.04	2.69	1.17	.62

$$SSR = \frac{.17 + 2.04}{2.32 + .40 + 2.21} = .40 \qquad RSR = \frac{5.15 + .62}{1.04 + 2.69 + 1.17} = 1.18$$

	Section B. Subjects Who Denied Possibility of ESP									
	Trials in Segment					Segments in Run				
	1	2	3	4	5	1	2	3	4	5
Deviation	− 8	−36	− 8	−22	+40	+10	−26	−24	−24	+30
χ^2	.37	1.68	.37	1.03	1.87	.22	1.46	1.25	1.25	1.96

$$SSR = \frac{.37 + 1.87}{1.68 + .37 + 1.03} = .91 \qquad RSR = \frac{.22 + 1.96}{1.46 + 1.25 + 1.25} = .55$$

	Section C. All Subjects Pooled									
	Trials in Segment					Segments in Run				
	1	2	3	4	5	1	2	3	4	5
Deviation	−11	+ 5	− 1	+17	+76	+50	− 8	+ 5	− 5	+44
χ^2	.16	.03	.00	.38	7.50	3.24	.08	.03	.03	2.51

$$SSR = \frac{.16 + 7.50}{.03 + .38} = 18.76 \qquad RSR = \frac{3.24 + 2.51}{.08 + .03 + .03} = 39.16$$

$P = .01$ $\qquad$ $P = .004$

CR (Cov. SSR *vs.* RSR) = 4.60
P = .0003

The point of principal interest is that although neither group of subjects produces significant salience ratios when the results of each group are taken separately, *both types of SR are significant* when the deviations of the two groups of subjects are pooled. The SR to be expected on a theory of chance is .67, and it is obvious that the SR's of both types, run and segment, for both groups of subjects

taken separately do not differ significantly from that value. The SR's of 18.76 and 39.16 for the pooled groups in Section C are obviously outstanding in comparison with those given for the separate groups.

The procedure for evaluating interrelations between the salience effects shown by the segment and those shown by the run is the covariation statistic (1). There would, of course, be no significant covariation between the very small SR's given by the separate groups, but when this statistic is applied to the SR's of the pooled data, a covariation CR of 4.60 is obtained. This is an unquestionably significant CR; the probability of its occurrence by chance would be approximately .0003.

As it turns out, this is the *first case* of its kind (that is, of significant salience emerging from the pooling of data from two groups of subjects having different attitudes), and it may best be taken as tentative until further research permits wider generalization. The importance of the finding which is stressed here is not the absolute value of the significant figures, but the *contrast* of the SR's of the pooled data with those obtained when the two groups of subjects are considered separately. The results suggest that perhaps in the earlier research on salience effects the outstanding CR's might have been due to the use of an undifferentiated group of subjects in regard to attitudes. The desirability of reanalysis of the earlier findings on the basis of separation of favorable and unfavorable attitudes toward the experiment is at once apparent; or, if such separation is not now possible, it would be worth while to reanalyze on the basis of negatively scoring, as contrasted with positively scoring, subjects, with a view to ascertaining whether the significant salience effects disappear when such a break-down of the total data is made.

References

1. Greenwood, J. A. The statistics of salience ratios. *J. Parapsychol.*, 1941, 5, 245-49.
2. Rhine, J. B. Terminal salience in ESP performance. *J. Parapsychol.*, 1941, 5, 183-244.

THE CARINGTON FREE-DRAWING APPROACH TO THE ESP PROBLEM

By C. E. Stuart
Stanford University

In 1940 W. Whately Carington presented in this Journal the first results of his experimental attack on the ESP problem. This research was later assisted by the grant of the Perrott Studentship in Psychical Research at Cambridge University. The most recent report under the general title of "Experiments in the Paranormal Cognition of Drawings," is subtitled "Steps in the Development of a Repeatable Technique."[1]

The technique differs in many important respects from the card-guessing methods used so extensively in this country. Its value, therefore, is greater than any immediate contribution that may be apparent, for it is a "different" approach. The ESP problem needs new approaches.

The Problem

Carington sets for himself the problem of developing a "repeatable" experiment. Two basic requirements of any such technique are that it be *valid* and *practicable*. By the first is meant that "given a normal degree of care, intelligence, and integrity, such as we may reasonably postulate in any responsible worker, it should be virtually impossible to obtain spurious results; by the second, that it should not make prohibitive demands in the way of unobtainable apparatus, special skill and knowledge, inaccessible 'sensitives,' or excessive expenditure of time and energy." A third necessity is that the technique successfully demonstrate the phenomenon in question. The demands of repeatability should not, however, be considered to imply invariable success.

The problem of repeatability grows out of the conviction that in its present status rigorous "proof" of paranormal cognition is

[1] *Proc. Amer. Soc. psych. Res.*, Vol. XXIV, January, 1944.

logically and experimentally impossible; the former because the phenomenon must be controlled before the laws of that control can be manipulated, the latter because the hostile critic has no social limitation upon the degree of implausibility permitted in his counter-hypotheses. These difficulties can be countered only by a greater degree of responsible scientific attention to the problem. A good repeatable technique would instrument the step from initial interest to effective research. Carington regards the need for this technique as of more importance than the attempt of individual investigators to press on to new discoveries in isolation.

The Method

The methods proposed are the outcome of Carington's several years of experience with experiments using "free material" as stimuli. In these experiments a series of ten words, chosen randomly from a dictionary, were each illustrated by a simple drawing denoted an "original." The originals were displayed, one each evening for ten evenings, in a room inaccessible to the percipients. The percipients were shown a photograph of the place where the original was displayed and were asked to attempt to reproduce the original in drawings of their own. The responses were then compared to the originals to see what degree of reproduction had been achieved.

It is impractical to list here all variations within this framework of method that have been actually employed in the eight experimental series Carington has reported. Several are definitive, however, with regard to his final suggestions and are therefore basic to understanding them.

One important technical step was the decision to score only "palpable hits," that is, response drawings which clearly and unambiguously reproduced the idea of the original. With this criterion an unbiased judge found a fairly large number of responses which seemed to reproduce some one of the originals of the given experiment.

It was observed immediately, however, that the "palpable hits" did not coincide consistently with the intended original, but seemed to occur randomly throughout the experiment in which the original had been used. When the responses were "cross-scored" against

the originals of previous and following experiments it was found that there were significantly more hits on the originals of the intended experiments than on those of the preceding and following ones. This observation led Carington to give up the traditional notion of a strict temporal relation between response and original. His technique of scoring as a hit a response which matches any of the originals of the experimental series is a new one.

The cross-scoring method was judged unsatisfactory, however, on several grounds. For example, the results varied markedly between judges. It was particularly inadequate, moreover, in dealing with a very evident variable of the free response method that should be one of its virtues, namely, the difference in common-sense value of specific reproductions. For example, if the original is an apple and if a percipient draws an apple, we are not particularly impressed, because we know that among free drawings we are certain to get a lot of apples. But if the original is an oilcan, and the percipient draws an oilcan, we are much more impelled to entertain the notion of a cause and effect selection, for we know that not many oilcans will be drawn. Yet the cross-scoring method scores such hits equally.

Calculation of Scores

Carington's paper is almost wholly devoted to exposition of a method of scoring which makes up for the defects of the cross-scoring method. The problem, as can be seen in the above illustration, is to give each hit a weight representing its likelihood of occurrence. If we knew how many people would draw an apple in a series of ten free drawings, then we could calculate the probability that any given set of these people would draw a certain number of apples in a particular experiment.

Carington approaches this problem by presuming that his large mass of data constitutes a good example of what people will draw in the free situation. In his Experiments I-VII, from 741 percipients, he got 10,006 drawings. Of these, 9,295 could be identified and listed in a Catalogue of 1,259 headings. Looking under APPLE we find that a total of 105 subjects drew an APPLE; under OILCAN we find that only two subjects drew an OILCAN; etc. The Catalogue is, in other words, a table of frequencies of objects drawn by the 741 subjects.

If the assumption is reasonable that these relative frequencies are representative of the population of subjects who will ordinarily partake in experiments of this kind, then we can predict that in any experiment sufficiently similar in conditions, the objects drawn will occur with the same relative frequency. If 14 per cent of the known subjects drew an APPLE, then we may expect 14 per cent of a new group of subjects to draw an APPLE, within certain limits of error of prediction. The relative frequency expected can be stated as a "probability." The "probability" that a given subject will draw an APPLE in a series of 10 responses is .142, or very nearly 1/7. The probability that a given subject will draw an OILCAN is .0027.

With the aid of one more assumption these probabilities provide a statistical tool that renders free material as directly assessable as the familiar card-guessing experiments. The assumption is that the predicted frequencies will be distributed according to binomial law.

If n is the number of subjects in an experiment, and p is the Catalogue probability that they will draw a given item, then the expected number of subjects drawing that item will be np. The error of this prediction will be the square root of npq where $q = 1 - p$.

If one of the originals of that experiment is an APPLE, then the number of subjects drawing an APPLE is counted as so many hits. Whether this number of hits is greater than chance expectation can be calculated by the familiar formula:

$$CR = \frac{D}{\sigma} = \frac{h - np}{\sqrt{npq}}$$

The empirical probabilities provide also a basis for a properly weighted "score" of the performance of a subject, of responses to a given item, of an experiment, or of a series of experiments. The "score" is derived by a method introduced by R. A. Fisher in 1924 and is essentially the same as that used by Saltmarsh and Soal, and by J. G. Pratt, in their evaluations of mediumistic material.

The Fisher score for any original is calculated by the formula:

$$f = (h - np) \log \frac{1}{p}$$

This has a variance:

$$V = npq \log^2 \frac{1}{p}$$

The score of a whole experiment is the sum of the scores for all the originals used, and the total variance is likewise the sum of the individual variances.

In actual practice the structure of the Catalogue requires some complication of these formulas. In the first place, when many of the drawings there tabulated were made, the item was actually an original in the experiment. So unless we want to ignore the hypothesis being tested, it is necessary to calculate the expected frequency from the observed frequency when the original was *not* available as a test stimulus. Thus if c is the number, given in the Catalogue, of the percipients who have drawn the object in question in the course of experiments in which that object was *not* an original, and N is the total number of percipients working in those experiments, then $p = \frac{c}{N}$.

Secondly, an object might well be used as an original which does not occur in the Catalogue at all. In this case, $p = 0$ and the score would be indeterminate. This is avoided by taking $p = k/M$, wherein $k = h + c$, and $M = n + N$, n being the number of subjects in the experiment being evaluated.

Finally, the Catalogue frequencies are based upon the fact that the subject, in a series of ten responses, drew an average of 11.42 relevant items. If, in a new experiment, subjects produced a higher average than this, then the probability of hits would be proportionately increased. To correct for this, it is necessary to find the actual average of items produced. If it is greater than 11.42, then in the above formulas, n should be replaced by n' equal to the first integer greater than $741\ r/R$, wherein r is the total number of drawings produced in the experiment under examination, and $R = 8{,}463$, the total number of relevant "items" in the Catalogue. If the average of items produced is less than that in the Catalogue, Carington advises that no correction be made.

The fact that the Fisher score can be calculated for the single subject with respect to a particular original provides an alternative method of scoring which is much less laborious than the above. For the single subject response to a single original, the n of the Fisher formula is 1, while h is either 1 or 0, depending upon whether the subject scores a hit or not. If he makes a hit, his score is $(1 - p) \log 1/p$ with variance $pq \log^2 1/p$. If he misses, his score is $-p \log 1/p$ with the same formula for the variance as for a hit. But the value of $p = (h + c)/(n + N)$ is not the same in the two cases. For a hit $p = (c + 1)/(N + 1)$. For a miss $p = c/(N + 1)$.

If the score and variance appropriate to a hit are denoted U and V, respectively, and those appropriate to a miss are denoted U' and V' (conveniently referred to as Unit Scores and Variances), then for any fixed

N, such as the 741 of the Catalogue, they can be calculated once and for all for the appropriate range of c values.

Carington provides Appendix tables of U, V, U', and V' for each of the four fixed N's of his Catalogue.

The total score and variance of an experiment is then equal to the sum of the unit scores and variances. If h percipients score hits on an original while m do not, the score for that original will be $hU - mU'$ with variance $hV + mV'$.

Designation of Hits

One of Carington's objectives is a method which will permit the experimenter himself to judge the hits made. Because the knowledge of the actual correspondence between original and reproduction may emphasize a similarity which would not be recognized if the true correspondence were unknown, the experimenter is not an unbiased judge. However, Carington believes that by following the principle of "literalism" the effect of bias is rendered negligible.

The principle is simply that of scoring as a hit only those responses that literally and unambiguously reproduce the idea of the original. Thus a drawn "cat" is a hit on an original "cat" despite any variation in form or posture, but would not be a hit on "tiger" no matter what the formal similarity or zoological association.

The principle of literalism is not a necessary feature of the method in general, however. Carington recognizes that for "informational" purposes one may decide (in advance, of course) to designate as hits a liberal range of closely associated objects. One must in this case, however, use the Catalogue with caution. If it is decided in advance to score any animal as a hit on "cat," then all the frequencies applicable to animals in the Catalogue must be used to determine the expected frequency.

Application

The method is applied first to Experiments I-VII, upon which the Catalogue is based, and then to a further experiment, number VIII. The following table is from Carington's Table Ia and II. Under "Expt." is the number of the Experiment, "n" is the num-

ber of subjects, "*s*" is the number of originals used, and "CR" is the critical ratio.

Expt.	*n*	*s*	CR
I	37	12	1.89
II	20	14	3.67
III	11	10	.19
IV	105	11	1.52
V	77	10	.11
(Total I-V	250	57	2.81)
VI	246	10	3.09
VII A	245	10	.39
VII B	245	10	2.72
VII C	245	10	.54
VII D	245	10	1.19
VII E	245	10	.40
(Total VII	245	50	2.33)
VIII	431	10	5.03

The range of CR's observed is considerable, but all are based upon positive scores. The CR's of the totals, 2.81, 3.09, 2.33, and 5.03 respectively, indicate the consistently extrachance nature of the results and support the contention that the method is repeatable.

Structure of the Catalogue

The Catalogue is essentially a list of the objects drawn by Carington's 741 subjects. These were all persons of the higher educational classes of "Western European culture." The great majority were British, and 80 per cent were students of college age. Carington warns that it would be injudicious to use the Catalogue in experiments with mental defectives, children, or other population groups grossly differing in cultural milieu from his subjects.

The instructions given the subjects varied somewhat as the experiments proceeded. In Experiments I through V the instructions were: "Certain drawings, all of which will be quite simple, will be displayed. . . . You are asked to reproduce these as well as you can." In Experiments VI and VII, the subjects were cautioned not to draw vague scenes, elaborate interiors, or geometrical diagrams. The intent of the later instructions was to try to eliminate useless and unassessable material.

The names of items in the Catalogue are presented in three sorts of type, SMALL ROMAN, ordinary, and *italic*. Those in small

roman are items which Carington judged to be suitable for originals because of their general familiarity and lack of ambiguity in representation; those in ordinary type are used as subheadings under the suitable items; and those in italic type are the ones he considers *unsuitable* for originals for one or another reason.

One kind of unsuitable item is that which forms the actual or implied background of a number of objects, like *sea* or *field*. This includes the large class of items which are part of the response but which provide no sure indication that the percipient intended to express that particular idea. Other criteria of unsuitability are undue rareness, technicality, obscureness, ambiguity, etc.

The frequency of occurrence of the item is then tabulated in four double columns. The first double column gives the frequency of occurrence in Experiments I-V; the second, the frequency in Experiment VI; the third, the frequency in Experiment VII; and the last double column, the totals of all Experiments. The double column lists the frequency, first, of the occurrence of the item as a single item of the response; and second, of the occurrence of the item as part of a more complex drawing. (A drawing of a HORSE would be listed as an item in the left of the double column, and a drawing of a horse and cart would produce an entry in the right of the double column for both HORSE and CART.)

Summaries at the end of the Catalogue give the totals for the various cross-classifications.

Certain rules to guarantee objectivity and order of entry are laid down. All objects drawn were entered; nothing of a given percipient was entered twice under the same heading; plural objects were entered as if they were singular except in such cases as it seemed reasonable to keep the two categories; parts were treated as equivalent to wholes, but wholes were not itemized in their parts; content rather than form was the determining factor; and so on.

Some Critical Remarks

An immediately practical question may be raised: What will be the effect of a real difference in population of subjects, say American college students, upon the validity of the Catalogue frequencies? For example, only one of Carington's subjects drew a baseball bat,

whereas thirteen drew cricket bats. Secondly, it is reasonable to suppose that there will be topical and seasonal effects in free drawings of this character. More turkeys will be drawn in November and more conifer trees in December than in June.

As Carington shows, in the long run these variations will cancel out. General preferences for one item will result in suppression of frequency of other items. If the originals are chosen with proper randomness, then the likelihood that a preferred item will occur as an original in the experiment where it favors positive scores is countered by the additive likelihood that unpreferred items will occur in equivalent experiments and favor negative scores. This is certainly true, in the very long run. However, a particular research project is apt to use as originals only a small percentage of those available in the Catalogue. And, accordingly, there is a good chance for spuriously high or spuriously low scores in any single project.

Carington counters this hypothesis with the evidence that his experiments are all positive in score, a result that could not be due to accidental correspondence of originals and topical preferences. (This is cogent evidence in regard to his experimental results, but it does not give us any information on the validity of the method.) He suggests further that, on the basis of his experience with the material, it seems that the effect is negligibly small (which is more in line with what we need to know).

Surely there is a much safer line of defense on this point in the use of common-sense judgment. Either scoring method gives total scores with respect to each original. If, perchance, a turkey was the original in a November experiment, then it would be the work of a moment to find out whether that score lent an unreasonable proportion to the total critical ratio.

The need for responsible critical appraisal of all scores on originals is not emphasized by Carington to the degree that it should be, unless we are to take the indefensible view that the method is a finished machine for transforming subject responses into valid answers to problems. An example will show the danger. The table illustrating the scoring techniques lists ten originals and the steps of arriving at the total evaluation of the experiment. I present it here greatly abridged:

Original	Score	Variance
1 Herring	1.03	1.71
2 Leaf	2.34	1.58
3 Turban	−.07	.21
4 Marigold	.28	1.70
5 Nail	−.32	.67
6 Shell	1.18	.98
7 Bird	1.92	1.65
8 Rays	2.26	.34
9 Steeple	−.44	.84
10 Mermaid	−.07	.21
Total	8.10	9.89

$D/\sigma = 2.575$; $P = .01$

The total score and variance give a significant positive result. But if we look at the individual contributions to that score, one stands out: original no. 8, RAYS. This item has an individual CR of 3.86. It contributes over one-fourth the total score and a negligible amount to the total variance. Since there is no psychological or parapsychological expectation for this unusual weight, it is reasonable to ask whether it might not be spuriously excessive. In the Catalogue, *rays* is italicized as unsatisfactory for an original. Anyone who has tried to catalogue items will recognize that the major difficulty is in deciding whether the subject intended the idea of *rays* or merely utilized them to identify the idea of *sun* or *moon* or other light source. The only way to be sure is to tally only those cases where the subject has made special mention of rays, usually by writing out the word. This seems to be what Carington has done in the Catalogue. The listed item reads, "Rays of the sun, specifically mentioned."

Of the original in question Carington says, "For RAYS the experimenter drew an arc of a circle with a number of radiating lines." Strictly, then, the original used is not the catalogued item, since any formally similar drawing of rays would satisfy the original, while the Catalogue frequencies have two definite qualifications. Further, the response in question was entitled "storm, with rain and sunshine between" with a drawing indicating rays. A judge could, with perfect reasonableness, either include or exclude this drawing from the Catalogue classification of *rays*.

None of these considerations counter the fact that the response to the original was a hit. But all suggest that the score and variance figures for this item are spurious and in excess of what they should

be. In this case the fault was in the use of an unsatisfactory original. In any case, however, at least until the method has been thoroughly tried, it would be wise to examine contributions to the score from any item, particularly if that contribution seems excessive.

In Experiments I to V the percipients were permitted fairly complete freedom of response, influenced of course, by their knowledge that the originals would be simple line drawings of readily drawn objects. After it was clear that scoring of only palpable hits seemed the best method of assessment, the natural course was to limit the response performance somewhat in order to reduce the amount of useless material such as geometric forms and vague ideas. The slight change in instructions for Experiments VI and VII was effective. The number of indeterminate items, geometric figures, numbers, and letters was reduced from 535, in Experiments I to V, to 54 in Experiment VII. At the same time the proportion of ROMAN type items to *italic* type items increased. In Experiments I to V there were 1,795 ROMAN type items to 1,196 *italic* type items. In Experiment VII there were 2,141 ROMAN type items to 959 *italic* type items. A chi-square test of these frequencies yields $P = .02$, suggesting that the difference is a true change of proportion.

It is apparent that a hit in Experiments I to V evaluated by the frequencies of Experiments VI and VII actually is undervalued. But, conversely, evaluation of an Experiment VII item by the frequencies of Experiments I to V overvalues that item. In Experiment VIII the instructions limited the response even more strictly (the response was named and then sketched), and one would expect another increase in proportion of the ROMAN type items with resulting overvaluation in scoring. In any given experiment the magnitude of this effect may be negligible. It would in the long run, however, lead to just such an excess of positive scores as Carington observes.

Return to the simpler unrestricting instructions would seem to be a "safe side" correction. A better one would be the use in assessment of only those frequencies derived from strictly similar sets of instructions.

Evaluation

It is evident that Carington's work has been exposed to much criticism. Carington himself is careful to indicate many of the areas of uncertainty in considerable detail. I think it would be unfortunate to emphasize these failings in disproportion to the fact that the technique is eminently *usable*. It does not provide an absolutely valid statistical result, but it does provide the only method so far available to approximate in statistical results the evident common-sense difference in value of hits among free responses. It is only by being used and improved that its ultimate value can be realized.

Carington seems to vary between recognition of the method as a step in the right direction and regarding it as a "standard" procedure to be generally adopted. The most serious objection to the latter is the use in scoring of only palpable hits, a procedure which is a convenience to the experimenter, but a limitation to the phenomenon under study. There is always the danger that the experimenter, because he has no technique for dealing with the unassessable response material left over, will ignore it as having no value for his problem. We do not yet have such understanding of the variables important to the problem to discount in advance *any* behavior in the response situation.

Let me emphasize again that Carington has given the ESP investigator a usable technique. The most serious limitation of this review is that it cannot reproduce in detail the Catalogue, the Appendix tables, and much of Carington's comment pertinent to them, all of which are essential to employment of the method. Even with them the attempt must be a careful one. My own effort required three sets of calculation before I felt the method had been properly applied to experimental material from one of my experiments. The resulting CR's, 4.5 and 4.3 for the two methods respectively, seem too high in terms of a common-sense appraisal of the quality of the responses. It is clear, however, that the proper basis for even a "common-sense appraisal" can be gained only through many further applications of the method.

PK TESTS WITH SIX, TWELVE, AND TWENTY-FOUR DICE PER THROW

By J. B. RHINE and BETTY M. HUMPHREY

ABSTRACT: This work of Lieutenant and Mrs. Frick represents one of the most interesting variations of the PK experiments, particularly in the comparison of two sizes of dice thrown from the same cup, 12 of each at a time. The results are statistically significant, and a fairly good case is made against the common alternative hypotheses. But the main feature is the way in which the rate of success was patterned on the record page. The larger dice showed a typical decline pattern—vertically, horizontally, and diagonally; the smaller dice showed the reverse. This is the only experiment of its type on record as yet, but it is one which will assuredly challenge repetition.—Ed.

IN THE summer of 1934 Harvey L. Frick, a young psychologist, and his wife, Ruthanna, visited the Parapsychology Laboratory at Duke and were incidentally initiated into the PK tests which were being conducted there—tests based on the hypothesis of the dominance of mind over matter. The two visitors became interested in the dice-throwing techniques and decided to conduct some experiments themselves on returning to their own institution. The present paper is a report of research which grew out of that resolution.

Beginning in December, 1934, these investigators carried out work on the PK problem for a number of years, completing three experimental series. Only the first of these three is covered here.[1] Throughout the research the investigators cooperated with the Parapsychology Laboratory and, on the termination of their experiments, they made out a full report which they placed, along with the original record books, in the laboratory files for eventual reporting when the time should come for public discussion of the PK researches. The Frick data were studied, when the analysis of the PK records was finally undertaken, along the lines already familiar to readers

[1] At the time this series was carried out, Frick was instructor in Physiology at the College of Medicine, Wayne University. He is now serving as Personnel Consultant, Adjutant General's Department, United States Army.

of the reports which have preceded this paper in publication. The analyses were carried out under the guidance of B.M.H. Both of the experimenters (hereinafter referred to as *H.* and *R.*) have read this manuscript and have agreed that the factual basis of the procedures and the test data is correct.

Outline of the Experiment

In this investigation the two experimenters were very evidently trying primarily to test the PK hypothesis for their own satisfaction. Like everyone else undertaking such experiments, they began with a normal degree of skepticism regarding the PK hypothesis, and this was modified only after long experimentation. Although the primary purpose of H. and R. in their experiments was to find out for themselves whether PK occurs, the main objective of the writers of this report is much less the establishment of PK than the accumulation of evidence that will help in the understanding of it. As evidence of the occurrence of PK, the report is not one of the very strongest cases. There can be no doubt, however, that it is one of the most interesting studies published up to this time in its contribution to our knowledge of the effect of experimental conditions upon performance.

H. and R. were the only ones who took part as subjects in this series. In most sessions, H. recorded while R. was throwing the dice, and R. recorded for H. while he threw. There were a certain number of sessions, however, in which H. threw the dice and took his own records. In view of the nature of the evidence, especially that arising from the analyses for position effects conducted years later at the Parapsychology Laboratory, this unwitnessed work is regarded as being on a par with the witnessed.

It has seemed best to chart the work of this first H. and R. exploration in four sections, each one representing a different stage in the development of the method. These four parts will be numbered in chronological order. The first three are of interest largely because they indicate the course of development which led to the main experimental procedure; most of the tests were done in the fourth section.

In Section I, 150 standard runs of 24 single die-readings each were made. Since there were six dice thrown at a time, every four

throws completed a run. The hit records were entered in a bound record book in a continuous sequence like the printed words in a paragraph. In all, 600 throws were made in this section. Out of these 150 runs, 48 were made by R. and 102 by H. All were made with the same six dice thrown from a cup on to the carpeted floor or blanket-covered table top. The dice were the medium-sized common commercial variety measuring 5/8 in. on the edge and made of plastic materials.

At first an effort was made to throw the dice an equal number of times for each face as target in order to control any possible inequalities. Preferences were respected on occasion, however, and in the course of time there was too much emphasis on certain preferred faces to make it worth while to equalize the number of throws per face. Moreover, the experiment in this initial form proved to be discouraging. A change of plan was decided upon, and exploratory Section I, with its unequal number of throws for the different faces, was abandoned.

Section II was carried out with 12 dice per throw, six—which we shall call the *medium* dice—being of the same size as those used in Section I. In addition, six others were used which were somewhat smaller in size and which measured 1/2 in. on the edge. We shall call these the *small* dice. Although the two sizes of dice, six of each, were thrown together from the same cup, each size was read separately and the records entered independently.

Horizontally, the records were made with 12 entries across the page from left to right; vertically, the lines were labeled *L* and *S* alternately down the page, *L* meaning larger dice and *S* meaning smaller dice. After the first throw the medium dice were recorded in the first space of the top line, labeled *L;* the small dice, in the space directly underneath, labeled *S;* and so forth across the page for 12 entries. The page was filled out solidly from top to bottom with these scores in rows of 12 entries, alternating medium and small.

There were 282 runs (or 564 throws) made for Section II under these conditions before it was felt that the method should be changed again because the results were not showing extrachance scores.

Using the same recording procedure in Section III, the experimenters decided to reverse the objective, assuming a negative attitude of mind and trying to avoid a designated target face. Whereas

they had previously been attempting to influence the dice to fall with a certain face turned up, the subjects now took exactly the contrary point of view. But again, after 216 runs (432 throws) with very little improvement over the preceding section, this third procedure was abandoned.

Section IV was undertaken with a still further increase in the number of dice per throw.[2] There were now 12 dice of the medium size and 12 of the small, all thrown from the cup at a single throw, counted separately for each size, and entered separately on the records. At the beginning of Section IV the record pages were filled out solidly as in Sections II and III; however, after a few pages the record was split into two equal blocks, an upper and lower, each with four double rows. (See Fig. 1.)

All the faces of the die were at one time or another used as target face, but in the end the targets were very unequally represented because of the freedom of choice allowed the subject. The predominant choice naturally was made for the faces affording the strongest stimulus—the larger denominations. (See Table 7.) Under these conditions, 2,292 runs (a run in this case is a single throw) were completed by March 15, 1935, when the PK experiments were interrupted.[3]

Results

General

A brief general survey of the results of all four sections of the H. and R. Series is shown below in Table 1. Specific figures concerning the actual number of runs and the total number of hits above expectation for each target face of the dice appear in Table 7. A total of 2,940 runs for the entire series gave an average score of 4.09, representing a total of 279 hits above expectation with the significant CR of 2.82.

[2] This increase in the number of dice per throw after the previous failures of H. and R. to score above expectation was not so radical a change as it might seem. As stated above, the experimenters were in touch with the Parapsychology Laboratory and knew of tests being made there with large numbers of dice—larger even than those tried in the present case.

[3] In his letter to J.B.R. dated March 14, 1935, Frick wrote: "I haven't figured the values of the figures mentioned in the previous letter but am aware that none can be very significant." Before he learned otherwise, the experiment had already been stopped. This fact is of interest in connection with the hypothesis of optional stopping.

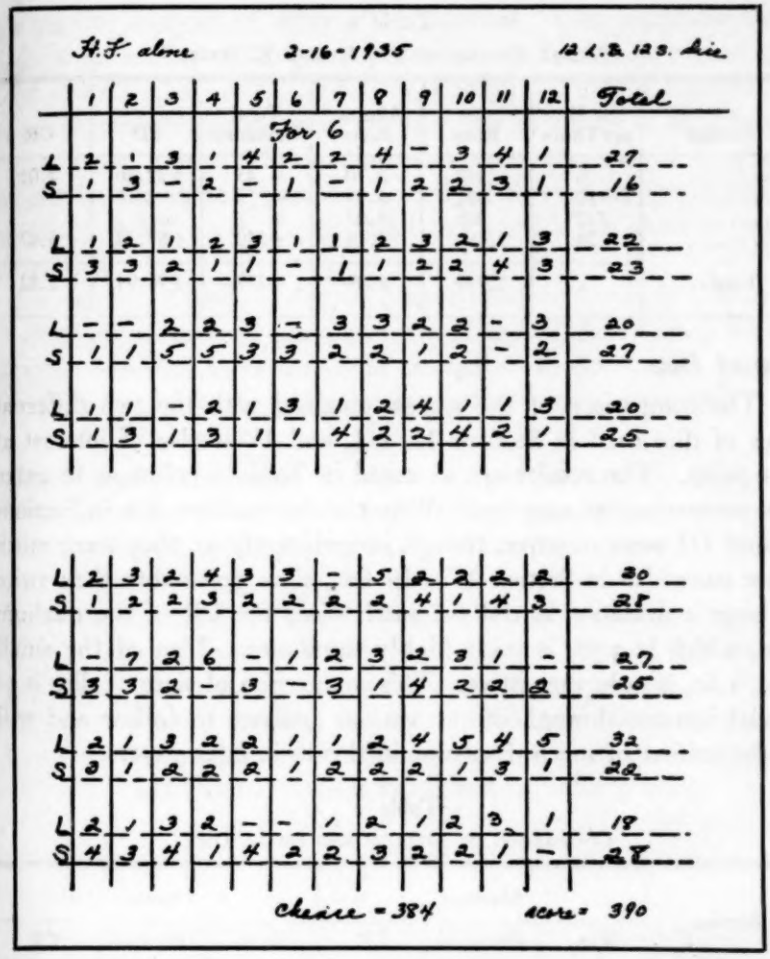


FIG. 1.—Sample of record sheet commonly used in Section IV showing the entries recorded in two separate blocks. Chance expectation for each throw is two hits.

The outstanding figure in the evaluation is the CR of 3.67 for Section IV. It is interesting to note the progressive increase in the average scores given in the fourth column. Beginning at 3.70 for Section I, increasing to 3.99, then to 4.03, and finally to 4.14, the average scores show a marked improvement along with the changes of method.

Table 1

TOTAL RESULTS OF THE H. AND R. SERIES

Section	No. Dice per Throw	Runs	Average Score	Total Deviation	SD	CR
I	6	150	3.70	− 45	±22.36	2.01
II	12	282	3.99	− 3	.....	....
III	12	216	4.03	+ 7	.....	....
IV	24	2,292	4.14	+320	±87.12	3.67
Total	..	2,940	4.09	+279	±98.94	2.82

Size of Dice

The comparison of the success obtained with the two different sizes of dice used in Sections II, III, and IV is also of interest at this point. The results are so stated in Table 2 below as to make this comparison an easy one. Whereas the medium dice in Sections II and III were negative, though insignificantly so, they were much more successful in Section IV. In fact, they gave more than twice as large a deviation as did the small dice; the CR of the medium dice, which is 3.57, is itself highly significant. That of the small dice, 1.62, is only suggestive. This comparison of sizes of dice is of special interest throughout the various analyses to follow and will be the subject of much discussion in this as in later papers.

Table 2

COMPARISON OF MEDIUM AND SMALL DICE

Section	MEDIUM			SMALL		
	Runs	Deviation	CR	Runs	Deviation	CR
II	141	− 15	....	141	+ 12	....
III	108	− 1	....	108	+ 8	....
IV	1,146	+220	3.57	1,146	+100	1.62
Total	1,395	+204	2.99	1,395	+120	1.76

Position Effects

In the analyses which were made of the distribution of hits on the record page in the H. and R. Series, there were found a number of effects on success of the position of trial in the run or page or series. Some of these effects are closely in line with the findings

from similar analyses of other PK researches. (See especially the report of the survey of the quarter distribution of the page appearing in the preceding number.) There is, however, a unique discovery, one which probably arises from the fact that two sizes of dice were thrown at one time. The medium dice show the normal type of distribution, the more typical one as judged by past reports of position effects in PK research; the small dice *consistently show the opposite trend.*

Though the records of Section I were carefully made in a bound account book, as were all records of the H. and R. Series, they were not entered in such form as to lend themselves to distribution analyses. Accordingly, analyses for position effects were made only on Sections II, III, and IV.

Horizontal Distribution. In the records of Sections II, III, and IV the row of entries across the page corresponds closely to the vertical column in the ordinary record page in the research reports made up to this time. Conversely, the totals of the vertical columns correspond to the usual horizontal distribution. For the analysis of position effects comparable to the usual column, then, it was necessary to break down the horizontal row into four segments of three entries each, literally dividing the record page from left to right into four parts. The number of hits above expectation for each segment was counted, and these deviations are shown in Table 3 below. The results and the number of runs are given separately for the medium dice and small dice for each of Sections II, III, and IV and for all three pooled.

Even for the smaller numbers of runs in the second and third sections the opposing trends of the two sizes of dice are fairly discernible. In the fourth section, with its large number of runs, they are quite apparent. The results obtained by pooling the three sections show a marked decline for the medium dice and an upward trend for the small dice. There is a CR of the difference of 2.49 between the total deviations for the first and fourth segments of the medium dice results and an insignificant one for the small.[4]

The extrachance character of the position effects shown in the

[4] There is sufficient evidence of the decline effect in the PK test data to warrant an *a priori* expectation that the first segment will be highest and the fourth lowest in scoring rate.

Table 3

DEVIATIONS IN THE SEGMENTS OF THE ROW FOR SECTIONS II, III, AND IV

Section	Runs	Deviation			
		1st Segment	2nd Segment	3rd Segment	4th Segment
MEDIUM DICE					
II	141	+ 16	− 5	− 7	−19
III	108	− 2	+15	+ 2	−16
IV	1,146	+ 92	+54	+53	+21
Total	1,395	+106	+64	+48	−14
		$CR_{d(1-4)} = 2.49$			
SMALL DICE					
II	141	+ 2	−16	+16	+10
III	108	+ 6	0	+ 7	− 5
IV	1,146	+ 2	+10	+15	+73
Total	1,395	+10	− 6	+38	+78
		$CR_{d(1-4)} = 1.41$			

hit distributions of these three sections is best brought out by considering the sections in the order analyzed. Section IV, being the largest, was analyzed first. The deviations by segments, as shown in the table, gave contrary trends for this section: +92, +54, +53, and +21 with the medium dice, as opposed to +2, +10, +15, and +73 with the small; or, in terms of half-rows instead of four segments: +146 to +74 (medium) and +12 to +88 (small). These figures were consistent and striking enough to warrant expectation of opposite trends in similar situations. In other words, a "prediction" could be made, and the results actually found in Sections II and III may well be regarded as having been "expected." With the sections pooled, the medium dice distribution based on left half *versus* right half of the row gave +24 to −40, a drop of 64; and the small dice, −8 to +28, a rise of 36. Since these opposite trends were to be expected, the differences could properly be pooled for statistical evaluation. The combined differences totaled 100 which, with an SD of ±40.73, gave a CR of 2.45. This CR is of slender significance, but

confirming, as it does, the trend of Section IV, it goes far toward justifying attention to the extraordinary position effects found.

Vertical Distribution. The trend of distribution shown by the two sizes of dice in the downward direction of the page in Sections II and III was obtained by comparing the deviations of the top and bottom halves of the page. Since the pages were not all of equal size, it seemed best to make only two subdivisions in the page in this distribution study. The full page was halved; but, where the record had been made on the half-page basis (see Fig. 1), each half was regarded as a "page" and halved accordingly. The top-bottom comparison for the pooled results of Sections II, III, and IV follows in Table 4.

Table 4

Comparison of Top and Bottom Halves of "Page" for All Three Sections Pooled

	Medium Dice			Small Dice		
	Runs	Total Deviation	Average Score	Runs	Total Deviation	Average Score
Top	684	+134	4.196	684	+17	4.025
Bottom	660	+ 67	4.102	660	+86	4.130

Here, too, the trend is obviously in opposing directions for the two sizes of dice, and the difference in average scores is greater in the small dice results than it is in the medium.

Quarter Distribution. Now with these two types of position effects apparent in the horizontal and vertical directions on the record page, it seemed worth while to note the combined effect of the two tendencies. It was therefore necessary to get the hit distribution in terms of the quarter page. In Section IV, where the page was broken into two subdivisions, each with its own experimental conditions, the half page was treated as a page. Table 5 gives the quarter distribution (QD) for Section IV and the pooled QD for Sections II, III, and IV. (Sections II and III had comparable conditions and too small a number of runs for independent consideration.) The squares are analogous to the record page so that the upper left quarter of the square represents the upper left quarter of the page (or half-page), etc. The squares contain the number of

runs for each quarter, the deviation from expectation, and the average run score.

Table 5

QUARTER DISTRIBUTIONS FOR SECTION IV AND FOR SECTIONS II, III, AND IV POOLED

MEDIUM DICE

Section IV

285 +83 4.29	285 +54 4.19
267 +58 4.22	267 +17 4.06

Sections II, III, and IV Pooled

342 +96 4.28	342 +38 4.11
324 +69 4.21	324 − 8 3.98

$CR_{d(1-4)} = 2.12$

SMALL DICE

Section IV

285 −19 3.93	285 +25 4.09
267 +42 4.16	267 +53 4.20

Sections II, III, and IV Pooled

342 −25 3.93	342 +42 4.12
324 +38 4.12	324 +59 4.18

$CR_{d(1-4)} = (-) 1.77$

The table is self-evident in showing that for the medium dice the general trend is a decline in all directions from the upper left quarter which is the beginning of the page. On the contrary, for the small dice the trend is exactly opposite, an incline toward the lower right quarter. It is apparent that as the subject went through the process of the test, his success was shifting negatively with the medium dice and improving almost proportionately with the small.

The differences created by these declines and inclines are not individually significant, but they are highly suggestive. Certainly they are striking enough to warrant further research on the problem of why these two sizes of dice should behave with such consistently opposite tendencies.

Chronological Distribution. One more distribution remains to be mentioned: the chronological distribution of hits within the four sections. The data of each subject in each section were divided into three temporal subdivisions; these are not equal, however, because it was not considered proper to divide a test session, inasmuch as the session itself might show a decline. Rather, the nearest approximation to equality of the subdivisions was made without splitting sessions. It will be seen that when Sections II, III, and IV were each

Table 6

CHRONOLOGICAL DISTRIBUTION OF HITS

Temporal Subdivision	Total, Sections II, III, IV*						Total, All Four Sections		
	Medium Dice			Small Dice			Both Sizes Pooled		
	Runs	Deviation	Aver. Score	Runs	Deviation	Aver. Score	Runs	Deviation	Aver. Score
1.....	525	+ 67	4.13	525	+124	4.24	1,102.5	+167	4.15
2.....	444	+120	4.27	444	+ 12	4.03	937.5	+106	4.11
3.....	426	+ 17	4.04	426	− 16	3.96	900.0	+ 6	4.01
	$CR_{d(1-3)}$ = .76			$CR_{d(1-3)}$ = 2.35			$CR_{d(1-3)}$ = 1.71		

*Section I is omitted here since it has no comparison of medium and small dice. It is included in the right-hand portion of this table.

divided into the three temporal subdivisions, both the medium and small dice showed a drop with time. In the case of the small dice there was a decline to approximately mean chance expectation after the first period. There is a significant difference between the first and third periods in the totals for the small dice data (CR=2.35; P=.01). In the case of the medium dice, there was an increase in the second period but this was followed by a great drop in the third subdivision. When the four sections are taken jointly, with both sizes of dice included, the decline is more consistent. In terms of deviation from run score expectation, the three subdivisions give

4.15, 4.11, and 4.01. The differences are quite suggestive but are not significant.

Another interesting observation arises when it is recalled from Table 1 that H. and R. increased the rate of scoring with each change from section to section. It now appears that at the same time within the sections there was a marked decline in scoring rate with time.

The point of greatest interest in Table 6 is perhaps the fact that the small and medium dice do not show trends contrary to each other in the chronological distribution as they do in the page distribution described above. This suggests that the chronological decline may not be related closely to the page distribution effects.

Discussion

Counterhypotheses

We emphasized in the beginning that this report is concerned primarily with other issues than the question of whether PK occurs, that it is not intended as a major contribution to the evidence on that point. It is important, however, to determine whether we are dealing here with evidence of PK before going on to the consideration of other findings.

The ever-present question of chance as an explanation of the results may be eliminated from the list of possibilities without hesitation. The CR of the deviation either of Section IV alone or of the total of all the sections disposes of argument on that score. The workings of chance could not reasonably account for the joint results obtained by H. and R. in these tests.

Nor is there any better prospect of explaining the results in terms of manner of throwing the dice. The use of a cup for throwing throughout the series left no opportunity for any kind of skilled manipulation, a point which is still further emphasized by the fact that so many dice were cast at a time. Twenty-four dice go far toward randomizing each other in the very act of falling out of a container even if they are not well shaken or vigorously thrown.

The most common counterhypothesis left is that perhaps the irregularities of the dice could account for the extrachance results. It is quite reasonable to suppose that the dice were not perfect, since even such approximations to perfection as are within the range of engineering could not be expected in the common commercial dice

which were used. The dice were not inlaid. The spots were marked by the use of paint inserted in small cavities, and presumably some inequality or shift of the center of gravity might have resulted from the difference between the material removed and the paint inserted. It seems most probable that the dice would therefore favor the faces with the largest number of spots (the "cavity hypothesis").

As we have already explained, the control on the hypothesis of faulty dice that was intended in the original plan of the experiment —that of making an equal number of throws for the six faces—was dropped because of the development of preferences for specific targets. The distribution of throws for the different target faces and for the different sections of the experiment may be seen in Table 7. There, a superficial glance might give the impression that the one- and two-faces were biased so as to appear less frequently than expected on a theory of chance and that the four-, five-, and six-faces were favored by inequalities of the dice. This would be expected on the cavity hypothesis. But a closer examination, especially of the grand totals, fails to bear this out. The six-face, which should have the highest average score, is second to the four-face; and the five-face, which should be next to the six-face, is practically on a par with the three-face. The three-face is not even on the negative side. The one-face should have the lowest average, but it is the two-face which occupies that position.

A survey of the sections themselves further interferes with any easy assumption about the role of inequalities in the dice. For example, in Section I all the faces but the three-face give negative deviations, and that one gives a deviation of only +1. The cavity theory is totally inapplicable here. Likewise, although the small dice in Section IV strongly disfavor the one-, two-, and three-face, all of these give positive deviations in Section II. The medium dice, on their part, do as good a job of reversing in these two sections, the one-, four-, and six-face showing reversed signs for the sections mentioned. In Section III the largest deviation is on the three-face of the small dice, but these went to the negative side on Section IV. The five-face, which should have been very much favored in the cavity hypothesis, was totally negative in Section III. These contradictory results—contradictory, that is, from the point of view of the dice defects hypothesis—are not to be attributed to the fact that 12

Table 7
Hit Distribution for Various Faces

Section	Subject	Dice	FACES OF THE DIE																	
			One-face			Two-face			Three-face			Four-face			Five-face			Six-face		
			Runs	Dev.	Aver.	Runs	Dev.	Aver.	Runs	Dev.	Aver.	Runs	Dev.	Aver.	Runs	Dev.	Aver.	Runs	Dev.	Aver.
I.....	R.		6	− 5		9	− 5		6	+ 4		6	+ 3		9	− 3		12	− 10	
	H.		41.5	−23		3	0		3	− 3		6	− 6		15	− 2		33.5	+ 5	
	Total.....		47.5	−28		12	− 5		9	+ 1		12	− 3		24	− 5		45.5	− 5	
II.....	R.	Medium	21	+ 5		21	−14		3	+ 3		30	− 6		3	+ 8		3	0	
	H.	"	27	− 9		21	− 6		3	+ 1		3	+ 1		3	+ 4		3	− 2	
	Total.....		48	− 4		42	−20		6	+ 4		33	− 5		6	+12		6	− 2	
	R.	Small	21	+ 2		21	− 7		3	+ 4		30	+ 8		3	− 1		3	+ 3	
	H.	"	27	0		21	+ 8		3	− 2		3	0		3	− 2		3	− 1	
	Total.....		48	+ 2		42	+ 1		6	+ 2		33	+ 8		6	− 3		6	+ 2	
III.....	R.	Medium							27	− 8										
	H.	"							27	+ 3		27	+ 6		27	− 2				
	Total.....								54	− 5		27	+ 6		27	− 2				
	R.	Small							27	− 6										
	H.	"							27	+18		27	+ 1		27	− 5				
	Total.....								54	+12		27	+ 1		27	− 5				
IV.....	R.	Medium				6	−12		30	0		126	+57		90	+ 4		126	+ 54	
	H.	"	12	+ 6		6	− 6								12	− 7		738	+124	
	Total.....		12	+ 6	4.50	12	−18	2.50	30	0	4.00	126	+57	4.45	102	− 3	3.97	864	+178	4.21
	R.	Small				6	− 7		30	− 9		126	+20		90	+15		126	− 5	
	H.	"	12	−10		6	− 3								12	+ 6		738	+ 93	
	Total.....		12	−10	3.17	12	−10	3.17	30	− 9	3.70	126	+20	4.16	102	+21	4.21	864	+ 88	4.10
Grand Total.....			167.5	−34	3.80	120	−52	3.57	189	+ 5	4.03	384	+84	4.22	294	+15	4.05	1,789.5	+261	4.15

other dice were added in Section IV, for the 12 dice added were of the same type as the originals in Sections II and III.

There are also other features of the results and their analysis that bear upon this important question. Recognizedly, the best control on the dice lies in the discovery of internal comparisons to be found in the distribution analyses. Going back to the portion of Table 3 which presents the results on medium dice, it will be noted that in the fourth segment of the pooled results totaling 1,395 runs, there is actually a negative deviation of 14. Now this occurs in spite of the fact that the primary statistical significance of the results is contributed mainly by the medium dice. This segment represents the equivalent of more than 348 runs and involves the readings of 8,370 dice. It is a sizable set of data. More than that, it is the segment which, according to general observation on the decline effect, is most likely to be depressed in scoring rate on the hypothesis of the PK effect. Since the first segment, with its positive deviation of 106, is significantly different from this fourth segment, as is shown by a CR of the difference of 2.49, there is adequate basis for interpreting the results for the medium dice as not attributable to faulty cubes. That is, we cannot suppose both the +106 and —14 to be due to imperfections in the dice.

In turning to the small dice, we have the less significant half of the data to deal with. So far as the case for the PK effect is concerned in this paper, we could quite as well ignore the small dice data as inconclusive and base the conclusions upon the medium dice. But the primary value of the report goes far beyond the question of evidence in favor of the PK hypothesis. Indeed, it is in the peculiarity of the results obtained with the *small* dice that the main contribution lies. It is therefore important that in the small dice data of Table 6 we have significant internal evidence that is unexplainable by the faulty-dice hypothesis. There, the third temporal subdivision in the chronological distribution shows a negative deviation of 16 for the 426 runs, whereas the first 525 runs give a positive deviation of 124. There is a CR of the difference of 2.35, which is marginally significant (P=.01). It is difficult to suppose that the small dice were faulty enough to have produced the significant results in the first subdivision and then to have continued through 426 runs to drop below expectation in the third subdivision. It should be added,

of course, that in making these temporal distribution analyses, the work done on a given target face was subdivided. Thus it is not possible to attribute the decline to any grouping of special target faces in one subdivision.

We may, then, safely reject the hypothesis of faulty dice as an explanation of the results. This hypothesis does not, in any case, fit in with the consistency of the position effects, the chronological decline, or the steady incline of the scoring average from the first to the fourth section as shown in Table 1.

We shall turn, then, from further consideration of the hypotheses of chance and physical imperfections of the dice to the interesting problems that arise with the assumption that the PK hypothesis is the best explanation of the results.

Number of Dice per Throw

The outstanding novelty of the tests we are reporting is the fact that in Section IV there were 24 dice thrown at a time. Unfortunately, it is not possible to compare this 24-per-throw work with the throwing of smaller numbers of dice at a time in Sections I, II, and III because undoubtedly other factors than the increase in the number of dice per throw came into the experiment as the methodology changed from section to section. Therefore it is not proper, for example, to compare the six dice per throw which gave negative deviations with the 24 per throw which were positive.

On the other hand, it would not do to say that nothing can be concluded from the findings with respect to numbers of dice per throw, for it is evident that increasing the number at least does not determine or limit the results. If we cannot say that it was because of the increase in the number of dice that the scoring improved, neither can we say that increasing the number of dice prevented an improvement. It seems fairly clear, at least as a suggestion, that H. and R. went ahead with their increased rate of scoring from section to section with considerable independence of the number of dice per throw. This is also the view that would be in line with what has been published to date on the numbers of dice per throw in relation to success in PK tests. Significant findings have been reported with one die per throw, two dice per throw, and six dice per throw. *Wherever comparison has been offered, it has thus far*

invariably been the larger number that produced the best results. But as we are now accustomed to realize in the light of the Gestalt experiments[5] on the "transposition effect," the number of dice per throw must not be considered alone; it is its relation to the background—that is, to the other numbers per throw that are involved in the same experiment—which is determinative. There is more evidence on this subject yet to be reported, and the question becomes increasingly involved as the reports multiply.

Comparison of Size of Dice

Another of the outstanding features of the H. and R. experiment is the comparison of the sizes of the dice. The ratio of the length of edges of the two sizes of cubes was .5:4 which would give an approximate 2:1 ratio of volume. From this it would be expected on the analogy of physical law, other things being equal, that there would be a deviation twice as large for the small dice as for the medium, but instead the *opposite is the case.* The small dice of Section IV have a deviation of +100, and the medium, of +220. *Here, then, is a case where heavier bodies are influenced more effectively than bodies half their size* though the materials in both are essentially similar in density.

A great deal more interest attaches to this comparison because of the way in which H. and R. carried it out. In every respect the medium and small dice were treated alike except that the medium dice were always counted first. They were thrown for the same faces, from the same cup, the same number of times. All variables presumably affected the two alike unless, of course, there was discrimination in the operation of PK itself.

Position Effects

As stated in the introduction, the most valuable contribution of the H. and R. research lies in the position effects discovered. If it were simply a matter of the gross supplementation of the con-

[5] The point of these Gestalt experiments is that apparent reaction to a perceived object is actually reaction to its *relation* to the total situation of which it is a part. For example, response to the darker of two stimuli is not a response to a stimulus of that order of darkness; if a still darker object is substituted for the lighter one, this still darker object receives the response. That the reaction is made to the comparative relation is manifest. The same might be expected to hold true of the number of dice per throw.

siderable number of such effects already found, we should not emphasize this series particularly. It is even a conflicting case, superficially; but as already indicated, it is the analysis of this conflict that is likely to be the most enlightening characteristic. Why did the scoring on the small dice increase in the row, down the page, and diagonally while the medium dice showed the usual decline? There is nothing in the analysis thus far discovered to warrant a conclusion on this question. All that can be done is to offer a suggestion which should not be regarded as of more than hypothetical value. It will serve perhaps to initiate further specific experimental investigation.

If the subject, in throwing two sizes of dice at once, is more interested in the medium dice, as we find the average person usually is, he is likely to give them his major attention. As already noted, they were invariably recorded first. The orientation toward the larger of two sizes of dice is likely to be better throughout an experiment, and the greater deviation of the medium dice bears out this conception. But as the medium dice declined on the record page, where did the PK process go? Upon what was it exerting its impulse? Presumably a kind of "displacement" occurred so that the smaller dice, normally ignored, came in for an increased measure of attention (or if "attention" carries a connotation of consciousness, perhaps the word "orientation" might be better). In other words, the subject, for reasons unknown, shifted his actual influence away from the medium dice and probably would have done so whether there had been any small dice or not. But whereas, in the ordinary situation, with one size of dice involved, there would have resulted only the usual decline, in the H. and R. Series there was present a second target for the absorption of the hypothetical displacement. We suggest this hypothesis with much more confidence in view of the findings from three other studies, to be reported later, which appear to show a kind of displacement effect in PK data.

Summary

The H. and R. Series of PK tests introduced a new feature in the number of dice thrown at a time—12 medium and 12 small, the medium dice giving larger deviations than the small. Chance is defi-

nitely ruled out as an explanation. The manner of throwing likewise is not regarded as a contributing factor, and the evidence against the hypothesis of faulty dice is moderately good. It seems reasonable, therefore, especially against the background of the PK work in general, to regard this series as evidence of the PK effect.

The analysis for position effects yielded the most interesting results of the experiment. In their vertical, horizontal, and diagonal declines, the medium dice showed typical distribution with regard to position on the record page. However, the small dice, thrown at the same time, showed distinct reversal of these trends, and a suggestion of a kind of displacement effect was offered as an explanatory hypothesis.

The results as a whole declined in scoring level over the brief chronological subdivisions of the experiment.

PRECOGNITIVE TELEPATHY: COMMENTS AND DEVELOPMENTS

By S. G. Soal and K. M. Goldney

We read with great interest in the March issue of the Journal of Parapsychology Dr. Pratt's able compression of the account of our "Experiments in Precognitive Telepathy"[1] carried out with our remarkable subject, Mr. Basil Shackleton (B.S.). May we say that we much appreciate the enthusiastic review and are very grateful for the invitation to contribute to the present issue an account of any further developments which have taken place since the experiments already described in our report.

It is with regret we have to state that owing to circumstances outside our control we have been unable to continue the experiments since April, 1943. Mr. Denys Parsons, M. Sc., using a machine of his own invention, has, however, carried out with B.S. some 882 trials in pure clairvoyance, with negative results. Mr. Parsons' apparatus is specially designed to exclude the possibility of telepathic precognition in pure clairvoyance tests; and since we found that B.S. consistently failed in our own clairvoyance experiments, we are not surprised that Parsons' results were also negative. Parsons also conducted about the same number of tests in GESP with either himself or Mr. Draper (B. S.'s secretary) as agent. These tests were also negative, a fact which appears to be in conformity with our own findings so far as they went; that is, that the role of agent was successfully played by only a minority of persons. Further, it should be noted that Parsons used as his presentation material *five colors;* and in Appendix H of our report, we show that B.S. obtained only a dubious and limited success with this kind of material.

Although we have no further experiments on which to report, there are a few points in Dr. Pratt's review on which we would like to comment.

[1] *Proc. Soc. psych. Res.*, 1943, **47**, 21-150.

Dr. Pratt says he gets the feeling that we did not realize just how unique B.S. was in his ESP performance: we did not, for instance, refer to the study by W. Russell, "Examination of ESP Records for Displacement Effects." As a matter of fact, Mr. Russell's interesting report[2] did not reach us until the draft of our own report was in the hands of the printers, and this alone prevented us from making any reference to his findings. These findings, however, are being discussed in a letter which we hope will appear in the March issue of the *Journal of the Society for Psychical Research.*

It must be borne in mind that only an infinitesimal fraction of the world's population has so far been tested for ESP; and we conjecture that as experiments on a larger scale are conducted, all sorts of surprising modifications of the normal ESP faculty will be encountered. Looking back on the last fifteen years of our own experimentation, we fully realize that B.S. possesses a faculty so distinctive and unique that we might search for another twenty years without discovering his equal. At the same time, the number of subjects with whom any individual experimenter or any existing group of parapsychologists can hope to deal is pitifully small, and many generations will probably pass before we are acquainted with the possible variations that exist in the manifestations of ESP. We certainly had no intention of giving the impression that the "displacement" effect had been a common occurrence in previous card-guessing experiments and that we were simply the first to draw attention to it.

If B.S. is "unique" in comparison with the relatively few genuine ESP subjects reported on in the literature of such experiments, that is surely not our fault but merely our good luck; and it suggests that in the past, experiments have been carried out on far too limited a scale. A great organization of trained experimenters will have to be created in the future so that individual and personal contact can be made with tens and hundreds of thousands of prospective subjects. Only by such means can the various types of ESP functioning be fully explored.

Dr. Pratt says he would have liked to have seen more effort made toward interpretation and discussion, and hopes we will find an early opportunity to go more deeply into the possible psychological significance of our work. Curiously, in a recent very favorable re-

[2] *J. Parapsychol.*, 1943, **7**, 104-17.

view of our experiments in the scientific journal, *Nature,* for March 11, 1944, we were congratulated on our "wisdom in not attempting facile explanations to describe the nature of phenomena the meaning and interpretation of which are likely to elude us for a long time to come." Clearly, therefore, all people do not agree with Dr. Pratt on the advisability of discussing the psychological and philosophical implications of these experiments at their present stage of development. In any case, problems dealing with the nature of psychological time are best left to the professional philosophers and could scarcely be dealt with in any satisfactory way by people who are untrained in philosophy.

Indeed, throughout these experiments our main object has been to ask such questions as could be answered by simple experiment, and then to let the facts speak for themselves. That there are many questions left unanswered there is no dispute, but nevertheless we succeeded in settling quite a number of interesting points. It must be remembered further that, whereas in America positive results in card-guessing experiments have been reported over many years, until the present series, the English experiments carried out at University College, London, and elsewhere registered a total failure. The present occasion was really the first in which success had crowned years of endeavor. Week after week, before witness after witness, we succeeded in demonstrating the *psi*-faculty[3] possessed by our subject. Our witnesses, people of high standing and repute, were allowed to take any active role they chose in the tests. They were permitted to shuffle the cards, to record the code, to assume the part of first or second experimenter, and to make themselves on many occasions entirely responsible for the decoding and entering of results (watched, of course, by ourselves).

Our first and main object was to demonstrate beyond any shadow of doubt that B.S. really possessed the unique precognitive faculty which we claimed for him. Psychological discussion was, at the moment, an entirely secondary matter from our point of view. When Dr. Pratt asks for psychological discussion and interpretation, we feel that he fails to realize just what is the attitude of orthodox science toward these obscure mental phenomena in this country. The

[3] The term "*psi*-faculty" is the term proposed by Professor R. H. Thouless to designate what has been called "ESP" or "extrasensory perception."—Ed.

slightest weakness in experimental method, the least disinclination on the part of the investigators to lay the experiments open to the fullest inspection, would be seized upon as grounds on which they might be discredited. Our critics will not question our *interpretations;* it is our *facts* and *methods* only that they will try to overthrow! No wonder, then, that our first and only concern was to conduct and record our experiments with such scrupulous care that no loophole should remain for adverse criticism.

But even if we had wished to attempt any detailed psychological discussion of the results, this would have been practically impossible owing to the paper shortage. We were originally offered one hundred pages of the *Proceedings of the S. P. R.* for our report and only at the expense of cutting out thirty pages of the Chronicle (detailing the procedure and minute happenings at each sitting) were we able to compress the report into its present one hundred and thirty pages. Every inch of space at our disposal was required to render the report convincing to an outside critic who had had no opportunity for being present at any of the experiments and who perforce had to form his judgment on the record itself.

It may be, again, that Dr. Pratt has not fully visualized the difficulties under which this investigation was carried out. "Remarkable forward strides in pure science" may indeed be made, as Dr. Pratt comments in opening his review, "against the background of national and world crisis," but wartime England is not an easy environment in which to carry out parapsychological researches. In our report we mentioned—and it might be as well to reiterate here—the many difficulties under which we labored. One of us (S.G.S.) was evacuated with his college to Cambridge, and it was possible for him to travel up to London for the experiments at week-ends only. The other experimenter (K.M.G.), in addition to running her household of five or six people, had an arduous full-time job which sometimes necessitated traveling to the north of England for days at a time. Our subject (B.S.) was of a very highly strung temperament, suffered frequently from ill-health, and found the severe air-raids sometimes experienced anything but an encouragement to the tedious and concentrated work the experiments involved. Our agents were busy people who did not find it easy to get to our experiments at the end of a tiring day of war work. Finally, we had to get the desired

witnesses able and willing to brave war-time travel in blitzed England and the night black-outs of London. As the dates in our report indicate, such circumstances caused our modest series of only forty experiments to be spread out over two and one-half years. In view of the fact that positive results were consistently maintained over this long period, we cannot but regret the still more striking and searching results we might have obtained had these experiments taken place in normal times which permitted regular "twice-weekly" sittings.

In explanation of why no experiments have been carried out during the past twelve months, there is one aspect on which we would like to comment. This is the reluctance shown by our remarkable subject to continue working at what now seems to him to be a monotonous, tedious, and pointless repetition. B.S., indeed, does not lean toward the methods of the scientific worker which consist in testing the effects of minute variations in the conditions of the experiment and patiently noting the results. He constantly asks, "What use is it all? What are you trying to achieve?" We can only reply that we ourselves do not know, that we are working in the dark, seeking for a clue. B.S. has what might be called the "artistic" temperament, and the only results he could appreciate would be some discovery that could be put to specific use or would revolutionize human life. To his practical temperament this guessing of the figures on cards for month after month must appear to be exceedingly small fry indeed. In our report we have expressed in no uncertain terms the immense debt we owe Basil Shackleton for the time he has placed at our disposal and which has enabled us to acquire valuable psychological data so far as we have gone; and we greatly hope that he may yet find himself willing to resume experiments which, though unlikely to produce results for which an immediate "use" can be claimed, may nevertheless lead to an advance in pure knowledge the value of which is, at our present stage, almost incalculable.

We will now touch upon another of Dr. Pratt's comments, as to whether precognition is the inescapable conclusion forced upon us by our experiments, or whether they can be accounted for by clairvoyance alone or by a combination of clairvoyance and telepathy. Dr. Pratt quite rightly points out that in the "counters" experiments the counters were extracted from the bowl or bag with alternate

hands, so that, as described by K.M.G. in her notes at the time, she "dipped each hand in [the bag] alternately and showed a counter at the screen aperture with one hand while the other hand was already delving in the bag for the next counter." Now, as Dr. Pratt suggests, there might be a possibility that the counter for the *next* trial had already been grasped by the hand which was stirring in the bag at the time that the subject was recording his guess, in which case it would have been quite possible for the subject to respond to this by clairvoyance. We were aware of such a possibility, but unfortunately it was only by using alternate hands that it was possible for K.M.G. to maintain the required speed and regularity; and even with the use of both hands it was by no means an easy feat to keep the eyes averted from the counters, to act by touch alone, and to present the counters at the small screen aperture with rapidity. But although the hypothesis of clairvoyance might be put forward in the "counters" experiments at "normal" rate, it is no longer tenable, as Dr. Pratt himself points out, when the same "counters" experiments were conducted at "rapid" rate; for at the "rapid" rate the subject is foreseeing TWO counters ahead (that is, the card precognitively guessed by the subject is governed by the "counter-after-next"). It might also be added that when Miss Jephson selected the counters from a bag at "normal" rate on March 14, 1941, she did *not* use alternate hands but drew the counters from the bag with one hand only; yet the results were highly significant on this occasion. Again, in our tests for clairvoyance—whether or not he knew that such tests were being substituted for telepathy—the subject invariably failed to score positively. In these circumstances it is, to say the least, *highly improbable* that B.S. uses clairvoyance in obtaining his "precognitive" results. That telepathy plays an important part seems evident from the fact that B.S. fails with certain agents and succeeds with others. We think it probable, though we do not assert it dogmatically, that "precognitive telepathy" affords the most economical hypothesis that covers all the facts observed.

It is true, as Dr. Pratt remarks, that the role of the agent was only superficially studied. But we had a very good reason for avoiding frequent changes of agent. We were extremely anxious to do nothing that might seriously interrupt the working of the *psi-*

faculty, as were it to peter out prematurely we might never have got it going again. For this reason, even when we did clairvoyance tests or experiments at "slow" rate (at both of which B.S. invariably failed), we were careful to sandwich them in among other tests in which the *psi*-function had unobstructed play. To have worked for any considerable period with unsuitable agents might have had fatal consequences which we were anxious to avoid. On the other hand, it would not have been possible to ascertain whether or no a given person was a good agent without doing a large number of trials spread over many occasions; and in view of the limited time available, we judged it far more profitable to get through as much work as possible by continuing with the two or three persons we knew to be good agents. Any experimenter can obtain negative results at any time, but sustained positive results are as rare as lunar rainbows and are to be guarded like pearls of great price.

LETTERS AND NOTES

Ommen
Sennen Cove
Cornwall
11.v.44.

Sir,

I have very little to say in reply to Dr. Stuart except to thank him very cordially for his friendly review of my paper. His account of my method seems to me a masterpiece of lucidity, while it is a real pleasure to be criticised by someone who has so evidently taken the trouble to read and understand one's work.

As for the criticisms themselves, I think they are for the most part perfectly sound, though I do not agree with every point raised—in particular, I am quite impenitent about the "rays" in the illustrative experiment.[1] But the last thing I want to do is to wrangle about points of detail. Let us at once admit that this item may reasonably be regarded as suspicious, for the sake of argument; the fact remains that the experiment provides good *prima facie* evidence for supposing that something other than chance is at work; and this, rather than an absolute assurance, is what any sensible experimenter wants.

Let us also admit that no catalogue method can ever, in principle, be quite so rigid as a method using cards or drawings cross-scored by a "blind" judge and assessed by a suitable "signs test"; that is to say, that any catalogue may occasionally be misleading, one way or the other—though not, I think, often, or to any great extent if proper precautions be observed. The seriousness of this will depend

[1] It seems to me that when a percipient writes the words "storm, with rain and sun*shine between*" (my italics), he could hardly indicate more clearly that the idea of sunbeams (i.e., "rays") was in his mind—much more clearly, in fact, than if he had merely said "sunshine"; the "between" almost irresistibly suggests that almost searchlight effect of sunbeams passing between masses of cloud which we have all observed. The picture evoked is, indeed, a much better illustration of "rays" than was the original itself.

almost entirely on the purpose for which we are using the method. It would be fatal if we were attempting a formal "proof" of paranormal cognition *de novo*, and in the face of hostile criticism; but the need for doing this is now virtually past. If, on the other hand, we are concerned with gathering information *about* the phenomena, with a view to understanding the kind of process that is going on and to formulating an explanatory theory, etc., then it is worth while to pay the price of risking an occasional false indication, if the method enables us to investigate points which we could not otherwise deal with.

I venture to believe that readers of my paper on the theory of the subject, now in press,[2] will agree that from this point of view the use of drawings handled in this sort of way[3] has enormous advantages as compared with cards or cross-scored drawings.

To take two examples: I find that when several experimenters (as in my Experiment VII) prepare and display originals contemporaneously, percipients score indiscriminately on all of them to a significant extent, showing no preference for the originals used by their *own* experimenter. Again, I find that they may score significantly on originals which have been drawn but not displayed, or on objects which have only been listed in connection with an experiment, though they do not score so heavily on these as on originals drawn and actually used. Points such as these are of great importance in working out a theory of the phenomena; but I think it would be virtually impossible to discover them by the use of cards.

For such purposes, which are extremely important in the general development of the subject, the value and practical validity of the method will be indicated by whether the results obtained fit together into a coherent pattern or merely present a chaotic mass of contradictions. If the first outcome be observed (as I find that it is), we need not worry much if the theoretically imperfect rigidity of the method leads to an occasional anomalous result—it only means that we must study the relevant point again; and this is an occurrence by no means peculiar to parapsychology.

[2] *Proc. Soc. psych. Res.*, Lond., 1944, **47**, Part 168.

[3] Necessarily by means of a catalogue or equivalent, since there are no *a priori* probabilities available; but very preferably, of course, by means of one based on the experimenter's own "local" data. My own catalogue is best used for a preliminary "sighting shot" and to stimulate interest.

To sum up: I should never claim that the drawings-and-catalogue technique is at all suitable for rigidly "proving" the occurrence of paranormal cognition; but, if used with intelligence and discretion, it is an uncommonly good way of studying it; and it has the merit of being repeatable in the sense that it does not depend on specially gifted subjects but can be used by anyone who likes to take the trouble.

(Signed) WHATELY CARINGTON

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. Words defined elsewhere in the glossary are italicized in the text of the definitions.

*For a simple description of those terms marked by an asterisk, as they apply to the ESP test data, see Chapter VIII and the Appendix of *A Handbook for Testing Extra-Sensory Perception* by C. E. Stuart and J. G. Pratt. A mimeographed copy of the relevant pages will be sent on request to subscribers who do not have the book mentioned. Further explanation may be found in any elementary statistical text.

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: *Extrasensory 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 (EXTRASENSORY 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.

EXTRACHANCE: Not due to *chance* alone.

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

GESP (GENERAL EXTRASENSORY PERCEPTION): A technique designed to test the occurrence of *extrasensory 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 paranormal—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: *Extrasensory 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*.