The Journal of Parapsychology

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

Volume 9 JUNE, 1945 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 9 JUNE, 1945 Number 2

EDITORIAL

THE QUESTION OF PRACTICAL APPLICATION OF PARAPSYCHICAL ABILITIES

From the very beginning of parapsychological investigation the question has often been asked, "Can any practical use be made of these parapsychical abilities?" And since the outbreak of the war, it has been asked even more frequently. The answer has invariably had to be, "No." No practical use can be made of them with our present state of knowledge. They are not reliable enough.

We can, of course, go on to say—and, indeed, it should be said—that practical application has never been the objective of the investigations. This is not because practical application is regarded as of no importance, but because the true goals of the research are so incomparably greater in importance that practical applications seem downright trivial in contrast. The search for an understanding of the fundamental nature of man and his place in the universe, the urge to follow these elusive but transcendent parapsychical powers to the end of the trail of causal explanation and to discover a true philosophy by which men can live better and more happily—these purposes must of necessity make the more common needs of practical life seem by contrast unimportant.

But this is not intended to belittle the question of the applicability of parapsychical capacities to practical uses. Rather, it will eventually be necessary to find out what part the abilities under discussion do play in the normal daily life of men and what greater

role can be assigned to them when better understanding and control over their exercise is attained.

Undoubtedly the greatest barrier to the further exercise of parapsychical abilities lies in their being *unconscious* and therefore not subject to ready volitional control. Possibly the use of unconscious or automatic muscular movement may facilitate a more reliable parapsychical response than the usual tests based upon conscious, volitional participation on the part of the subject. But when, for example, we consider the evidence of parapsychical abilities obtained through automatic writing and the ouija board, we are not encouraged to think that merely making the response automatic insures its success in utilizing the ESP capacities.

The high percentage of success claimed for the location of water by the dowser or water-diviner appears to be a different matter. There, if we may take general reports such as that contained in the interesting article by Kenneth Roberts in the *Country Gentleman* of September, 1944, as a basis for judgment, ESP ability is working infallibly. We say "ESP ability" because if the success in locating water described is not attributable to luck or to knowledge otherwise gained, ESP is the only possible explanation.

What is clearly needed is an exhaustive experimental study of dowsing on a scale and with a thoroughness that has not yet been attempted. We need to know the degree of accuracy obtainable under favorable test conditions which allow no possibility of erroneous interpretation. Here, perhaps, is the parapsychological phenomenon which presents the greatest challenge for the practical application of ESP. The testimony and the claims are at least good enough to warrant investigation. The official status given the dowser in some countries and the preliminary scientific investigations that have already been made are sufficient to justify a well-designed and thorough research.

It is not only in the field of dowsing, however, that we come upon testimonials of the role of parapsychical abilities in practical affairs. In the more intimate and confidential statements of business and industrial leaders, prominent political figures, and especially inventors, we find instances of belief in some cognitive power not yet understood. Commonly the descriptive term used is "intui-

tion" or "hunch." Sikorsky, the aviation engineer and designer, boldy asserts his deliberate utilization of what he calls the "mysterious faculty." He devotes a chapter to it in his autobiography[1] written in 1938. In John J. O'Neill's interesting biography of the electrical engineer and inventor, the late Nicola Tesla,[2] which has just appeared, a great deal is made of the role which parapsychical abilities played in his life and work. Although Tesla himself found these capacities somewhat embarrassing to his attempts at a mechanistic philosophy of life, he recognized their existence. We know of others who speak only in private of their recognition of the role of these parapsychical capacities in their professional or other practical affairs, and it is highly probable that in still others similar powers may be exercised without being recognized as such.

Dr. Schiller, the Oxford logician, once spoke of the importance of the practical application of parapsychical abilities in establishing widespread conviction of their reality. But until we can either get these abilities more under conscious control or find an unconscious mode of response that utilizes them successfully, we cannot take the idea of applied parapsychology very seriously.

J. B. R.

[1] Sikorsky, Igor I. *The Story of the Winged-S.* New York: Dodd Mead.
[2] O'Neill, John J. *Prodigal Genius: The Life of Nicola Tesla.* New York: Ives, Washburn.

AN EXPLORATORY EXPERIMENT ON THE EFFECT OF CAFFEINE UPON PERFORMANCE IN PK TESTS

By J. B. Rhine, Betty M. Humphrey, and Richard L. Averill

ABSTRACT: A box containing 96 dice was tipped onto an inclined runway so that the dice poured out upon a padded table while subjects tried to influence them mentally to fall with the six-face turned up. After a preliminary control series four subjects each drank a bottle of Coca-Cola for the caffeine it contained and were tested soon after as to their ability to score on the tests. They did significantly better after the caffeine than before. There is some question whether the effect was due to the physiological action of the drug or to the psychological effect of taking it. The significant differences found, however, leave no doubt of the presence of the PK factor. One of the strongest features of the evidence comes from examining the record sheets to discover how the hits were distributed on the page. The scores were recorded in columns of five entries. Almost all the success in the pre-caffeine period was on the first two trials and very little on the last two. The effect of the caffeine seemed to consist of raising the scoring on the last two trials, counteracting the decline effect.—Ed.

In an earlier report two of the present writers submitted the results of a PK experiment in which two subjects attempted to influence the fall of dice without physical contact before and after taking a strong dose of ethyl alcohol (**1**). The results of the comparison of the two conditions showed in one subject a mild reduction of the score level, and in the other, a striking reduction which brought the score average down from 4.79 (expectation is four hits per run) to a negative score of 3.87 per run. The difference is quite significant. It was pointed out that this lowering of the score is in line with the effect which strong doses of narcotic drugs have upon ESP ability.

The use of stimulant drugs in ESP tests has been reported in two investigations (**2, 6**) to have been favorable to higher scoring, especially in overcoming conditions such as fatigue and the effect of narcotics. It was regarded as likely that a favorable influence upon scoring would be found in the PK experiments as well, provided the pre-drug state of the subjects was one in which a stimulant was more or less clearly needed.

While the experimental routine had been fairly well worked out so far as safeguards were concerned, the all-important matter of maintaining an adequate mental state on the part of the subjects was far from being under control. We were, at the time, and still are, unable to induce at will in our subjects—or for that matter in ourselves—the proper state of mind for producing the best PK test results of which we are capable. As is the case with a great many other mental capacities, a simple request or a simple resolution are not all-determinative in producing it.

At the time the experiments here reported were conducted, we felt that the first steps involving drug treatments should be exploratory, almost casual, and should be introduced into such situations as afforded opportunity. This opportunity arose on May 14, 1936, two days after the alcohol experiment. Again, as in that instance, two of the authors, J.B.R. and R.L.A., together with Mr. A. J. Linzmayer, the Laboratory secretary, met together for PK research as we had been doing in a series of experimental investigations for some time previous. On the morning in question, it was generally agreed that since we did not feel as alert as usual, particularly R.L.A. and J.B.R., perhaps the time had arrived for the test of the effect of caffeine. It was therefore decided that all of us would undergo a preliminary test to ascertain whether our scores would be as low as we anticipated from our feelings; and that following this control series, we would each have a bottle of Coca-Cola, which contained approximately three-fourths of a grain of caffeine. It was understood that if any of us scored well above expectation, he would not enter into the experiment, since such scoring would suggest that no stimulant was needed and there was no reason to anticipate that high scoring could be raised to a still higher level. It turned out, as in the alcohol experiment, that A.J.L. was eliminated (this time by scoring the highest), while R.L.A. and J.B.R. continued as subjects.

Two days later, a similar performance was undertaken by subjects C.D.C. and J.B.R. On this occasion a longer preliminary series was undertaken by both, and the experiment therefore was somewhat more formally balanced than the previous one in respect to the pre- and post-caffeine conditions.

Conditions of the Experiment

In this experiment, as in the alcohol experiment, 96 small white dice, 7/16 of an inch on the edge, were thrown at a time. (The same dice were used in both series.) By pulling a string, the subject released the dice from a box situated about 2½ feet above a specially constructed dice table with a padded surface three feet by six feet in area over which the dice bounced and rolled. The dice were all thrown for the six-face; that is, the subject attempted to influence the dice volitionally in the act of throwing so as to cause as many as possible to fall with the six-face uppermost. The observer, but not the subject, picked up the sixes one at a time, keeping them apart from the other dice until all were removed and both the subject and the experimenter (as well as the third person present in the first day's work) were satisfied that all the sixes had been found. The keenest interest and greatest alertness were exercised in this act of counting, and errors are fairly improbable. The danger of knocking over a die was appreciated fully so that extreme care in picking up the dice was exercised.

When the sixes were counted and the score was agreed on and recorded, the dice were returned to the box, which was put in readiness for being tripped again in the next trial. The release of the dice by means of a string eliminated any manual contact, during the throw, with either the dice or the box which contained them.

The 96 dice used were not "perfect" and were not subjected to a control run to test their degree of imperfection. The objective in the experiment was to make a comparison of two conditions: before and after caffeine administration. If, as was anticipated, important differences were obtained, the pre-caffeine condition would serve as a control on the dice. Other controls are available also. The problem of dice bias is more fully dealt with in the presentation of the results below.

Again, as in the alcohol series, the subjects knew what the drug was to be and knew what effect to expect from it. We were not prepared as yet for the more elaborately planned investigation of the effect of these drugs by the method of disguising the drug and more carefully apportioning the dosage. We were, in fact, prepared to be quite satisfied if we obtained, through the administration of the drug, any significant effect whatever. We were still

at the stage where it was enough to be able to manipulate our hypothetical process, causing it to come on and off at will, however little we might understand the causal processes intervening.

Results

It is of interest, first, to note the total deviation for the runs in this experiment. There are 680 runs of 24 die throws, which mean 170 actual throws of 96 dice each. These gave 2,936 hits on the six-face, which is equivalent to a positive deviation of 216. The SD for 680 runs is ±47.59. A highly significant CR of 4.54 is given by this series as a whole.[1] Such results would occur by chance alone but once in 300,000 such series. The above total includes all of the work done by any of the subjects who had anything to do with the caffeine experiment.

We turn now to a chronological mode of reporting these results to get a clearer picture of what occurred. After the three subjects on the first day's session had each made five throws of 96 dice—that is, the equivalent of 20 runs each—the results stood thus: A.J.L. had obtained a positive deviation of 16 in 20 runs (which is an average of 4.80 where 4.00 is chance), whereas R.L.A. and J.B.R. had obtained average scores per run of 3.95 and 3.30, respectively. It was obvious, then, that it was the latter two who needed the stimulant. R.L.A. and J.B.R. each drank one bottle of Coca-Cola and resumed their experimentation approximately 20 minutes later, alternating as subject in units of five throws each. R.L.A. averaged 4.47 through the next 120 runs (24 throws) while J.B.R. averaged 4.63 through the next 60 runs (12 throws). A.J.L., who had not taken the stimulant, was allowed to continue, and while he scored positively, he did not hold to his high initial level; for, over the next 80 runs, he scored at an average of 4.29, as compared to his initial rate of 4.80. In other words, he fell from the highest to the lowest, whereas, R.L.A. and J.B.R., who took the stimulant, rose to a level well above A.J.L.'s average for the day, which was 4.39.

In the second session, in which C.D.C. and J.B.R. participated,

[1] This estimate is based on the binomial method, which is intended only to be approximate here. But the evaluations upon which the conclusions of the paper rest are supported by application of the arc sine method, as will be appropriately indicated below.

the subjects agreed to do 80 runs (20 throws) each before taking a stimulant. The results of C.D.C.'s 80 runs was an average of 4.13, slightly above expectation. J.B.R. scored very slightly below expectation, giving 3.99. Here again, the effect of the stimulant, one bottle of Coca-Cola, was quite marked in the case of J.B.R., whose score rose to an average of 4.64—very close, as may be noted, to the average (4.63) which he reached in the preceding session. C.D.C., however, after making ten throws (40 runs) and obtaining a score of only 3.7 per run (which was even lower than his pre-stimulant series), commented that he had felt no effect—indeed, he was a large man and might well have required a more than ordinary dosage of any drug—and suggested that he be given another drink. Whether it was due to this second Coca-Cola, to the greater lapse of time which permitted better absorption of the first dose of the stimulant, or, indeed, to a purely psychological effect, cannot be determined; but at any rate, in the next ten throws (40 runs) his average score rose to 4.88. He did ten more throws at a somewhat lower average, which brought his post-caffeine total of 120 runs to an average of 4.32. This, however, compares quite favorably with his pre-caffeine average of 4.13 on 80 runs.

Table 1

Results of PK Tests Before and After Caffeine

Time	Subject A.J.L.		Subject R.L.A.		Subject C.D.C.		Subject J.B.R.		Total		CR	CR_d
	Runs	Av.	Runs	Av.	Runs	Av.	Runs	Av.	Runs	Av.		
Before	20	4.80	20	3.95	80	4.13	100	3.85	220	4.05	.37	2.78
After	..	....	120	4.47	120	4.32	140	4.64	380	4.48	5.14	
Total									600	4.32	4.32	

We are now ready for a comparison of the three subjects who participated in the caffeine phase of the experiment, and we include A.J.L.'s preliminary series of 20 runs for completeness. The results are summarized in Table 1 where it may be seen that each subject (except, of course, A.J.L.) showed a marked difference between the before- and after-stages. In fact, when results for the

four subjects are pooled, the pre-caffeine series totals 220 runs with an average of 4.05, while the post-caffeine work gives 380 runs with an average score of 4.48. There is a significant CR of the difference (2.78) between these two groups of data.[2] The odds against its occurrence by chance are well over three hundred to one.

The actual throw scores for the two series are given in the Appendix table.

Position Effects

The growing interest in the effects of position of the trial in the test structure (as, for example, on the record page) led B.M.H. to the examination of the data of the caffeine experiment for evidence of such effects. The score records allowed only the analysis for vertical distribution in the column of five entries, each entry being the score for a throw of 96 dice.

The results of the analysis are most interesting and are reproduced in Table 2. There it will be seen that a very marked vertical

Table 2

VERTICAL DISTRIBUTION OF HITS IN THE COLUMN IN TERMS OF DEVIATION FOR EACH THROW

ORDER OF THROW	PRE-CAFFEINE* (300 runs)	POST-CAFFEINE (380 runs)	TOTAL* (680 runs)
1	+25	+53	+78
2	+38 +63	+56 +109	+94 +172
3	+11	+24	+35
4	−17	+32	+15
5	−24 −41	+18 + 50	− 6 + 9
	CR_d=3.68	CR_d=1.85	CR_d=3.83

*Including all of A.J.L.'s supplementary data.

decline occurred in the pre-caffeine control results (which include *all* of A.J.L.'s records) giving the highly significant CR of the difference (3.68) between the first two and the last two entries of the column. The post-caffeine data likewise show a decline, but it is not nearly so great, the CR of the difference being insignificant (1.85). These differences in the vertical distributions are shown graphically in Figure 1.

[2] By the arc sine method, the CR of the difference would be 2.74.

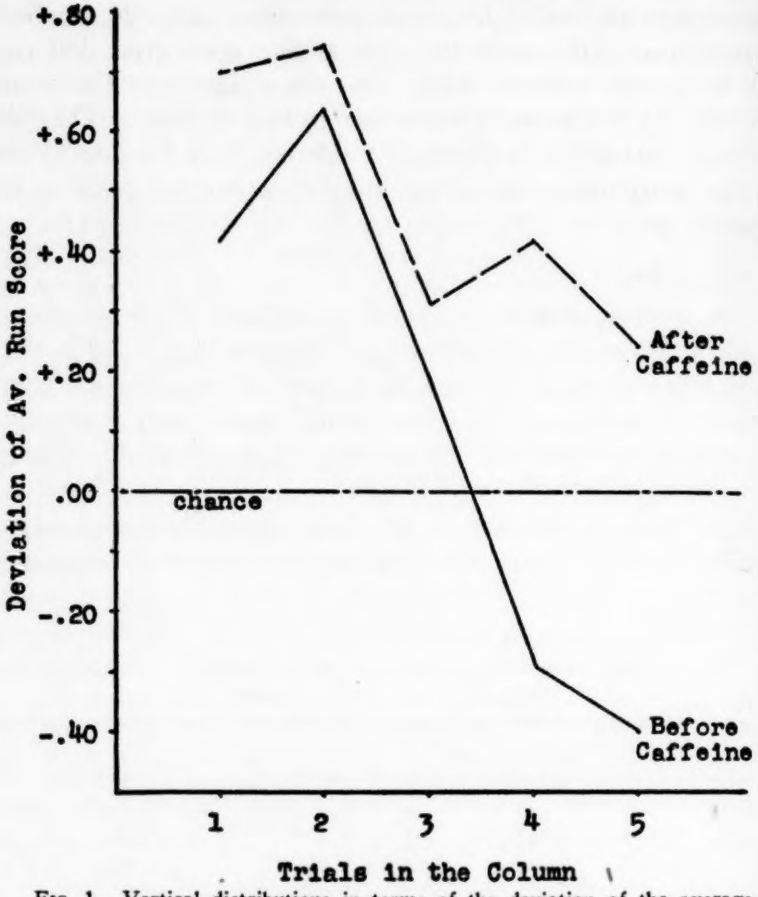

FIG. 1. Vertical distributions in terms of the deviation of the average run score for each of the five trials.

A clue to what is occurring in some of the position effects and what the caffeine is accomplishing may be found in the fact that in the pre-caffeine work the scores dropped well below expectation at the lower end of the five-entry column while in the post-caffeine data they remained well above. The major effect of the caffeine apparently was to counteract the factor producing the decline.

When the vertical distribution is found for the entire body of data (680 runs) which are reported in this paper, the deviations per trial in the column are as follows: +78, +94, +35, +15, −6. This decline has a highly significant difference between the first two

and the last two trials. The CR of the difference is 3.83, and the odds against the chance explanation here are ten thousand to one.

Discussion

The Case for PK

The CR's in this report are such as to render any further discussion of the chance hypothesis unnecessary. Likewise, the semi-mechanical way of throwing the dice makes it possible to dismiss the question of skilled throwing. We may therefore turn to the more outstanding alternative hypotheses and first of all to the ever-present concern of whether the dice are adequately true for the purposes of the experiment. On this question a definite position may be taken, a much more positive one than was possible in the report on the alcohol scores. The scoring after caffeine was significantly better than before. This could not have been due to any hypothetical defects of the dice. If that were not enough evidence for rejecting the hypothesis that the results are due to faulty dice, the even larger difference between the upper and the lower trials in the five-entry column would confirm it. This very significant CR of 3.83 for the total data could not have been produced by faulty dice of any kind. The dice may well be imperfect to a degree, but their imperfections could not under the circumstances of these tests have produced the significant differences described.

Thus there seems to be no reason for hesitating to accept these results as affording another confirmation of the PK hypothesis. The three types of evaluation which have been utilized all give significant CR's. One of the measures, the CR of the difference in the vertical distribution, was not contemplated when the work was done. It constitutes an entirely independent check, not only on the statistical aspect, but also on other features of the experiment as well, for this difference based on the position effects could not reasonably be ascribed to any of the common counterhypotheses such as skilled throwing or biased dice.

Caffeine or Suggestion?

There is always something to be desired, of course, in any exploratory experiment, and there is much indeed left for investigation at the point where the present research leaves off. For ex-

ample, how much of the striking contrast obtained with the caffeine was due to the effect of suggestion and how much to the subject's belief that his scores would be improved by the stimulant; or to the subject's desire to show improvement in order to obtain a significant difference? Sufficient for one research, however, is the fact that an effect *was* introduced—whether physiologically or psychologically—which showed up in the statistical analysis of the PK test results. As was said in the report of the alcohol study, it is something to be able to turn on and off anything which we understand as little as we do the PK phenomenon.

Ninety-Six Dice per Throw

The effect of PK upon a large number of dice released at a time is again demonstrated, and the after-caffeine average of 4.48 on 380 runs, when compared with the pre-caffeine average, gives a difference of .43, one that stands up well in comparison with the results obtained with other numbers per throw. Thus it is further confirmed that the mechanical analogy of mass action does not apply to the PK effect and that the subject, if he wants to influence 96 dice at a time, can do so as effectively as he can a smaller number.

The physicist, when confronted with this success in throwing large numbers of dice, thinks of field effects, like gravitation, which affect large numbers of objects as forcefully as one. *But in PK each effect has to be individual;* each die has to be turned a different amount at different points of the time-space continuum. Rather, what is called for is both a knowledge of the die's position and a current contact of the mind that goes beyond the sensory range and into ESP itself. Moreover, such ESP has to be what Foster (4) called a "diametric function"; that is, ESP which is capable of apprehending a total situation instantaneously rather than perceiving its elements one by one, for often there are many of the 96 dice involved at once in the PK effect operating in a successful throw.

The Vertical Decline

The persistently lawful appearance of these position effects is one of the major phenomena in parapsychology. Their first, more immediate, importance lies in the special high quality of the confirmatory evidence they contribute concerning the existence of PK.

This is something that may well enable parapsychology to "turn the corner" in its progress.

But there is a great deal more in the position effects than evidence of PK. Their value in suggesting relationships between ESP and other mental processes has often been referred to. Here in the caffeine experiment a new line of thought is suggested by the fact that the post-caffeine results showed up mostly as a modification of the decline, especially an elimination of the tail of the curve where it dropped below expectation. Perhaps the drug, in counteracting the decline, is acting somewhat as it normally does in counteracting fatigue. Although Jephson (5) and Estabrooks (3) have suggested that declines may be due to fatigue, we do not see how that is possible in view of the recurring rise and fall of scoring within the smaller units of the record page. Yet there may be something in common between the factor of fatigue and the cause of the decline—something which is counteracted by the stimulant. To continue in this hypothetical vein of thought, we might add that since the effect of a stimulant drug on the decline is like its counteractive effect on fatigue, it may follow that fatigue will accentuate declines. Since alcohol and fatigue affect the subject in much the same way, we should expect that alcohol too would exaggerate declines. Now it happens we already have some data on the effect of alcohol on the vertical distribution in the column. In the paper on the alcohol series we reported that the pre-alcohol (control) series gave a vertical distribution (for a total of 240 runs) of +7, +29, +20, +11, +26, while the post-alcohol series of 200 runs gave +15, +4, −5, +1, −4. Here a ratio of +36 to +37 (first two trials to last two) was changed to one of +19 to −3. It is not a significant change; but at least it fits into the hypothetical picture and should serve to justify much more investigation of the effect of drugs and fatigue as registered, among other measures, in position effects.

There is a minor problem raised by the fact that the pre-alcohol series did not show any vertical decline whereas the pre-caffeine did. Further investigation will be needed to explain this difference with finality, but off-hand it seems plausibly ascribable to the normal difference in attitude on the part of the subject in approaching the two experiments. In the alcohol experiment the subjects to be

selected were those obtaining the *highest* scores while in the caffeine series the subjects were to be those scoring the *lowest*. The suggestion given here—that the subject's motivation is a factor in determining whether a decline occurs—may have some importance in the eventual explanation of decline effects.

Appendix

Scores per Throw of 96 Dice Made in PK Tests Before and After Taking Caffeine

(Chance expectation per throw is 16 hits.)

	Before					After			
	A.J.L.	R.L.A.	J.B.R. First Session	J.B.R. Second Session	C.D.C.	R.L.A.	J.B.R. First Session	J.B.R. Second Session	C.D.C.
	21	12	10	22 19	21 13	21 20 25	17 17	24 16	17 15
	25	18	17	17 16	20 17	15 19 17	24 23	16 20	11 18
	16	20	11	15 19	22 16	13 12 15	15 13	22 19	11 21
	21	16	14	13 16	14 13	19 12 19	17 26	19 16	17 9
	13	13	14	19 11	14 14	10 16 22	19 10	16 21	14 15
				18 17	21 20	17 22	22	21 16	22 20
				14 17	20 22	17 18	18	18 25	15 19
				13 18	16 23	19 15	20	14 18	16 30
				12 16	7 15	18 19	17	15 18	17 18
				15 12	11 11	24 13	20	15 22	17 21
						18			14 13
						27			20 20
						15			17 23
						21			22 17
						18			17* 12
Throws.	5	5	5	20	20	30	15	20	30
Runs...	20	20	20	80	80	120	60	80	120
Dev....	+16	− 1	−14	− 1	+10	+ 56	+38	+51	+38
			Runs = 220 Dev. = + 10 CR = .37				Runs = 380 Dev. = +183 CR = 5.14		

Total pre-caffeine *vs.* caffeine: $CR_d = 2.78$

Total both sections: CR = 4.32

*A second bottle of Coca-Cola.

References

1. Averill, R. L., and Rhine, J. B. The effect of alcohol upon performance in PK tests. *J. Parapsychol.*, 1945, **9**, 32-41.

2. CLARK, C. C., and SHARP, V. The effect of sodium amytal and caffeine upon ESP test performance. [Unpublished MS.]
3. ESTABROOKS, G. H. *A Contribution to Experimental Telepathy.* Bull. 5, Boston Soc. psych. Res., 1927.
4. FOSTER, A. A. Is ESP diametric? *J. Parapsychol.,* 1940, **4,** 325-28.
5. JEPHSON, I. Evidence for clairvoyance in card-guessing. *Proc. Soc. psych. Res.,* Lond., 1929, **38,** 223-68.
6. RHINE, J. B. *Extra-Sensory Perception.* Boston: Bruce Humphries, 1935.

Parapsychology Laboratory
Duke University
Durham, North Carolina

Washington State Penitentiary
Walla Walla, Washington

A CLASSROOM ESP EXPERIMENT WITH THE FREE RESPONSE METHOD

By C. E. Stuart[1]

ABSTRACT: A class of students in experimental psychology were respondents in a GESP experiment. The teacher acted as agent. The agent looked at a series of eight simple stimulus drawings. The respondents made eight responses each. For the first two they were asked to "concentrate"; for the second two, to make an automatic response; for the third two, to free associate; and for the last two, to limit their responses to a single idea. The responses were matched by an independent judge.

The total scores of the matching were not significantly different from chance. But there was significant variation among the responses. The automatic and free association responses showed significant avoidance of the stimulus. The difference in scoring between these and the "limited" responses was also significant.—Ed.

Group tests of ESP have not been considered a fruitful experimental method. The group situation seems to lack the possibilities of good experimenter-subject rapport that is apparently necessary for ESP demonstration. This plausible generality, however, may not apply to an effective teacher-and-class group. Classroom activity is a particularly normal life-situation for the student. He has been conditioned by many years of training to accept the instruction of the teacher. While not all students cooperate willingly and understandingly in projects led by the teacher, and while perhaps no student cooperates all the time, good student-teacher rapport is likely to be the rule rather than the exception in an effective class. This line of speculation would suggest that as far as good adjustment to the task is concerned, the sudent making a response in a familiar classroom at the direction of an accustomed teacher should perform at least as well as he would in a strange laboratory under the guidance of a strange experimenter.

The literature on teacher-conducted ESP tests is remarkably

[1] The research was carried out while the writer held the Thomas Welton Stanford Fellowship in Psychical Research at Stanford University. This report has been approved by the Psychical Research Committee and is Communication No. 13 from the Psychical Research Laboratory.

scarce. Bond, reporting number-guessing by her class of retarded children, found a significant excess of correct guesses (1). Clark and Sharp report group work (3), but do not specify the classroom conditions in such a way as to indicate when the experiment was teacher-conducted. It seems probable that the class group tests were conducted wholly by the assistant. An unpublished note by R. W. George reports successful prediction by his class of whether or not they were to have a brief test on certain days, the test days being randomly determined. Carington elicited the cooperation of class groups for his experiment with free drawings, but his data are not itemized with respect to teacher-directed experiments (2). Even from this scant background it appears that teacher-class experiments have not been wholly unsuccessful.

The experiment reported here was generally an attempt to apply the free response method (4) to the group situation and specifically to demonstrate ESP method in a laboratory course on experimental methods in the study of sensation and perception.

The "free response method" is the term I have used to designate experiments with relatively unlimited stimulus material, to distinguish them from experiments of the card-calling type in which the response is limited to one of a known, fixed number of choices. In previous use of the method one of the working principles was to let the respondent develop his own ways of response with as little instruction as possible. It became evident, however, that while this freedom made for more expressive response by some subjects, others seemed merely frustrated by not knowing what to do. The present experiment permitted some investigation of the effect of instructions.

The instructions suggested four different limitations upon the subject's freedom of response. The first was the attitude of "concentration" upon the agent's situation; the second was that of relaxed and automatic response; the third, that of free association; and the fourth, that of decisive selection of a single idea. All but the fourth represent attitudes conventionally advocated as favoring an ESP response.

The respondents were 18 students in the Psychology 61 class at Stanford University during the winter quarter of 1943. The agent was the teacher of that course, Professor Roland C. Travis.

The experimenter directing the respondents was Mr. Edward L. Walker, the teaching assistant.[2]

Procedure

Instructions

The students were each given mimeographed directions for the experiment. The laboratory assistant discussed briefly the nature of ESP as an experimental problem. He then read over the procedure to make sure that the instructions were clear. The instructions were as follows:

> In this experiment an "agent" (Professor T.) will look at a "stimulus drawing" projected on a screen in a room separate from the respondents. He will look at eight stimulus drawings in succession, looking for four minutes at each one, with a two-minute pause between presentations. The instructor (Mr. W.) will time these presentations and tell you the beginning and end of each. He will also remind you before each response of the following attitudes to be assumed.
>
> You are to make your responses on the half-lettersize sheets provided. Use the carbon to make a duplicate of each response. *Write at the top of each sheet your name and the number of the response. Be careful to work alone. Do not be influenced by your neighbor's responses.*
>
> For the responses, take the following attitudes as well as you can: Responses 1 and 2: Close your eyes for a moment and visualize in imagination the screen at which the agent is looking. "Concentrate" a moment and try to draw it as your response. If your imagination produces nothing, draw anything at all on the response sheet.
>
> Responses 3 and 4: Imagine momentarily the scene of Prof. T. looking at the screen and then draw anything at all on the response card. "Doodle" as carelessly as you like.
>
> Responses 5 and 6: Visualize momentarily the screen, and then sketch quickly any series of things that come to mind. The task here is to get a train of free associations. If you can't sketch your ideas, write them out.
>
> Responses 7 and 8: Visualize momentarily the screen at which the agent is looking; then try to think of some single concrete object or idea to draw, and sketch it. Change your mind if you like, but make the response a single object or idea.

Stimulus Situation

The stimulus drawings were simple line drawings on cards about the size of a bridge card. These were mounted on 8" × 5" cards

[2] I gratefully acknowledge my indebtedness to these colleagues who made the experiment possible, and to Professor E. K. Strong, Jr., who originally suggested the experiment.

for easy insertion in a balopticon projector. The order of presentation was determined by rearranging an already shuffled order according to a pattern derived from a table of random numbers. The cards were placed in an enevelope. The order was unknown to anyone at the time of the experiment.

In their ultimate order of presentation the stimulus drawings were: 1. a CANDLE; 2. a RABBIT; 3. an ARROW; 4. a BOOK; 5. a FLOWER; 6. a mathematical EQUATION; 7. a cartoon figure, KAYO; and 8. a STAR inscribed in a circle (an aircraft insigne). I had carried out all the preparatory steps, so neither the students nor the experimenters knew what the stimuli were to be.

The agent was in a room about 75 feet distant from the respondents' room. Both doors were closed. Signaling was done by a system arranged with the key in the respondents' room and the buzzer in the agent's room. The assistant in the respondents' room timed the presentations.

Upon signal the agent took the first stimulus drawing from the envelope containing the drawings in their randomized order, placed it in the projector, and looked at the image thrown on the screen. At the signal indicating the end of the presentation he took the picture out of the projector, selected the second stimulus drawing, and placed it in the projector in readiness for the second presentation. This procedure continued for the eight presentations.

The agent kept a list of the order of stimuli. This list he retained until the final check-up. After the eight presentations he shuffled the stimulus drawings. The assistant then came into the agent's room and took the randomized stimuli to the respondents' room.

Scoring

The respondents handed in the first copies of their responses to the assistant and retained the carbon copies. It was planned that the latter should be matched to the stimulus drawings and evaluated by the method of Correct Matchings. A misinterpretation of the instructions resulted in a matching task that could not be evaluated.[3]

[3] In the method of Correct Matchings the subject arranges N responses to N stimuli in the way that they seem best to fit. In this case the misinterpretation consisted in requiring the subjects to match each stimulus, as they were exposed one by one in random order, with one of the responses, without permitting any ultimate rearrangement when all the stimuli were exposed. This would result in a valid judgment of similarity for the first pair only.

It was necessary, therefore, to have the responses scored later by an independent judge.

The problem of scoring was complicated by the systematic variation in attitude required in the experiment. For example, Responses 5 and 6 were characteristically different from Responses 7 and 8 for all subjects. An attempt was made to separate the pairs of responses according to the four attitudes and have a number of judges evaluate them separately; that is, all Responses 1 and 2 were matched to Stimuli 1 and 2 by the judge. This method was statistically too crude as a measuring device, mainly because there were only 18 pairs in each category to be matched.

The method finally decided upon was that of scoring each subject's responses by preferential matching in two sets of four responses each. The responses could be grouped into larger sets of four without any evident systematic differences in the responses in the set. The instructions for trials 1 and 2 asked for "concentration," and an attempt to draw what was on the screen. The instructions for trials 7 and 8 asked for a single concrete object or idea. The responses produced under these two sets of directions were virtually the same in character. In a comparable way instructions for trials 3 and 4, calling for "automatic" response, and those for trials 5 and 6, calling for free association, elicited responses that were very similar in character, or at least not so dissimilar that one would infer from their form that they had been produced from different instructions.

The responses were therefore grouped into two sets: Responses 1, 2, 7, and 8, which may be characterized as "limited," and Responses 3, 4, 5, and 6, which may be characterized as "unrestricted." Each subject's responses in each set were coded and the order randomized. The sets were then preferentially matched by Mrs. D. H. Pope of the Parapsychology Laboratory. Mrs. Pope was thoroughly familiar with the method and needed no instructions. She was, however, completely unaware of the details of the experiment. The material was matched in two ways: first, by ranking the stimuli with respect to each response; and, second, by ranking the responses with respect to each stimulus. The preferential matching technique is described in detail elsewhere (**4**, p. 33).

Results

The numerical data of this experiment consist of two ranks assigned by the judge to each response. The ranks (denoted by the numbers 1, 2, 3, and 4) represent the relative degree of similarity of the given response to the stimulus picture corresponding to it. The first assigned rank represents the degree of similarity of the response to the stimulus relative to three other *stimuli*. The second assigned rank represents the degree of similarity of the response to the stimulus relative to three other *responses*. Table 1 displays the assigned ranks in detail.

Table 1

Ranks Assigned to Responses

Subject	Limited Responses				Unrestricted Responses								Limited Responses				Mean Rank
	1st Response		2nd Response		3rd Response		4th Response		5th Response		6th Response		7th Response		8th Response		
	Matching		Matching		Matching		Matching		Matching		Matching		Matching		Matching		
	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	
Bac.....	4	3	4	3	1	2	2	3	2	1	3	2	4	2	2	1	2.438
Cam.....	2	1	3	3	2	4	4	3	3	3	2	4	2	3	2	1	2.625
Edw.....	1	1	4	2	3	3	3	1	2	3	4	4	1	1	3	4	2.500
Fox.....	2	2	3	1	2	1	2	4	4	3	4	4	1	1	2	4	2.500
Gra.....	2	1	1	1	3	2	1	2	2	1	4	4	3	1	1	4	2.063
Gro.....	2	3	2	2	3	3	2	1	3	2	4	4	2	1	3	2	2.438
Han.....	2	1	4	4	2	3	2	2	1	1	4	4	3	1	2	3	2.438
Har.....	1	2	1	1	4	3	1	4	4	3	4	4	4	2	4	4	2.875
Lam.....	2	1	4	4	2	1	2	2	4	3	2	1	1	4	1	1	2.188
Mar.....	2	4	4	2	4	2	2	3	3	4	2	2	3	1	3	2	2.688
Mil.....	1	1	4	2	3	2	4	4	4	4	4	4	3	2	1	2	2.813
Por.....	2	4	3	1	2	4	2	4	4	3	3	2	2	2	4	4	2.875
Sch.....	2	4	4	2	3	3	3	4	4	3	4	2	4	1	4	4	3.188
Sci.....	1	2	1	2	4	4	4	2	2	4	4	4	1	1	2	2	2.500
She.....	1	1	4	4	2	2	2	4	3	4	4	4	1	1	4	3	2.750
Sti.....	1	1	1	2	2	3	1	1	1	2	3	3	2	2	3	1	1.813
Whi.....	4	3	3	2	4	3	2	1	4	1	4	3	1	1	4	1	2.563
Wic.....	1	1	4	4	2	3	2	4	4	4	3	2	2	2	4	4	2.875
Mean Rank...	1.917		2.667		2.667		2.500		2.861		3.306		1.917		2.667		2.563

If no consistent similarity occurs between responses and stimuli, then the mean of a series of ranks assigned in this way should be 2.50. The observed mean ranks for the total experiment for the individual subjects and for the individual stimulus presentations are given in Table 1. In order to keep the statistical issues from obscuring the experimental questions, the evaluative methods are discussed in the Appendix.

The mean rank given all responses was 2.563. This is not significantly different from the chance expectation of 2.500. It is slightly greater than the chance mean and so gives no reason to believe that a greater number of cases might have given better-than-chance results.

The mean rank for most of the subjects was not noticeably different from the chance mean. Subject *Sti* was the best, with a mean of 1.813 ($P = .014$). Subject *Sch* was strikingly negative, with a mean rank of 3.188 ($P = .002$). Neither of these extreme cases is adequate to establish a valid subject difference in performance.

Variability within the experiment is the next point of consideration. The eight responses of the experiment may be grouped with respect to the ways in which we might expect the performance to vary.

1. Modes of Response. A natural grouping is that into which the responses were grouped for matching; namely, into Limited and Unrestricted modes of response. The mean rank of the Limited responses is 2.292; the mean rank of the Unrestricted responses is 2.833. The difference is significant ($P < .001$).

The mean rank of 2.292 for the Limited responses is in the direction we would expect if an ESP factor produced some recognizable similarities between responses and stimuli. The mean is not significantly different from the chance mean of 2.500, however, so that on the basis of this experiment alone we cannot conclude that the limited conditions produce effective ESP responses.

The mean rank of 2.833 for the Unrestricted responses is greater than 2.500, and with tests of significance yielding $P = .011$ and $P = .001$ it is evidently significant. That is, there was a significant absence of readily recognizable similarities between the stimuli and responses of this group.[4]

2. Attitudes. A second natural grouping is that of the four attitude conditions of the experiment. The mean ranks were in order: Responses 1 and 2 (concentration), 2.292; Responses 3 and 4 (automatism), 2.583; Responses 5 and 6 (free association), 3.084; Responses 7 and 8 (single object), 2.292. Of these only the free association mean rank is significantly different from chance,

[4] This phenomenon is analogous to significant negative scoring in card-calling ESP tests.

but that is highly significant ($P = .0001$). Again, the relationship established is that of a significant absence of similarity.

3a. Individual Trials. A third grouping is that with respect to the individual trials. The mean ranks vary from 1.917 for the first and seventh trials to 3.306 for the sixth trial. The trial-to-trial variation is highly significant ($P < .0001$) when tested with the hypothesis that only chance variations in rank occur among trials.

Here, however, it is necessary to consider a possibly spurious source of variation. The stimulus pictures are themselves different in the richness or amount of association they stimulate. There may simply be more things that look like a candle than look like a mathematical equation. The judge himself may have an emotional bias that would lead to favoring one or the other stimulus. In the first matching, when four stimuli are ranked with respect to each response, the favored stimulus would get more 1's and 2's, the unfavored stimulus more 3's and 4's. The second matching, however, wherein four responses are ranked to each stimulus, forces an equal number of first, second, third, and fourth ranks for each stimulus, thus ruling out any preference factor.

The trial-to-trial variation of the ranks assigned in the second matching can be considered separately. It is significantly different from chance, with $P = .006$.

These last two groupings (attitudes and individual trials) have not been independent of the first grouping into Limited and Unrestricted responses. The difference there was so marked that the question arises whether the trial-to-trial variation is simply a result of the trials falling into those two very different groups. The problem is whether there is a significant trial-to-trial variation *within* the Limited and Unrestricted categories. When all ranks are considered, there is a significant variation, with $P = .002$. When attention is restricted to the second matching ranks, the resulting variation is negligibly different from chance ($P = .32$).

The results here are ambiguous. The trial-to-trial variation can be fully accounted for by the observed variation between the Limited and Unrestricted groups, plus the hypothesis of variation due to judge's bias. On the other hand, there is no evidence to show that the Limited and Unrestricted difference is not merely an accidental

result of trial-to-trial variation. The evidence of the next section lends weight to the latter view.

3b. Stimuli. A prevalent question in ESP investigation is the difference in effectiveness of stimulus objects. Might not the difference between the mean ranks of the first trial and the sixth trial reflect the fact that the CANDLE was a better ESP stimulus than the equation? From the structure of this experiment we cannot get an answer. But the stimuli used in the experiment had been used in Series S1 (7) in which there was no variation of attitudes. A comparison of the mean ranks of the various stimuli gives the order of "effectiveness" displayed in Table 2.

Table 2

EFFECTIVENESS OF STIMULI

STIMULI	SERIES S2		SERIES S1*		
	N	Mean Ranks Second Matchings	N	Respondents' Mean Ranks	Ind. Judge's Mean Ranks
Kayo	18	1.61	13	2.00	2.38
Candle	18	2.00	15	2.40	2.33
Rabbit	18	2.33	17	2.65	2.18
Star	18	2.61	12	2.42	2.50
Arrow	18	2.67	12	2.75	2.58
Flower	18	2.72	12	2.75	2.75
Book	18	2.72	14	2.79	2.79
Equation	18	3.17	12	3.00	2.08

*In Series S1 the second matching of the responses was made by the respondents and by an independent judge (not the independent judge of this series). From the standpoint of objectivity of the relationships used to justify the matching, the independent judge's ranks would seem to be the more valid. But in this case internal patterns in the judge's matrices suggested that the responses were matched in many cases as a group instead of independently. Since this tendency was apparently unconscious and since the judge is by instruction free to use any basis for matching, I have no real grounds for questioning the validity of these rankings. But since I observed the respondents' matchings myself and thus know that they were done conventionally, I am inclined to accept them as the more valid of the two sets of matchings.

The similarity of order between the results of this series and the respondents' ranks of Series S1 is apparent. Rank correlations of the effectiveness in order in this experiment (Series S2) to the two measures of order in Series S1 are +.96 and +.23, respectively. The correlation between this series and respondents' rankings in Series S1 is significant even when devaluated as one of two possibilities ($P < .01$).

The "best" stimuli, KAYO and CANDLE, were the two in which shading or background was used in the drawing to give an impres-

sion of a third dimension. The "worst" stimuli were BOOK and EQUATION. The BOOK was drawn in pencil and obscured by lines drawn across the picture in such a way that it was momentarily difficult to identify. The EQUATION was chosen for its obscurity of meaning. (It was $x^{2/3} + y^{2/3} = a^{2/3}$.) Of 24 subjects in Series S1, none was able to identify it otherwise than as a mathematical expression.

4. Position Effects. Because of the changes of instruction during the experiment the effects of position of the trial in the series of trials cannot be observed separately. The mean rank for the first half of the experiment is 2.441. The mean rank for the second half is 2.688. The difference is in the direction of decline-of-performance hypotheses but is not significant. There are two trials with each instruction, however, and performance in the first and second trials may be compared. The mean rank of first trials (Responses 1, 3, 5, and 7) is 2.340. The mean rank of second trials (Responses 2, 4, 6, and 8) is 2.785. The difference is significant ($P = .004$).

DISCUSSION

The first question to be asked concerning this experiment is whether ESP has been demonstrated. The usual expectation of a total mean rank significantly smaller than a chance rank of 2.50 has not been observed. However, significant variation between the mean ranks of two attitude categories demonstrates that chance does not account for these data.

Certain customary counterhypotheses are readily excluded. Sensory cues were unlikely in the experimental situation, which provided closed rooms on opposite sides of a building. No person in contact with the stimulus pictures was ever in the respondents' room before the response period. Signaling was from respondents to agent. The original responses were turned over to the experimenter before the stimuli were brought into the room. The final evaluation was carried out by an independent judge ignorant of the conditions or personnel of the experiment. The important fact that dissimilarities between stimuli and responses constitute a major factor in the significant results further invalidates the customary counterhypotheses.

The methods of statistical evaluation, if not as precise as those applicable when "hits" are counted, are conventional for populations of unknown character. The statistical assessment only verifies a score difference obvious to inspection.

There remain two notable areas of vulnerability. The first is the hypothesis of a general psychological seriality of response. Suppose, under the instructions, students naturally drew more light sources or elongated objects at a first trial, avoided animal references on a second trial, avoided mathematical associations on a sixth trial, and favored human associations on a seventh trial. There is no complete refutation of this view possible. The only argument against it is a certain logical unlikelihood: one would expect seriality of this sort to operate most strikingly at the beginning of a series, rather than near the end.[5] Yet it will be noted that the most deviant responses are the sixth and seventh. Experimental control of the hypothesis would require repetitions of the experiment, with all factors held constant except the stimulus order, which would be permuted systematically.

The second area under question is the objectivity of the ranking by the judge. To what extent do these rankings reflect variation in objective stimulus-response relations? Might not another judge rank the same material so as to give significant variation in an opposite direction? Except for one possibility these differences would not concern the ESP case greatly, since the significant difference would be sound evidence that *some* undefined but objective relations must support it. But suppose ESP were exercised, not by the subjects but by the judge. The observed differences would then be attributable wholly to the vagaries of the judge's ability to rationalize as perceptual judgments what were really ESP judgments. Appeal may be made again to reasonable likelihood. The judge tried to make objective judgments. The variations in the rankings coincide well with reasonable rationalizations of stimulus-response relations. It is easy to find many reasonable similarities in responses to the CANDLE and KAYO, and difficult to rationalize the responses to the BOOK and the EQUATION.

[5] Analysis of patterning in a two-alternative choice, which is the simplest kind, reveals that the statistical *effect* after the third trial is practically negligible. This fact is evident in Goodfellow's study of the Zenith Radio Telepathy Experiment. (*J. Exp. Psychol.*, 1938, **23**, 601-32.)

To establish the objectivity of the relations scored, it would be necessary to have repeated ratings of the material by different judges. Under present manpower conditions it is not easy to find or train competent judges. But when it becomes possible to undertake a thoroughgoing study of the most effective criteria of judgment (actually an investigation of what constitutes definitive ESP relations), the rescoring of this and other experiments will be part of that project.

Thus, although this purely statistical case presents a number of unresolved ambiguities, the likelihood that the variation in the results reflects in some way variation in ESP performance by the subjects is great enough to justify the method as a fruitful approach to the general problem of the ESP response.

The objective of this paper has been to report an observation. Attempts to explain the variations at this point would be relatively unsupported speculation. But after reports of further experiments, I hope to be able to take up the interesting lines of hypothesis suggested.

Appendix on Statistical Evaluation

The mean chance expectation of the preferential matching method is 2.5 for the null hypothesis. The variance is more difficult to establish theoretically. In the original application two hypotheses were proposed: the hypothesis of independence which supposed that the judge matched each response independently according to directions; and the hypothesis of dependence, which supposed that the judge arranged the matchings in a pattern (4). These had a variance per matching of 1.250 and 1.500, respectively. In practice neither has been *generally* satisfactory in that observed variances seemed to fall between these values.

Instead of attempting to establish a theoretical population variance, I have here used the observed variance to get a "best estimate" of the population variance.

The chance hypotheses presented for comparison are based upon the hypothesis of independence. Where scores are the sum of both first and second matchings of the same material, the standard deviation is corrected for a correlation of +.40, the observed correlation between the first and second matchings. (The consistency of this

correlation is indicated by the fact that for the Limited responses the correlation was +.39, for the Unrestricted responses the correlation was +.41.)

Probabilities cited are those of "Student's" t distribution when the number of items is less than 30. For N greater than 30 the regular probability integral tables were used.

Trial-to-trial variation was evaluated by a chi-square method. Ranks of 1 and 2 were grouped together as low ranks and ranks of 3 and 4 were grouped together as high ranks. The null hypothesis was that the proportion of high and low ranks was constant and equal to the observed total proportion.

Table 3

STATISTICS OF DISTRIBUTIONS OF THE RANKS GIVEN IN TABLE 1

A. Subject is the unit and score is sum of ranks of all responses in category.

	N	MEAN SCORE	SD	MEAN RANK	CR	2P
Limited	18	18.333	3.400	2.292	2.02	.060
Unrestricted	18	22.667	3.829	2.833	2.87	.011
(Chance)	..	(20.000)	(3.742)	(2.500)	....	....
Diff. U-L	18	+ 4.333	.....	+.541	3.49	.001

B. Response is unit and score is sum of two ranks given that response.

	N	MEAN SCORE	SD	MEAN RANK	CR	2P
Limited	72	4.583	1.927	2.292	1.82	.070
Unrestricted	72	5.667	1.716	2.833	3.27	.001
(Chance)	..	(5.000)	(1.871)	(2.500)	....	....
Diff. U-L	72	+1.083	.....	+.541	3.54	.001

C. Response is unit and score is rank given that response on the second matching only.

	N	MEAN SCORE	SD	MEAN RANK	CR	2P
Limited	72	2.139	1.146	2.139	2.65	.008
Unrestricted	72	2.819	1.072	2.819	2.51	.012
(Chance)	..	(2.500)	(1.118)	(2.500)	....	....
Diff. U-L	72	+ .680	.....	+.680	3.65	.001

In Table 3 the ranks of the two matching groups, Limited and Unrestricted, as given in Table 1, are summarized in three ways. The first gives the statistics of total scores per subject for the two matching groups. The total score is the sum of all ranks given responses in the Limited or Unrestricted categories. The second summary is with respect to the responses as a unit, the score being the sum of the two ranks given the response. If the judge's rankings are intercorrelated, the validity of the CR's of this second distribution may be questioned on the grounds of violation of the assumption of simple sampling. However, the observed SD's are negligibly different from the binomial SD; so the assumption of independence is warranted. The third distribution is that of the second matchings alone. As noted in the discussion of trial-to-trial variation, it is probable that the second matchings have a greater validity than the first matchings when the individual responses are the unit.

The probabilities listed are doubled to correct for the fact that the direction of deviation from chance has not been predicted by any hypothesis.

References

1. Bond, E. M. General extra-sensory perception with a group of fourth and fifth grade retarded children. *J. Parapsychol.*, 1937, **1**, 114-22.
2. Carington, W. Experiments on the paranormal cognition of drawings. *Proc. Amer. Soc. psych. Res.*, 1944, **24**, 1-107.
3. Sharp, V., and Clark, C. C. Group tests for extra-sensory perception. *J. Parapsychol.*, 1937, **1**, 123-42.
4. Stuart, C. E. An ESP test with drawings. *J. Parapsychol.*, 1942, **6**, 20-43.
5. ———. The free response method in ESP testing. [Unpublished MS.]

Parapsychology Laboratory
Duke University
Durham, North Carolina

MINOR ARTICLES AND NOTES

[Under this heading will occasionally appear briefer publications having value and interest for our readers but being in some respects less complete or less conclusive than our major articles.—*Ed.*]

EARLY PK TESTS: SEVENS AND LOW-DICE SERIES

By J. B. RHINE

ABSTRACT: A pair of dice was thrown for combinations of seven in an experiment in which the dice were rolled by hand in one section and by machine in another. The scores are clearly not explainable by chance. A good case, though not the best one possible, is offered against such counterexplanations as biased dice or skilled throwing. The main finding, however, is a typical QD (quarter distribution) of the hits on the record page. Readers of the reports on the QD studies will recognize this typical QD as having the highest scoring in the upper left quarter and the lowest in the lower right.

In another series in which the pair of dice was thrown for low dice, the results were below chance, but examination showed that the sevens combinations for the same series were significantly above chance. This appeared to the author to suggest a displacement effect, somewhat like that of hitting the wrong target. Such displacement has been found in ESP work, and ESP and PK are known to be closely associated. Two similar cases of displacement toward sevens in low-dice tests are discussed.—Ed.

WHEN my wife (Dr. Louisa E. Rhine) and I reported our first PK tests (8), we described only the main experimental work known as the High-Dice Series. In that series, the subject threw a pair of dice and attempted to influence them to fall with the uppermost faces totaling eight or above. This High-Dice Series yielded significantly high scoring which, under the circumstances, was regarded as evidence of a process known as psychokinesis.

In this earlier experiment two subordinate series of tests were conducted as controls on certain aspects of the High-Dice Series to ascertain whether the hypothetical PK effect might be exerted upon low-dice and sevens combinations as well as upon high dice. They are known as the Sevens and Low-Dice Series. Since these series were designed as controls for the High-Dice Series, they are not complete experiments in themselves. It seems best, therefore, to

take the two short series for what they are worth, to state the circumstances under which they occurred, and to regard the conclusions as tentative.

It will clarify the appraisal if it be stated in advance that the results are not offered primarily as adequate evidence for the establishment of the PK hypothesis. In view of the advanced state of the PK research today such evidence is available now in quantity in the literature on the subject. Rather, it is in the interesting side-effects which have emerged in the course of the analyses that the contributions of the report are to be found. To the student of parapsychology, sometimes these incidental observations are fruitful products of the research.

I. The Sevens Series

The Sevens Series consists of two sections, the first carried out in 1934 and 1935, and the second in 1936. The first section was supervised mainly by myself; the second was carried out principally by J. L. Woodruff, who was my research assistant at the time.[1]

Strictly speaking, it is only the first section (91.3 runs, each comprising 12 throws of a pair of dice) that makes up the control connected with the original High-Dice Series. The second section of 116 runs was done later and under sufficiently different circumstances to require separate treatment. The two sections together comprise all the tests for sevens which have been conducted.

Procedures

The first section of the Sevens Series was the work of seven subjects. They used the ordinary medium-sized commercial dice, made of plastic materials, which had been used throughout the High-Dice Series. There was not, however, any adequate effort made to equalize the number of throws for sevens and high dice. Therefore no claim can be made that the same dice were used to an equal extent in both series, nor can it even be assumed that all of the dice used in the High-Dice Series were duly represented in the Sevens. The informal, exploratory character of the testing at that time and the absence of a developed and standardized test accounts for this state of affairs.

The recording was not made in a uniform manner in the earliest

[1] Dr. Woodruff is now serving in the United States Army.

runs, though it came to be so for the main part of the work. For the most part, the run was recorded in 12 entries to the column, each entry representing the reading of the pair of dice.

The dice were hand-thrown, shaken between cupped hands and bounced either upon a blanketed surface or against an upright backstop. (A comparative series on "mechanically released" throwing of the dice was conducted at the time in connection with the High-Dice Series referred to above, but this pertained only to the High-Dice Series. It is of interest to remember, however, that it gave higher score averages than the hand throwing.) Two dice were thrown at a time throughout the series, the purpose being to turn up faces totaling seven.

The second section differed principally from the first in the fact that by 1936 the completely mechanical throwing of the dice had been introduced. The electrically driven machine described in my report on the comparison of cup and machine throwing (7) was used for all the dice of this section. The records in this section were taken on standard forms prepared for the purpose.

The work reported here, like that of the High-Dice Series, was rechecked in 1942, and the analyses to be described were carried out at that time. The rechecking and analyses were done under the direction of Miss Betty M. Humphrey.

Results

In the entire Sevens Series, both sections, there was a total of 481 hits in the course of the 207.3 runs. This is 66.33 hits above the number expected on the chance hypothesis. The standard deviation is ±18.58 and the CR is 3.57. The likelihood of this occurring by chance alone is one in five thousand and is quite significant. The average score for the entire series is 2.32 hits per run as against an expectancy of 2.00.

Table 1 gives the results for the two sections of the Sevens Series as well as those for the two sections pooled. It will be seen that the work of the first section is independently significant, with a CR of 3.60, while that of the 1936 section, although positive and suggestive, is not.

On the strength of the findings just presented in the table, the chance hypothesis may be set aside and the questions of other alter-

Table 1

RESULTS OF THE SEVENS SERIES BY SECTIONS

YEARS	RUNS	AVERAGE SCORE	TOTAL HITS	TOTAL DEVIATION	SD	CR
1934-35.....	91.3	2.49	227	+44.33	±12.33	3.60
1936.....	116	2.19	254	+22	±13.89	1.58
Total.....	207.3	2.32	481	+66.33	±18.58	3.57

natives to PK considered. There are, however, no results which deal with the other alternatives as simply and conclusively as the above-mentioned CR of 3.57 deals with the chance hypothesis.

But the fact is that for the principal contribution which this Sevens Series has to make, it is not important that we be concerned with the usual counterhypotheses of imperfect dice and skilled throwing. Even the exclusion of the chance hypothesis is a secondary, though certainly a favorable, circumstance. We do, however, regard it as worth mentioning that the dice used in the Sevens Series could hardly have been biased in favor of sevens, for *when the High-Dice Series itself, which used these dice, was checked for the occurrence of sevens incidentally found therein, there were fewer sevens than the number expected by chance.* In fact, there were 34 sevens below expectation in the 543 runs. Even the 60 runs of the High-Dice Series that were conducted in the same sessions in which 21 runs for sevens were made gave a negative deviation of 6. The further fact that the 21 runs made for sevens gave a positive deviation of 16 and a CR of 2.71 adds importance to the comparison. There is thus a fairly good case against the hypothesis of faulty dice in the evidence that the dice did not produce sevens when the target was high dice.

But it is the results of the analyses for position effects which are of special interest in the light of the repeated finding of such effects in PK research data. So important is the evidence of PK contributed by the position effects that even the less striking findings of the Sevens Series are of value in that they contribute to the general case.

The hit distributions of the Sevens Series must be presented

separately for the two sections. The first section, with its fewer runs, was not recorded throughout in a sufficiently uniform fashion to permit some of the analyses that are now customary. Accordingly, there is less to be derived by way of distribution analyses from the first section than from the second. The vertical distribution analysis, however, was possible, and it yielded a vertical decline—that is, a drop in scoring in the run from the first half to the second—in the 71 runs of the first section on which the analysis could be made. This gives a drop from a total deviation of +24 on the first half of the run to one of +11 on the second.

The 81 runs of the first section on which a three-run or horizontal distribution analysis could be made show the following averages for the first three runs in the sequence or experimental session: 2.36, 2.42, and 2.21. The fourth and other runs averaged 2.69. There is a decline, then, only between the second and the third runs.

It is in the second section that the distribution pattern which has now become familiar in the PK literature was found to be clearly marked. Here a quarter distribution was obtained for the record page, and it was found that the scoring from the beginning of the record page falls away to the right and downward at the same time. The total vertical decline in the column is from a positive deviation of 17 for the top half to one of 5 for the bottom, while the horizontal decline is from a deviation of +24 on the left half of the page to a −2 on the right. However, the number of runs varies in the left-right comparisons and the deviations are not strictly comparable. Converted to average scores for comparison they give 2.32 on the left and 1.95 on the right where 2.00 is expected by chance. The quarter distribution (QD) for the page summarizing these trends and showing the diagonal decline from the upper left to the lower right quarters of the page is shown in the square below. Each quarter gives the deviation of the average score (the amount above or below 2.00). The upper left is the highest in score average and the lower right the lowest, dropping below expectation from chance.

The difference between the highest and lowest quarters is not a significant one (CR of the difference = 1.60), but it is suggestive, and the QD is in line with the general trend of QD's as they have

been found in previously published researches. It is a *typical* QD of the page.

+.46	.00
+.19	-.10

Discussion

It is especially interesting to find a more strikingly typical QD in the second or machine-thrown section of the Sevens Series. This is not a new thing, for in the Gibson Machine Series (3), in the Cup-Machine Series (7) earlier reported, and in the Small-Medium-Dice Series presented in this number (5), an equally typical effect was obtained. There is something impressive about the fact that with all conditions mechanically controlled, these QD patterns still emerge. What could an electrical machine alone do to a pair of dice to produce such a QD pattern?

There is a point of psychological interest in the fact that for PK to work on a combination such as sevens it must achieve a more definite calculation than in the case of any other target—that is, some computation involving the two dice while they are on the roll. The subject's "action" is a "diametric" response dealing with the two dice as a unit. He cannot wait until one die lands and then turn upon the other to compute which face is needed to complete the seven. Rather, an appreciation of both the dice is needed, and this could only be through ESP and the view of ESP formulated by Foster in his "diametric function" hypothesis (2).

Even in throwing a single die for a single face as the target the supposition of ESP is required if intelligent direction of the rolling die is to be accounted for. With combinations involving a running addition of the faces of two dice the case is seemingly complicated to a much greater degree. The computation required for the purposive direction of these two moving bodies so as to produce the specified combination of *sevens* heaps further bafflement upon the mystery we already experience in dealing with the simplest instance of PK or ESP.

II. Low-Dice Series

Results

The Low-Dice Series was carried out in very much the same general way as the first section of the Sevens Series. There is no complete uniformity of recording, and, accordingly, analyses for position effects are possible only in a portion of the runs. The importance of this series, however, rests on quite a different basis than does the Sevens Series. The 104 runs that were made, all of them in 1934, do not give statistically significant results. The deviation of 24 was negative. For the 89 runs on which it was possible to work out a top-bottom or vertical distribution, the deviations are +4.5 for the upper half and −13.5 for the lower half of the column.

Discussion

The interest in the Low-Dice Series stems from the fact that in these data the total number of hits for high dice as well as for low dice was below expectation. This series was begun primarily as a control for the High-Dice Series. The subjects were aware of the fact that if high-dice combinations were obtained to an excessive degree, it would indicate that the dice naturally favored the higher combinations and it would render the High-Dice Series valueless as evidence of the PK hypothesis. They were therefore motivated *to avoid high-dice combinations.* They did not, however, succeed in getting low-dice combinations. While the low-dice total was, as stated, 24 below expectation, the high-dice combinations themselves were 18 below. This means, of necessity, that the only other possible combinations, the sevens, were being favored to the extent of the sum of 24 and 18, or 42. *This is a significant deviation of sevens,* with a CR of 3.19, which would not be expected from chance alone but once in 1,000 such series. Yet sevens were not the objective or goal of the subjects participating in the throwing (except in the case of seven runs in which sevens and low dice were both targets).

In attempting to explain this significant deviation in sevens—unsought as they were—the easiest solution, if it were a possible one, would be to conclude that the dice were simply loaded for sevens; but the fact is that the same dice, when thrown in the

High-Dice Series, had given a *negative* deviation on the sevens combinations. And even though we cannot regard this as a perfectly adequate control, since we cannot affirm that the dice were equally represented in both series, it carries a great deal of weight. What is closer to the point is the fact that at the same session in which the 104 runs for low dice were thrown, there were 102 runs for high dice made with the same dice and under the same conditions, and these gave 72 high-dice hits above expectation. This deviation is independently significant since it would occur by chance but once in 100,000 such series.

Thus, when the dice were thrown for high dice, the high-dice scoring was significantly positive and the sevens below "chance." When they were thrown for low dice (and high dice were not wanted), significant scoring in sevens resulted and the high-dice (and low-dice) hits were below expectation from chance. We cannot easily suppose that loading is the explanation of these results, for loading does not shift from sevens to high dice, and the dice could not have been loaded for both or midway between and be expected to give results as significant as those presented here.

Rather, this appears to be a case of a curious psychological effect, an avoidance of less desirable targets. For some reason, after throwing for high dice for a long period, the subjects, who were all veterans by the time the Low-Dice Series was begun, found it difficult to reverse; and, in avoiding high dice, they "displaced" in favor of sevens instead of obtaining the smaller combinations of low dice.

This will recall to the general student of parapsychology the displacement effect reported by Carington (1) and later by Soal (9) and Thouless (10) and will add another probable link between the ESP and PK phenomena to the considerable number which have already appeared in the analyses for position effects.

Here, as may be the case in the ESP displacement referred to, we have to do with a "bad aim" effect, a hitting of the *wrong* target rather than a complete missing as indicated by a mere chance score; and there is, at least in the case of the PK displacement, a plausible, rational hypothesis which might be put as follows: High dice were to be rejected by the subject and low dice were to be aimed at; but long previous throwing for high dice and rejecting

of low dice left a residual antagonism which worked unconsciously against low dice. PK is effective on the unconscious level anyhow, and under conflict over unconscious rejection of low dice and conscious rejection of high dice, the subject unconsciously favored sevens, a neutral outlet.

But what is needed more than an explanatory hypothesis for this case of the displacement-toward-sevens effect is more evidence from comparable situations. It appears that we have some data from a nearly identical situation in the Reeves work on high-dice and low-dice tests (**6**). When Mrs. Reeves threw for high dice, she obtained a deviation of +290 in 492 runs. But she also obtained 26 sevens above expectation. When she threw for low dice, in her 435 runs, she obtained a positive deviation of +89 and a deviation of sevens which was *nearly twice that obtained in the High-Dice Series.* The actual figure is +49. With the standard deviation of ±26.91, this rise from 26 sevens above expectation in the High-Dice Series to +49 in the Low-Dice Series is a very suggestive increase. It does not, it is true, prove the point, but it is confirmatory in its bearing. For here again the turning to low-dice throwing, while yielding more low-dice hits than in the tests reported here, also shows a big increase of sevens over the deviation of sevens given in the high-dice tests. (The CR of the 49 sevens occurring with low-dice tests is 1.82.)

There is still another related item. The Hilton and Baer work (**4**) was based on high-dice tests. On the supposition that this hypothetical displacement would be most marked when the subject was avoiding high dice, I decided to examine a section of the Hilton and Baer Series done by subject W.S., who had gone consistently and significantly negative on his high-dice tests. In his 20 runs his deviation for high dice was −22, with a CR of 2.88. He was somehow avoiding high dice. Would he, too, not be expected to show a disproportionate shift to sevens? In his 20 runs there were 16 sevens above expectation. The CR is 2.77. This is significant, showing that he had displaced to sevens to a degree that cannot be ascribed to chance. (There were six low-dice hits above expectation, which is insignificant.) This goes far toward confirming the displacement hypothesis in PK performance.

Needless to say, further tests are in order before too much is made of the point. The present report on the Low-Dice Series with its significant displacement toward sevens serves a purpose by introducing the question into the literature of parapsychology. In the present stage of our explanations, every possible link we find between our unknowns helps toward the eventual making of a rational scientific picture. Displacement in PK linked with displacement in ESP may be only a play on words, or again it may represent a common property of the two phenomena.

References

1. Carington, W. Experiments with paranormal cognition of drawings. *J. Parapsychol.*, 1940, **4,** 1-129.
2. Foster, A. A. Is ESP diametric? *J. Parapsychol.*, 1940, **4,** 325-28.
3. Gibson, E. P.; Gibson, L. H.; and Rhine, J. B. The PK effect: mechanical throwing of three dice. *J. Parapsychol.*, 1944, **8,** 95-109.
4. Hilton, Jr., H.; Baer, G.; and Rhine, J. B. A comparison of three sizes of dice in PK tests. *J. Parapsychol.*, 1943, **7,** 172-90.
5. Humphrey, B. M., and Rhine, J. B. PK tests with two sizes of dice mechanically thrown. [In this issue.]
6. Reeves, M. P., and Rhine, J. B. The PK effect: II. A study in declines. *J. Parapsychol.*, 1943, **7,** 76-93.
7. Rhine, J. B. Dice thrown by cup and machine in PK tests. *J. Parapsychol.*, 1943, **7,** 207-17.
8. Rhine, L. E., and Rhine, J. B. The psychokinetic effect: I. The first experiment. *J. Parapsychol.*, 1943, **7,** 20-43.
9. Soal, S. G. Fresh light on card guessing—some new effects. *Proc. Soc. psych. Res.*, Lond., 1940, **46,** 152-98.
10. Thouless, R. H. Experiments on paranormal guessing. *Brit. J. Psychol.*, 1942, **32,** 15-27.

Parapsychology Laboratory
Duke University
Durham, North Carolina

AN EXPLORATORY CORRELATION STUDY OF PERSONALITY MEASURES AND ESP SCORES

By Betty M. Humphrey

ABSTRACT: Ratings on six personality traits for the ESP subjects in three different experiments were correlated with their ESP scores. No significant relation was found, but positive indications were given that in general are in line with such impressions as experimenters through the years have reported concerning the personality make-up of good subjects.—Ed.

In a previous paper (1) the possible relationship between ESP ability and intelligence was discussed in the light of the results of five correlation studies. There it was shown that the ESP scores and the intelligence test ratings of the subjects in two ESP series were significantly correlated. On the basis of these results it was concluded that a positive relation between ESP and intelligence was suggested though not established.

In an exploratory study carried on incidentally in connection with the intelligence-ESP research, I was able to get certain personality test ratings for subjects in three ESP series. Since the number of cases involved in this study is small, no conclusions can be reached on the basis of the material at hand; but the trends found may serve as background data in planning an inquiry on the important problem of whether the personality characteristics of the high-scoring ESP subjects differ from those of the low- or chance-scoring subjects.

Before a study of personality trait ratings and ESP test scores can properly be made, it is desirable to have adequate assurance that ESP has been functioning in the series involved. The three series in the present study have all been shown in previous reports (2, 3, 4) to be significant in some measure. The measure of significance is not the same for the three series nor were they carried out under comparable procedures and conditions. In a more thoroughgoing study it would be desirable to have strictly comparable data from all series involved, but in view of the exploratory nature of the present research, it seemed worth while to take advantage of every

opportunity to gather relevant material. This opportunity presented itself in connection with the Earlham College Series I, the Humphrey-Pratt Chutes Series, and the Humphrey-Pratt Precognition (PDT) Series, each of which will be described in more detail below.

The first Earlham College series (3) was conducted under the GESP procedure with the subject attempting to identify the card looked at by one of the experimenters. The total of 1,690 runs in this series gave a significantly positive deviation which would be expected by chance only once in 5,000 times. In addition to this, further evidence of the extrachance nature of the data was found in the chronological distribution of hits. The scoring level of the first half of the experiment was significantly different from that of the second half ($CR_d = 2.75$).

The next series on which it was possible to get subjects' personality test ratings was the Chutes Series (4). In this experiment the tests were made by a matching procedure in which cards enclosed in sealed opaque envelopes were dropped through chutes. The deviation given in the 490 runs of this series was significantly negative, having a CR of 3.73 and a probability of .0001.

The Humphrey-Pratt Precognition Series (4) was one in which subjects recorded their calls for ten runs of ESP cards before the decks were shuffled. The deviation given in these tests was positive but insignificant (CR = .95). In subsequent analyses of the data for position effects a quarter distribution (QD) of hits on the record page was worked out. This revealed that most hits had occurred in the third (or upper right) quarter of the record page, while the smallest number of hits were made in the second (or lower left) quarter. The CR of the difference between these two quarters was 2.96. Since this same pattern of hit distribution had been found in the two earlier Earlham series conducted by the same experimenter, this CR does not need to be corrected for selection, for this trend is just what one would have predicted on the basis of the other ESP series. The measure of ESP evidence which will be correlated with personality test ratings in this series, therefore, will be the CR of the difference between the second and third quarters of the page for the work of each subject.

The personality test given to the subjects in each of the three series was Bernreuter's Personality Inventory. When the items in

this questionnaire are scored on the six different scales, measures of personality adjustment are found for the following traits: neurotic tendency, self-sufficiency, introversion-extroversion, dominance- submission, self-confidence, and sociability. Experimentation with the inventory has shown that these personality traits are not independent of each other; for example, correlation studies indicate that the scale for neurotic tendency is measuring the same trait as is that for introversion-extroversion. For practical purposes one may consider *either* the scales for neurotic tendency, self-sufficiency, dominance-submission, *or* the scales for self-confidence and sociability. The two latter scales are generally considered to give purer measures of the traits involved in the first three scales. In the subsequent presentation of the results of the correlations of ESP scores with the ratings on the personality scales, coefficients for all six scales will be given; one should keep in mind, however, that these figures are not independent.

Results

For the Earlham Series I and the Chutes Series, the *average run score* for each subject was correlated with his percentile rank on the various personality scales. For the Precognition Series, the *CR of the difference between the scores for the second and third quarters of the record page* was taken as the measure of ESP for each subject. If the difference followed the pattern given by the whole series (and also found in the Earlham series)—that is, if

Table 1

Correlations of the ESP Scores and the Personality Test Ratings for Subjects in the Three ESP Series

Trait	Earlham College Series I		Chutes Series		Precognition (PDT) Series	
	N	*r*	N	*r*	N	*r*
Neurotic tendency	14	−.14	22	−.21	19	−.38
Self-sufficiency	14	−.32	22	+.35	19	+.38
Introversion	14	−.30	22	−.17	19	−.35
Dominance	14	+.14	22	+.46	19	+.36
Self-consciousness	14	−.17	22	−.35	19	−.50
Nongregariousness	14	−.34	22	+.20	18	+.06

the third quarter was higher than the second—the CR was given a positive sign. If, however, the reverse was true, the CR was given a negative sign.

The correlation coefficients, together with the number of cases involved in each, are presented in Table 1. (All coefficients were obtained by Pearson's product-moment method.) Although some of the coefficients are fairly large, the standard deviations for such small numbers of cases are also large; thus none of the correlations are significant. Probabilities in the suggestive range between .05 and .01 are associated with two of the coefficients: that for the Chutes averages and dominance ratings ($r = +.46$), and that for the Precognition CR's of the difference and self-consciousness ratings ($r = -.50$).

Although the coefficients are not significant, we may find trends which suggest that certain personality traits may be more frequently associated with high-scoring ESP subjects than with low-scoring subjects. Any such trends would be more obvious, perhaps, if these coefficients were translated into words. This is done in the following table by means of which the areas of agreement and

A Positive ESP Series	A Positive ESP Series	A Negative ESP Series
In Earlham College Series I evidence of ESP (+deviation) was given mainly by subjects who were:	In PDT Series evidence of ESP ($CR_{d(2-3)}$) was given mainly by subjects who were:	In Chutes Series evidence of ESP (−deviation) was given mainly by subjects who were:
r	*r*	*r*
Stable.....(.14)	Stable.....(.38)	Unstable.....(.21)
Dependent.....(.32)	Independent.....(.38)	Dependent.....(.35)
Extroverted.....(.30)	Extroverted.....(.35)	Introverted.....(.17)
Dominant.....(.14)	Dominant.....(.36)	Submissive.....(.46)
Self-confident....(.17)	Self-confident.....(.50)	Self-conscious.....(.35)
Sociable.....(.34)	.....	Sociable.....(.20)

disagreement can be appraised more quickly. From this arrangement we can see that evidence of positive ESP scoring[1] was given

[1] In regard to the Precognition Series, the reader may well wonder whether the factors producing a positive total deviation are similar to those producing a decline from the third to the second quarter of the record page and whether it is therefore appropriate to compare the Earlham Series I and the Precognition Series since different measures have been used in evaluating the two. In this connection the following facts concerning the Precognition Series may be helpful: 1. The average run score and the $CR_{d(2-3)}$ for each subject correlate positively ($r = .39$). 2. The average run score for each subject and his various per-

mainly by subjects who were stable, extroverted, dominant, and self-confident, while the negative deviation of the Chutes Series was given mainly by subjects who had the opposite traits: instability, introversion, submissiveness, and self-consciousness. The measures for dependence-independence and for sociability show no consistency of trend in the present study.

Discussion

If the large negative deviation in the Chutes Series is regarded as the opposite of a large positive deviation in the other two (positive) series, four out of the six coefficients are compatible for the three series. Positive deviations were given mainly by stable, extroverted, dominant, and self-confident subjects in the positive series; negative deviations were given mainly by unstable, introverted, submissive, and self-conscious subjects in the negative Chutes Series. Stated in this manner, the findings from the three series are preponderantly in agreement.

On the other hand, we know from other series that high-scoring subjects may at will score negatively and that negative deviations may be produced by unconscious frustration or by adoption of a consistently misleading manner of choosing symbols ("poor aim"). In discussions of these negative deviations it has been assumed that some knowledge of the correct target must be possessed by the subject in order to enable him consistently to reject the correct symbol. If he has no knowledge as a basis for these rejections, he could be expected to score only at chance. If the ability to score both positively and negatively is normally possessed in equal degree

---

sonality trait ratings give coefficients very similar to those given when the $CR_{d(2-3)}$ is used as the measure of ESP. Both sets of coefficients are given below for comparative purposes:

Trait	Correlation with Av. Run Score	Correlation with $CR_{d(2-3)}$
Neurotic tendency	—.32	—.38
Self-sufficiency	+.17	+.38
Introversion	—.33	—.35
Dominance	+.36	+.36
Self-consciousness	—.31	—.50
Nongregariousness	+.13	+.06

From these facts it seems safe to conclude that in this case at least it makes no difference whether we use the CR of the difference or the average run score for correlation purposes.

by the same individual, then our coefficients showing opposite trends for positively and negatively scoring subjects are inconsistent. But it may well be that the person who is unconsciously thrown into rejection of the symbols when he is aiming for high scores has quite a different personality profile from the individual who can score high or low at will. It is quite possible that the submissive, self-conscious, introverted individual is easily thrown into an unsuccessful manner of calling by the demands of the test situation. We do not have sufficient data at hand at present to solve this complex problem, but it promises to be a fruitful line of inquiry.

One of the strongest suggestions from the present study is that self-confidence is associated with high scoring in ESP tests. Now self-confidence, as measured by the test given, refers to characteristic attitudes to situations in general; whereas self-confidence, as mentioned in the descriptions of good ESP subjects, usually refers to the subject's attitude only toward the ESP test situation. In spite of this difference it is interesting to note that there are a number of instances in which it has been recorded that the high-scoring ESP subjects have had a great deal of confidence in their own ability.

Soal mentions that his able subject, B.S., came to take the tests "not to win the prize, but because he felt confident he could accept the challenge" (**9,** p. 157). "He had come, he said, 'not to be tested,' but 'to demonstrate to us the reality of telepathy.' . . . His manner was extremely assured and confident . . ." (**9,** p. 183-4).

Similarly, several of the outstanding subjects of Rhine had a great deal of confidence in their ability. Hubert Pearce, for example, in his first meeting with Dr. Rhine, said frankly that he possessed extrasensory capacities but that he was "afraid of them." The medium, Mrs. Garrett, believes confidently that she has extrasensory powers. Several of the other high-scoring subjects reported in the early Duke work presented themselves for testing because they believed they possessed ESP.

Martin and Stribic stated that their outstanding subject, C.J., at the University of Colorado "presented himself as a subject, eager to try the tests, and confident that he would be able to score above chance. This confidence seemed to spring from several 'notable coincidences' in his own experience and from his trust in his mother's reliable 'intuitions' " (**5,** p. 173).

A series of experiments by Schmeidler (**6, 7, 8**) seem also to confirm the importance of confidence for high-scoring. Dr. Schmeidler separated subjects into two groups: those who believed they could succeed and those who expected to fail in ESP tests. Throughout the five series published to date, those who believed they could succeed have scored positively, while those who expected to fail have scored negatively. The results of the "believers" are independently significant and the difference between the scores of the two groups is highly significant.

There is little comparable material in the literature in regard to the suggestion given in the results of the present paper that dominance may be associated with high scoring. Soal and Goldney, however, do mention that B.S. was "assertive" (**10,** p. 64), which is probably akin to the trait measured by the Bernreuter scale.

It is difficult to discuss the matter of the stability of good ESP subjects. Experimenters may be hampered by not being able to put in print that such-and-such a good subject was "neurotic." Thus we do not have any body of literature with which to compare the suggestion given in the present study—that high scores are given mainly by stable individuals, those low in "neurotic tendency."

It is well to keep in mind the possibility that in regard to the exceptional subjects cited above we may be dealing with a different personality problem from that represented by the contributors to the three series dealt with in this paper. We must remember that here we do not have the high order of scoring obtained by the exceptional subjects cited in the literature. For example, Riess's subject (who had a "nervous breakdown") average approximately 18 hits per run when 5 is expectation, while the highest of the subjects in the present paper averaged only 5.95 hits per run.

In conclusion: The correlation coefficients given here suggest certain trends which may be worth following up in more extensive studies. They also suggest that a more intensive attack on the problem of the personality of the negatively scoring subject may give some clue to the nature of "negative ESP."

References

1. Humphrey, B. M. ESP and intelligence. *J. Parapsychol.*, 1945, **9,** 7-16.

2. ———. Further position effects in the Earlham College series. *J. Parapsychol.*, 1945, **9**, 26-31.
3. ———. Patterns of success in an ESP experiment. *J. Parapsychol.*, 1943, **7**, 5-19.
4. HUMPHREY, B. M., and PRATT, J. G. A comparison of five ESP procedures. *J. Parapsychol.*, 1941, **5**, 267-92.
5. MARTIN, D. R., and STRIBIC, F. P. Studies in extra-sensory perception. III. A review of all University of Colorado experiments. *J. Parapsychol.*, 1940, **4**, 159-248.
6. SCHMEIDLER, G. R. Predicting good and bad scores in a clairvoyance experiment: a preliminary report. *J. Amer. Soc. psych. Res.*, 1943, **37**, 103-10.
7. ———. Predicting good and bad scores in a clairvoyance experiment: a final report. *J. Amer. Soc. psych. Res.*, 1943, **37**, 210-21.
8. ———. Separating the sheep from the goats. *J. Amer. Soc. psych. Res.*, 1945, **39**, 47-49.
9. SOAL, S. G. Fresh light on card guessing—some new effects. *Proc. Soc. psych. Res.*, Lond., 1940, **46**, 152-98.
10. SOAL, S. G., and GOLDNEY, K. M. Experiments in precognitive telepathy. *Proc. Soc. psych. Res.*, Lond., 1943, **47**, 21-150.

Parapsychology Laboratory
Duke University
Durham, North Carolina

PK TESTS WITH TWO SIZES OF DICE MECHANICALLY THROWN

By Betty M. Humphrey and J. B. Rhine

ABSTRACT: Two small dice were thrown mechanically in PK tests with subjects who were trying to influence the dice by direct mental action. For comparison, a medium-sized pair was thrown in similar fashion. The tests with both sizes of dice gave significantly high scores, with very little difference between the two. Some evidence is offered to prove that the results were not due to biased dice. The most important finding, beyond that of the comparison of the scores for the two sizes, lies in the position effects for the distribution of hits on the record page. A typical QD (quarter distribution of hits) on the page was found.—Ed.

When, in the fall of 1936, there was introduced into the PK investigation at Duke University the use of an electrically driven, rotating cage for the throwing of the dice in PK tests, a special experiment was undertaken for the purpose of comparing the effects of PK upon two different sizes of dice. The apparatus used for this experiment has already been described in a report on the comparison of cup and machine throwing in PK tests (3).

There had been comparisons of different sizes of dice in earlier investigations, the most extensive being those of the two Hilton series carried out in 1934 (1, 2). The results of these earlier comparisons did not conform to what would be expected if mechanical principles alone controlled the dice. In fact, it was indicated that something other than physical law applied to the rates of scoring obtained from the different sizes of dice. But there was need for more investigation of this problem under the conditions of mechanical throwing afforded by the apparatus referred to. With this dice-throwing machine, the impulse given to the dice was that of gravitation, and there was no possibility that preferences or attitudes on the part of the subject might introduce differences in manner of throwing.

Description of the Experiment

The experiment was carried out in one day, October 31, 1936, in the Parapsychology Laboratory at Duke University by J. L. Woodruff.[1] A summarizing report of the records was turned over to J.B.R. at the close of the day.

The data consist of the work of four subjects: Woodruff, J.B.R., A. J. Linzmayer (then secretary of the Laboratory), and Mrs. Linzmayer.

Sixty-three runs were made with two different sizes of dice, a run consisting of 24 die readings, or 12 throws of a pair. Common commercial dice were used, those of medium size being approximately 11/16 of an inch square and the smaller size, 7/16 of an inch. Both pairs were of white plastic material and were similar in all respects but size. The dice were thrown in the rotating cage at a speed which was regulated to suit the subject.

The observer recorded the hits after the dice had fallen to the lower end of the cage as it reached the vertical position. Thus a reading was made for every half-revolution of the cage. The recording was done on standard mimeographed record sheets such as that shown in our report on the quarter distributions of the page in the March, 1944, issue of the JOURNAL (**4**, p. 24). The uppermost faces of the two dice were read and recorded separately, one underneath the other, in the column on the record page. Thus 12 throws utilized 24 spaces in the column. Most of the runs were grouped into sets of four columns each. Frequently two sets would be recorded on the same record sheet but with space between to indicate a break. This is important in considering the analysis of the results for position effects.

An observer was present to read the dice and make the records during the tests, with the exception of 40 runs out of the total of 126. Woodruff was the observer except when he was the subject, and then either A. J. Linzmayer or J.B.R. took his place. In the 40 unwitnessed runs, J.B.R. was the subject and made his own records. No distinction is made between these and the other runs: first, because the total results are significant without them, and also because, as will be clear later on in the report, the principal con-

[1] Dr. Woodruff was Research Assistant to J.B.R. at the time. He is now serving in the United States Army.

tributions resulting from the special analyses could in no way have been affected by the absence of a witness.

All the subjects preferred the six-face as the target in the tests, and accordingly all subjects attempted throughout the experiment to influence the dice to fall with the six-face turned up as often as possible. No record of preferences as to size of dice was taken, but the larger size was preferred except in the case of Mrs. Linzmayer. The fact that the dice were not handled probably tended to reduce the effect of preference since it is mainly (though not entirely) in the picking up and handling of the dice that preferences become emphasized.

The analyses for distribution effects which have become a routine procedure in the handling of the data from PK experiments were carried out in 1943 by B.M.H. and will be reported under the heading of "Results."

Results

Total Hits

The total of 126 runs represents 1,512 throws of the pair of dice, or, altogether, 3,024 die readings. From these an expectation of four hits per run would give 504 hits on a theory of chance. There were actually 573 hits or 69 above expectation. This gives an average score of 4.55.

The total positive deviation of 69, when measured by the standard deviation of ±20.48, gives a CR of 3.37. The probability of obtaining such a CR is approximately .0004. In view of these results there can be no doubt that the experiment gave evidence of the operation of factors other than chance in the fall of the dice.

Table 1

Comparison of Results for Two Sizes of Dice

Size	Runs	Average Score	Total Hits	Deviation	SD	CR	P
Small.....	63	4.54	286	+34	±14.49	2.35	.01
Medium....	63	4.56	287	+35	±14.49	2.42	.008
Total.....	126	4.55	573	+69	±20.48	3.37	.0004

Comparison of the Two Sizes of Dice

As we stated above, the experiment consisted of 63 runs for each size of dice. These gave approximately the same average score, and in both cases the CR's are significant. (See Table 1.) The probabilities of the two CR's are .01 and .008.

One thing, then, is clear; namely, that whatever the factors controlling the dice, they made no appreciable distinction on account of size.

Results of the Analyses for Position Effects

The question that remains to be cleared up before conclusions can be drawn may be stated as follows: Granted that the results are not due to chance, how can we distinguish among such remaining possibilities as skilled manipulation, imperfect dice, and the PK hypothesis? We can reject the hypothesis of skilled throwing outright, since the mechanical handling of the dice allowed no such possibility. This places the issue between faulty dice on the one hand and the PK hypothesis on the other. The results, as given above, do not permit a distinction between these hypotheses. It is important, then, to introduce the analyses for position effects before attempting to draw any conclusions. These analyses in many earlier PK reports have provided internal evidence that has ruled out the hypothesis of faulty dice. This has been so generally true that the analysis for position effects might almost be regarded as in itself a method of evaluation.

Vertical Distribution. The vertical distribution of hits in the record column in this series of data represents also the vertical distribution on the record page, since the column extended the full length of the page. Table 2 shows the vertical distribution of hits on the page as a whole for the small and medium dice separately and also for the two sections of data pooled. The figures are given for four segments of the column, each segment representing six entries. The distribution shows a decline from the top half to the bottom half of the column in both sections of the data. The differences are not significant, but they are considerable, the deviation of the first segment being nearly twice that of the fourth. The two upper segments have almost twice the deviation of the two lower.

Table 2

VERTICAL DISTRIBUTION OF HITS IN TERMS OF TOTAL DEVIATION FOR THE SEGMENT OF THE COLUMN

SEGMENT OF THE COLUMN	SMALL DICE (53 runs)*	MEDIUM DICE (53 runs)*	TOTAL (106 runs)*
1	+16	+ 3	+19
2	+ 5	+12	+17
3	+ 3	+ 8	+11
4	+10	0	+10
Total	+34	+23	+57

*Ten runs with each size of dice are omitted from this analysis because they were not recorded in the standard manner.

Horizontal Distribution. The distribution of hits for the left and right sets on the record page also shows a decline. In this instance the comparison has to be stated in terms of the deviation of the average scores because the numbers of runs in the two halves of the page are unequal. The results of the left-right comparison of the record page are given below:

	Left Set	Right Set
Small Dice	+.74	+.29
Medium Dice	+.64	+.44
Total	+.69	+.36

Here it will be seen that the score averages declined to the right for both the small and medium dice. The decline in the total is represented by average score deviations of +.69 to +.36. This, again, is not significant, but it represents a noteworthy ratio.

Thus far we have been speaking about the page as a whole. An interesting effect is observed if we look into the horizontal distribution of the set. As stated above, most of the sets were made up of four columns each. These sets are pooled regardless of position on the page, and the score total is obtained for each run in the set.[2] The deviations from expectation for the columns of the set are as follows:

	First Column	Second Column	Third Column	Fourth Column
Small Dice (40 runs)	+14	0	+ 9	+7
Medium Dice (40 runs)	+ 9	−3	+ 4	−1
Total	+23	−3	+13	+6

[2] Some of the sets had more than four columns, but only the first four of each were included in this analysis.

This distribution is peculiar, both sizes of dice showing a "double decline."

The Quarter Distribution. In the light of the vertical and horizontal declines described above, the QD (quarter distribution) of the page would be likely to show a decline from the upper left-hand quarter to the lower right. Such a diagonal decline does

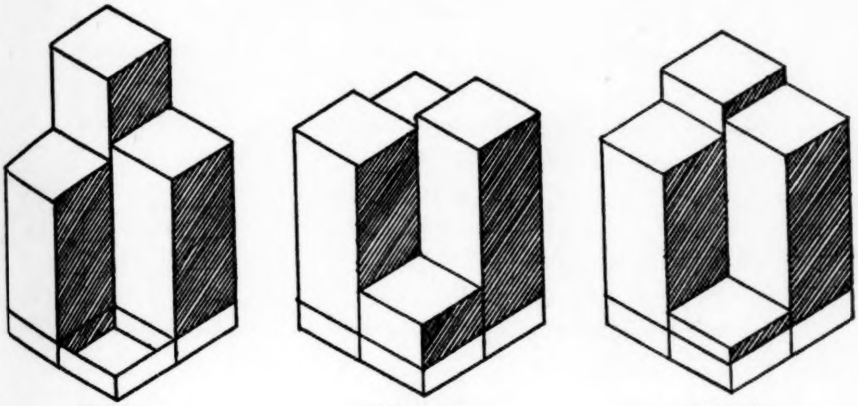

FIG. 1. Quarter distribution of hits on the page for the small and medium dice separately and for their pooled total.

appear in the results of the QD analyses of both the small and the medium dice. When the quarter distribution of the total data is found, it turns out to be a typical QD as judged by the results obtained in general QD studies (4); that is, the upper left quarter has the highest scoring level and the lower right the lowest. Even so, the difference between these two quarters is not significant, having a CR of 1.55. (See Fig. 1 and Section A of Table 3.)

To a certain degree the QD's of the smaller subdivisions of the record page show the trend to diagonal decline given by the QD of the page as a whole. The QD of the set, while not typical in all respects, shows a drop from a deviation of +19 for the first quarter to +10 for the fourth. (See Section B of Table 3.)

DISCUSSION

As we stated in introducing the analyses for position effects, the explanation of the results lies between the PK effect and the

Table 3

QUARTER DISTRIBUTIONS OF PAGE AND SET

(Within each quarter of the QD's are given the number of runs represented in the quarter, the deviation from mean chance expectation, and the average run score.)

A. QD's of the page

Small		Medium		Total	
17.5	14	18	13.5	35.5	27.5
+16	+9	+11	+10	+27	+19
4.91	4.64	4.61	4.74	4.76	4.69
17.5	14	18	13.5	35.5	27.5
+10	-1	+12	+2	+22	+1
4.57	3.93	4.67	4.15	4.62	4.04

$CR_{d(1-4)} = 1.55$

B. QD's of the set

Small		Medium		Total	
15	15	15	15	30	30
+14	+11	+5	+12	+19	+23
4.93	4.73	4.33	4.80	4.63	4.77
15	15	15	15	30	30
+6	+2	+6	+8	+12	+10
4.40	4.13	4.40	4.53	4.40	4.33

hypothesis of faulty dice. In order to take account of the question of faulty dice, the experimental plan called for the running of a control series under the same conditions with each pair of dice used; that is, a series in which all faces of the dice were to be recorded by the observers, without mental preference as far as this was possible. This series was undertaken by Woodruff with the assistance of Dr. Louisa E. Rhine but was interrupted before completion. The results obtained for the six-face in this interrupted control series are as follows:

	Runs	Deviation
Small Dice	30.4	−2.0
Medium Dice	11.6	−0.4

As far as they go, these figures show no favoring of the six-face.

Further evidence against the biased-dice hypothesis is provided by the data on position effects. While these effects are not in any single case so marked as to give significant differences, the general trend of the declines and their consistency with typical decline patterns previously reported gives the entire distribution of the data a lawful appearance. The vertical and horizontal declines as well

as the diagonal decline show typical patterns. The lower right quarter in the QD of the page for the total results so clearly approximates expectation with its average 4.04 that it is not easy to suppose that the dice are biased. Even the interesting though peculiar double decline effect shown in the horizontal distribution of the set adds to the argument against the biased-dice hypothesis, for this decline appears in both sections of the data, that for the small and that for the medium dice.

Unlike most of the PK reports, however, this is not a paper in which the data leave no room for argument. Recognizedly, the evaluations presented do not with finality rule out the question of imperfections of the dice as a positive factor. They only render it relatively unlikely as an explanation. For this reason it is not concluded here that this body of data furnishes sufficient proof for an independent establishment of the PK hypothesis. Nevertheless, it has a distinctive value. In view of the advanced stage of the PK experimentation, it is now possible to credit the PK hypothesis with much more a priori likelihood than formerly as against the hypothesis of faulty dice. And we do have here a reasonably good comparison of two sizes of dice, mechanically thrown and consequently free from any possible differentiation in manner of throwing due to preferences for size. The very close similiarity of average score for the two sizes of dice makes this experiment therefore of importance in considering explanations of the PK effect. If two sizes of dice with an approximate ratio of one to four in total weight do not make any difference in the rate of scoring in the PK tests, it is at once indicated that a set of governing principles is manifested which is different from the mechanics with which the movement of bodies has in the past been explained. The implication of this finding for the general understanding of the nature of mental activity is, of course, profound.

Another point which justifies the present report is the fact that it contributes to the cumulative evidence of position effects, particularly to that patterning of distributions summarized by the QD. A typical QD of the page has been encountered in the series total (Fig. 1). The addition of this contribution to the general evidence for QD's does not in any way depend upon the settlement of the question of whether the dice were or were not perfect. The very

superiority of the evidence for the PK effect which, it has been pointed out in earlier papers (**4, 5**), derives from these QD findings makes of every recurrence of the typical QD effect a contribution of special importance at this stage of the PK research.

References

1. Hilton, Jr., H.; Baer, G.; and Rhine, J. B. A comparison of three sizes of dice in PK tests. *J. Parapsychol.*, 1943, **7**, 172-90.
2. Hilton, Jr., H., and Rhine, J. B. A second comparison of three sizes of dice in PK tests. *J. Parapsychol.*, 1943, **7**, 191-206.
3. Rhine, J. B. Dice thrown by cup and machine in PK tests. *J. Parapsychol.*, 1943, **7**, 207-17.
4. Rhine, J. B., and Humphrey, B. M. The PK effect: special evidence from hit patterns. I. Quarter distributions of the page. *J. Parapsychol.*, 1944, **8**, 18-60.
5. ———. The PK effect: special evidence from hit patterns. II. Quarter distributions of the set. *J. Parapsychol.*, 1944, **8**, 254-71.

Parapsychology Laboratory
Duke University
Durham, North Carolina

FALLACIES IN A CRITICISM OF ESP ASSESSMENT[1]

By Donald J. West

In a book entitled *Beware Familiar Spirits* by John Mulholland (published 1938, Charles Scribner and Sons, New York and London), there occurs a criticism of the statistical basis of Rhine's ESP experiments, which seems to have escaped notice and to have remained unanswered.

In collaboration with Professor Pitkin, Mulholland begins his attack on theoretical grounds (p. 221). The main argument seems to be that, since runs of successes of any size may be found anywhere in an infinitely large chance series of trials, no run of success which is observed in a limited number of trials is incompatible with the hypothesis that the observed trials form part of an infinitely large chance series. Theoretically, this proposition is undoubtedly true, but it can be shown by sampling statistics with what frequencies varying degrees of success will be expected to turn up in finite random samples, and some of these frequencies are so small that they can be safely neglected for all practical purposes. Thus, if an experimenter assumes that all experiments yielding a .01 level of significance are not the result of chance, he will be right 99 times out of a hundred, which is good enough for practical purposes. The same argument could be applied to every scientific experiment, for every observation, whether evaluated statistically or not, could, in the last resort, be due to coincidence.

In an endeavour to obtain some experimental support for their argument, Mr. Mulholland and Professor Pitkin commissioned the International Business Machines Corporation to produce a random sequence of 200,000 cards bearing the numbers 1 to 5 in equal proportions. These were divided into two groups of 100,000 and paired off so as to produce a series of 100,000 "mechanical E.S.P."

[1] The Pitkin and Mulholland experiment evaluated here did not escape notice, but it has not previously been commented on in this Journal. Mr. West has skillfully reconstructed what was done and has made possible a more complete statistical evaluation. This note appeared in the privately circulated *Journal of the Society for Psychical Research* for October, 1944. We reprint it here with the permission of the editors of that *Journal*.—Ed.

trials. As Mr. Mulholland says, "Just as with Dr. Rhine's test there was one chance in five of the pairs of digits in any given line being the same—that is matching. But with our test there was no possible chance of mind reading or clairvoyance as a factor."

Statisticians say that it is exceedingly difficult to produce a pure random sequence mechanically, especially if the method entails the shuffling of cards, so that we have some reason to distrust the reliability of Mr. Mulholland's shuffle, no details of which are revealed. However, we can accept provisionally that the series really represents 100,000 chance trials, and proceed to examine Mr. Mulholland's figures.

Here again we are confronted with the difficulty of insufficient information. Instead of the raw figures being presented in the usual form, all that is given is a series of incomplete statements, from which the reader has to deduce what the observations really were. The statements may be dealt with one by one.

(1) p. 225. "There were as many as 32 lines of figures in sequence without one matching pair." This statement means that a run of 32 failures has been found somewhere in the series. The expectation of runs containing r failures in sequence in a series of N trials, where the probability of an individual failure is p, is given by the formula $E = Np^r(1-p)^2$. Substituting, we find that the expectation of runs of 32 failures in the present series is 100,000 $(4/5)^{32}(1/5)^2 = 3.2$. It is surprising, therefore, that Mr. Mulholland should be surprised to find one such run in his series.

(2). "Runs of 5 matching pairs in sequence fell 25% below theoretical frequency, while runs of 6 rose to 25% above theoretical frequency. Runs of seven jumped still higher to 59% above chance expectancy, and with runs of eight we went to 780% above theoretical frequency." Now it is possible to calculate, from the formula already given, the expectations of runs of successes of different sizes and to deduce from the given percentages what the actual deviations were:—

Size of Run	Expected Frequency	Observed Frequency
5	20.5	(20.5 × .75) = 15.4
6	4.1	(4.1 ×1.75) = 5.1
7	.82 } 5.08	(.82×1.59) = 1.3 } 7.8
8	.16	(.16×8.8) = 1.4

Fractions of runs are meaningless, so we should expect the figures in the observed frequency column to be whole numbers. The fact that they are not whole numbers leads me to suspect that Mr. Mulholland has, without mentioning it, calculated his expectations from the formula Np^r, which gives the expectation of runs of r including those contained in larger runs, i.e., a run of $(r+a)$ successes is counted as $(a+1)$ runs of size r. Preparing a fresh table on this basis we obtain :—

Size of Run	Expected Frequency	Observed Frequency
5	32	24.00
6	6.4	8.00
7	1.28	2.03
8	.256	2.00*

*Taking 780% above chance as 7.8 x chance expectation.

This certainly brings the observed frequencies nearer to whole numbers, but whichever table is correct it becomes clear that Mr. Mulholland's percentages give a very false picture, because the chance expectations are so small. Moreover, instead of considering runs of all sizes, he has picked out the larger ones and examined only the tail of the frequency distribution, an utterly unjustifiable procedure as the "tail" is known to be statistically unreliable. To apply a valid χ^2 test to the first table, it is necessary to combine the last three classes, and when this is done a value of $\chi^2 = 2.8$ is obtained, with 2 degrees of freedom, which is insignificant.

(3) p. 226. "In the first forty thousand pairs there were almost three times as many runs of five as there were in the next sixty thousand." It would seem that Mr. Mulholland has quite arbitrarily divided his trials into these unequal groups. One can produce any effect one likes by such a procedure; it is a wonder Mr. Mulholland could not devise something more startling, but such results will never be comparable with Dr. Rhine's experiments, in which all forms of arbitrary selection were most carefully avoided.

(4). "When we arbitrarily selected segments for their high frequency of matching pairs, we could find twenty-five and twice twenty-five with half the pairs matching." For myself, it is impossible to tell by inspection whether this observation has any sig-

nificance, nor do I know of any statistical method to test the point, and I strongly suspect Mr. Mulholland is in the same position.

(5) p. 226. Lastly, Mr. Mulholland divides the trials into 100 groups of 1,000 trials each. In 24 of these groups the matchings came within 2% of expectation, in 30 the expectation was exceeded by more than 2%, while in the remaining 46 the successes were more than 2% below expectation. Apparently the reader is intended to find it surprising that only 24 came within 2% of chance expectation.

The expectation of successes in a group of 1,000 trials is 200; a 2% deviation therefore equals 4, which is .316 times the standard deviation. We can find, from normal distribution tables, what is the expected proportion of deviations falling outside this range, and the following table results:—

Deviations	Expected Frequency	Observed Frequency
$> +4$	30	37.5
± 4	24	25.0
< -4	46	37.5

It will be seen that Mr. Mulholland's result is in close agreement with chance.

To conclude, there is little doubt that Mr. Mulholland's figures, despite their superficial impressiveness, show no evidence of any extrachance effect. The only question is whether they were cited as evidence through an extreme ignorance of statistical method, or in a deliberate attempt to mislead the reader. It is unfortunate that so public a figure as Mr. Harry Price seems to have fallen into the trap, for, in his book *Fifty Years of Psychical Research* (pp. 182-3), he quotes Mr. Mulholland's figures at length as a cogent argument against ESP experiments.

LETTERS AND COMMENTS

A SUGGESTION FOR A PK TEST AND ITS BEARING ON THE QUESTION OF SURVIVAL

Dear Sir:

Outstanding even among the remarkable recent developments of parapsychology are the reports of the psychokinetic influence on the fall of dice. The importance of such discoveries both for our understanding of the nature and power of the human mind and for their possible practical applications can hardly be exaggerated.

I have been wondering whether, as we introduce parapsychological inquiry into the arena of matter and motion which is the especial field of physics, ways of testing our theories might not be devised other than the statistical methods which have been carried over from the ESP research, methods that might be more directly convincing to scientific experimenters. In PK, if it is genuine, we have a power of mind to influence matter directly, though in a very feeble degree. But science is abundantly familiar with devices by which it is possible to step up even the weakest mechanical operation to any desired degree of intensity.

The most spectacular example of such a procedure was presented to us by the Chicago World's Fair of 1933. At a certain moment one evening the widespread exposition grounds burst into life and light and the gates of the exhibition were thrown open. It was no touch of a Presidential finger in Washington that produced the change, but an incredibly feeble light impulse which for forty years had been travelling across the two hundred and forty trillion miles from far Arcturus. Reaching the earth, the rays passed through the tubes of four great telescopes in the East and Middle West, fell at the eyepieces on photo-electric cells, and there engendered an extremely weak current of electricity. These electric impulses, amplified millions of times, were sent over telegraph lines to the exposition grounds and there set the machinery of the great exhibition in motion.

This achievement at Chicago suggests a new possibility for exploring the psychokinetic effect. It would be easy to construct a machine in which a delicately hung bit of metal or a slender wire rested near a metal plate in such a way that the slightest motion of the former would close an electric circuit. The current so set in motion could be amplified to any desired degree and be used to ring a bell or to perform any physical feat which the experimenter wished. It would also be easy to seat around the instrument ten or fifty or a hundred people, simultaneously willing the desired result, thus cumulating the operating force.

If by such methods no positive results were attained, grave doubt would be cast on the reality of the PK phenomenon. If, on the other hand, the outcome was affirmative, proof would be given that could hardly be ignored. If such power could be developed till it was adequate for practical application—especially if, as seems to be the case with ESP, the new force is independent of distance—an almost indefinite power of mental control of suitably designed mechanisms at a distance would be conferred, and such instrumentalities would be incorporated into the pattern of our daily living. From the theoretical point of view also great gain would be registered, for by an experiment of this sort the whole elaborate paraphernalia of calculation of chances and the complicated mathematics with which we have become familiar from the beginning of the ESP work would be short-circuited, since with a properly constructed machine sufficiently guarded against air motion, or earth or building tremors, the number of times that a bell would be rung by chance would be zero. Assuming only that positive results were secured often enough to render the experiment repeatable, the number of failures set off against the number of successes would be irrelevant.

A similar method might be applied to the solution of that great question that will always loom before workers in parapsychology until it is finally solved, the question of human survival after death.

The strongest argument against survival is based on the intimate relation subsisting between mind and body, many holding that the mind is only one form of functioning of the whole psychophysical organism, so that the mental processes could no more continue after the body has decayed than the motion of an automo-

bile could continue after the automobile had been destroyed. This argument, very strong, though by no means conclusive—for we are very far from knowing the nature of the mind so completely as to justify categorical statements of what can and cannot be on the basis of that knowledge—has been substantially weakened by parapsychological findings in regard to ESP which ascribe to the mind a power and an independence which academic psychological theories deny to it. But the more important answer to the critic is found in the fact that elaborate experimental work on this subject by highly competent students of psychical research has amassed an amount of relevant evidence, not sufficient indeed to establish an affirmative conclusion, but amply adequate to render further prosecution of the inquiry not only justified but mandatory.

The most effective approach to this problem has been through the phenomena of mediumship. Now here the mental phenomena have taken the most important place. These link up very naturally with telepathy. It has been suggested that if the human personality survives death, and the various mechanical processes of communication which have been laboriously evolved by men on earth, by voice and by pen, were no longer feasible, it is very likely that communication would be telepathic. If so, all the knowledge which we have gained by the ESP investigations would be germane to the inquiry. But spiritistic literature is full of reports of physical evidence for survival, and while much of this testimony is of a pretty shady sort, some of the evidence for the physical phenomena of mediumship seems to be fairly strong. If the authenticity of the psychokinetic effect is finally established, psychokinesis will be given a new prestige as a possible means of the intercommunication in question. If the mind can influence the motion of matter in some small degree, then the surviving personality, if such there be, ought to be able to make its presence known through such instruments as have been suggested above.

It is a very interesting fact, though one known, I think, to very few, that Thomas A. Edison set to work at one time on exactly this task, and by the construction of an instrument of such a sort as has been described above.

"I am proceeding on the theory," Mr. Edison said, as reported by Mr. B. C. Forbes, "that in the very nature of things, the degree

of material or physical power possessed by those in the next life must be extremely slight; and that, therefore, any instrument designed to be used to communicate with us must be superdelicate—as fine and responsive as human ingenuity can make it. For my part I am inclined to believe that our personality hereafter will be able to affect matter. If this reasoning be correct, then, if we can evolve an instrument so delicate as to be affected, or moved, or manipulated—whichever term you want to use—by our personality at it survives in the next life, such an instrument, when made available, ought to record something."

Several points are very interesting to me in this incident. One is the mental attitude which Mr. Edison showed. He did not try to settle the question by ridicule as so many are content to do. He did not rule it out of court by saying on theoretical grounds prior to experiment that such survival was impossible. His attitude was: "It may be. Perhaps it is. I do not know. Nobody knows. Let's experiment to see if we can't find out." I submit that there speaks the true spirit of the scientist far more than in the vehement denials of the critics. If all thoughtful people took the same attitude in regard to this class of facts, progress in psychical research would be more rapid and the outlook more encouraging.

Interesting also is the way in which Mr. Edison envisaged the problem. For most people the issue has been so deeply enswathed in double and treble and quadruple wrappings of prejudice and superstition and gruesome fables that even intelligent people find it difficult to face the question at issue in a clean-cut way. Mr. Edison saw the inquiry simply as a problem in communication, and Mr. Edison had had a good deal of practical experience with communication devices. How extremely realistic his approach to the inquiry was he showed to Mr. Forbes when he ventured the opinion that he would not be surprised if the responses on his invention "should first come from telegraphers, or scientists, or others thoroughly understanding the use of delicate instruments and electric currents."

Edison dropped the project after no long time though what he had done about it all or why the venture was abandoned, I have been unable to learn.

It is also interesting to observe that Edison, absorbed predominantly during his life, though he was, in the manipulation of

material things, nevertheless cherished the hope that through mechanism we might rise to something that was higher still. He said to Forbes:

"If the apparatus I am now constructing should provide a channel for the inflow of knowledge from the unknown world—a form of existence different from that of this life—we may be brought an important step nearer the fountainhead of all knowledge, nearer the intelligence which directs all."

CHARLES E. OZANNE.

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.

HIGH-DICE TESTS: Tests of *PK* in which the aim of the *subject* is to try to influence a pair of dice to fall with the two upper faces totaling eight or above.

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.*

LOW-DICE TESTS: Tests of *PK* in which the aim of the *subject* is to try to influence a pair of dice to fall with the two upper faces totaling six or below.

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.

SEVENS TESTS: Tests of *PK* in which the aim of the *subject* is to try to influence a pair of dice to fall with the two upper faces totaling seven.

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 2.33 or greater.

SINGLES TESTS: Tests of *PK* in which the aim of the *subject* is to try to influence dice to fall with a specified face up.

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.

TARGET FACE: The face on the die which the *subject* tries to turn up as a consequence of direct mental action.

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.*