*The* Journal *of*
# Parapsychology

Volume 10 SEPTEMBER, 1946 Number 3

## EDITORIAL

### ARE PSI PHENOMENA PARANORMAL OR NORMAL?

How *normal* are the psi phenomena, ESP and PK? This question is important in our consideration of these unusual capacities, and we need to see how far we can go toward answering it.

There are, of course, different ways of interpreting the question as to whether psi phenomena are normal. It might refer to the mental health of the persons manifesting such phenomena—the degree of abnormality; or it might bear upon the mental efficiency of the subject—that is, he might be subnormal. Another usage would involve statistical normality, the distribution of psi abilities in the general population. For parapsychology, however, the word "normal" has a special meaning as evidenced by such prefixes as "super," "extra," "meta," and "para." Although the definition of these prefixes is seldom expressed, general usage implies that when they are attached to the word "normal," the resulting word refers to that which is so peculiar as to be beyond explanation in terms of principles recognized at present. The word "supernormal" was most often used in this sense until, in recent years, "paranormal" became the more prevalent term.

The more we study psi phenomena themselves, however, the more *normal* they appear. There has been no reason whatever to think of them as associated with the *subnormal*—that is, with men-

tal deficiency of any type. Both published and unpublished studies correlating intelligence and ESP even favor to some degree a positive relationship between intelligence and ESP capacity.

So far as the *abnormal* is concerned, the case is not so clear. But certainly no abnormal state has yet been found to be associated with outstanding success in psi test performance. The ESP researches conducted with psychopathic patients compare favorably with those carried out on normal groups under comparable conditions. It is significant that no experimenter today would think of turning to the abnormal in looking for able ESP subjects.

We do not know as yet whether or not psi capacities are normally distributed among people. We do have ample reason to think they are very *widely* distributed. Much depends upon the conditions of the test, and the capacities themselves are so delicate that it could well be that everyone possesses psi potentialities while only a limited number are able to demonstrate them under the conditions imposed. The much larger percentage of high-scoring subjects obtained with children's groups suggests, too, that the more "natural" and uninhibited we are, the more likely we are to exercise the abilities in question. Such knowledge as we have, then, favors the view that psi abilities are general. There is, however, no way of knowing yet whether subjects who do demonstrate ESP or PK ability would show a normal curve of distribution of scoring rate or of any other measure of the ability. Certainly they do differ widely in their ability, and it could well be that the curve of distribution would be as normal as those reflecting height of trees or size of skulls.

The question of normality which is of greatest interest at this stage concerns the *para*normal. It is time to inquire whether psi abilities are not now ready to be recognized as part of the standard equipment of the normal individual personality, and whether they are any longer to be regarded as the freakish, isolated, and incomprehensible phenomena they once seemed to be. This is a different sort of question from the others, a question of how we look at psi capacities. Do we know enough about them yet to accept them as part of the natural mental system of the individual? Can we properly admit them to orthodox psychology?

General Library (P.R.R.)
Psych.
Durham
12-4-46
5619



The answer is affirmative. Any new claim is recognized, a new phenomenon accepted as normal, if and when it is found to follow currently familiar laws. This is the test which psi phenomena too have to meet. History furnishes ready examples. Hypnotism was mysterious and could not be accepted until a plausible theory of it was provided. When it was found that hypnotism was a phenomenon of suggestion, which was of course already familiar, hypnotism took on the aspect of normality—that is to say, familiarity—and could be recognized.

The conflict of psi phenomena with the principles of sensorimotor functions makes them seem at first exceedingly strange and unfamiliar. Extrasensory perception differs radically from sensory perception, which is the familiar form. Accordingly ESP seems far from a normal process—hence, paranormal. The same is true, of course, of the relationship between PK phenomena and the ordinary kinetic phenomena of motor response.

The sensorimotor processes are, of course, necessarily bound up closely with the physiology of the organism. On the other hand, wherever it has been possible to get at the question of whether or not psi phenomena are physical, the answer has been negative. The contrast between psi activity and sensorimotor activity stems from this difference.

It is only when we turn to the more complex mental life above and beyond the sensorimotor level of activity that we find the proper setting for psi. It has been found to have strikingly close relationships with the so-called higher thought processes—reasoning, attitudes, interests, judgment, esthetic experience, humor, and the like. The latter themselves differ from *sensorimotor* functions. They show many parallels to the findings regarding *psi* capacities. The effect of narcotic and stimulant drugs upon these latter capacities is quite in line with their effect upon the higher thought processes, but differs radically from their effect upon the elementary sensorimotor functions. The various psychological conditions found to affect performance in ESP and PK tests, conditions such as subjects' attitudes, subject-experimenter relations, incentives, novelty, distraction, tempo of response—all these are conditions likely to affect

the higher mental processes, but not the more purely sensori-motor functions. Similarities between curves of performance in psi tests and those in memory and learning experiments have been pointed out, and much more could be assembled in support of the relationship indicated—namely, that psi phenomena belong to a *level* (to use the term figuratively) of the mind above that of sensori-motor functioning.

The relationship thus found between psi phenomena and higher mental processes goes far toward making them familiar, understandable, and therefore normal. It would seem advisable, then, for students of ESP and PK to recognize from now on that *psi phenomena are normal*. Certainly, as far as we know, they fit in with our conceptions of normality; and the evidence is now considerable.

The point is more than a merely academic one. It is, after all, in man as a normal, typical individual that we are most interested. Only if the findings of parapsychology are generally applicable to everyone can we make the most of their bearing on life and society. If we are correct in concluding that psi factors are normal, the new discoveries regarding such factors are really discoveries as to the nature of the mind in general; they represent the operation of the mind under the conditions of the tests. Thus, when ESP, for instance, transcends space and time in the experiments, it is not just an isolated function but the human mind itself that is doing so—the *normal* mind.

However, we do need some kind of characterizing term to distinguish psi capacities from others, at least until they are better and more generally understood. Therefore they may purely for convenience continue to be called "parapsychological" as long as the term "parapsychology" is of use in identifying this more venturesome branch of the field of psychology. We believe our fellow-inquirers in this field will approve our tentative position that psi phenomena, then, may be called "parapsychological" or, more briefly, "parapsychical" merely for identification purposes; but they are no longer paranormal.

To continue to use paranormal (supernormal, extranormal, or metanormal) longer would be to perpetuate a misconception which the various contributors to psi research have labored long, and successfully, to correct.

J. B. R.

# AN INTEREST INVENTORY RELATION TO ESP SCORES

By C. E. Stuart

**ABSTRACT: An interest inventory, consisting of 60 items which the subject checked on a scale varying from "like very much" to "dislike very much," was given to subjects after they had participated in an ESP experiment. In the ESP experiment, four free drawings were made in response to four sealed stimulus pictures. Among the interest test items were four which referred to the stimulus pictures. ESP scores for each stimulus picture were then compared with the subject's expression of like or dislike for the topic of the picture. There was no evidence that the subject's specific attitude toward the topic of the stimulus picture influenced the degree of success of his response.**

**The interest inventory supplied, however, a general measure of the comparative "affectability" of the subjects; that is, the extent of general liking and disliking. From the interest test data two measures of affectability were derived. Both of these separate the subjects significantly into two groups, one in which the direct ESP scores were positive and the "displacement" scores negative, and another in which the direct ESP scores were negative and the "displacement" scores positive. The degree of affectability of the subject, as expressed in his interest test, appears to be directly related to the kind of ESP response he will make.—Ed.**

In an article in this Journal for December, 1945, I reported an ESP experiment in which subjects made free drawing responses to stimulus pictures sealed in envelopes. Despite a number of apparently good hits, comparison of the responses with the intended stimulus pictures yielded scores that were indistinguishable from chance. The responses did, however, show a significant backward displacement; that is, they corresponded consistently with the stimulus drawings presented just before the intended ones.

Later study of the responses of these experiments demonstrated that when an unbiased judge noted the degree to which the response drawings showed "expansive" or "compressive" behavior on the part of the subjects, the subjects could be separated into two groups: those who characteristically responded to the intended stimulus, and those who characteristically displaced to the just-previous stimulus. These findings are reported in detail by B. M. Humphrey in the June number of this Journal (1). The phenomenon of displace-

ment observed in the first report could now be attributed largely to the responses of the "compressive" subjects.

The observations on which these findings are based consisted of four experiments closely similar in design. Informal inspection of the responses of the subjects in the first series suggested that the character of the stimulus drawing might have something to do with success. A picture of a dog might be a more effective stimulus than a picture of a soap box simply because of the greater number of personal associations bound up with dogs than with soap boxes for most people. The dog would be more interesting and attention-demanding. This line of speculation led to the incorporation in the last experimental series (Series Z) of an interest summary to be checked by the subject after he completed his drawings. The interest test was intended to provide a direct indication of the subject's attitude to the stimulus object itself. For example, one stimulus was a drawing of a revolver, and one item in the interest test is "guns."

The ESP procedure of Series Z has been reported previously in detail (2). Twenty-four subjects made free response drawings to each of four stimulus drawings sealed in envelopes. The drawings were ranked by two experimenters with the preferential matching method. Three hypotheses were considered: (1) that the subjects responded to the intended stimulus; (2) that they responded to the just-previous stimulus; and (3) that they responded to the stimulus next to be presented.

While the total work was significant, none of these hypotheses yielded significant results for this series alone. The figures are summarized in Table 1. All the probabilities in the column "2P"

Table 1

ESP Results of Series Z

| | Number of Subjects | Mean Score per Subject | Mean Weight per Response | CR | 2P |
|---|---|---|---|---|---|
| Direct | 24 | 40.875 | 2.555 | .57 | .57 |
| Displaced Back | 24 | 31.792 | 2.649 | 1.74 | .08 |
| Displaced Forward | 24 | 28.333 | 2.361 | 1.52 | .13 |

are above the .01 level of significance. These results, however, are sufficiently consistent with the other experimental series to contribute appreciably to the significant backward displacement observed when all were pooled.

Since a direct score for each subject is based upon all four of his responses and a displacement score is necessarily based upon only three of the responses, the mean score per subject for these two conditions is not directly comparable. The direct scores have a chance mean of 40 and the displaced scores have a chance mean of 30. To display them comparably, the mean weight per response is given. Each stimulus-response pair is given a weight of 1, 2, 3, or 4 by each judge. A chance mean weight is therefore 2.5. Mean weights greater than 2.5 indicate a degree of association greater than chance. Mean weights below 2.5 indicate a degree of association poorer than chance matching would provide.

The interest test given each subject is shown in Figure 1.

The stimulus pictures used in the experiment were (1) a pattern of Flags; (2) a flying Bird; (3) a revolver (Gun); and (4) a comic-strip character, Kayo. The four items in the interest test which seemed most associated with these stimuli were "Parades," "Birds," "Guns," and "Comic Strips."

The scores made on the individual stimulus pictures were grouped with respect to the five attitudes recorded on the corresponding interest test item. The results are given in Table 2. The mean weights do not vary significantly from a chance mean of 2.50.

The interest test yields a more general variable, however, in that

Table 2

ESP Scores Compared to Attitudes Expressed for Corresponding Item

| Attitude | N | Mean Weight | *t* |
|---|---|---|---|
| LL | 13 | 2.673 | .69 |
| L | 39 | 2.320 | 1.35 |
| I | 32 | 2.703 | 1.24 |
| D | 5 | 3.050 | .81 |
| DD | 7 | 2.500 | .00 |

People differ in their interest in activities and objects. Please record *your* personal reactions to the following common items by writing after each:

LL if you *like it very much.*
L if you *like it.*
I if you are *indifferent* in interest.
D if you *dislike* it.
DD if you *dislike it very much.*

| | |
|---|---|
| Algebra | Movies |
| Chemistry | Parades |
| Geography | Comic Strips |
| History | Musical Comedies |
| Nature Study | Concerts |
| Dramatics | Animal Zoos |
| Art | Lectures |
| Military Drill | Football Games |
| Music | Libraries |
| Public Speaking | Church |
| | |
| Tennis | Cameras |
| Hunting | Stamps |
| Fishing | Books |
| Driving | Guns |
| Boxing | Toys |
| Hiking | Coins |
| Bridge | Birds |
| Chess | Dogs |
| Reading | Pin-ups |
| Skiing | Antiques |
| | |
| Teachers | Dancing |
| Doctors | Arguing |
| Policemen | Formal Affairs |
| Salesmen | Dinners |
| Soldiers | Initiations |
| Movie Stars | Traveling |
| Athletes | Night-clubbing |
| Children | Parlor Games |
| Cowboys | Poker |
| Mechanics | Picnics |

FIG. 1. Interest inventory.

it shows differences between subjects in the range or span of their likes and dislikes. The items are common activities, events, and things familiar to the college students who acted as subjects. In general one could expect more "like" (L) responses than any other; and this was the case. The observed attitudes were distributed: LL, 25 per cent; L, 32 per cent; I, 27 per cent; D, 10 per cent; and DD, 6 per cent. But the subjects varied markedly from this pattern. Some expressed many strong likes and dislikes. Others recorded practically none.

By converting likes and dislikes into numbers by the formula: LL = 2, L = 1, I = 0, D = −1, and DD = −2, the subject's interest test may be given a total score that expresses in a rough way the amount he likes these everyday things. When the tests were scored in this way they yielded a distribution ranging from −12 to +78 with a median of +35.5.

The four-item study shown in Table 2 is warning enough that the ESP scores are unlikely to have any direct correlation with interest scores. But a suggestion comes from an analysis of ESP card tests reported in 1941 (3). In this research subjects estimated their scores before the card runs were checked. It was found that "affectable" subjects, those who consistently altered their estimates with respect to their previous performance, scored at a virtually chance level, while "unaffectable" subjects, those whose estimates did not correlate with their performance, scored significantly above chance. The concept of affectability seems useful to the present problem. The person who is extreme either in the number of things he greatly likes or the number of things disliked can be characterized as more affectable than the person whose expressions are moderate and average.

The interest test scores may be separated into two groups: the 12 subject scores about the mean of 35.5, and the 12 extreme scores (that is, the six highest and the six lowest). The extreme group may be considered to represent the more affectable subjects, the mid-range group, the unaffectable subjects. This categorization is

not perfect by any means. A subject who marked all items with LL or DD might easily have a score in the mid-range group, yet be notably affectable. But there is not enough evidence here of such extreme variability to raise any serious question of this sort. The ESP scores of the subjects as separated into the mid-range and extreme groups are given in Table 3.

Table 3

ESP Score Deviations from Chance Associated with Mid-range and Extreme Interest Scores

| Interest Score Group | Number of Subjects | ESP Score: Direct (Deviation of Mean Weight from 2.5) | *t* | ESP Score: Displaced (Deviation of Mean Weight from 2.5) | *t* |
|---|---|---|---|---|---|
| Mid-range (Unaffectable). | 12 | + .203 | 1.33 | − .069 | .54 |
| Extreme (Affectable)... | 12 | − .094 | .87 | + .368 | 4.86 |

The ESP scores are expressed in terms of average deviation above or below the chance mean weight of 2.5 for each stimulus-response pair, and the *t* values test the significance of the deviation. It will be noted that for the direct ESP scores (the comparison of the response with the intended stimulus) the unaffectable group is positive and the affectable group, negative. Neither deviation is significant, nor is the difference. For the displaced ESP scores (comparison of the response with the just-previous stimulus) the results are just the opposite. The unaffectable group is slightly negative and the affectable group is significantly positive. The positive deviation for the affectable group has a probability value of about .0002. The difference between the affectable and unaffectable groups is, of course, also significant.

Another measure of affectability may be considered. Since affectability may be thought of as related to the number of likes and dislikes expressed by the subjects, we may count all likes and dislikes as of equal weight. Or more simply, we may count the

number of items marked "I" for "indifferent." The affectable person would be indifferent to few items, the unaffectable, to relatively many. When the indifferent responses are thus used to test affectability, the scores range from 8 to 26, with a median of 16.5.

The ESP scores of the subjects as separated into the 12 highest and 12 lowest in number of items marked "indifferent" are summarized in Table 4. Here, as in Table 3, the unaffectable group

Table 4

ESP Score Deviations from Chance Associated with High and Low Indifference Interest Scores

| Interest Score Group | Number of Subjects | ESP Score: Direct (Deviation of Mean Weight from 2.5) | *t* | ESP Score: Displaced (Deviation of Mean Weight from 2.5) | *t* |
|---|---|---|---|---|---|
| High Indifference (Unaffectable). | 12 | + .135 | .79 | − .083 | .69 |
| Low Indifference (Affectable)... | 12 | − .026 | .28 | + .382 | 4.91 |

scores positively and the affectable group scores negatively for the direct ESP comparison. The deviations and the difference between them are not significant, however. The displaced ESP scores again show the affectable group scoring at a highly significant rate ($P = .0002$), and there is a significant difference between the unaffectable and affectable groups. The close agreement of the two ways of scoring the interest test is to be expected, of course, since the two methods separate the subjects in virtually the same way.

The major importance of these observations is in their supplementation of prior research. The interest test has done essentially the same thing as the Humphrey "expansion and compression" categories in revealing two ways that subjects characteristically respond in the ESP test. And it brings an older research on affectability directly into the present-day successful line of direct psychological attack on the ESP problem. Extensive application of the interest

test will be necessary before we can be sure that it is a generally effective prediction device for control. But it provides another piece in the jigsaw pattern of our understanding of ESP.

## References

1. Humphrey, B. M. Success in ESP as related to form of response drawings: I. Clairvoyance experiments. *J. Parapsychol.*, 1946, **10**, 78-106.
2. Stuart, C. E. An analysis to determine a test predictive of extra-chance scoring in card-calling tests. *J. Parapsychol.*, 1941, **5**, 99-137.
3. ———. An ESP experiment with enclosed drawings. *J. Parapsychol.*, 1945, **9**, 278-95.

*Parapsychology Laboratory*
*Duke University*
*Durham, North Carolina*

# THE SOURCE OF THE DIFFICULTIES IN PARAPSYCHOLOGY

By J. B. Rhine

Research in parapsychology is extraordinarily difficult. It offers so many drawbacks that the young explorer looking over the field could not be blamed for being disheartened if he could but see them all. It is perhaps fortunate for every field of inquiry that this type of perspective is commonly achieved only after considerable experience in the field.

On the other hand, such difficulties are usually relative. Their gravity depends, for example, upon how we look at them. The more fully they are anticipated and the reasons for them understood, the better the inquirer is prepared to meet them, to endure them if he must, and eventually to overcome them if he can. It will be in order, then, to review some of the main hardships that go with the work in parapsychology and to make some suggestions about the source of these difficulties. If the picture drawn should seem gloomy to anyone, let him remember that it could be only on a foundation of strong confidence and optimism that one could bring himself to advertise his troubles in this fashion.

The difficulties in this research field generally fall under two headings—experimental and social. Under the first heading belong the troubles encountered in the actual research itself. These naturally spring from the character of the phenomena. The social difficulties have to do with the hardships the experimenter encounters when he seeks an opportunity to work in parapsychology—matters of academic freedom, financial support, outlet for publication, and the like. We can most profitably discuss these problems in the light of the research which has had the greatest amount of experimental development. This means the work done with the familiar psi phenomena, ESP and PK.

## Experimental Difficulties

Research with psi phenomena is difficult because they are so elusive. Few if any of the phenomena of nature have given so much experimental difficulty in the mere elementary establishment of their occurrence.

A number of experimenters have failed in their attempts to find psi phenomena. Their experiments gave only scores that could be explained by chance. It is true these unsuccessful experimenters are in the minority, but there have been enough of them to mark the ESP process as especially difficult to demonstrate.

There is even some evidence that experimenters who have success in obtaining evidence of ESP in their earlier researches may lose their capacity to obtain evidence of ESP as they continue in the research. This statement will not as yet bear wide generalization, but there is some basis for making it. Very probably the initial curiosity and enthusiasm of the earlier experiments are lost as the tests are monotonously repeated over and over. Something disappears from the experimental atmosphere that is vital to its success. Whether or not this hypothesis is correct, the phenomenon is a most sensitive one to be affected by such subtle changes in the experimenter.

It is, however, in the subjects themselves that the greatest variability occurs. Subjects who succeed well in the beginning tests may sooner or later completely lose their ability to score above chance. This has happened time and again. The scoring rate may even go down rapidly in the course of a single test session. The subject's attitude, his relation to the experimenter, the general atmosphere of the test situation, the type of test procedure—these and many other delicate influences can determine completely whether there is or is not evidence of ESP in the results.

It seems to be the very nature of the psi test to be depressing in its effect upon scoring rate. The first few trials in a run are the most likely to be successful, and if the subject knows when the end of the run is coming, he is likely to do well again toward the end.

While these position effects, as represented by curves of distribution, are extremely interesting and scientifically very valuable,

nevertheless they represent more of the same elusiveness of ESP and PK. The declines are evidences of the failure of psi phenomena to function under the conditions. Why mere repetition of trials in succession should cause such failure is a most important question, but the fact is, ESP and PK are escaping us none the less.

Sometimes these psi phenomena elude the experimenter by displacement. The process is directed to another target than the one intended, and the high scoring is on the adjacent card or drawing. Sometimes, too, there seems to be a definite avoidance of the target stimulus so that a negative deviation results; that is, the scoring rate falls below the mean expected from chance alone. And there may be still other ways in which the capacity can elude the experimenter and yet produce lawful and significant effects.

These troubles are not by any means trivial ones. Such difficulties are responsible for keeping psi phenomena from successful demonstration under control and on demand. They are the main bar to advancement toward complete understanding of and control over psi abilities. It is true a certain amount of repetition of what has been discovered is possible. Reliable confirmation has been produced time and again, but it has all been at great cost, with an enormous amount of waste and with a very slender yield. There has been an exceptional amount of hazard to the success of every parapsychology experiment. If the cause of these experimental difficulties can be brought into sufficiently sharp focus, we can at least understand the situation better. We will know more of what we have to expect and can tell better where to direct our further exploratory efforts.

The root of these difficulties is the *unconsciousness* of psi phenomena. The subject in the test has no introspective awareness of the experience of either ESP or PK activity. He does not know *when* it is that ESP or PK works; he does not know *how it feels* to exercise either function successfully. Success and failure are the same. The subject feels about as sure he is succeeding when he misses as he does when he hits. Therein psi phenomena differ greatly from sensorimotor activities. It is not that psi phenomena are involuntary; rather, they are, so far as consciousness allows,

volitionally directed. But it is volitional action operating in the dark for lack of the light of introspective insight.

Nor is the elusiveness of psi phenomena a matter of weakness of the capacities. We have, of course, no way of knowing what the real strength of these abilities may be until we can get them under better control, but it is safe to say that if conscious awareness of success or failure were to be achieved, ESP tests could be made to yield approximately one hundred per cent success; a subject would not give his response until he was confident that he was right. Practice with psi phenomena would then be possible too. Learning, of course, depends upon some knowledge of success operating in some way. In this case it would be knowledge of the means of producing the psi effects—awareness of how it felt to do it—that would require introspection. But there is no introspective window open as yet where psi phenomena are concerned.

It may be of interest in passing to note what the *practical* consequences would be of the acquisition of consciousness of psi activity. It is at once obvious that the high reliability that would result would put ESP into practical use as a means of acquiring knowledge that would transcend the common physical barriers that limit the senses. The ultimate effect of such practical extension of reliable knowledge upon human society ought in general to be a healthy one, but it would be revolutionary in the extreme.

It is important, therefore, to concentrate research upon the problem of getting conscious control over psi phenomena. It would have profound significance in human life. But this does not mean that at the same time there should be any neglect of the possibility of developing reliability in psi activities through unconscious modes of expression and response such as the various kinds of automatisms.

There is, however, some encouragement already offered us on the main line of inquiry just indicated. In spontaneous psi experiences the individual sometimes *is* conscious of what is happening to him. He is very sure of the knowledge that comes to him extrasensorially, knows when it came, and cannot be shaken out of his belief by reason or reassurance. Experiences of this character are sufficiently common to be familiar and are in no need of illustration.

We should be able to find out through the careful study of such experiences and the persons who report them something about the special conditions that allow for this introspective certainty. Is the percipient in such cases simply less resistant by nature to the affective accompaniment of the psi experience? Did the occurrence of the experience happen to come at a fortunate moment in his mental life, or was it the result of a specially powerful purpose attaching to the experience?

Occasional flashes in the laboratory experiments offer a bit of encouragement too; but these are very rare. Yet altogether the prospects for ultimate conscious control over psi activity offer enough hope to spur any interested explorer into pointing his utmost efforts at research in that direction. No one, of course, can reliably estimate the difficulty or the time or the work involved in reaching the objective. But the closer one comes to the "front line" of research, the more he is drawn toward a confident view of the prospects.

## Social Difficulties

It is well known that there are other difficulties in parapsychology besides the experimental ones. There has been a certain amount of professional stigma attaching to working with psi phenomena, although this has greatly declined as the work has spread and the findings have become better established. Evidences of this type of difficulty are usually not the most pleasant to recall, and no good point would be served by reviewing them here. Words of ridicule, whispered suspicion, charges of incompetence and other weaknesses are hurled about all too plentifully in the beginning stages of any new research developments such as those of parapsychology.

Closely related to this professional disapproval of the type of research is the difficulty of securing editorial acceptance of published reports of such unorthodox types of investigation. The psychologist who a few years ago was bold enough to undertake an ESP investigation was as likely as not to find that if his *experiments* met with success, his efforts with the editor of a psychological journal did not. The chances of securing financial aid from research foundations for work in parapsychology were extremely poor, prospects for the ad-

mission of a parapsychologist to scientific societies were not good, and the teaching opportunities open to him were very limited. The more one ventured into the parapsychological field, the more he became isolated from the orthodox professional group in psychology.

What is back of these unusual handicaps of a social nature from which parapsychology suffers? Indirectly it is the very same elusiveness of psi phenomena that causes the difficulties of experimental character. It is plainly the *uncertainty* of demonstration of these capacities that puts the investigator and his whole field in the various predicaments that parapsychologists have had to experience. This uncertainty is of course a function of the elusive character of the psi process.

It is true psi phenomena are by nature generally unwelcome to those who dislike the challenge they offer to a physicalistic view of human personality. But, generally speaking, in the sciences we do not argue long with the facts, once their certainty is beyond dispute. We give up our discordant theories and make our peace with the new findings, however unpalatable they are at first.

Right there, then, the uncertainty of the demonstration of the psi phenomena comes in. Their undependable character allows a wide zone for argument and delaying tactics for those who maintain an opposing viewpoint. The poor control over the psi functions requires the continual use of statistical methods to get evidence of them. The uneasiness that exists about statistics in general serves to enhance the feeling of uncertainy, especially among those whose philosophy could not accommodate psi phenomena in any case.

But it is safe to say if our research progresses to the point where even a single subject can acquire conscious control over his psi capacities, the resulting reliability will at once dispel the uncertainty that prevails. The consequent utility of ESP in practical affairs would alone render it immediately respectable in the highest degree. From then on, every university would welcome research in parapsychology, and any philanthropic organization with funds available for research would respond to an appeal for aid. Editors of science journals would reach out eagerly for any crumb of a research report regarding the new findings. Parapsychologists would cease to be

professional lepers and would be showered with honors and liberal stipends.

It must seem exaggerated to attach so much consequence to the solution of a problem that has not even been given conspicuous emphasis in the literature of parapsychology in the past. To say that all these matters depend upon the basic fact that psi phenomena are unconscious, and that all the troubles would disappear if only a few subjects could be found in whom conscious control could be developed doubtless seems like a very excessive claim. There is, however, one thing that need not be contested, and it is the main point of the principal difficulties—the lack of subjective awareness of psi activity. This point must be made the main target of research efforts in the future. And if we have any appreciable success at all, the consequences themselves will be so astonishing that no one, not even the historian, will find it important to look back and see whether a fair or an exaggerated picture has been drawn of the difficulties, their source, and the way toward their solution.

At the beginning I mentioned that the difficulties greatly depended upon how we viewed them. I have tried to make clear that the main social as well as experimental handicaps seem to spring from a single specific characteristic of psi phenomena. It may make the trials and troubles of the research parapsychologist somewhat less burdensome to realize this much of their source and character. The difficulties are bound up with the nature of psi effects themselves; hence every step of progress toward a better understanding of them will help to lighten the gravity of the hardships. The troubles are best faced, then, as problems that are natural to this challenging field.

*Parapsychology Laboratory*
*Duke University*
*Durham, North Carolina*

# A PK EXPERIMENT WITH DISCS

By ELIZABETH A. MCMAHAN

**ABSTRACT: Herein is reported another PK experiment carried out with the use of discs instead of dice. As in the earlier report by Miss McMahan, the subjects are adolescents and children tested in a social situation, a PK party. A long rotating cage was used for the throwing of the discs, and each subject was given three rounds, or turns, at the apparatus in a given session. The results of greatest interest consist of the position effects in the data. There was a decline throughout the three turns, although the differences were not significant. But there was a significant decline in the distribution of hits within the turn or set. The quarter distribution within the set gave a significant difference between the upper left and lower right quarters. When this experiment is combined with those reported earlier by Miss McMahan, the quarter distribution for the entire work to date is also significant. The quarter distribution given by the discs tests is similar to the type given predominantly by the PK experiments with dice.—Ed.**

IN THE SPRING of 1944 I began a line of experimentation which was designed to test the PK hypothesis through the use of two-sided objects. The beginning experiments, I and II, were reported in the JOURNAL for December, 1945 (2). Some of the subjects who took part were tested individually and others in a social atmosphere of party-like gaiety. The experiment conducted under the latter conditions gave the more interesting and significant results and initiated a series of similar investigations, one of which forms the basis for the present paper.

Before going into the details of the investigation to be reported herein, it is necessary to outline briefly its background and logical development.

In the first group experiment with discs, one of those already reported, the subjects were from the lower age ranges, mostly children and adolescents, and the experiment was carried out in the form of four "PK parties," to each of which six or seven individuals were invited. The sessions were held in the home of one of the subjects, and prizes for good scores, together with refreshments at the end of the session, helped preserve the party-like atmosphere.

The discs used were of plastic material, 7/8 of an inch in diameter and approximately 1/10 of an inch thick, with target colors or symbols painted on the sides. These were thrown either by means of stationary cardboard chutes or in a revolving cardboard tube.

When a subject's turn came, he threw the discs while the other subjects gathered around the apparatus and attempted to "help" him by concentrating on the target face and willing the discs to fall with that face uppermost.

The subjects had two turns apiece, each consisting of 12½ runs. (A run for disc experiments is defined as the throwing of 20 discs.) Since 10 discs were thrown at a time, two throws made a run, and 25 throws composed a turn. The trials were recorded in five columns across the page, with five entries in each column.

From the first session onward it was plainly evident that there was a difference in scoring between the first and the second turns when the results for all the subjects were pooled. The scoring, which was above chance in the first turn, dropped so consistently in the second turn that after four sessions the CR of the difference between the two was 3.45, with a probability (after due corrections were made) of .0009.

At this point, just before the beginning of Experiment III, which is being reported here, it seemed desirable to vary the procedure somewhat in order to throw more light on the consistent drop in scoring found in Experiment II. Hence it was decided that, instead of two turns, each subject should have three, and that this third turn should be made sufficiently different in external appearance from the first two to cause it to stand out. The hypothesis was that by making the third turn unlike the other two, it might be given the essential qualities of *firstness;* that is, it might be made to resemble the first turn in those essential psychological respects which favor PK performance.

Experiment III, then, was planned as follows: The social situation and age groups were not to be changed, and the first two turns of each subject were to be done under conditions identical with those in the previous experiment. After the completion of the second turn, however, a new type of apparatus was to be brought in and a new system of awarding prizes introduced.

## Procedure

*Apparatus*

The apparatus for the first two turns consisted of two cardboard chutes which have been described and diagrammed in the paper already published (2). The discs were dropped into the mouth of the top chute. From there they tumbled through the criss-crossed wires inside the chutes, bounded out onto a corrugated cardboard baffle, and then fell into a padded tray where they were counted. The discs used in the first two turns in the present experiment were yellow, with blue circles and crosses.

For the third turn an entirely new piece of apparatus was employed, a revolving cage in which white discs with red circles and crosses having blue centers were contained. The revolving cage was a boxlike tube, 5½ feet in length, three sides of which were of wood. The fourth side was made of wire screening to keep the discs inside while making them visible as they fell. A wire basket-like structure with a flat bottom was fastened to each end of the tube, and into these receptacles the discs fell and were counted. Wooden, triangular blocks were fixed at intervals along the inside of the wooden tube, and corrugated cardboard was tacked over these, making a series of baffles over which the discs fell. The apparatus, as it was being used by one of the groups of subjects who participated in the experiment, is shown in Figure 1.

In throwing the discs a subject would make sure that they were all in the basket at the bottom. Then he would rotate the tube so that it was in a vertical position, bringing the basket which contained the discs to the top. Gravitation caused the discs to come tumbling down over the baffles and end eventually on the flat bottom of the lower basket. Thus a half-revolution was required for each throw.

At the first session a new type of record sheet was introduced. It was different from that used in previous experiments in that it had only two columns across the page, with 10 entries in each column. This type was not used again because of its unpopularity with the subjects. They were accustomed to being awarded prizes by columns and complained that they did not have enough chances

to win. The total results obtained at this session were below expectancy.

Therefore, the record sheets for the remaining sessions were like those used in the previous disc experiment, having five columns across the page with spaces for five entries in each column. Because of the additional time required for the third turn in the present experiment, however, it seemed advisable not to use the last column. Thus each subject had 20 trials at a turn rather than the 25 which had previously been customary.

In Sessions 1, 3, and 5 the subjects were adolescents; in Sessions 2 and 4 they were children. All the subjects were girls.

The first two turns for each subject were, as has been noted, quite similar to those in the previous experiment. At the beginning of a session the subjects drew slips of paper to decide their order of throwing. For every session except the first a die throw determined the target face for each turn, with the following reservation: Neither face was used as the target for more than two subjects in succession, and the proportion of trials for the two faces was the same in all three turns.

In the first session, too, a die throw determined the target for the first turn, and in this session the opposite face was the target for the second. A die throw then determined the distribution of targets for the third turn.

It had been intended that both faces should be used as target an equal number of times for the experiment as a whole. However, a sizable difference between the two faces developed (with 340 runs for crosses and 290 for circles) as a result of both an unequal number of subjects at each session and the primary need to make the proportion of target faces the same for the three turns (except for the first session). The inequality, however, is of no consequence to the evaluation of the results or the conclusions and may therefore be ignored. (In addition to the regular results, the data are analyzed with the target faces equalized.)

For the first two turns prizes of toys, candy, and movie tickets were given for certain scores above chance on the column. In the third turn the requisite score for a prize was lowered, prizes were

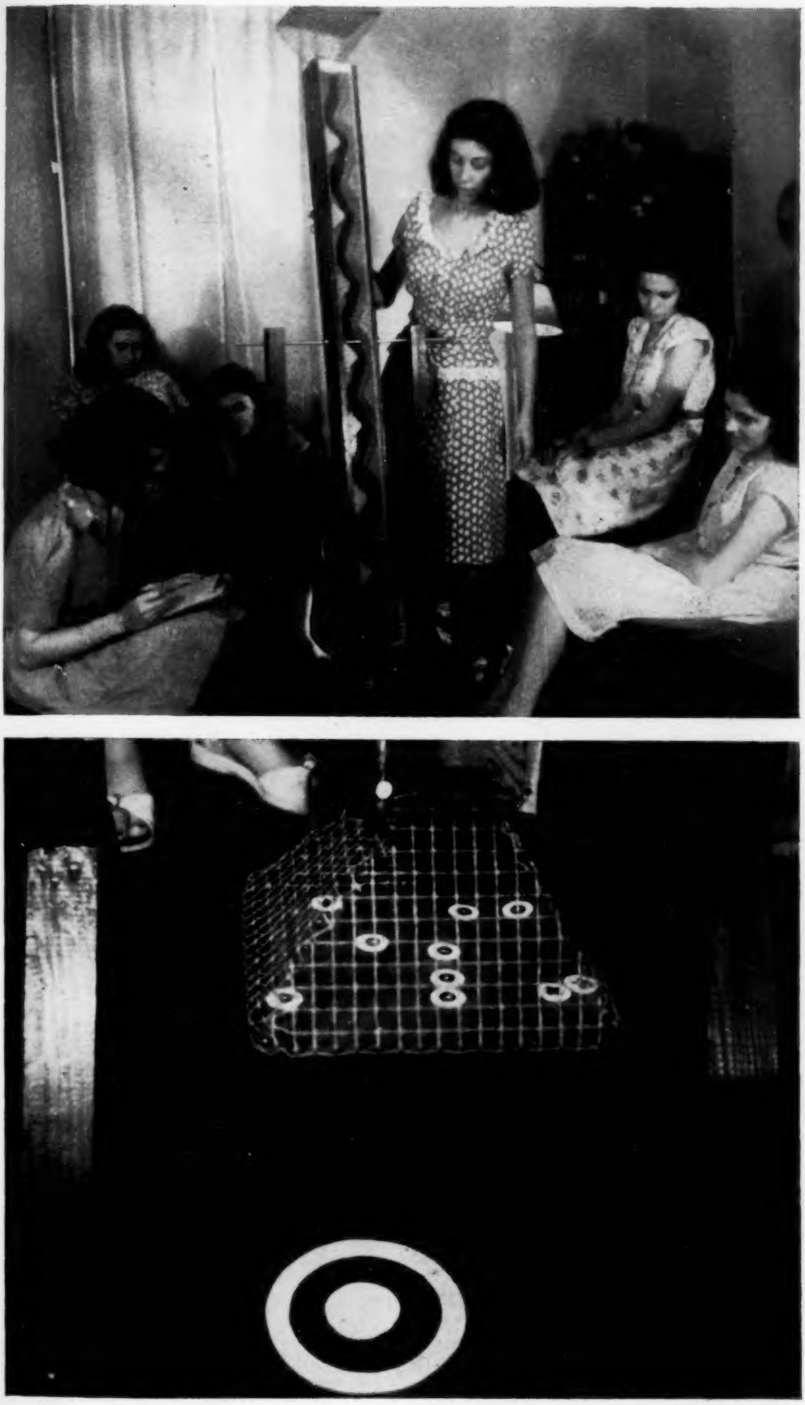


FIG. 1. In the upper picture an adolescent subject turns the rotating cage while the other subjects help her concentrate on the target face. The lower picture shows a close-up of one of the baskets of the cage.

given for individual trials, and the "helpers" as well as the subject received small prizes of candy when a good score was made.

As in the previous disc experiment two experimenters were always present, one counting and recording, the other double-checking for accuracy. In all of these group experiments Dr. Betty M. Humphrey cooperated with me, and the success of the experiments has been due in very great degree to her counsel and assistance.

## Results

### *Decline by Turns*

Two sessions were completed under these conditions of introducing novelty in the third turn. It was then apparent that the drop in scoring which was so much in evidence in the two-turn disc experiment continued also in the third turn, in spite of our attempts to raise the level. This decline may be seen in the data shown in Table 1.

Table 1

Results by Turns for Sessions 1 and 2

| Session | First Turn | | Second Turn | | Third Turn | |
|---|---|---|---|---|---|---|
| | Runs | Dev. | Runs | Dev. | Runs | Dev. |
| 1 | 30 | − 1 | 30 | −10 | 30 | −12 |
| 2 | 50 | +11 | 50 | + 2 | 50 | − 3 |
| Total | 80 | +10 | 80 | − 8 | 80 | −15 |

Although the scoring decline continued through the third turn, this fact was not taken as conclusive evidence that the change in procedure had had no effect. Perhaps the introduction of novelty actually arrested the decline to some extent, while not being sufficient to raise the scoring level. (There is a difference of 18 between the first and second turns, while there is a difference of only 7 between the second and third.)

In an attempt to test this supposition the experimental procedure was again varied. Each subject was to have three turns as before, but now the third turn was to be in no way different from the others.

In fact, all turns were now to be conducted in a manner similar to the third turn of the two just-completed sessions. On the basis of the supposition mentioned above, an even greater drop would be expected for the third turn. The revolving cage with the white discs was selected as the apparatus to be used.

Under these circumstances three more sessions were conducted. The distribution for the different turns showed no consistent trends, as can be seen from Table 2. There is a rather large total drop from the second to the third turn, due to the third session, but the average score of the second turn is higher than that of the first.

Table 2

RESULTS BY TURNS FOR SESSIONS 3, 4, AND 5

| Session | First Turn | | Second Turn | | Third Turn | |
|---|---|---|---|---|---|---|
| | Runs | Dev. | Runs | Dev. | Runs | Dev. |
| 3 | 30 | + 1 | 30 | +32 | 30 | −25 |
| 4 | 50 | + 5 | 50 | + 4 | 50 | +18 |
| 5 | 50 | +16 | 50 | − 4 | 50 | + 3 |
| Total | 130 | +22 | 130 | +32 | 130 | − 4 |

The results for the five sessions, pooled without regard to the change in procedure, are shown in Table 3. The decline in scoring is not significant.

Table 3

RESULTS BY TURNS FOR ALL SESSIONS POOLED

| First Turn | | Second Turn | | Third Turn | |
|---|---|---|---|---|---|
| Runs | Dev. | Runs | Dev. | Runs | Dev. |
| 210 | +32 | 210 | +24 | 210 | −19 |

As has already been mentioned, there was an unequal number of runs for the two target faces. In addition, two different sets of discs were used in each of the first two sessions. These facts make it difficult to interpret the distribution between turns because of the

hypothesis of biased discs, and therefore the results shown in Tables 1, 2, and 3 may not present a true picture. If the runs for the targets are equalized for every turn in every session, then biased discs cannot enter in as an explanatory hypothesis. This was done by omitting in each session all the runs of the last subject whose target had unequalized the faces in that session. Table 4 shows

Table 4

RESULTS WHEN THE TARGET FACES ARE EQUALIZED IN THE THREE TURNS

| | Turns | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Sessions 1 and 2 Combined | 60: + 4 | 60: − 3 | 60: −23 |
| Sessions 3, 4, and 5 Combined | 100: +41 | 100: +23 | 100: − 7 |
| Total | 160: +45 | 160: +20 | 160: −30 |

the result of such an equalization, first for Sessions 1 and 2 combined; then for Sessions 3, 4, and 5 combined; and finally for all sessions pooled. The decline between turns is accentuated, but it is still not independently significant.

*Other Results*

It is already evident that the total deviation for this experiment is not significant. Table 5 gives the total deviations by sessions and then pools these to reach a grand total of 630 runs with a deviation of +37.

Table 5

TOTAL RESULTS BY SESSIONS

| Session | Runs | Deviation |
|---|---|---|
| 1. Adolescents | 90 | −23 |
| 2. Children | 150 | +10 |
| 3. Adolescents | 90 | + 8 |
| 4. Children | 150 | +27 |
| 5. Adolescents | 150 | +15 |
| Total | 630 | +37 |

CR = .67

The data were next analyzed for vertical and horizontal position effects within the turn (or set). As will be recalled, the record sheets used in the first session were different from those used in the other four sessions. Instead of having four columns with five entries per column in a set, these first session sheets had two columns with 10 entries in each. In order to pool the results of this session with the other results so as to get a total vertical distribution, the deviations of Session 1 were combined by twos so that five deviations resulted. The following figures show the pooled vertical distribution for all five sessions, each of the five figures representing 126 runs:

$$\begin{matrix} +54 \\ +11 \end{matrix}\Big\} +65$$
$$-1$$
$$\begin{matrix} -29 \\ +2 \end{matrix}\Big\} -27$$

The CR of the difference between the first and second halves of the column (omitting the middle deviation) is 1.83, which, while it is not significant, is suggestive, with a probability of .034.

The horizontal distribution for Session 1 with its two columns across the page is:

45: −9 45: −14

The other sessions pooled show the following deviations for successive columns across the page:

| | Column 1 | Column 2 | Column 3 | Column 4 |
|---|---|---|---|---|
| Total Deviation | −1 | +63 | −9 | +7 |
| | +62 | | −2 | |

Each of the four deviations represents 135 runs. When the horizontal distribution for Session 1 is combined with that of the other sessions, the resulting left-right decline is:

315: +53 315: −16

The CR of the difference between the left and right halves of the set is 1.23, which is not significant.

The rather large decline in scoring which resulted from the

analyses for both the vertical and horizontal distributions made it certain that an analysis for the quarter distribution would yield a diagonal decline from the first to the fourth quarters. In carrying out this analysis, every turn or set was quartered (making it necessary to omit the middle entry in each column for Sessions 2 through 5). The following QD was obtained by pooling all five sessions regardless of the fact that Session 1 was recorded differently.

QD OF THE SET

| | |
|---|---|
| 130.6 +83 | 130.6 +16 |
| 130.6 +2 | 130.6 -36 |

$CR_{d(1-4)} = 2.44$

The CR of the difference between the first and fourth quarters has a probability of .007 and is significant. The QD is like that typically obtained in dice experiments: the upper left quarter contains the most hits, while the lower right contains the fewest.

## DISCUSSION

The chief evidence of PK in this experiment lies in the analysis for the quarter distribution. The QD is typical and the CR of the 1-4 difference has a probability of .007. The distribution by turns, while showing a decline, is not significant. Before any conclusion regarding PK can be reached on the basis of this experiment, it is necessary to consider these results in relation to the possible alternative explanations or counterhypotheses.

### *Chance*

If proof of PK depended on this experiment, the results could not be regarded as conclusive. The significant QD probability of .007 would have to be corrected, since it was the only significant finding after three analyses. The resulting P would be .021, which is not significant. This experiment, however, does not have to stand alone. Its results confirm those of the dice experiments, where the position effects noted are already established.

These results also bear out those of the already published disc tests. When the data for all three of the disc experiments which I have completed up through the present paper are combined, the case against chance based on the QD receives additional support. The pooled QD for the three experiments shows a 1-4 difference that is not significant but is suggestive. The figures given below present the runs and deviations for each quarter of the set when the three experiments are pooled. The CR of the 1-4 difference is 1.85, which has a probability of .032. Each of the three experiments, however, shows a diagonal decline. Evaluation of the three QD

QD of the Set for Experiments I, II, and III Pooled

| | |
|---|---|
| 606.5<br>+109 | 606.5<br>-5 |
| 606.5<br>-25 | 606.5<br>-35 |

$CR_{d(1-4)} = 1.85$

analyses together by a chi-square combination of the CR's from the three experiments, corrected appropriately for consistent sign of deviation, gives P = .01 (1, pp. 217-18; 3, pp. 386-87).

It may be of interest to note incidentally how Experiment III lines up with the earlier disc work in terms of decline between turns when the chi-square method of evaluation is used to combine the data of Experiments II and III. (Experiment I cannot be analyzed in this way because the subjects had only one turn.) In Experiment III only turns 1 and 2 will be considered in order to make the data comparable to those of Experiment II, where each subject had only two turns. A significant probability of .003 is obtained when the decline between turns 1 and 2, based on all the data in the two experiments, is considered. When the 1-2 decline is based on the data after the target faces have been equalized in the two turns, the probability is .002.

The hypothesis of chance may therefore be considered inapplicable in explaining the results of the disc experiments.

*Skilled Throwing*

The baffled chutes and revolving cage made manipulation of the discs impossible, and therefore the skilled-throwing hypothesis cannot be applied.

*Biased Discs*

As has already been noted, the evidence for PK in this paper is based chiefly on the QD of the set. Since the target face was never changed within the set, biased discs could in no case have accounted for the results. (When sets were compared, in the distribution by turns, the results were stated alternately in terms of an equalized number of target faces.)

*Optional Stopping*

After the first two sessions had been completed, a limit of three more sessions under the revised conditions was set. This setting of a limit before half the experiment was finished adequately answers the optional-stopping hypothesis.

*Errors in Observation and Computation*

Errors in observation and computation cannot account for the results. Dr. Humphrey observed at every session and, together with the subjects, checked on the accuracy of the records. Both hits and misses were counted and called aloud before a record was made. The data have been rechecked and recomputed, and no error has been found.

## Conclusion

The counterhypotheses are not adequate, either singly or in combination, to explain the results. The conclusion is that this experiment supports the previous disc work as well as the investigations using dice in showing that PK does occur.

Disc work is comparatively new in PK literature. Thouless' report of his coin-spinning tests (4) reported in the Journal for September, 1945, was the first dealing with two-sided objects to be published.

It is interesting to see how the results of the disc experiments compare with those of dice experiments. Position effects have occu-

pied a prominent place in the reports of the latter, and they have now been shown to occur in the same lawful manner in the data obtained from disc-throwing.

## References

1. Greenwood, J. A., and Stuart, C. E. Mathematical techniques used in ESP research. *J. Parapsychol.*, 1937, 1, 206-25.
2. McMahan, E. PK experiments with two-sided objects. *J. Parapsychol.*, 1945, 9, 249-63.
3. Rhine, J. B., Pratt, J. G., Stuart, C. E., and Smith, B. M., with Greenwood, J. A. *Extra-sensory Perception after Sixty Years.* New York: Henry Holt, 1940.
4. Thouless, R. H. Some experiments on PK effects in coin-spinning. *J. Parapsychol.*, 1945, 9, 169-75.

*Parapsychology Laboratory*
*Duke University*
*Durham, North Carolina*

# SUCCESS IN ESP AS RELATED TO FORM OF RESPONSE DRAWINGS:

## II. GESP EXPERIMENTS

By Betty M. Humphrey[1]

**ABSTRACT: A new development in parapsychology was reported by Dr. Humphrey in the June number of the JOURNAL. The report described a way of separating high ESP scorers from low ones by grading the drawings they made in the ESP tests according to the form qualities of expansiveness or compressiveness. Dr. Humphrey found that in the clairvoyance tests the expansive drawings were made by the high-scoring subjects, and the compressive drawings by the low scorers. The more compressive subjects were also found to be those responsible for the displacement effect—that is, the tendency of a subject to hit the just-previous target instead of the one he is attempting to hit.**

**The article below describes a second step in this development, one in which the same procedure is applied to drawings obtained in GESP tests (tests in which telepathy as well as clairvoyance is a possibility but in which telepathy is especially emphasized). Again the grading of the drawings for expansiveness and compressiveness separates the scorers into high and low groups with a significant difference. In addition, the separation worked consistently for all the subseries. In the GESP tests, however, the compressive subjects were the ones that scored high and the expansive ones scored low. While, therefore, the form quality of the drawings is again shown to be related to the ESP success of the subjects, this reversal of relationship raises a problem that has to be left unsolved for the present. The report represents another step of progress toward fitting the ESP process into the general personality picture. Further installments reporting on additional studies with this method are to follow.—Ed.**

### Introduction

Much effort today is being directed toward finding some correlate of ESP in the personality of the subject tested. In the former article in this series (2) it was reported that when response drawings which subjects made in ESP tests were divided into two groups on the basis of the form quality or type of expression used, the groups were found to have strikingly different scoring levels on

[1] This is the second article based upon a thesis entitled "Discrimination Between High- and Low-scoring Subjects in ESP Tests on the Basis of the Form Quality of Their Response Drawings," which was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Arts and Sciences of Duke University, April, 1946. The first article was published in the June issue of this Journal.

the ESP test. In the analysis of the four available clairvoyance drawings series the response drawings were separated into two categories: drawings rated as "expansive" and drawings rated as "compressive." It was found that the subjects whose drawings were rated as expansive scored well on the target pictures they were aiming at, while the compressive subjects avoided the target pictures. The difference between the mean scores of the two groups was significant ($2P = .0026$). It was also found that the compressive subjects gave highly significant evidence of displacement; that is, instead of hitting the target picture at which they were aiming, they drew responses which showed marked similarities on the average to the target picture presented just previously. The mean score of the compressive subjects was highly significant in terms of these backward displacement data, and there was a significant difference between the mean scores for the displacement data of the expansive and compressive subjects ($2P = .0001$).

In the original plan for the study of the effectiveness of the criterion of expansion-compression in separating ESP response drawings, all available ESP series using the free response method of Stuart (5) were to be analyzed. In the former report the results of the analysis of all available clairvoyance series were reported. The present report, therefore, will deal with all available tests of general extrasensory perception, or GESP, in which the free response method was used.

In GESP series the subject is asked to make drawings in response to pictures ("stimuli") being looked at by another person, an "agent," in a distant room. As the name implies, the free response method is one in which the subject is asked to draw anything that comes to mind in regard to the picture being looked at. He is not urged to try to "reproduce" the picture, for Stuart has pointed out that responses other than reproductions may also be valid responses; that is, the subject may respond by drawing associations which the stimulus picture calls up. This method, therefore, allows for a great degree of individual expression of personality in the drawing medium, and for that reason series using this method were chosen for the study of whether differences in *ways* of responding in ESP tests are associated with different levels of scoring.

In this study of the GESP series the method of analyzing form quality—*how* a picture is drawn rather than *what* is drawn—was exactly the same as that used in the former report on the clairvoyance series. Details of the method, the definition of the expansion-compression criterion, and the scoring method were given in the first report and need not be repeated here. It may be recalled, however, that the expansion-compression criterion was defined by Dr. Paula Elkisch (1) for use in her study of the drawings of adjusted and maladjusted children. In brief, the drawings rated as "expansive" are those which are judged to be "spacious," "free," "adventurous," and "extraverted," while "compressive" drawings are those which are "meticulous," "fearful," "timid," and "compulsive." Examples of the two types of drawing were given in the former report.

In Elkisch's scoring procedure, which was used here, certain scores are assigned to the judgment of each quality present. For example, if the judge decides that a given drawing is compressive, he will give it a score of 0. If, however, he thinks it is indicative of expansion, he assigns it a score of 1. If the quality present is not obvious or easily decided, he gives the drawing a score of .5. Since in the free response method four drawings are made by each subject, the range of scores for expansion-compression in these four drawings is from 4 to 0. All sets of four drawings which receive a form score of 2 or above are put in the expansive group, which simply means that there is no evidence of compression in the drawings. Sets of drawings which have scores of less than 2 are put in the compressive group.

There were six GESP series in which the free response method had been used, and Stuart was the experimenter in all of them (**5, 6, 8, 9, 10**). The essential details of these various series, together with the procedure of analysis, will be given below.

## Description of Series Analyzed

A summary of the series analyzed is presented in Table 1 where the details of each series may be seen at a glance. All of the series were done with the subject and agent in separate rooms and at least one room between. The available series fall into two general

classes: group tests carried out in the classroom situation, and individual sessions conducted in the laboratory.

Of the three classroom series, two were done at Stanford University (Series T-1 and T-2) and one at Duke University (Series R) under similar conditions (6, 9). In all three the stimuli were projected upon a screen in the agent's room. In each of these, eight stimuli were used instead of the four customary in the individual tests; thus each subject made two "sets" of four drawings each.

Table 1

GENERAL SUMMARY OF ALL GESP SERIES ANALYZED FOR FORM QUALITY

| Series | Condition | Type of Stimuli | Number of Subjects | Number of Sets of Four Drawings |
|---|---|---|---|---|
| T-1..... | Group | Line drawings | 18 | 36 |
| T-2..... | Group | Photographs | 14 | 28 |
| R ..... | Group | Line drawings | 31 | 62 |
| S ..... | Individual | Line drawings | 45 | 45 |
| D ..... | Individual | Line drawings | 29 | 29 |
| Q ..... | Individual | Photographs | 39 | 39 |

The three remaining series consist of sessions carried out with individual subjects in the laboratory. Series D consists of all sessions carried out at Duke University with the use of line drawings as stimuli, while Series S represents a similar series conducted at Stanford University. Series Q is made up of a number of smaller projects, but for our present purposes it may be characterized as consisting of all first-session results of a given subject in all tests in which photographs were used as stimuli.[2]

In Series T-1 (for the original report of this series see 6) the stimuli were line drawings on cards suitable for insertion in a balopticon projector. The same stimuli were used, too, in Series R. For Series S and D, line drawings were also used. Since in these series subjects were tested individually, the stimulus pictures were chosen from a stock of line drawings, and thus the selection varied from one subject to another.

[2] Since this paper was written Stuart has revised the filing numbers of his Stanford series. His new numbers, S1, S2, S3, and S4, correspond to the series here designated respectively as S, T-1, Q, and T-2.

The stimuli in Series T-2 were eight simple pictures cut from magazines and mounted on 8" x 5" cards. The pictures showed such things as a package of cigarettes, a violin, a bus, and two children. For Series Q a large stock of pictures cut from magazines were used. Since the subjects were tested individually, a different set of pictures was selected for each subject.

## Procedure of Analysis

All of the response drawings in the GESP series were obtained from Dr. Stuart who gave me no indication of the ESP score attained by any subject. This precaution was necessary in order that I should not be influenced in my judgment of the degree of expansion-compression shown in the drawings. I took each drawing, recorded the quality recognized in it, and gave it the appropriate score. When a set of four was completed, the total score was found, and the subject was labeled as either expansive or compressive. For Series T-1, T-2, and R there were two such sets for each subject, since each made eight responses. The expansion-compression criterion was applied separately to the two sets, so that it is possible for a subject to receive a rating of compressive in one set and expansive in the other.

After the responses in an entire series had been judged in terms of expansion-compression, the ESP scores were obtained for each group separately. The score indicative of ESP performance in tests with drawings is that obtained by the preferential matching method of Stuart (5). This score is composed of weights assigned to the similarities between stimulus pictures and responses.[3] A judge who does not know which drawings were made in response to which pictures is asked to assign weights to the degree of association which he perceives between stimuli and responses. These weights may range from 4 to 1, the larger number indicating the greater degree of similarity or association. The weights assigned to four pictures (in each set) would have a chance expectation of 10.00. The judge makes two matchings of each set of responses: each response is rated with respect to every stimulus, and each stimulus is rated with re-

[3] The actual matching technique may be obtained from Stuart's reports on its use (5, 6, 7, 8) or from the previous article in this series (2).

spect to every response. Thus, since each judge makes these two matchings, the score obtained from each will have a chance expectation of 20.00. In all of the work here reported, two independent judges did the matching, so that the score for each subject is made up of four such matchings, with a chance expectation of 40.00.

The rule set down in advance to govern the judgments used in evaluating the stimulus-response correspondences is as follows: Wherever possible, the judgments of the subject plus those of an independent judge will constitute the ESP score.[4] Where the experiments did not include subjects' matchings (as in the classroom series), the judgments of two independent judges will be utilized. In any case in which more than two independent judges shall have matched the drawings, only the judgments of the first two will be considered here.

At the time this research with form quality was begun, these GESP series had not yet been evaluated and prepared for publication. The use of them in this study necessitated the collection of additional judgments and the completion of all statistical work on them. In many cases only one judge had done a given series, and it was necessary to obtain the judgments of another. Thus, *after* I had analyzed each series in terms of the expansion-compression criterion, I secured the services of another judge to complete the ESP scores. Much of the checking of the true correspondences on the judge's record was done by me, always after the form analysis had been completed. Let it be stated unreservedly that after I had had access to the ESP scores, no expansion-compression score was ever changed!

After the ESP scores were completed for each subject, the totals were computed for the expansive subjects as a group and for the compressive subjects as another group. The mean score for each group of subjects is the basis of comparison used throughout this study. The test of significance applied to the means and to the differences between means was "Student's" *t*-test, as explained in

[4] The agent's matchings are omitted since it is theoretically possible for an agent to use time clues in judging the drawings. If, for example, the time interval during which the agent looked at a given stimulus picture was unusually short, the agent might be tempted to match the simplest response of the subject with that stimulus.

the former report. The same corrections for the degrees of freedom were used here as were discussed in that paper.

All checking and computations connected with this research have been carried out independently by at least two of the following three persons—Dr. Stuart, Miss Elizabeth McMahan, and myself. These analyses include the complete rechecking and recomputation of all ESP scores for all series.

## Results

The mean score of the compressive subjects is 40.96, while that of the expansive subjects is 38.23. (Mean chance expectation in each case is 40.00.) The $t$ of the difference between these two means is 2.95, which has a probability of .0016. Since there was no hypothesis as to the direction of this difference, it is appropriate to double the probability. The corrected figure in this case would be .0032, which is significant.

The conspicuous thing about this difference is that here it is the expansive subjects who show evidence of avoidance of the stimuli, while the compressive subjects give mean scores above chance. This is just opposite to the results of the clairvoyance series. Moreover, it can be seen from the presentation of the results in Table 2 that this direction of difference in the total is reflected in every

### Table 2

Total Direct Matching Data of the GESP Series as Separated by the Expansion-Compression Criterion

| Series | Expansive Subjects | | | Compressive Subjects | | | $t_d$ | 2P |
|---|---|---|---|---|---|---|---|---|
| | $N$* | Mean Score on ESP Test | $t$ | $N$* | Mean Score on ESP Test | $t$ | | |
| T-1..... | 16 | 36.12 | 2.79 | 20 | 41.60 | 1.14 | 2.67 | .0096 |
| T-2..... | 8 | 36.50 | 1.40 | 20 | 40.30 | .28 | 1.66 | .1092 |
| R | 29 | 38.52 | .97 | 33 | 39.82 | .15 | .68 | .4964 |
| S ..... | 14 | 39.36 | .28 | 31 | 40.19 | .14 | .33 | .7414 |
| Q ..... | 24 | 37.79 | 1.35 | 15 | 39.87 | .09 | .87 | .1921 |
| D ..... | 14 | 40.36 | .21 | 15 | 45.33 | 2.69 | 1.90 | .0670 |
| Total... | 105 | 38.23 | 2.41 | 134 | 40.96 | 1.64 | 2.95 | .0032 |

*$N$ refers to the number of sets of four drawings each.

series. Consistently the mean score of the compressive subjects is higher than that of the expansive subjects. In one series (T-1) the difference is independently significant. These results are presented graphically in Figure 1.

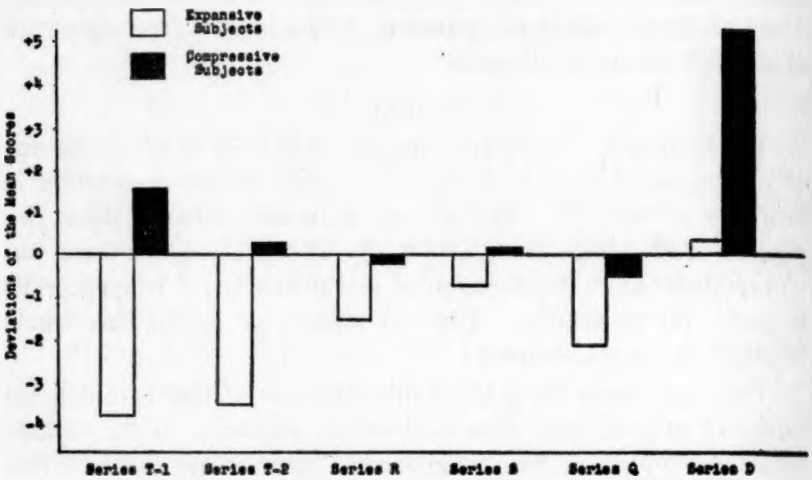



FIG. 1. Deviations of mean scores for the expansive and compressive subjects in each GESP series. (The deviations are found by subtracting the mean chance expectation, 40.00, from the mean score per set of drawings.)

The mean score for the expansive subjects has a $t$ of 2.41, which, when its probability is doubled, can be considered as only suggestive ($2P = .016$).

One of the most striking findings in regard to the clairvoyance series was the marked backward displacement; that is, the marked resemblance of the second responses to the first stimulus, and similarly the resemblance of the third responses to the second stimulus, and so on. In the GESP series, however, no evidence of displacement was reflected in the total deviation for the displacement analysis. Neither does the expansion-compression criterion reveal any significance here. In the backward displacement data the mean score for the expansive group was 29.92, while that of the compressive group was 30.07—both very close to the chance mean of 30.00.

## Secondary Results

### *Stability of a Subject's Expansion-Compression Score*

How stable is a given subject's expansion-compression score from one session to another? The answer to this question is vital to an understanding of what this measure means in terms of the subject's general personality constellation.

Some hints as to the answer may be found in the present data. Series Q is made up of the *first* sessions of a number of different subjects. Six subjects, however, returned for other sessions until they had completed five sessions each. Upon this small sample I did an analysis of variance to test the hypothesis that, in regard to expansion-compression, the scores of each subject vary as much from session to session as the scores of *different* subjects vary—or, briefly, that a subject's scores are not stable from session to session. The ratio of the variances (Snedecor's F) is 2.38. With the degrees of freedom here involved (five for the larger variance and 24 for the smaller), it would take an F of 2.62 to reach the low confidence level of five per cent (see 3, p. 145). Since the F observed here is not significant, we have no evidence that a given subject maintains a consistent expansion-compression level from one session to the next in the same type of experiment. Admittedly, this sample (six cases) is not adequate for the large number of subjects involved in the series reported here (approximately 165 different subjects), but it does give us a suggestion that expansion-compression as shown in ESP tests is not a stable characteristic of a given subject.

From the three classroom series in which a subject did two sets of drawings, we can obtain another bit of information as to what happens when a subject does shift in expansion-compression.[5] Let us take those subjects who were rated as compressive in one set of

[5] It is not feasible to do an analysis of variance or a correlation on the expansion-compression scores for the two sets of drawings done by each subject, since the judgment of expansion-compression was made on the two sets of each subject successively, and it is quite likely that the judgments on one set may have influenced those on the other. Thus a spurious correlation might obtain. The drawings from sessions after the first which were mentioned above were done individually in a random order. They were identified only by code numbers, so that I, as judge, did not know which drawings constituted a set. Neither did I know which drawings belonged to a given subject. Later on, then, by means of the code key for the series, an assistant was able to assemble my scores for each picture so as to give the total score for the set of drawings.

drawings and who received a rating which showed more expansion in the second set,[6] and see how their ESP scores changed. Of the 42 subjects rated as compressive in the first set, 25 became more expansive in the second set (although they still might be classified as compressive). The mean ESP scores for these subjects shifted from 39.88 on the first set to 37.08 on the second.

The eight compressive subjects who became even more compressive on the second set showed a shift in ESP scores of from 39.13 on the first set to 42.63 on the second.

Since both of these differences are in the direction expected on the basis of the hypothesis provided by the main results of this study (that compressive subjects will give higher scores than expansive subjects), it is proper to pool them and obtain a $t$ of the difference. This $t$ is 1.85, which has a probability of .03. This figure is not significant, but it does carry suggestive value.

It is not possible to present comparable figures for the expansive group since time was not taken to decide the degree of expansiveness when it was seen that a subject would receive a score of at least 2 (enough to put him in the expansive group). Thus there are no *gains* in expansiveness noted from one test to the next. Altogether five expansive subjects received a compressive rating in the second set. The four who lost .5 of a point gave shifts in ESP score according to hypothesis; that is, their scores on the second set were better than those of the first. The fifth subject—one who lost a whole point toward compression—gave a reversal in his shift from one set to the other.

The results of this comparison have served only to suggest that when subjects change in their degree of expansion-compression, their ESP scores change inversely.

*Other Problems*

Although the various types of procedure in the GESP series were not used with a view to comparing them, it may be of passing in-

[6] The use of the terms "first set" and "second set" here is not necessarily meant to indicate chronology, but rather to separate the two different sets done at one session. In Series T-1 and T-2, Set 1 indicates Responses 1, 2, 7, and 8, while Set 2 includes Responses 3, 4, 5, and 6. This division was made on the basis of different instructions as to mode of response. (See 6, 9.) In Series R all of Set 1 was completed before Set 2 was begun, but from the subjects' point of view there was no break or difference between the two sets.

terest to note any differences between the group tests and the individual tests.

In the three classroom series there is a significant difference between the expansive and compressive groups of subjects ($P = .007$). This is contributed mostly by the expansive subjects who gave a significantly negative result ($t = 2.50$; $P = .006$).

The difference between the expansive and compressive subjects for the individual tests is less striking. Here the $t$ of the difference is 1.60 ($P = .054$), which is not significant. Neither of the two groups is significant either, although the expansive subjects are negative and the compressive ones are positive.

It is interesting to note in passing that there was, on the average, a larger number of compressive subjects in the group experiments than there was in the individual test experiments. For example, the ratio of the number of compressive subjects to the number of expansive subjects for the group tests is 1.70, whereas that for the individual tests is 1.17. Such a result fits in well with logical expectations, for one would naturally suppose that a subject would feel much more at ease and uninhibited working alone in a room than he would working in the same room with a group of subjects.

## Discussion of Counterhypotheses

The first article of this series has already presented in detail the counterhypotheses to the conclusion that an analysis of expansion-compression in drawings does effectively separate subjects of ESP tests into two groups with significantly different scoring levels. It does not seem worth while to repeat them and the arguments against them in detail here, since they are essentially the same in both cases. For the detailed discussion the reader is therefore referred to the first article (2) and to the original thesis deposited in the Duke University Library. Only a brief summary of the counterhypotheses is presented below, with the arguments against them.

1. The original experiments were carried out in such a way that sensory cues were ruled out. The stimuli were always separated from the subjects by at least one intervening room.

2. The matching of the stimuli and responses was also done under carefully guarded conditions so that there were no cues by

means of which the judges could have inferred the true correspondences (see Stuart's original reports of these series for details of procedure).

3. All checking and statistics pertaining to these ESP series have been independently carried out by at least two persons. Thus the results could not be due to errors in computation. Likewise, all statistics pertaining to the division of the series by the expansion-compression criterion have been independently checked.

4. The judgments of expansion-compression were made before I ever knew the ESP score of a single subject. Stuart retained the ESP scores of all subjects until all the judgments of expansion-compression had been completed for a given series. Thus my judgments could not have been influenced to correspond with the ESP scores. As stated before, the judgments by the preferential matching method had not been obtained on some of these series, and it was necessary for me to obtain an independent judge to make these in order to complete the ESP scores for the subjects. This was done always after the expansion-compression judgments were completed, and none of the latter were ever changed after I had access to the ESP scores.

5. It was pointed out in the first report that in that research and in later experiments there has been a fairly high degree of correlation between the judgments of expansion-compression by two different judges. Although no special study of this was made in regard to the GESP section of the data, the evidence seems strong enough from other series to conclude that it is quite unlikely that my judgments are merely reflections of my own extrasensory perception of the subjects' ESP scores rather than true judgments of the form quality of the subjects' responses. Indeed, if I had been using ESP to divide subjects into high- and low-scoring groups, it seems improbable that I would have unconsciously produced a reversal of the effect noted in the four clairvoyance series.

6. The previous report met in detail the hypothesis that the form quality of the targets is the determining factor in producing the striking differences between the expansive and compressive groups of subjects. It was shown there that compressive and ex-

pansive subjects hit both those targets which were judged to be expansive and those judged to be compressive; that is, the quality judged to be present in the stimuli did not seem to be bound up with the type of subject group which obtained the best scores on that particular kind of picture.

Thus the only reasonable explanation of these results is that an analysis of expansion-compression in drawings does effectively separate subjects of ESP tests into two groups with significantly different scoring levels.

## Discussion of Findings

In the GESP series, as in the clairvoyance series, the expansion-compression criterion has proved to be a measure capable of separating high-scoring subjects from low-scoring subjects in ESP tests with enclosed drawings. In the clairvoyance series the expansive subjects obtained the high ESP scores and the compressive subjects scored below chance expectation. But in these GESP series it was the compressive subjects who scored above chance, while the expansive ones made low scores. There are many factors which make it hard to explain this reversal. For one thing, these two different types of experiment were not done with the intention of comparing the results. Therefore no attempt was made to provide similar conditions. Such unknown quantities as this difference make it hazardous to speculate about the reversal.

One difficulty in considering explanations for the results of the GESP series is that the function of the agent is unknown. Granting, however, that the agent does play some role in the GESP tests, either directly by his mental processes or indirectly because the subject believes he exerts an effect, let us try to understand how expansion and compression fit in. The compression shown by the successful subjects may be an indication that these subjects were more socially responsive and were searching uncertainly for fragmentary impressions which might be related to the contents of the agent's thoughts, and thus their drawings are more compressed, inhibited. On the other hand, the expansion shown by the negatively scoring subjects suggests that they were not so concerned to orient themselves toward the agent, that they perhaps became

involved in their own phantasies and so produced expansive, imaginative pictures which bore little resemblance to the stimuli.

Such lines of thought are pure speculation and are completely worthless unless they serve to stimulate experimentation. The reversal of trends of scoring for expansive and compressive subjects in the two different types of test offers a challenging problem for research. The lines of speculation drawn above are obviously incomplete, and yet in trying to find some explanation for the results, we can see more clearly where experiments are needed. For example, it is necessary to find out whether the same subjects respond differently with regard to expansion and compression in GESP and clairvoyance tests. We need to find out whether they change in degree of expansion or compression in the two tests, or whether they maintain a fairly constant level, with the expansive group succeeding at clairvoyance and failing at GESP, and the compressive group succeeding at GESP and failing at clairvoyance. The latter alternative does not seem very probable in view of the ESP researches based on card tests in which it has generally been found that the same subjects are able to succeed at both GESP and clairvoyance (4).

In addition to noting the performance of the same subjects in clairvoyance and GESP series, it also might prove very interesting to have the ratings of these subjects on personality questionnaires or on some other personality measures. Conceivably, one might find that subjects with certain constellations of traits tend to become expansive or compressive in the ESP test. Or certain personality "types" may be more easily changed in their degree of expansion by the experimental conditions suggested above.

Another type of approach to the understanding of the relationship between expansion-compression and ESP is one designed to find out whether certain experimental conditions or instructions could radically change the expansion or compression of subjects with a resulting shift in ESP scores. For example, can the degree of expansion or compression be changed by instructions given hypnotically? Or can the experimenter by his attitude or manner of

handling the subjects change their expansion *and* their ESP scoring level?

Another possibility brought to mind by the results of the present study is that the method may be adapted as an adjunct to other ESP tests. It is conceivable that ESP card runs made in the same session as ESP drawings might be effectively separated on the basis of the expansion-compression in the drawings. The affect evoked by the card-calling test might carry over in the drawing task, leaving clues as to whether the subject was responding expansively or compressively in the ESP test. This particular line of research has already received considerable attention, and the results of several of these experiments will be published in the JOURNAL in a later issue.

Several other experiments have been conducted since the initial discovery that the expansion-compression criterion separates ESP results into two groups. Thus far, at least, this initial finding has proved to be fruitful of several significant series. These should not blind us to the possibility, however, that we may be able to find other similar criteria which may be even more discriminative of high-scoring and low-scoring ESP subjects. The discovery of this one form quality which does discriminate opens up a whole new area for research. In the literature on the characteristics of the art productions of both normal and abnormal subjects, there are descriptions of differences between the drawings of the two groups. In many points these descriptions closely parallel the criteria of Elkisch but some points in those studies are not covered by Elkisch's criteria. A study of these and perhaps an adaptation of them to numerical scoring may reveal additional discriminative measures as well as insights into the subjects' performance.

## REFERENCES

1. ELKISCH, P. Children's drawings in a projective technique. *Psychol. Monogr.*, 1945, **58**, 1-31.
2. HUMPHREY, B. M. Success in ESP as related to form of response drawings: I. Clairvoyance experiments. *J. Parapsychol.*, 1946, **10**, 78-106.
3. KENNEY, J. F. *Mathematics of Statistics.* Part II. New York: Van Nostrand, 1939.

4. RHINE, J. B., PRATT, J. G., STUART, C. E., and SMITH, B. M., with GREENWOOD, J. A. *Extra-sensory Perception after Sixty Years.* New York: Henry Holt, 1940.
5. STUART, C. E. An ESP test with drawings. *J. Parapsychol.*, 1942, **6**, 20-43.
6. ———. A classroom ESP experiment with the free response method. *J. Parapsychol.*, 1945, **9**, 92-105.
7. ———. An ESP experiment with enclosed drawings. *J. Parapsychol.*, 1945, **9**, 278-95.
8. ———. GESP experiments with the free response method. *J. Parapsychol.*, 1946, **10**, 21-36.
9. ———. A second classroom experiment. [Unpublished MS.]
10. ———. Unpublished research.

*Parapsychology Laboratory*
*Duke University*
*Durham, North Carolina*

# MINOR ARTICLE

## AN EXPLORATORY INVESTIGATION OF PK POSITION EFFECTS

By J. G. Pratt and J. L. Woodruff

**ABSTRACT: In this PK experiment, Drs. Pratt and Woodruff carried out 300 runs of dice-throwing in which each served as subject for the other. Two dice were thrown at a time in mechanically rotated tubes. The total deviation of the experiment was positive but not significant. There was a pronounced top-bottom decline in scoring in the record column. This decline was quite significant when its magnitude and consistency were considered together. Also, it was concentrated largely in the runs made with ones, twos, and threes as target faces, so that the top-bottom difference for those faces alone is significant. The subject who always led off in throwing for a new target face gave suggestive evidence of a "lag effect," or a tendency to score for the face just previously used as target.**

**These analyses are presented, not as conclusive evidence, but in support of findings of similar nature already published or about ready to be published.—Ed.**

### Introduction

In the summer of 1940 the writers became interested in testing the PK effect under conditions that would permit more sudden tumbling of dice than that generally obtained from the use of motor-driven, rotating cages. We felt that some method of releasing the dice mechanically so they could complete their fall relatively quickly under the force of gravity would encourage the subject to feel that he could influence the results by direct mental action. Also, by having the dice complete their fall more quickly, we could shorten the time interval during which the subject would need to concentrate upon controlling them. It seemed to us that using an instantaneous mechanical release of the dice would make the situation more like that of throwing them from a cup by hand, which most subjects seem to prefer to the rotating cages.

### Apparatus

As a result of speculation along the above lines, we designed and constructed a suitable apparatus by means of which dice are tumbled

in two closed celluloid tubes about 1½ inches in diameter and 18 inches long, one die in each tube. The tubes are connected to the apparatus at one end instead of at the middle, so that the free ends describe a circle of about 36 inches in diameter when the machine is in operation. They are engaged by the driving pulley in such a manner that when the motor runs at a constant speed they are lifted gradually, at the speed of the pulley, through a 180-degree arc to the top of the turn. After the tubes pass the highest point in the circle, they fall freely under the force of gravity through approximately 180 degrees to bump and come to rest against a padded barricade at the end of the complete turn. The padded barricade is connected off-center with the pulley so that it is depressed out of the way of the ends of the tubes by the time they are again engaged by the motor. Each tube has randomizing baffles which consist of a spiral of ten turns of rubber-coated electric wiring threaded alternately inside and outside the celluloid wall through holes at approximately one-inch intervals.

The dice tumble through the tubes as they complete their free swing and come to rest against the barricade. In testing PK with this apparatus, the observers read the dice once for each complete rotation as they come to rest in the free ends of the tubes at the low point of the turn. There is ample time to read and record the dice after they have been "thrown" and before the driving pulley, continuing at its own speed, completes a full turn and catches the tubes for the start of another rotation.

The apparatus may be driven at the speed desired. In our series it was set to take approximately eight seconds for each rotation. A five-digit counter attached to the apparatus was read at the beginning and at the end of each record page to provide a check upon the number of turns of the apparatus.

The dice used in this series had inlaid spots and the cubes measured approximately 9/16 of an inch on each edge.

## Procedure

The writers served as their own subjects in this experiment, and the PK singles test was used (see Glossary).

A plan for the work covered in the present report was prepared

and placed on file with the Record Librarian of the Parapsychology Laboratory before the experimental series was started. This provided for the use of officially stamped and numbered record sheets and set the length of the series at 150 runs for each subject, a total of 7,200 die observations. Each sheet was designed to record 12 runs or two "rounds" of the dice, one run by each subject with each face of the die as the target. Each run of 24 trials consisted, in this series, of a column of 12 two-digit figures showing what faces were on top in 12 successive falls of the dice.

The target faces were always used in ascending numerical order, the subjects completing one run apiece for a specified face before proceding to try for the next target. J.G.P. always led off in throwing for a face, and his results were always recorded in the top half of the page. Each completed page thus consists of top and bottom sections, each subject's half comprising six parallel columns, or one run for each face. No hits were marked on the record sheets at the time of the experiment, but the investigators independently counted the number of hits in each run and kept a permanent personal list of scores for the entire series. The results were independently rechecked by a member of the Parapsychology Laboratory staff to insure complete accuracy in counting hits.

In keeping with the general policy of the Laboratory, all subsequent analyses and computations in this report have been completely and independently rechecked.

## Results and Discussion

The experiment was started on September 14, 1940, and continued through September 18, 1940. During that time 25 "rounds" of the die by each experimenter, as called for in the experimental plan, were completed. No further work was deemed advisable at that time because the results as a whole were not significant. Many PK experiments had been completed before this one was done, but the final analysis and publication of the results had not yet been undertaken. We were therefore not able at the time we did this work to evaluate the results in the light of evidence of PK based upon scoring trends and position effects which subsequent analyses showed to be present in most of the PK experiments. As it turns out, there

are some interesting features about our results which come to light when such analyses are made upon these data. None of these is statistically significant in this series alone, but the results are in the direction that would be expected from previous findings and they are closely enough related to other studies to make their presentation here worth while for whatever confirmatory and suggestive value they may have.

There is another reason for reporting these results, one which goes beyond the fact that they lend some support to the published evidence for PK from position effects. The authors of those reports have repeatedly emphasized the importance of pushing analysis of existing data beyond the points already reached. They have recognized right along that their analyses of position effects were made chiefly from the point of view of adducing a new and particularly forceful kind of evidence that PK *occurs*, and that the important problem of the psychological significance of these scoring trends would require further study and analysis of the data at hand, and perhaps ultimately a fresh experimental approach along specialized lines.

One of us (J.G.P.) has become interested in the further investigation of position effects in tests of psi phenomena, and he took the opportunity to use the present experimental series as a pilot study in the exploration of possibly fruitful avenues along which such a research might be developed. This series is well suited to be the testing-ground of some exploratory analyses which may give clues to the psychological significance of scoring trends, because it is so short that different evaluations of the data, many of which may yield nothing of interest, will not be too time-consuming. Those analyses which do seem to raise important questions can then be extended to cover other PK and ESP series with greater insight and with more assurance of success.

*General Totals*

The entire experiment comprised 300 runs, or 7,200 individual die observations. In 1,221 instances the face which turned up was the target for that trial. The number of hits is therefore only 21 more than mean chance expectation. In the 150 runs in which

J.G.P. was the subject, 596 hits, or four below the chance average, were recorded; and in J.L.W.'s 150 runs the score was 625 hits, or a positive deviation of 25. None of these figures based upon the general totals is even suggestive of the presence of other than random factors in the fall of the dice.

*Rate of Scoring through the Series*

Several reports of PK experiments have presented evidence that the average number of hits decreases throughout a long series. This series is comparatively short, but it seemed worth while to compare the two chronological halves. Here are the results in terms of runs and deviations for the indicated divisions of the data:

| | J.G.P. | J.L.W. | TOTAL |
|---|---|---|---|
| First Half ..... | 75: −7 | 75: +13 | 150: + 6 |
| Second Half ..... | 75: +3 | 75: +12 | 150: +15 |
| Total ..... | 150: −4 | 150: +25 | 300: +21 |

It is evident that no series decline occurred, as these figures are well within the range of expected chance variation.

*Rate of Scoring in Relation to Target Faces (across the Page)*

Many reports of PK experiments in which different target faces have been used have shown a better rate of scoring on the higher faces (four to six) than on the lower ones (one to three). The same tendency was found in the present series, as the following deviations from the chance average show:

| | One-face | Two-face | Three-face | Four-face | Five-face | Six-face |
|---|---|---|---|---|---|---|
| J.G.P. ..... | −2 | −15 | + 1 | − 3 | +10 | + 5 |
| J.L.W. ..... | −6 | − 1 | +10 | +19 | −13 | +16 |
| Total ..... | −8 | −16 | +11 | +16 | − 3 | +21 |
| | | 150: −13 | | | 150: +34 | |

Diff. = 47
$SD_d$ = 31.61
$CR_d$ = 1.49
P = .07

The first impression created by these results is that the dice were biased to favor the higher faces, an effect that is theoretically expected when dice with excavated spots are used. In this instance, however, *inlaid* dice were used, and it does not seem reasonable to

suppose that any bias in the dice would be directly related to the numerical order of faces.

The above difference in favor of the higher faces may be due to one of three factors or to some combination of the three: (1) chance fluctuation; (2) psychological preferences for target faces affecting PK; or (3) position of runs on the record page affecting PK. The third possibility arises, of course, from the fact that each target was always used in the same position on the page. For ordinary purposes of evidence, we accept the explanation of chance fluctuation without question. But in the present report, where the primary interest lies in making trial analyses of these records for exploratory purposes, it is legitimate to proceed on the assumption that this increase in scoring across the page may reflect something in the experimental conditions which affected the rate of PK success across the page. One of the most consistent trends reported from other PK series has been a decline in scoring across the page when the same target is used for all the runs. The fact that we found a fairly consistent incline suggests, therefore, that in this instance the higher faces were preferred and that this factor was strong enough to offset the expected score decline. It must be emphasized, however, that this is only a suggestion of an effect of target preference.

*Vertical Decline in the Run*

Another scoring trend which has been found consistently in other PK series is the tendency for subjects to show a decrease in their rate of success as they progress through a short sequence of trials which are recorded in a column on the record page and which usually comprise a run, or 24 die readings in a singles test. The same target

| J.G.P. | | | J.L.W. | | | TOTAL | | |
|---|---|---|---|---|---|---|---|---|
| Quarters | Halves | | Quarters | Halves | | Quarters | Halves | |
| + 3 | +14 | | +14 | +29 | | +17 | +43 | |
| +11 | | | +15 | | | +26 | | |
| −17 | −18 | | − 4 | − 4 | | −21 | −22 | |
| − 1 | | | 0 | | | − 1 | | |
| Diff. = 32 | | | Diff. = 33 | | | Diff. = 65 | | |
| SD = 22.35 | | | SD = 22.35 | | | SD = 31.61 | | |
| CR = 1.43 | | | CR = 1.48 | | | CR = 2.06 | | |
| P = .08 | | | P = .07 | | | P = .02 | | |

face is always used throughout a run, so the present series is as well suited as any for the study of this kind of scoring trend. The run in this instance consisted of 12 rolls of a pair of dice, and the preceding figures show the results in quarters and halves of the run.

A decline in scoring as large as that between the first and second halves of all runs combined would occur by chance once in 50 such series.

*Run Declines in Relation to Run Scores*

A question which may prove to have bearing upon the psychological significance of vertical declines in scoring is the relation between the general level of success in the run as a whole and the distribution of hits within the run. The runs were separated into three groups according to the run scores and were checked for the decline effect in each group. It was found that all of J.G.P.'s decline, a difference of 32 hits between the first and second halves, was concentrated in his 64 runs which had a score of 3 and below (SD = 14.60; CR = 2.19; P = .01). He had 32 runs with a chance score of 4 and 54 runs which gave scores of 5 and above, and these showed no difference between the first and second halves. In J.L.W.'s case, the decline effect was about equally distributed among the three groups of runs: differences of 14, 8, and 11, respectively, for 46 "low" runs, 37 "chance" runs, and 66 "high" runs.

*Run Declines in Relation to Target Faces*

The fact that all of J.G.P.'s decline in scoring occurred in those runs with scores below chance led to speculation as to the meaning of this relationship, if indeed it indicates anything beyond an unusual chance coincidence. It will be recalled that the three lower target faces—which in this experiment were always the first half of the record page—gave lower scores than the three higher faces. It seemed possible, therefore, that the decline in run scores was occurring in the three runs comprising the left half of the page, and the apparent relationship between low scores and run decline might have resulted from the fact that fewer hits were obtained in these runs, either as a result of their position on the record page (which is contrary to what one would expect on the basis of the QD studies) or because of factors of psychological preference related to the use of

different target faces. In spite of the difficulty of distinguishing between these possibilities on the basis of the present results, we decided to divide the runs into two groups on the basis of target faces and compare the vertical declines in the two groups.

Targets One to Three, First Three Runs on Page

| J.G.P. | | J.L.W. | | COMBINED | |
|---|---|---|---|---|---|
| Quarters | Halves | Quarters | Halves | Quarters | Halves |
| + 6 | + 9 | + 7 | +19 | +13 | +28 |
| + 3 | | +12 | | +15 | |
| −10 | −25 | −12 | −16 | −22 | −41 |
| −15 | | − 4 | | −19 | |
| Diff. = 34 | | Diff. = 35 | | Diff. = 69 | |
| SD = 15.80 | | SD = 15.80 | | SD = 22.35 | |
| CR = 2.15 | | CR = 2.22 | | CR = 3.09 | |
| P = .02 | | P = .01 | | P = .001 | |

Targets Four to Six, Last Three Runs on Page

| J.G.P. | | J.L.W. | | COMBINED | |
|---|---|---|---|---|---|
| Quarters | Halves | Quarters | Halves | Quarters | Halves |
| − 3 | + 5 | + 7 | +10 | + 4 | +15 |
| + 8 | | + 3 | | +11 | |
| − 7 | + 7 | + 8 | +12 | + 1 | +19 |
| +14 | | + 4 | | +18 | |
| Diff. = 2 | | Diff. = 2 | | Diff. = 4 | |
| SD = 15.80 | | SD = 15.80 | | SD = 22.35 | |
| CR = .13 | | CR = .13 | | CR = .18 | |
| P = .45 | | P = .45 | | P = .43 | |

The striking feature here is that all the decline occurs in the first three runs of the page, which are also the runs with the three lowest faces as targets. In future series it will be important to see whether this effect is a real one in PK results and to find out whether it is causally related to the position of the run on the page or to the use of different target faces.

A start toward answering this question might be made with the present data by finding out how the decline in scoring within the run is distributed for each run position on the page. The figures shown below give the differences between the top and bottom halves of the runs. Those given with a plus sign indicate a decline in scoring, while the minus sign represents an increase in hits in the bottom half of the run.

| | Column 1 (One-face) | Column 2 (Two-face) | Column 3 (Three-face) | Column 4 (Four-face) | Column 5 (Five-face) | Column 6 (Six-face) |
|---|---|---|---|---|---|---|
| J.G.P. ..... | +12 | +11 | +11 | +1 | +2 | −5 |
| J.L.W. ..... | +18 | +15 | + 2 | +7 | −1 | −8 |
| Total ..... | +30 | +26 | +13 | +8 | +1 | −13 |

It is evident from these figures that the rate of scoring within the run for both subjects combined is very closely related to the position of the run on the page, as the probability that the six runs would show a progressively diminishing rate of decline across the page is $(\frac{1}{2})^6$, or approximately three in 100. By comparing these results with the run scores for the different targets as given earlier in this report (see p. 201), it is equally evident that the run declines are not so closely related to the rates of scoring with the different faces. This suggests that run position on the page is the more important factor in determining the vertical declines in scoring.

The extent of lawfulness shown in the vertical decline is evident when both the top-bottom difference for all runs combined and the distribution of these differences across the page are considered together. These are statistically independent effects, as the presence of a vertical decline itself tells us nothing about how the differences will be distributed among the run positions of the page, and the two probabilities may therefore properly be combined. The likelihood that the present results might occur by chance is the probability for the observed top-bottom decline (.02) *times* the probability for the observed manner in which that decline was distributed on the page (.03), or six in 10,000. This finding gives considerable reason for concluding that the position effects found in the present study are genuine PK results in spite of the shortness of the series and the exploratory character of the analyses.

*Lag Effect*

Each subject, it will be recalled, did one run for each face on each record page, starting with ones and taking the faces in numerical order. J.G.P. always threw his run for a target face before J.L.W. threw his. It is of interest to check each run for the number of hits made upon the *preceding* target face, for there has arisen among investigators of PK a general impression that subjects may show psychological inertia in changing targets, in that they continue

to score above chance on the just-preceding face. The expectation in the present instance would be that, if a lag effect occurs, J.G.P.'s data would show it more clearly, for he was the one to make a run immediately after the target was changed.

There are only 10 runs on the record page, five for each subject, that can be scored for lag effect, as the first run is ordinarily the beginning of a session and there is no reason to expect the motivation in throwing for the six-face to persist from the end of one page to the beginning of the next. The procedure used is to check the second run for ones (the target of the first run), the third for twos, etc. This results in a total of only 250 runs in the entire series to be checked for lag effect. The following results were obtained:

| | J.G.P. | J.L.W. | TOTAL |
|---|---|---|---|
| Lag Score | 125 : +42 | 125 : + 7 | 250 : +49 |
| Corrected Lag Score[1] | 125 : +41.4 | 125 : +13 | 250 : +54.4 |
| SD | 20.40 | 20.40 | 28.85 |
| CR | 2.03 | 0.64 | 1.89 |
| P | .02 | .26 | .03 |

This result is not statistically significant, but it certainly has some suggestive aspects. The fact that J.G.P.'s runs show the higher lag score is in line with expectation, as he was the first to throw for a new target. While we do not wish to strain the evidence by placing weightier interpretations upon it than the facts will bear, we do feel that this finding adds to the importance either of planning experiments to minimize the possibility of a lag effect or of taking this factor into account in the evaluation of the results.

A check upon the distribution of "lag hits" within the run shows that J.G.P. scored two more successes in the top half of his runs than in the bottom, and J.L.W. scored 29 more hits in the bottom half, a net *incline* of 27 hits. (SD = 28.85; CR = .94; 2P = .35. It is appropriate to double the probability value here, as there existed no basis for predicting the direction of difference in this situation.) These lag effect results are not significant, but they seem worth recording in that they raise the question whether *displacement* of

[1] This correction of the lag score takes account of the degree to which the deviations observed for the target faces affect the expected number of hits on the other five faces.

scoring in PK tests and direct hits on the target will show similar internal scoring trends.

## Concluding Remarks

In view of the exploratory nature of these analyses and the consequent factor of selection which enters into the more outstanding results, no definite conclusion regarding the operation of PK in producing these results is justified. This report of the findings will have served its purpose if it stimulates the thinking of research workers and suggests fruitful avenues for further studies into the psychological significance of position effects.

*Parapsychology Laboratory*
*Duke University*
*Durham, North Carolina*

*College of the City of New York*
*New York, New York*

# RESEARCH NOTES

## THE SCHWARTZ PK EXPERIMENT

THE FOLLOWING brief report of a PK experiment, carried out by Mr. Marvin Schwartz with the use of the dice-throwing technique, is of interest mainly for two features: first, for a significantly *low* scoring effect obtained in one section of the experiment, and, second, for the distribution of hits with respect to position of the trial on the record page and in the set. When Schwartz turned in his records to the Parapsychology Laboratory at Duke, he did so with the statement: "In my unauthoritative opinion the results contribute nothing." This statement was based on the fact that the total deviation for the experiment was quite insignificant. However, later analyses carried on at the Parapsychology Laboratory under Dr. Betty M. Humphrey brought out points of interest which fully justify the following brief description.

The experiment was carried out while Schwartz, who is now a student at Siena College, was serving in the United States Army. The experiment was planned in correspondence with the Parapsychology Laboratory and was recorded on numbered record sheets obtained from there. The main purpose in the design of the experiment was to compare position effects in data where the recording was done in sets of different size. One hundred and forty-two runs (of 12 throws of a pair of dice) were to be made with each of three types of record sheet. The first type of sheet was to have eight sets on the page, with each set consisting of six columns of six entries each. These sets were arranged on the page in two columns of four sets each. A second type of record sheet consisted of four sets to the page, one to each quarter, with each set consisting of six columns of 12 entries each. The third type of sheet had two sets, one on the left and one on the right, each consisting of six columns of 24 entries each. It was thought that with all other conditions of the experiment essentially similar, the differences in structure of the sets would produce differences in distribution of the hits in the set.

Schwartz himself did all the throwing of the dice and all the recording, and there were no witnesses. Plain commercial dice, 5/16 of an inch in size, were used. All six faces were equally represented as target face to eliminate any effect, on the results, of bias in the dice. The dice were thrown from a cup, after being shaken, and were bounced on a padded surface against a backwall, the whole situation being improvised for the purpose under the varying conditions of camp life.

## Experimental Results

A total of 432 runs gave a positive deviation of only 2. However, when the data are presented separately for the three types of sets quite a difference is found, as may be seen in Table 1. The six

Table 1

Total Results for the Three Different Sizes of Sets

| Type of Set | Runs | Dev. | Av. Score | CR | $\chi^2$ |
|---|---|---|---|---|---|
| Six by Six | 144 | −60 | 3.58 | 2.74 | 7.51 |
| Twelve by Six | 144 | +32 | 4.22 | 1.46 | 2.13 |
| Twenty-four by Six | 144 | +30 | 4.22 | 1.37 | 1.88 |
| Total | 432 | + 2 | 4.01 | | $\Sigma\chi^2=11.52$<br>$P=.01$ |

by six set shows a total deviation of −60, which gives a significant CR of 2.74. The twelve by six and the twenty-four by six sets both give positive deviations, but neither is significant. The table also shows the chi-square combination of the CR's of all three sections, with its significant probability of .01.

If the tests had been conducted with alternation of the order of different types of record sheets, we would be better able to make comparisons between the three types of sets. But since Schwartz carried out the experiment by completing first the six by six section, then the twelve by six, and finally the twenty-four by six, the difference may be due to the chronology of the series rather than to the type of set. The first Frick series began like this one, with a negative deviation (1).

Nevertheless, it is of interest to note that in spite of using the same dice, the same manner of throwing, and similar experimental conditions, the six by six set should have given so large a deviation in comparison to the other types of sets. It may be worth while to mention in this connection that Schwartz's attitude toward the PK hypothesis was one of considerable skepticism, although it was accompanied by an open-minded curiosity and a willingness to experiment with the problem. We do not know enough yet about the role of attitude to say to what degree the intellectual conflict involved might unconsciously have been a score-depressing factor in the experiment, making for a negative deviation. But we know from Schmeidler's work in ESP (3) that attitude toward the hypothetical ability being tested is most important in its effect on scoring rate in psi experiments.

From the point of view of *size* of deviation and consequent CR the smallest set has the advantage. However, more information is needed to permit generalization on this point.

### Position Effects

As shown below, all three types of sets gave a horizontal decline from the left half to the right in terms of size of deviation (though not in scoring rate). Each figure represents 24 runs.

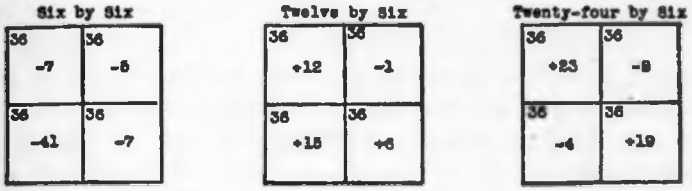
| Type of Set | Left Half | Right Half |
|---|---|---|
| Six by Six | −48 | −12 |
| Twelve by Six | +26 | + 6 |
| Twenty-four by Six | +18 | +12 |

The vertical distribution in the three types of sets is comparable only if we divide the column into fractions, since the columns for the three types of sets are of different length. On the basis of com-

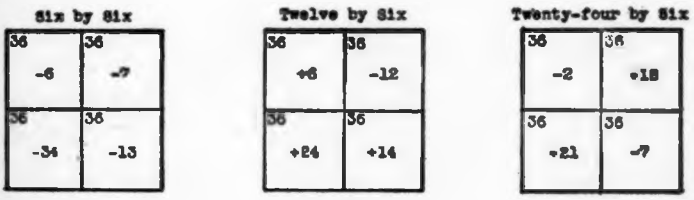
| | Six by Six Sets | Twelve by Six Sets | Twenty-four by Six Sets |
|---|---|---|---|
| 1st third | − 7 | + 6 | + 5 |
| 2nd third | −20 | +11 | +10 |
| 3rd third | −33 | +15 | +15 |

parison of thirds of the column, however, we can see that there is in all three sets a steady incline in size of deviation, regardless of scoring rate.

Turning now to the quarter distribution of the set, we have the following figures for the three types of sets.




It is particularly interesting to compare these QD's of the set with the QD's of the three types of record pages. The latter are as follows:




These QD's are not typical of QD's in general as found in other PK researches. They all represent distinct departures from type. The usual thing has been to find the upper left quarter giving the highest scoring rate, and the lower right the lowest. Here we find both in the set and in the page that in the six by six and the twelve by six sets the lower left quarter gives the highest deviation, disregarding signs. Since the page QD and the set QD for the six by six and the twelve by six types of sets are mutually independent it is especially interesting to see the striking similarity in QD's between the set and the page.

By way of a concluding word, it may be said that the experiment suggests that the six by six type of set is the best of the three used, from the point of view of the size of deviation and correspond-

ing CR, and that it shows the most marked range of distribution of hits in the vertical and horizontal, as well as the quarter distribution. It would seem worth while, then, to conduct such experiments with the smaller type of set. This conclusion is in line with the comparative observations made by Rhine and Humphrey in their article on the QD of the set (2).

### References

1. Rhine, J. B., and Humphrey, B. M. PK tests with six, twelve, and twenty-four dice per throw. *J. Parapsychol.*, 1944, 8, 139-57.
2. ———. The PK effect: special evidence from hit patterns. II. Quarter distributions of the set. *J. Parapsychol.*, 1944, 8, 254-71.
3. Schmeidler, G. R. Progress report on further sheep-goat series. *J. Amer. Soc. psych. Res.*, 1946, 40, 34-36.

J. B. R.

# BAILEY'S COMPARISON OF A COIN AND A DIE IN PK TESTS

THE TWO experiments described here were performed informally in his own home by Mr. Wilber E. Bailey of Palo Alto, California. He conducted them for his own satisfaction by way of testing his PK ability, and with no intention, at the time, of reporting the results; but they offer suggestive data on the comparative success of dice and discs in PK experiments and are reported with that in view.

## I. THE PENNY-TOSSING EXPERIMENT

The first experiment was carried out in 10 different sessions between February 1, 1945, and March 13, 1945. Mr. Bailey, who was the only subject, thoroughly shook a penny in a small drinking glass which was approximately two inches across and four inches deep. He tossed the penny out upon a large floor rug, trying by PK to make it land with the selected face uppermost. Each session lasted about half an hour and there were 100 trials at each one, a total of 1,000 throws in the whole experiment. With a view toward avoiding the results of fatigue or boredom, the 10 sessions were prolonged over six weeks. The target face for the first session was heads; for the second session, tails; and so on through 10 sessions, with the target face alternating in order to offset any bias in the coin. The same coin was used throughout the experiment.

The scores for the 10 sessions are as follows:

| Session | Target | Number of Heads | Number of Tails | Deviation |
|---|---|---|---|---|
| 1 | Heads | 51 | 49 | + 1 |
| 2 | Tails | 41 | 59 | + 9 |
| 3 | Heads | 63 | 37 | +13 |
| 4 | Tails | 49 | 51 | + 1 |
| 5 | Heads | 50 | 50 | 0 |
| 6 | Tails | 42 | 58 | + 8 |
| 7 | Heads | 45 | 55 | — 5 |
| 8 | Tails | 47 | 53 | + 3 |
| 9 | Heads | 48 | 52 | — 2 |
| 10 | Tails | 43 | 57 | + 7 |
| Total | | 479 | 521 | +35 |

From the 1,000 trials there were 535 hits, which is 35 above mean chance expectation and which gives a CR of 2.21. This result would be expected by chance about once in 70 such series. The excess of hits was distributed fairly evenly throughout the experiment: in seven of the 10 sessions, the face thrown for was the face which received a majority of hits; in two, it was negative; and one session was a tie.

## II. The Single-Die Experiment

In the period between December 13, 1945, and February 20, 1946, nine months after completing the penny-tossing experiment, Mr. Bailey made a total of 216 throws of a single die. The same die was used throughout the experiment. It was the common commercial type, measuring approximately three-fifths of an inch and was marked with excavated spots. It was shaken in the same glass used in the previous experiment and rolled onto a linoleum-covered floor. Mr. Bailey again was the only subject and did his own recording.

The die was thrown in series of 36 throws, one such series at each of six sessions, and with the sessions anywhere from a day to a month apart. The one-face was the target for the first series; the two-face, for the second; and so on.

In a total of 216 throws there were 47 hits, where mean chance expectation would be 36. This gives a CR of 2.01. Such results would occur by chance about once in 40 such series. As may be seen from the following figures, the scores were fairly well distributed throughout the experiment.

| | One-face | Two-face | Three-face | Four-face | Five-face | Six-face |
|---|---|---|---|---|---|---|
| Hits | 8 | 9 | 8 | 3 | 10 | 9 |
| Dev. | +2 | +3 | +2 | −3 | +4 | +3 |

Since there were an equal number of throws for each face of the die, possible bias in the die could only affect the total results adversely as far as obtaining otherwise significant deviations from the theoretical mean is concerned.

When the results of this experiment and the preceding one are combined by the chi-square method, they give $\chi^2 = 8.92$, which

has a probability of .012. This is close to the criterion of significance (.010).

Aside from the fact that these two experiments are marginally significant, they are of interest because they show that a single subject achieved an approximately similar order of success in throwing two-sided objects as in throwing cubes.

D. H. P.

# 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.

*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. 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.

CALL v.: To attempt to identify by ESP a stimulus object or a mental state of an agent.

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 expressed as an expected score, generally in terms of average per run.

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.

CLAIRVOYANCE: Extrasensory perception of objective events as distinguished from telepathic perception of the mental state of another person.

*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 average scores of two samples of data divided by the standard deviation of the difference.

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 elementary unit of data in a dice-throwing test represented by the throwing and reading of a single die, no matter how many dice are thrown at a time.

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 the hypothesis being tested.

ESP (EXTRASENSORY PERCEPTION): Response to an external event 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.

EXPECTATION; see CHANCE.

EXTRACHANCE: Not due to chance alone.

FREE MATERIAL: The stimulus objects of experiments in which an unlimited or unspecified range of stimulus objects is employed (as contrasted with methods such as card-calling in which the subject knows that the stimulus object is one of a known range).

FREE RESPONSE METHOD: An ESP test method in which a relatively unlimited range of stimulus objects is used and in which the subject is instructed to express freely as his response any range of thoughts and ideas.

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 TEST: A PK technique 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 more.

LOW-DICE TEST: A PK technique 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 less.

MEAN CHANCE EXPECTATION; see CHANCE.

*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 those psychical effects which appear not to fall within the scope of what is at present recognized law.

PERCIPIENT: The person who makes the calls in an ESP test.

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

PRECOGNITION: Cognition of a future event which could not be known through rational inference.

PREFERENTIAL MATCHING: A method of scoring free responses. A judge ranks the stimulus objects (usually in sets of four) with respect to their similarity to, or association with, each response; and/or he ranks the responses with respect to their similarity to, or association with, each stimulus object.

PSI: A general term to identify personal factors or processes in nature which transcend accepted laws. It approximates the popular use of the word "psychic" and the technical one, "parapsychical."

PSI PHENOMENA: Occurrences which result from the operation of psi. They include the phenomena of both ESP (including precognition) and PK.

QD (QUARTER DISTRIBUTION): The distribution of hits in the record page (or in a logical subdivision thereof, such as the set or the half-set) as found in the four equal quarters formed by dividing the selected unit horizontally and vertically.

RUN: A group of trials, usually the successive 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.

SCORE: The number of hits made in one run.

TOTAL SCORE: Total of scores made in a given number of runs.

AVERAGE SCORE: Total score divided by number of runs.

*SD (STANDARD DEVIATION): Usually 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 $= 2\sqrt{\text{no. of runs.}}$) Sometimes the SD used is that estimated from the observed variability in the scores.

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 $\sigma_s$ is the SD of a single run and $R_1$ and $R_2$ are the number of the runs in the respective samples compared. This gives the SD of the difference for run score averages.

SERIES: Several runs or experimental sessions that are grouped in accordance with a stated principle.

SESSION: A unit of an ESP or PK experiment comprising all the trials of one test occasion.

SET: A subdivision of the record page serving as a scoring unit for a consecutive group of trials, usually for the same target.

SEVENS TEST: A PK technique 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 TEST: A PK technique in which the aim of the subject is to try to influence dice to fall with a specified face up.

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

SUBJECT: The person who is experimented upon. In ESP tests, most commonly the percipient (though also the agent in GESP and telepathy). In PK tests, any individual whose task it is to influence the objects thrown.

TARGET: In ESP tests, the stimulus object; or in telepathy, the mental state of the agent. 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: In ESP tests, a single attempt to identify a stimulus object. In PK tests, a single throw of the dice or other objects thrown.