The Journal of Parapsychology

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

Volume 7 SEPTEMBER, 1943 Number 3

Contents

DUKE UNIVERSITY PRESS
DURHAM, N. C.

The Journal *of* Parapsychology

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

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

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

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

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

DUKE UNIVERSITY PRESS

The Journal *of*
Parapsychology

Volume 7 SEPTEMBER, 1943 Number 3

EDITORIAL

SIGNIFICANCE OF THE PK EFFECT

WHILE THE publication of the PK research continues, let us assume, for purposes of discussion, that the PK hypothesis will be established, and proceed to ask, "What difference does it make?" We raise this question now because the average individual needs to have the bearing of an investigation in his perspective while consideration of the evidence is going on. For that matter, the experimenters themselves, both new and old, are, from the very outset, greatly concerned with this question of the significance of their findings. It is, after all, only from such glimpses of the larger meaning of their work that the inspiration and encouragement to undertake their uncertain labor must largely derive.

At the present stage of published knowledge of the PK research, only a few of the main points of significance can profitably be indicated without undue speculation, although, as the series of reports continues, more and more factual material will be made available, and this will considerably enlarge the scope of interpretation and the application of the findings. But even now, enough may be said about what the PK effect means to warrant this brief introduction.

Quite naturally, the minds of many lay-readers turn first to the more objective and practical applications of the PK work. This work comes first to be thought of merely as "dice-throwing tests" and the bearing upon other kinds of dice-throwing and comparable pastimes is at once suggested. But, like the ESP work, the PK studies are in too unfinished a state to permit any practical use in the dice-throwing

"arts" in general. And we need hardly add that the experimenters themselves have not been greatly concerned with this field of application.

For Medicine

A more thought-worthy field of application, one that is still within the range of the practical, is that of medicine. For many years now, medical literature has been taking with increasing seriousness reports of mental effects upon the organic system. Whether these effects arise as a result of the magical practices of a primitive medicine man or the slightly more familiar suggestion of the hypnotist, these faith-healings and other psychogenetic effects on the tissues of the body have been more thoughtfully regarded than ever before. Psychophysiological hypotheses, such as that which Professor W. B. Cannon offered as an explanation of voodoo death, have come into consideration. But clearly we are at a loss as yet for an adequate explanation for the effects as reported. We simply do not know of any intermediating mechanism that could account for some of these reported organic effects of the mind.

Now, however, the PK research has provided a specifically applicable explanation. It is not a difficult logical jump at all to say, "If a subject can influence the fall of dice by his thinking, he can surely be expected to influence the physiological processes of his tissues, such as the movement of corpuscles and foreign organisms, the functions of healing and growing, and the operation of disease and repair in general." Sufficient impression has already been made upon some members of the medical profession by the evidence for mentally determined organic effects to justify this suggestion of the application of the PK hypothesis. We do not at all draw any conclusions here; this is rather to point to a reasonable possibility regarding the PK hypothesis in a sphere of great practical importance.

For Evolutionary Theory

Less utilitarian, but surely as specific, is the bearing the PK research will properly have upon the question of the causes of organic evolution. The origin of species is not by any means fully understood. We are still in the unproved-hypothesis stage as to *how* those changes have come about which have survived in the various species as evolutionary change. One of the oldest hypotheses, and one of the

most unpopular among biologists today, is the suggestion of Lamarck that in the animal kingdom, where purposive effort is conspicuous, the striving of the organism to adjust to his environment succeeds in producing heritable changes that make for its survival in the struggle for existence.

While this hypothesis has never been in any way disproved (Weismann's experiment was not a relevant test), the evidence in favor of the Lamarckian view, on the other hand, has likewise never been decisive. The antagonism toward this view has been aggravated by our lack of any plausible explanation as to *how* the effort of the organism could be brought to bear upon the tissues to be affected, and above all upon the germ-plasm. If, however, the mental system of the organism can to some degree dominate the physical world about it, why not suppose that its own bodily processes are within the scope of its influence? No matter, then, what one's favorite speculation about organic variation may be, the PK hypothesis, if taken at its full significance, gives new meaning and plausibility to the Lamarckian conception of evolution. If this be correct, a major problem in biology may become closely identified with what promises to be a much-discussed issue in psychology.

For Psychology

It is, however, in psychology itself and its branches that the major effect of the establishment of the PK hypothesis should be felt. To take PK seriously in the world of the mind is to do for modern psychology what the Copernican revolution (from geocentrism to heliocentrism) did to the beginnings of modern astronomy. It shifts the "center of the personal universe"! All who are familiar with even the vaguest outlines of the history of psychology are aware that for several centuries the trend has been to decentralize the mind in the personality system and to make the nervous system the real determinant of human conduct. The great progress of the physical sciences has required that this physical system, primarily the brain, be considered the actual center of causation in the personality. The great development of physiology and, in psychology, of its physiological branch, has so far outdistanced progress in understanding of the processes of the mind in its other aspects as to offer the student very little but the physicalistic approach to the interpretation of per-

sonality. The patterns of thought and even the terminology of physics have dominated the science of psychology to the extent that one might well hesitate to contest these modernistic versions of the older materialism. For what, indeed, has one had to offer to the contrary except the evidence of common sense and tradition?

Now, however, one has either to refuse to accept the PK effect or else undergo a complete revolution in his mental philosophy. For the PK principle indicates the mind to be a real force, able to go effectively beyond its own physical organism. And, along with ESP, it points to an order of psychical causation which by present conceptions is clearly nonphysical, yet one which is capable, as the experimental data show, of really influencing the physical world in a purposively intelligent manner.

If the reader is not sufficiently familiar with the ESP research and has not been led to the conclusion mentioned regarding the extraphysical character of the mind, he may nevertheless reach the same viewpoint on the basis of the PK research alone. It will, in fact, be more obvious by far than it was with the studies in ESP. For it must be immediately clear to everyone that if rolling dice can be significantly influenced by the subject who is aspiring to make them turn up with a specified face, this could only be done if the influence is a *knowing* one. No mere "ignorant" physical force of any type known to physics could be expected to follow up these moving, bounding bodies and wilfully change their course in the right direction at the strategic instant and point. Indeed, some cognitive rather than physical force is in action; some intelligent purpose is in operation. This is the human *mind* doing *physical work*, but with its own peculiar forces. It is as simple as that, so long as we do not ask, "How is it done?" But that question we have long been used to being unable to answer in matters concerning the mind.

Psychology, then, if it has to accept the PK investigation, will have to return to the mind, and this mind to which it must return is not merely an abstraction. It is a mind which has real energy, which does real work, actually influences moving bodies. Curiously enough, this "mind" is fundamentally much the same old conception as that from which the infant science strayed off, away back in the eighteenth century, when it followed the exciting but (for psychology) misguiding lead of the physical sciences.

Questions for the Future

Is all this too revolutionary for conservative minds? Perhaps this is the place, then, to raise some questions which will further test the entire issue. What, for instance, does the student of psychology say at this point who has been brought up on the "cerebrocentric" view (which makes the brain the center of psychology)? He will likely demand a further sign, some verifying tests, before making the great concession of an extraphysical principle in personality. He might appropriately inquire, "Does this so-called PK effect itself follow the familiar lines of physical law, or does it anywhere definitely depart from it? To point up this question further, for example, what is the effect when large and small dice are used in the PK tests? How about dice with different densities? What happens when different distances are interposed between the subject and the moving bodies, or different numbers of dice thrown at a time?"

These are specific, usable questions for research and, one by one, they will be taken up in the reports of PK experiments appearing in the future. Some little indication of the trend is already available in the articles on PK now in print. In this issue, for example, there is one report of tests in which different numbers of dice were thrown at a time. In the preceding number, another appeared. A comparison of *size* of dice, too, has been made. These are at least beginnings, whether or not they are decisive.

What if time after time the research should show that PK defies mechanical principles and follows psychical laws instead? Perhaps after all, physicalistic psychology is a tissue of untested hypotheses that we have never applied in practical life. As it is, we may very well be watching a persistent and complete transition from a cerebrocentric to a psychocentric view of man as we see the PK reports issue through the medium of these pages.

EFFECT OF INCENTIVES ON ESP AND VISUAL PERCEPTION[1]

By J. L. WOODRUFF and GARDNER MURPHY[2]

ABSTRACT: In the tests reported here, the attempt was made to discover whether certain incentives would improve performance in ESP tests with college students as subjects. Along with the ESP tests, the subjects were given parallel tests in near-liminal vision in which faintly illumined symbols, like those used in the ESP tests, were employed. This was done with the purpose of comparing sensory with extrasensory perception under a variety of conditions. The results of the extrasensory perceptive tests are conclusive, the sensory tests not. The tests with incentives gave significantly higher score averages than those without. (Statistical difficulties prevent similar comparison of the vision test results.) A moderate improvement in scoring was obtained merely by informing the subject of his scores. A larger improvement accompanied the giving of small money rewards for high scores.—Ed.

THE PROBLEM

AN INVESTIGATOR in extra-sensory perception is confronted with a peculiar difficulty. The phenomenon which he is purporting to test is by no means recognized by all scientists to be a real one. An individual may study the effects of variations in motivation on learning, on memory, and on near-liminal perception, and the interest of his colleagues will center on whether the variations mentioned were ac-

[1] Thanks are due the Hodgson Fellowship Committee, and especially to its Chairman, Dr. E. G. Boring, for their understanding cooperation during our period of research work under the Committee's auspices. This spirit of cooperation was manifested in part in their willingness to prepare many thousands of random numbers for our use during the course of the research.

[2] The research was conducted by Dr. Woodruff and the report is from his hand; Dr. Murphy agrees to the use of his name provided it is made clear that he contributed nothing to the work except encouragement and counsel. This article was originally submitted as a research report to the Hodgson Fellowship Committee of Harvard University. Since the first writing, additional research work has been completed. That part of it bearing especially on the ESP question has been incorporated in the tables, and those conclusions affected by the additional data have been rephrased. Otherwise, the report remains essentially as presented to the Committee. Dr. Woodruff's entrance into the armed services has prevented incorporation and proper analysis of all the data concerned with the section on near-liminal perception.—Ed.

companied by variations in performance. No one will deny that learning, memory, and near-liminal visual perception are real functions of the organism. The attitude taken in the present report is this: It is recognized that since ESP has not been universally accepted as a real phenomenon, any investigation purporting to study it will be judged mainly in the light of the question whether it constitutes positive evidence of the phenomenon. It is urged, however, that an investigator may also be interested in the *psychological relations of ESP, if it is a real phenomenon.*

The language in the present report may suggest that the writer is convinced that ESP is an established phenomenon. This is primarily a matter of convenience. It is impossible to reiterate the phrase, "if ESP is a real phenomenon," each time the term is used. However, the results obtained will be discussed, first in the light of the question as to whether they constitute positive evidence for the existence of ESP as a real phenomenon; and second, as they bear on the psychology of ESP, if it actually occurs.

Generally speaking, experimental variables which are effective in producing variations in one mode of behavior will be effective in producing variations of a similar nature in some related type of behavior. In this respect, the functions of the organism are apparently closely related. Variations in the motivational aspects of a situation will probably be accompanied by variations in the rate of learning, or by variations in the amount remembered. ESP investigators have been interested in the relation of ESP to a number of variables; thus the effects of variations in motivation brought about by competition, or by individual money rewards, have been studied. But these investigations have not as a rule been conceived with the thought of making a definite quantitative comparison with one other closely related function of the organism. The present project is an attempt at a comparison of this kind.

As stated, the problem investigated here is: What effect do variations in motivation have on extra-sensory perception and on near-liminal visual perception? The question arises as to why the comparison is made between ESP and near-liminal visual perception. As contrasted with ESP, near-liminal visual perception is a "known," in the sense that its occurrence is not questioned among psychologists. We have a certain, though surprisingly limited, amount of knowl-

edge about its mechanism and about certain psychological factors which affect it. In comparing a "known" with an "unknown," one is guided by the possibility that through such a comparison his knowledge of the mechanism of the "unknown" will be increased.

Also, we are in both cases dealing with a type of perception. The *task* of the subject is the same in the two situations. He must respond to the presentation of a given stimulus by naming it. The frequency with which he is able to do this is the measuring stick by which we decide that perception has (or has not) taken place.

Third, it was possible to make the objective experimental situations for the visual perception and for the ESP situation very similar. Such similarity would not have been possible had memory or learning been one of the functions involved in the comparison.

Motivation is the parameter involved in this comparison. It was chosen as the experimental variable for certain reasons. First, ESP seems especially sensitive to changes in the motivational aspects of an experimental situation. Many of the abrupt declines in scoring have been attributed to certain changes in motivation. As an example, Pratt and Woodruff (**5**) found that scoring decreased with the length of time a specific type of stimulus material was used for a given subject but increased when a "new" type of material was introduced. Miller (**3, 4, 7**) found, in a series of papers, that sub-liminal learning varied with changes in motivation; and Woodruff (**8**) found that near-liminal visual perception was susceptible to certain objectively gross variations in the motivational situation. The special motivations used in the present study are *a.* rewards; and *b.* knowledge of results.

Experiments involving changes in the objective motivational aspects of a situation are, in a sense, circular. We can be sure that we have subjective changes in motivation only when the efficiency of the performance changes. On the other hand, we are interested in determining whether a change in motivation will bring about a change in the efficiency of performance. In the present case, we shall argue that the changes in the objective aspects of the situation will constitute a change in motivation. Thus, if we in one case tell the subject to do the best he can in the performance of some task, and in another case offer him a monetary reward for improvement, we shall assume that the motivation has been changed. This has some justification since monetary rewards have been found to be effective in producing im-

provement in certain types of performance as compared with instructions alone. On the other hand, it is recognized that should the monetary reward be too small or should the subject have all the money he wants, he will actually not be motivated to a greater degree.

Experimental Conditions

A two-room set-up was used in the present research. The subject sat in a small alcove, the walls and floor of which formed an isosceles triangle. At the back of the alcove was a small window against which the subject's chair was placed in such a way that a dim light filled the room. Normally the alcove opened into a larger room in which the experimenter was stationed. For purposes of the experiment, however, the opening from the alcove into the larger room was completely closed, partly by a permanent partition, and partly by a screen, hinged in such a manner that it could be swung aside to allow entrance into the alcove. By means of the screen and the partition, the subject and the experimenter were completely shut off from each other. A small ground-glass screen was constructed in the center of the screen, and the subject's chair was placed at a distance of approximately eight feet from it in such a position that his eyes, with the head held level, were in the same horizontal plane as the center of the glass. The subject's head was held in a fairly stationary position by means of a rest constructed of elastic and leather belts.

Located in the experimenter's room was a table, the surface of which was approximately two by three feet. Rollers were fastened to the legs of the table so that it was movable. Upon this table, another table approximately two by two feet was constructed. At the back of this second table (the point farthest from the ground-glass screen), a small table approximately six by ten inches was built and on this, a projector (Kodaslide projector, Model 2, Eastman) was placed. On the second table, and approximately a foot in front and six inches below this projector, another projector of the same type was placed. Both projectors were fastened to their supports and could not be moved. The projectors were placed in such a position with relation to the ground-glass screen that their barrels were in the same vertical plane and both were focused on the center of the glass screen. The upper projector was placed in a circuit in which the voltage was

kept constant by means of a voltage stabilizer. The purpose of this projector was to furnish a constant light upon the glass screen. For use in the lower projector, slides, two inches square, were constructed. The slides were made from 35 *mm.* positive film strips. The slides bore the five ESP symbols: circle, square, cross, star, and wavy lines. The lower projector was placed in a circuit with a rheostat in such a way that the voltage of the current could be varied from 0 to 110 volts. By means of this projector, the stimuli on the slides could be shown on the glass screen at different intensities. They could be made to appear subjectively from "very vivid" to "non-existent."

In front of the projector which presented the stimuli, an Ilex shutter was fastened about six inches in front of the projector; and across the intervening open space, a small fan blew. This was necessary to prevent warping of the shutter leaves. By the use of an automatic slide changer, 25 slides could be placed in the machine at once; and *there was no necessity for the experimenter to have knowledge of the order in which the slides were presented.* His ignorance of the order was further insured by a hinged screen fastened to the table which prevented him from seeing the ground-glass screen.

A paid assistant banded the slides together in groups of 25. These had been selected according to random number sequences which were provided by the Hodgson Fellowship Committee. Since the slides were operated in a slide-changer, there was no necessity for the experimenter's handling each slide individually.

In the tests of sense perception at near-liminal values, each slide was presented on the ground-glass screen for one second. The length of time was controlled by the shutter which was worked by means of a foot pedal. Attached to the pedal was a bell which sounded each time the shutter was opened. This provided a means of signaling to the subject that a figure was being presented. The slide-changer was provided with a plunger arrangement which operated in such a manner that the slides were pushed through the projector one at a time and kept in the same order.

For the tests of extrasensory perception, a wooden screen was constructed in such a way that it could be swung down over the glass screen. The wooden screen was slightly smaller than the glass one

so that a thin outline of light showed around it when it was placed in front of the glass screen. This wooden screen differentiated the ESP situation from the near-liminal situation. In the near-liminal situation, the figures were shown on the glass screen and were, therefore, available for the subject to see visually, while in the ESP situation the wooden screen was lowered into position so that the figures did not shine on the glass screen, and the subject had no means of seeing them. The stimulus figures were made considerably smaller than the wooden screen, so that there could be no danger they would overlap the edges.

The Experiment

The Control Experiment

Two groups of subjects have been employed in this experiment up to the present time. The Control Group (without the special motivational factor of *rewards* and of *knowledge of results*) consisted of 20 individuals. The majority were volunteer students from psychology classes at City College. Each subject served for two experimental sessions which were usually about a week apart. The subjects in this group were given no indication as to the success of their "guessing" in the ESP situation.

The detailed procedure followed for each subject in this group was as follows. The subject was brought into the experimental situation and the apparatus was explained to him briefly. It was pointed out that the experimenter could present the stimuli in the visual situation as vividly or as dimly as he wished. The ESP and the visual situations were differentiated. He was told that while he would be unable to see any of the figures in the ESP situation, he was to attempt to call the figures which would be presented on the back of the wooden screen. The shutter speed of one second was demonstrated and the use of the bell as a presentation signal was made clear. Further, it was indicated that in both the visual and the ESP situations the slides were arranged in a random order.

After it was clear that the subject understood his duties, he was taken into the alcove and fastened into the chair facing the glass screen. A practice period then started. During the practice period, the experimenter was aware of the nature of each figure presented.

The results from the practice period are not included in the tables. When the subject signified his readiness to begin, the slide-changer was operated in such a manner that a slide was placed ready for presentation. The pedal operating the shutter was pressed. This rang the signal bell simultaneously with the presentation of the figure on the screen. During the initial stages of the practice period, the rheostat governing the vividness with which the figures were presented was maintained at a high level in order that the subject might become familiar with the figures. The vividness with which the figures were presented was gradually reduced during the practice period until the figures were sufficiently dim to permit approximately one-third of the figures to be called correctly by the subject. The subject's calls were made orally and were written down by the experimenter. The practice period consisted of four runs of 25 figures each. Thus, 100 figures were presented during the practice period, and the illumination level to be maintained during the rest of the session was determined.

Following the practice period a series of four runs (100 presentations) was given to the subject under the near-liminal conditions. Following this, four runs were given under the ESP conditions; that is, the small wooden screen was lowered in front of the ground-glass screen. The subject was then given a rest period of five minutes, after which four more runs were given under the near-liminal conditions, followed by four runs under the ESP conditions. This constituted one "session" for the subject. The same procedure was followed during the second session. The illumination level was again determined during the practice period and two series each under near-liminal and ESP conditions were run.

The important motivational aspects of the control situation are as follows. Subjects were instructed to do the best they could in both the near-liminal and ESP situations. They were not informed concerning their scores. They were offered no rewards.

The Motivation Experiment

In the Reward Group section, 20 subjects began the experiment and continued through it to the end of the first session. At this point, two dropped out, leaving 18 to complete the second session of that group. The failure of the two subjects to participate in the last session was unavoidable by the experimenter.

The conditions for the Reward Group differed from those of the Control Group in the following manner. The subjects were informed concerning their scores after each four runs in both the near-liminal and ESP situations. They knew that this would be done before the experiment started. During the second session, before the final two series (one near-liminal and one ESP) the subjects were informed that for each point which they were to better the total score made on the previous two series they would be given five cents. The amount of improvement then determined the amount that each subject won.

The scoring of the subject's calls was done in the following manner. His calls, as recorded by the experimenter, were first checked against the order of the slides as they appeard after the completion of the experimental session. The calls were then rechecked against the random numbers that had been employed in making up the order of the slides. Where the call was the same as the slide, a hit was scored. Since the sheets of random numbers and the subject's calls are both available a further recheck is possible.

In the reward situation, as has been indicated, it was necessary to tell the subject his scores after each series of four runs. To facilitate matters, the assistant who arranged the slides recorded the order on regular data sheets. These sheets were folded and given to the experimenter who did not look at them until each series of calls had been made and it was necessary to check the calls to determine the score on the series.

Results

In this experiment, 31,200 trials[3] have been completed. In other words 15,600 trials have been completed under near-liminal conditions and an equal number under ESP conditions. Twenty-five trials compose one run. Results are given in terms of the number of runs. The deviation from the expected number of five hits per run is given for the ESP conditions. The average number of hits per run is given for the near-liminal conditions.

[3] The tables given for the ESP results are based on these complete figures. However, because of the circumstances indicated in footnote 2, the results of the near-liminal work as given are based on a total of 13,400 trials instead of the full total of 15,600 trials as given above. But no essential difference in result is involved.

Table 1

I. Results of ESP Tests

A. Control Group: no reward offered, subjects kept in ignorance of scores.

Session	Series	No. of subj.	Runs	Dev.	Av. per run	SD	CR
1st	1st	20	80	+ 10			
1st	2nd	20	80	— 21			
Total			160	— 11			
2nd	1st	20	80	— 24			
2nd	2nd	20	80	— 4			
Total			160	— 28			
Control Group Total			320	— 39	4.88	35.77	1.09

B. Reward Group: monetary reward on last series for improved scores; subjects informed of score at the end of each four-run series.

Session	Series	No. of subj.	Runs	Dev.	Av. per run	SD	CR
1st	1st	20	80	+ 14			
1st	2nd	20	80	+ 36			
Total			160	+ 50			
2nd	1st	18	72	0			
2nd	2nd	18	72	+ 60			
Total			144	+ 60			
Reward Group Total			304	+ 110	5.36	34.87	3.15

CR of the diff. between Control and Reward Group totals is 3.00.

II. Results of Near-Liminal Vision Tests*

A. Control Group: same motivational conditions as for ESP experiment.

Session	Series	Av. per run	SD	SD_m	$CR_{diff.\ means}$
1st	1st	9.42	4.26	.48	
1st	2nd	9.61	4.09	.46	.29
2nd	1st	8.50	3.63	.41	
2nd	2nd	8.35	3.71	.41	.83

B. Reward Group: same motivational conditions as for ESP experiment.

Session	Series	Av. per run	SD	SD_m	$CR_{diff.\ means}$
1st	1st	8.21	4.29	.58	
1st	2nd	8.71	4.14	.56	.57
2nd	1st	10.73	4.16	.58	
2nd	2nd	10.37	3.44	.45	.41

* Based on 13,400 trials out of a total trials completed of 15,600.

Two questions concerning the results are of interest. First, did the offering of the monetary reward affect the efficiency of performance in either the ESP condition or in the near-liminal condition? A

comparison of the results obtained in the first series of the second session with the results of the second series of the second session is necessary to determine the answer to this question. Second, was ESP performance for the entire Reward Series *(involving both knowledge of results and reward)* significantly different from that of the Control Series which *lacked knowledge of results and reward?* This same question cannot be answered in the case of the near-liminal situation, since the illumination levels were determined separately for each subject at the beginning of each session.

Discussion

Results of Near-Liminal Vision Tests

The differences between the means in the case of the vision experiment are so small that it can safely be concluded that the reward had no effect on the performance at this level of illumination. This result is in line with that found by Miller (**3, 4, 7**) and by Woodruff (**8**) in earlier attempts at the same type of research. It is possible that the monetary rewards offered were too small to provide much incentive. The next step in the procedure will be to try threat of shock for failure to improve as a motivation force.

It is not possible to compare the Control Group and the Reward Group under the near-liminal visual situation since the level of illumination varies with each subject and is partly a function of the visual acuity of the subject and partly a function of certain fortuitous factors. This is due mainly to the fact that the level of illumination was determined from only four practice runs.

Results of ESP Tests

It can be seen that the ESP results present a quite different statistical picture from the vision test results. The results of the ESP Control Group are definitely insignificant. On the other hand, the total results of the ESP Reward Group, as well as the difference between the means of the Control Group and of the Reward Group, are clearly significant, with CR's respectively of 3.15 and 3.00. On the basis of the results as they stand, it is difficult, therefore, to escape the conclusion that they constitute evidence in favor of the ESP hypothesis. The odds against a critical ratio of the nature of that obtained in the Reward Group are greater than 1000 to 1. The critical

ratio of the difference between the averages of the two groups yields odds of 800 to 1. There is no available alternative to accepting such probabilities as evidence of the operation of an ESP function and of the limitation of its appearance to the results of the Reward Group alone.

As has been pointed out, the motivational aspects of the Reward Group differ from those of the Control Group in at least two obvious ways. First, a monetary reward was offered to the subjects of the Reward Group during the last part of the second session. Second, after each series of four runs, subjects were informed of the scores they had made. This was not done for the Control Group. Inspection of Table 1 will show that the significant difference between the means of the Control and the Reward groups is to be attributed largely to the second series of the two sessions, those in which incentives were given. These two "second series" gave a total deviation of + 96 in the 152 runs, and the CR of 3.89 is quite significant.

Of these two incentive series, the first was the lower, with + 36 in 80 runs. However, this deviation is more than twice the standard deviation (of 17.88) and is not negligible. It is thus strongly suggested that even the promise of information on scores was, taken alone, a determining factor in the situation.

But the considerably larger deviation of the money reward session, + 60 in 72 runs, gives a CR of 3.54, and leaves little doubt that some non-chance principle is involved, presumably the incentive to win the rewards offered.

These results indicate an effect of incentives on ESP performance, a mild one with promise of score information, a larger one with the promise of money rewards. They show, therefore, a further lawful relation between the process of ESP and the motivational system of the individual.

Relevant Findings by Others

In earlier ESP research, there have been several observations reported concerning the effect of special incentives, some consisting of general impressions such as the importance of telling the subject his score in order to promote interest in the tests or of associating candy with the test situation for child subjects. Others are based on specific experimental work intended to measure reward effects.

Woodruff and George (9) introduced a small prize (movie ticket) during one month of a four-month period of experimentation. They found that the scores in general were improved to a significant extent during the month in which the prize was offered. However, the prize was won only two weeks out of the four; and yet, the two weeks in which the prize was not given produced scores approximately as high as those in which it was. Apparently a generally stimulating effect, presumably one of competition (since the ticket went to the high man), had been introduced during the period in which the prize was offered. A somewhat similar effect is reported by Rhine (6) in connection with precognitive matching tests. Sessions in which money and other rewards were used produced markedly larger deviations than sessions in which non-reward tests alone were conducted. When non-reward tests were alternated with reward tests at the same session, it was found that the children who participated in the experiment produced approximately twice the deviation on the reward test that they gave on the non-reward. Even so, these non-reward tests that were alternated with reward tests were still considerably above the score averages obtained in the sessions that were made up entirely of non-reward tests. There apparently was, then, the same type of spread of incentive effect which Woodruff and George observed. In the case of the adult subjects, however, although the mixed sessions of reward and non-reward tests gave considerably larger deviations than the purely non-reward sessions, the tests in the mixed sessions did not favor the reward tests. The effect seemed to be a more dispersed one with adults than with children.

In the field of learning, Book and Norvell (1) report experiments with college students leading to the definite conclusion that knowledge of results of scores in test performance improves the rate of progress substantially. The four types of tests upon which the conclusions were based ranged from a simple muscular skill to the multiplication mentally of two-place numbers. There is, of course, a difference between this finding on learning experiments and the ESP test results reported in this paper. We do not report an improved rate of *progress,* whereas it is directly the rate of progress that is involved in the learning experiment just mentioned. It is a question whether there is a relation between the two findings, whether there is basically

some element in common between the single ESP judgment and the single increment or step in the learning process.

There may be some relevance to the present work in the investigation by Chapman and Feder (2) in which boys and girls in the fifth grade were given external incentives to ascertain whether these led to improvement in simple addition, cancellation, and substitution tests. Individual scores were published daily as one possible incentive, and the progress of the class was presented graphically. Stars were given to the upper fifty percent on the succeeding day, with prizes at the end for the individual with the greatest number of stars. A control group without the incentive showed little if any improvement. The incentive group showed marked improvement on the addition test, some improvement on the substitution test, though little on the cancellation test.

The findings from the ESP investigation described in this report are such as to justify the further utilization of such incentives as were used in sections of the ESP tests; that is, wherever it is desirable to provide the most favorable conditions for high scoring. They confirm the general impression of ESP workers that informing the subject of his score is conducive to better performance and they support the earlier work from which it was found that higher scores followed the use of rewards. There remain, of course, many further questions regarding the use of incentives and the role of motivation in general.

Concluding Remarks

The significant findings from these experiments are entirely restricted to the ESP section of the work. It is not possible to compare the Control Group and the Reward Group for the section on visual acuity.

The ESP experiments gave results not attributable to chance: first, in the total results of the Reward Group; and second, in the difference between the means of the Control and the Reward groups. This is interpreted as evidence of the occurrence of extrasensory perception in the Reward Group tests.

There is strongly suggestive evidence that it raises the score average to tell the subject his scores and more definite evidence that small money rewards will improve his performance. These findings are in

line with earlier experiments involving the use of incentives, not only in the field of ESP investigation but in learning experiments as well.

References

1. Book, W. F., and Norvell, L. The will to learn: an experimental study of incentives in learning. *Pedagog. Seminary*, 1922, **29**, 305-362.
2. Chapman, J. C., and Feder, R. B. The effect of external incentives on improvement. *J. Educ. Psychol.*, 1917, **8**, 469-474.
3. Miller, J. G. Discrimination without awareness. *Amer. J. Psychol.*, 1939, **52**, 562-578.
4. ———. The role of motivation in learning without awareness. *Amer. J. Psychol.*, 1940, **53**, 229-239.
5. Pratt, J. G., and Woodruff, J. L. Size of stimulus symbols in extra-sensory perception. *J. Parapsychol.*, 1939, **3**, 121-158.
6. Rhine, J. B. Experiments bearing upon the precognition hypothesis. III. Mechanically selected cards. *J. Parapsychol.*, 1941, **5**, 1-57.
7. Sterling, K., and Miller, J. G. The effect of hypnosis upon visual and auditory acuity. *Amer. J. Psychol.*, 1940, **53**, 269-276.
8. Woodruff, J. L. The effects of certain factors on visual form discrimination at near-liminal levels. Thesis for Ph.D. degree, Duke University, 1941.
9. ———, and George, R. W. Experiments in extra-sensory perception. *J. Parapsychol.*, 1937, **1**, 18-30.

THE PRESENT POSITION OF EXPERIMENTAL RESEARCH INTO TELEPATHY AND RELATED PHENOMENA

[A DIGEST IN THE SPEAKER'S OWN WORDS OF THE 1942 PRESIDENTIAL ADDRESS TO THE SOCIETY FOR PSYCHICAL RESEARCH, LONDON.*]

By ROBERT H. THOULESS
Cambridge University, England

[The address by Professor Thouless is a most timely and significant contribution to the literature of parapsychology. Professor Thouless is the first experimental psychologist since William McDougall to have been elected President of the S.P.R. We believe that the contents of this Presidential Address are of interest to a much wider reading public than would be reached here in America by the *Proceedings of the S.P.R.*, especially under wartime restrictions. A lengthy review was first considered; but the article was found to be too rich in content for extreme condensation and it seemed to lend itself best to "digesting," i.e., to abbreviation in the medium of the author's own words. This digest, prepared by Dr. J. G. Pratt, furnishes the reduction required for republication in the *Journal of Parapsychology* and preserves the advantages of Professor Thouless' own expression. An effort has been made to present faithfully the thesis and main content of the address and to do so without change of emphasis or context.—Ed.]

IF, AS AN EXPERIMENTAL psychologist, I limit myself in my present address to the experimental study of extra-sensory perception, this is not intended to imply that I think that other methods of studying extra-sensory perception, or that fields of study in psychical research other than extra-sensory perception, are unprofitable or unimportant. The special function of the study of spontaneous cases is to serve as a guide to the problems to be investigated by experimental methods. In psychical research, the choice is not between statistics and experiment on the one hand and observation of spontaneous cases on the

* Printed in the *Proceedings of the S.P.R.*, Volume 47, July, 1942.

other. Let us have much more of both. No one would have thought of asking a subject to guess what card has been turned up in a pack, if someone else had not reported spontaneous observations of telepathy or clairvoyance. New problems for experimental investigation may be suggested by new observations of spontaneously occurring phenomena, although also, of course, new problems for experimental investigation may be suggested by experiments themselves.

Still less reason has the student of the experimental approach to extra-sensory perception for throwing doubt on the relative importance of other questions for psychical research, such as that of personal survival after death. I think that few questions can be more important than that of whether we survive the death of our bodies. On the other hand, the experimental techniques acquired in a psychological laboratory do not seem at present to provide a useful way of studying that question. Even if in itself relatively unimportant, the study of extra-sensory perception may have an added importance as a preliminary to the solution of the more difficult question of the evidence for personal survival.

It is generally agreed in discussion of psychical research that if any person can produce correct information on any topic there are three possible explanations in order of increasing intrinsic improbability: (a) the information has been gained by normal sensory channels or by rational inference, (b) it has been gained by extra-sensory perception, and (c) it has been communicated by some discarnate intelligence. It is also generally agreed that evidence for any of these explanations can only be regarded as sufficient if the possibility of the preceding ones (presumed to be less intrinsically improbable) has been excluded. It is obvious that, wherever one adopts such a scale of explanation and such a principle of exclusion, no explanation higher in the scale can ever be established unless the limits of all those lower in the scale are already known. Explanations of the phenomena of extra-sensory perception by sensory hyperaesthesia have never been generally accepted by psychologists (however sceptical they may have been of the E.S.P. explanation on other grounds) because it has seemed to them that the limits of sensory perception were sufficiently well known. Since there is no good evidence for sensory hyperaesthesia in any case, and since, in some cases, well

attested facts of extra-sensory perception (such as succesful card guessing experiments over long distance) cannot be explained by any extension of hyperaesthesia, the possibility of sensory hyperaesthesia is no longer felt to be a serious obstacle to the acceptance of evidence for extra-sensory perception.

There seems to be no corresponding certainty that we know the limits of extra-sensory perception sufficiently well for it to be possible to regard any evidence for spirit communications as sufficient if we admit the principle that nothing is to be regarded as evidence that might be explained as an effect of extra-sensory perception. It is extremely likely that the possibilities of obtaining information by extra-sensory perception have their own definite limits although these limits are obviously very different from those of sensory perception or rational inference. There also seems every reason to hope that experimental study may one day give us a clear idea of what those limits are. Certainly we do not know yet, but the experimental discovery of the limits of extra-sensory perception has, amongst other things, the value that it may be a necessary preliminary to a satisfactory assessment of the evidence for survival.

I have so far used the term "extra-sensory perception" because that is a name now generally understood. It is not with any enthusiasm for introducing novelties in terminology that I propose now to drop that term and to suggest a new one. The objection to the term "extra-sensory perception" is that it suggests a theory of the nature of the phenomenon in question, and I see no reason to suppose that this is a true theory and some reason for suspecting that it is false.

It would be pedantic to object to a misleading terminology unless it, in fact, misleads. I think that this term may be seriously misleading since it leads us to put these effects within a framework of expectations in which ordinary perception is placed. They may belong to a totally different framework, requiring quite different expectations and quite different modes of thought to deal with. Perception lies, for example, within the system of scientific causation, but it may be necessary for our thought to abandon this system of expectations if we are to understand paranormal determination of correct responses.

I suggest that we should use a term proposed by Dr. Wiesner, and call this group of effects the "psi phenomena," a term which has the important negative merit that it implies no theory as to their nature.

If experimental psychologists have, on the whole, been slow to accept the reality of the psi phenomena, this is to be explained partly by the difficulty of repeating at will the successful results which others have reported, partly by the intrinsic improbability of the phenomena themselves. It is to be hoped that the first of these difficulties will be got over by further experimentation on the conditions favourable to the appearance of the phenomena. The literature of the subject contains many hints as to favourable conditions, but these seem generally to be based on the impressions gathered by experimenters in the course of their work rather than on exactly controlled experiments establishing significant differences in scores obtained under one condition and another. This situation poses, as an important problem to be determined by experimental methods, the finding by exact methods of the optimal conditions for the appearance of the psi phenomena. When we know this, it is possible that we shall no longer be dependent on occasional subjects or on large scale experiments for successful results. It ought to be possible by suitable training of our subjects and by suitably arranged conditions to produce successful results anywhere. When this is possible, the main rational defence for rejection of the experimental evidence for the phenomena will disappear. Anyone will be able to produce the evidence for himself by carrying out the experiments in the right way.

On the intrinsic improbability of the psi phenomena, there seems to be some difference of opinion. Against those who think the intrinsic improbability so great that they find themselves unable to be convinced by a weight of evidence far in excess of what would be regarded as decisive in any other field of research, there are others who seem to find nothing intrinsically unlikely in the facts reported. I should like to suggest that there is a real intrinsic improbability in the psi phenomena, and it may perhaps be worth while to consider for a few moments what this assertion of intrinsic improbability means.

If we ask what sort of causes were accepted and what rejected by the scientific system of expectations, the answer in its broad outlines is sufficiently obvious. The essential characteristic of the scientific type of explanation was that one thing could only cause another if there were a continuous chain of physical events between the two. The facts of ordinary perception, although they may be called won-

derful and mysterious, fit very well within the system of expectations of physical science. In visual perception, electro-magnetic waves come from the object to the eye, on the surface of the retina they produce chemical changes which start an impulse along a nerve fibre to the visual area of the cerebral cortex. If we ignore the problem of how a material change in the cerebral cortex is related to the conscious process of perception, and confine ourselves to the physical processes between the emission of rays from a material object and the completion of a muscular or glandular response on the part of the organism, the continuity of the chain of physical events is complete.

If ordinary visual perception falls within the system of scientific expectations because it preserves the principle of continuity of material causation, the psi phenomenon does not. It is, therefore, inevitable that the first reaction of the scientific mind should be to relegate it to the class of things which science teaches us do not happen. I think it does belong to that class. If so, there are three things to be done about it: (1) to say that the psi effects do not take place and that those who report them are either incompetent experimenters, or frauds, or the unfortunate victims of a vast system of chance coincidences; (2) to postulate some unknown cause which would reduce the phenomena to a kind admitted by science, as, for example, by postulating some unknown form of radiation; or (3) to accept the phenomena as genuine and as not fitting into the system of scientific expectations, and as requiring, therefore, a modification of our system of expectations.

Before this audience, it is not necessary to argue the impossibility of accepting the first alternative. Even before the laborious and extensive investigations of Rhine, Soal, and Tyrrell had produced experimental and statistical evidence of an overwhelming character, it was arguable that the case was sufficiently proved by sporadic evidence alone. There is, however, one important advantage for proving the reality of psi that experimental evidence controlled by a proper use of statistical methods has over the observation of spontaneous cases and over experimentation to which statistical checks can be less easily applied. This advantage lies in the fact that it is possible to give a figure indicating how unlikely it is that we are being misled by chance coincidences. It remains true that some sceptics do say that the odds of billions to one against a chance interpretation of modern

experimental results are merely a lucky accident. I do not think that this should be found surprising or particularly disturbing; no evidence can compel belief in those who are sufficiently determined not to believe.

Empirical methods of research can never give certainty; wildly improbable coincidences are a possibility, and human fallibility is also a possible source of error. All that we can reasonably demand is that evidence for the psi phenomena should be strong enough for reasonable conviction even for an intrinsically unlikely effect. That point has, I think, been passed, and the reality of the phenomena must be regarded as proved as certainly as anything in scientific research can be proved.

Accepting the reality of psi, we may seek to fit it into the existing framework of scientific explanations. We may postulate some form of unknown radiation received by some unknown sense organ. Such explanations have had the support of great names in the past (such as Crookes and Ostwald). The reason for our confidence in the scientific type of explanation is that the entities we postulate to fill such gaps [in our knowledge] can be shown to have other properties by which their reality can be confirmed. We do not know light waves only by their action on the eye; if we did there might still be reason for doubting their reality. We know them also by their action on photo-chemical substances, by their heating effects, and so on. An unknown form of radiation of which we have no knowledge except its action in the psi effect and an unknown sense organ similarly known only in this way and undetected by anatomical investigation, lack plausibility as scientific hypotheses and can only be regarded as desperate expedients to save the system of scientific expectations.

There is, of course, also the argument which has often been urged that we know of no kind of radiation which does not lose its effect with distance, and that even if we explain such facts as the success of Rhine's subjects in reading packs of Zener cards "down through" by supposing that some form of radiation proceeded from the cards, it is impossible to see how any form of radiation could enable the subject separately to perceive the cards low down in the pack. Still less is it possible to conceive how any form of radiation could enable a subject to discriminate at a great distance a particular pack from all

other packs and all other objects which must be supposed to be sending out similar radiations.

We seem then to be forced to accept the third possibility—that the psi phenomena are real and that they are not explicable in terms of the scientific expectations based on the necessity for physical continuity in chains of cause and effect. This would not, of course, mean that we suppose that the system of scientific expectations is wholly mistaken. That would be absurd in view of the remarkable success which has been achieved in building up a means of controlling the outside world by following that system of expectations. It is clear that over a very large field of phenomena, this system of expectations has proved a trustworthy guide. It remains possible that there is also a field in which it is not a trustworthy guide. The psi phenomena appear to be such a field. There may be others. Possibly the human will is one.

There is also the very difficult question of how we are to think of psi phenomena if we are not to try to fit them into the framework of the scientific system of expectations. Are we to postulate a type of continuity between cause and effect different from that in the kind of causation recognised by natural science, and if so how are we to think of this continuity? I do not think we are in a position yet to answer these questions. If the way of thinking appropriate to the psi phenomena is to become clear, it must be by more research and by new ranges of speculative thinking. I have only the negative suggestion to make that the first step is to eradicate from our minds the influence of the deeply ingrained habits of thought which make up the scientific system of expectations. These are deeply ingrained, and they are influencing us when we start thinking in terms of unknown radiations. I think we unnecessarily tempt them to exert their influence when we talk of these phenomena as if they were a kind of perception. As to what are the appropriate ways of thinking about these phenomena, I have no idea. I do not think that we have them yet.

Already I think there are indications in the results of experiments which would guide us in our thinking if we knew how to interpret them. One of the most encouraging signs in the experimental research on this subject at the present time is that we have not merely shown that psi phenomena can be demonstrated under experimental

conditions, but that our results are showing odd, unexpected and (at present) inexplicable uniformities which are cropping up independently in different laboratories. No scientific research worker can feel quite so well satisfied with results that come out exactly as he expected them to, as with results which persistently come out as he had no previous reason to expect. Here he feels he is against the hard rock of a reality independent of his expectations; here is a challenge to his powers of constructive speculation.

For example, Rhine found that if he made his subjects continue their experiments beyond the point at which they were thoroughly bored with them, they began to score persistently below mean chance expectation. At least two other investigators working independently of Rhine have reported the same result. A drop to mean chance expectation would be easy to understand, but a drop that is significantly below mean chance expectation is entirely unexpected.

There is also the discovery, originally made by Mr. Whately Carington, of what may be called "temporal dislocation" of responses. He found in his experiments of pictures exposed on successive nights that subjects might not draw a picture resembling the one exposed during the night on which they made the reproduction but one resembling some picture exposed on some other night. This observation was not in itself of sufficient statistical significance to carry overwhelming conviction to those who would regard such temporal dislocation as too improbable for belief without coercive evidence. It led, however, to Mr. Soal's re-examination of the results he had obtained in his repetition of Rhine's experiments, and, although his intervals were so different from those of Whately Carington (a second or so instead of a day), he found, with some subjects, evidence of correct guessing with temporal dislocation that was of unquestionable significance.

I think that Whately Carington's discovery is of the greatest importance. On the theoretical side, it shows that psi reactions (already proved to be relatively independent of space by Rhine and other experimenters who have obtained successful results over great distances), show also an indefiniteness with respect to time, and that successful psi results may be determined by a future event even when the intention of the experimental subject is to make them refer to the present time.

Another practical point is the desirability, if temporal localisation is possible, of adopting some system of regional scoring, as, for example, by counting hits, not on the card turned up, but on any card within a region of five about the one turned up. This is the scoring method that I am now using.

Another unexpected feature of experimental results is the tendency of temporal dislocation, on the whole, to be in the direction of guessing ahead. I understand that Soal in his latest work has found his subject guessing ahead. In my own experiments, that also is the tendency. If this is general, we must again ask "Why?" Still odder was Soal's finding of a tendency for the guesses just before the card turned up to be below mean chance expectation. I have found indications of the same tendency, although my minimum is not itself significant, and I should attach no importance to it if it had not been more adequately shown by Soal. It looks as if the probability of the response to a card increases to a maximum some seconds before it was turned up, then declined to a point at which it was less likely that that response was given than any other response, as if the subject at the minimum point both knew the card and also knew that it was not a correct response but belonged to the past. Again if we could explain why this is so, we might hope to advance in understanding of the phenomena.

Even the experimental results achieved so far have given us many unsolved problems which give hope of future more complete understanding.

I wish now to turn to another question—that of the aims which experimentalists should now set themselves in studying the psi phenomena. Rhine complained in his first book that, in this field, every experimenter regards it as his task to prove the reality of the phenomena all over again as if it had never been done before. If it was arguably necessary when Rhine started his work, it surely is so no longer. By different methods, a number of workers have obtained under stringent experimental conditions positive results which cannot reasonably be attributed to chance or to experimental error. The work of obtaining these results has been laborious, and great credit is due to those who have undertaken it. A mere repetition of that work now would be a great waste of time. Let us get on to other problems, to be solved by other methods. If we meet with sceptics as

to the reality of the phenomena we are studying, let us refer them to the researches of Rhine, of Soal, and of Tyrrell and not succumb to the temptation of trying to satisfy them ourselves.

The reason for calling this a temptation is that the methods appropriate to a research intended to establish the reality of the phenomena are not generally appropriate to a research intended to elucidate the character and the conditions of the phenomena. The investigator seeking to establish the reality of the phenomenon repeats his experiments a large number times under identical conditions. He aims at getting enormous odds against a chance explanation of his results and is unwilling to introduce variations in method which may be unfavourable to positive results and may therefore reduce the significance of his total score. He also feels it necessary to safeguard himself against critics who will attribute his results to dishonesty or incompetence, so he has one or more impartial witnesses as observers of his experiments whose testimony can establish that he has obtained the results he says he has under the conditions he has described.

Indefinite reduplication of witnessed experiments has been valuable in the past when the primary object of experimenters was to establish the reality of the phenomena. If we agree that this reality has now been sufficiently established, the need for these methods has passed and they should not be allowed to become standardised methods for future experiments. When experimenters have as their primary purpose the understanding and control of the phenomena, frequent variation of conditions is necessary, and those variations will be most fruitful of results which lead to reduction or extinction of the phenomena. Workers must be content with such moderate standards of significance as are used in other branches of scientific research, since the time available for any course of experiments is limited, and time spent in reduplication of experiments under identical conditions is time lost for the more important task of working under variable conditions. Time is limited and possible witnesses with expert knowledge of experimental methods are busy with their own tasks. Insistence on the presence of witnesses must have the effect of reducing frequency of experiment and freedom of variation of condition. Once the reality of the phenomena is taken as established, witnessing is no more necessary than in any other branch of scientific investigation. Undoubtedly there will be error and incompetence in experimental

research in this as in all other topics. It will be subjected to the usual check that an erroneous finding by one worker will be corrected by the confirmatory work of others. The check of independent confirmation is working to a very considerable extent.

If we agree that the type of experimental research in the psi phenomena now necessary is one in which we try to find out as much as we can about the phenomena by experiments in which conditions of working are varied as much as possible, we are immediately faced by the practical difficulty which is perhaps the principal obstacle at present to fruitful research along these lines. In any experimental research, mere random variation of conditions of experiment is not enough. We must vary our conditions in such a way as to try to obtain answers to definite questions. The most pressing need to furnish fruitful research is that we should know what questions to ask. The questions partly arise out of the research itself; they are partly initiated by the propounding of suitable hypotheses.

Let us not be deterred from bold speculation by the fear lest our speculations should be wrong. The process of speculation and experimental testing is a self-correcting one. It does not matter if a speculation is wrong; if so it will be proved wrong by experiment, and that will be a step forward. The caution which we properly observe in drawing conclusions from our experiments is out of place in the preliminary task of devising hypotheses to be tested.

I cannot, of course, suggest what these fruitful hypotheses may be; I can only state what has been the hypothesis that has guided my own researches in the psi phenomenon. This is a point of view which has been reached in discussion between Dr. Wiesner and myself and, in what follows, it would be impossible to disentangle the contributions made by us both.

The essential point from which we start is that expressed by Bergson in his Presidential address to this Society. To Bergson, the brain was not an organ whose function is that of transforming material vibrations into mental states but an "instrument of selection charged with choosing, in the immense field of our virtual perceptions, those which are to be actualised." "I think," he says, "that we perceive virtually many more things than we perceive actually, and that here once more the part that our body plays is that of shutting out from our field of consciousness all that is of no practical interest to us, all

that does not lend itself to our action." In other words, Bergson thought of the psi function as a possible kind of reaction normally suppressed because unserviceable for the demands of practical life, and kept suppressed so long as the organism maintained the attitude of attention to life, but liable to manifest itself under conditions in which that controlling mechanism was thrown out of action by the development of an attitude of inattention to life.

Let us adopt Bergson's speculation as a starting point and consider its general plausibility, how we may profitably amplify it, and how it may be used as a guiding hypothesis for experimental research. Let us consider a hypothetical primitive organism possessing the psi capacity in its most extreme imaginable form, like Leibniz's monads mirroring all facts in the universe without any limitation with respect to space or time. It would possess a capacity of no possible biological value to it, since acquaintance with facts is only serviceable so far as it discriminates between near and distant facts and between present and past or future ones.

Our imagination of an organism with completely unlimited psi is, of course, purely fantastic. We can suppose that such an organism never existed, but that any primitive psi function always possessed some degree of limitation. On the other hand, observation and experiment both seem to make it clear that the psi function is less limited in these respects than is sensory perception.

Let us now take a step not taken by Bergson, although it seems to be a natural extension of his thought, and suppose that psi is the primitive way by which organisms oriented themselves to the outside world and that the evolution of the sense organs and of sensory perception was a later acquired means, of greater biological usefulness because more limited. Perhaps we may take a simple example of what is meant by saying that the more strict limitation in space and time of sensory perception makes it more biologically useful than the relatively unlimited psi function. Let us suppose that each one of us present here were so open to psi influences that his thoughts were effectively determined by those of everyone else in the room. One is perhaps wondering whether his clothing coupons will be sufficient for a new suit, another is thinking that the speaker is going on too long and that she will be too late to cook the fish for supper, and so on. If we had such capacity, it would not be a wonderfully effec-

tive extension of our powers; it would, on the contrary, produce a very ineffective confusion. We are clearly better off if our knowledge of the thoughts of others is limited to the thoughts of the one person who is speaking and so is in auditory communication with us. As compared with the possibility of unlimited telepathic powers, this restriction of knowledge of the thoughts of others to those who are producing the sound waves of speech in our immediate vicinity is a limitation, but clearly a limitation which increases efficiency of intercommunication. If we had to choose between the obtaining of information by psi and obtaining it through our senses, we should unhesitatingly choose the latter.

This leads us to consider the possibility that the psi function is the earlier in evolutionary history and that it may have been suppressed by the development of the special senses which, with their limitation to what is here and now, serve much more efficiently the biological end of securing the survival and efficient adaptation of the organism. The view that I am suggesting may, of course, not be right; it is put forward as an example of speculation that imposes a number of clear problems for experimental research. So far the expectations it raises seem to be fulfilled. If psi is a more primitive function, normally suppressed by the higher mental activities of perception and reasoning, we should expect, on the whole, that conditions favourable to the higher mental activities would be unfavourable to psi and vice versa. The attitude of inattention to the practical demands of life induced by muscular relaxation is reported to favour psi, although most people find muscular tension favourable to ordinary cerebral activities. I also have the impression (although I have no experimental proof) that the absence of any effort to guess right is also favourable to psi, although efficiency in any intellectual or perceptual activity requires sustained conation.

Clarification of the conditions for psi success is of considerable practical importance since a knowledge of them may enable us to train subjects to give successful results and, therefore, make psi experiments possible to every research worker who wants to investigate them without it being necessary for him to wait for the lucky chance of finding a suitable subject. The suggestion I am making is that the favourable conditions are those summed up in Bergson's phrase "inattention to life," because these conditions are those in which the higher

mental functions are reduced in efficiency so that the more primitive psi function is no longer effectively suppressed. The spontaneously successful psi subjects seem often to be more or less dissociated, and, therefore, to have a pathological attitude of inattention to life. The non-pathological psi subjects may be those who have the power of voluntarily adopting an attitude of inattention to life, a power which probably most people could acquire although we should reasonably expect more difficulty in those subjects who are habitually tense.

The speculations contained in the latter part of this address are not to be taken as claimed by me to be true. Perhaps they are flights of fancy with no foundation. Primarily they are intended as guides to the sort of questions we may submit to the test of experimental research. They can be proved to be well or ill founded, not by argument, but by the results of such research.

Finally, I should like to suggest that the general result of our survey is that the present position of experimental research in these topics is a hopeful one. Existing researches have not merely proved the existence of the psi phenomena; they have also found out some odd and unexplained things about them that are a challenge to further research. Let us now give up the task of trying to prove again to the satisfaction of the sceptical that the psi effect really exists, and try instead to devote ourselves to the task of finding out all we can about it. With fuller knowledge of its nature, the difficulties of believing in its existence may appear less formidable than they do now.

A COMPARISON OF THREE SIZES OF DICE IN PK TESTS

By HOMER HILTON, JR., GEORGE BAER, and J. B. RHINE

ABSTRACT: This is a report of an early exploration in the PK field. It is, in fact, the earliest confirmatory research to be completed. Hilton and Baer undertook to test each other's capacity to throw "high dice," attempting to "will" the dice to produce combinations of eight or above. They succeeded and obtained much higher scores than chance expectation. In a large section of the tests, the dice were thrown in semi-mechanical fashion, and this section gave a higher average score than the rest. A control series was carried out to test for possible bias those dice which were most often used in the experiment. Neither faulty dice nor skilled throwing is regarded as offering an adequate explanation of the results.

Most of the testing was done with throws of a pair of dice, but in a section of the work only one die was thrown at a time. The scores of the tests with two dice per throw averaged higher than those of the one-die tests. Also there were three sizes of dice used; and, to a certain limited extent, the smaller the dice, the better the scoring. It was not much better, however, and not at all proportionate to the volume of the dice.

The most exceptional feature of this research is the absence of a marked decline either in the four-run sequences or vertically within the run. There is, however, a hit distribution in the run called the "double decline" which appears to be characteristic of this test condition.

Mr. Hilton, now serving in the United States Army, has been associated with the law firm of Root, Clark, Buckner, and Ballantine in New York, and Dr. Baer is practicing medicine in Detroit.—Ed.

INTRODUCTION

THERE HAVE BEEN three reports of experiments on what is called the PK effect, that is, the hypothesis that in some way the movement of bodies (dice) can be influenced by the volitional attitude of a subject, apart from any known means of contact. These reports concern experiments in the psychophysical control of the fall of dice either when thrown by hand, by cup, or by mechanical methods. They have to do with the throwing of dice either for a specified face or for a combination of faces.

Other variations have been introduced in the researches reported to date. For instance, there have been two different numbers of dice

thrown at a time. In most of the tests, two dice were thrown; but in the Gibson and Rhine series (1), six were used and gave a somewhat higher rate of scoring. This result seems to go contrary to what would be expected from analogous physical laws.

The primary importance of discovering the degree to which any new phenomenon is related to the world of physics is well recognized in scientific procedure. It seems necessary, therefore, to push this point of inquiry as vigorously as possible. For if, on the one hand, the mass of the bodies to be influenced should be found to be inversely proportional to the success with which they can be influenced in the PK test, that fact at once would align the effect with many well-known phenomena of physics. It would immediately "place" the thing we are investigating and go far toward identifying its character. If, on the other hand, the results should continue, as they have begun, to show a relative independence from the mechanical relations such as numbers per throw and similar criteria, there would clearly be evidenced a phenomenon of a new order—new, at least, to the world of measurement; in fact, an effect requiring a new set of principles peculiar to human personality. Such a finding would be highly significant for the interpretation of the place which personality occupies in the causal scheme of things we call the universe.

With this concern in mind, we conducted these experiments involving three sizes of dice—sizes that were available on the local market. We threw them under two different sets of conditions, each involving a comparison. The first contrasted one die and two dice per throw, and the second involved throwing by hand as compared with mechanical throwing.

The experiment was undertaken in the spring of 1934 while two of us, H.H. and G.B., were senior pre-professional students at Duke University, H.H. in the field of law, and G.B. in medicine. The research was planned primarily as an investigation as to whether the PK effect actually occurs, and the experimental variations were introduced to further the value of the experiment if it demonstrated satisfactorily the presence of the experimental principle. In other words, if evidence of PK were obtained, it was hoped that the special comparison of numbers of dice per throw, of the different manners of throwing, and of the three sizes of dice would all add something

toward the understanding of the phenomenon beyond the evidence of PK itself.

The Experiment

While the experiment was planned by the three of us together, the conduct of it was entirely in the hands of H.H. and G.B. It was carried out in G.B.'s room in one of the university dormitories. H.H. and G.B. each participated as subject in the Main Series, always with the other acting as observer and recorder. Three student friends of H.H. and G.B. took a minor part in these tests, following the routine procedure. But they did not participate in the main part of the experiment comparing size of dice, manner of throwing, and number of dice per throw. Their results will be stated separately as a Miscellaneous Series.

The objective in the test was to cause a pair of dice, when thrown, to come up high dice, that is, with faces totaling eight or above. The dice were usually thrown two at a time, but in 67 runs (of 24 throws each) they were thrown one at a time. When this was done, the first two throws made a pair for purposes of recording and counting, and each successive two throws thereafter, the same.

In order to avoid the inhibiting effects of over-routinizing the experiment, we allowed the subject to choose which of the three sizes of dice he wished to work with and allowed him to choose, also, whether he would throw one die or two. While such a practice works against nicety of design in the experiment, it must be remembered that we are working primarily with the human beings and not with the dice. According to the hypothesis under test, the degree of interest of the subject is all-important and the frequent exercise of choice on his part serves to maintain interest.

The three sizes of dice used throughout were 7/16, 11/16, and 1 inch on the edge, respectively. The large and medium dice were made of plastic material and the small ones of bone. The spots were marked with paint set in small pits.

The medium-sized pair of dice, which was the pair most often used, was tested in a control series of throws totaling 128 runs of 12 throws of the pair. In this series, the dice were thrown by hand and the two subjects who participated were told that the test series was being made to verify the laws of chance. This was done to prevent their attempting to influence the dice. The subjects, J.H.A.

and B.C.W., did not take part in the experimental series. H.H. witnessed and recorded. This series adequately endorsed the dice as sufficiently true for the purposes of the experiment; the results will be given and discussed later on in the report.

The throwing of the dice was done about three fourths of the time by hand; but in order to insure the elimination of any effect of skill, either actual or ascribed, the dice were bumped on a hard-surfaced desk and made to bound against a rack of books before readings were taken. In the other fourth of the throws, totaling 128 runs, the dice were mechanically thrown. They were released from a position about two feet above the desk and were allowd to roll down a steeply inclined board, bouncing on the desk against the rack of books. The release of the dice is called "mechanical release" because in the initial position, they rested upon a small ruler laid horizontally upon two nails across the end of the board. To release the dice, the subject removed the ruler and allowed the dice to roll. In the strict sense, there was no throwing. There was no other impulse or mechanical effect that could be imparted to the dice other than from gravitation and the path of their fall. After the readings were taken, the dice were replaced on the ruler in readiness for another trial. They were picked up in the same way each time, that is, with the same faces uppermost as when they lay on the table. The purpose of this routine was to eliminate any possible "fixing" on the part of the subject by picking up the dice and placing them in a special manner calculated to favor the face for which he was aiming. These conditions of the mechanical-release method will be recognized as identical with those reported in the Rhine and Rhine article on the PK effect (5).

The records were made in composition books in the presence of the subject. They were checked at the end of each four-run page and the subject was told his score. Later, the records were completely rechecked. The summaries prepared at the end of the experimental series, as well as the fact that the record books were bound, assure against any loss of records; the series is complete.

Results

With Reference to Chance

In addition to the 128 control runs, there were 484 runs in the Main Series. Of primary consideration is the question of the sig-

nificance of these taken as a whole. As Table 1 indicates, the 484 runs of the Main Series averaged 5.27 hits per run as against an expectation of 5.00. Compared with the 2,420 expected for the entire series, 2,550 hits were obtained. This gives 130 above expectation which, with an SD of ±37.62, has a significant CR of 3.46. The primary question of whether the results are explainable by chance can be said to be settled, then, on the basis of these figures. The Control Series of 128 runs averaged very close to chance expectation of five hits per run. (The total deviation was + 4.) The Miscellaneous Series, with its three subjects, totaled 88 runs with 423 hits, or 17 below expectation. This is an average of 4.81 hits per run and is insignificant.

Table 1

TOTAL RESULTS OF PK TESTS AND CHANCE CONTROL SERIES*

Series	Runs	Average Score	Expectancy	Total Hits	Deviation	SD	CR‡
Main	484	5.27	5.00	2,550	+130	±37.62	3.46
Control	128	5.03	5.00	644	+ 4	±19.34	Insig.
Miscellaneous	88	4.81	5.00	423	− 17	±16.04	Insig.†

*This Control Series was done with only the medium dice.
†Pooled with the Main Series, this would reduce the CR from 3.46 to 2.76, which would still be significant.
‡The total deviation from expectation, divided by the SD. In this case, the SD is found by the following process: $\sqrt{R \times 35/12}$ where R = number of runs. This is the familiar $\sqrt{npq}$ formula.

The results constitute a sample adequately large, totaling, for the Main Series, 5,808 throws of two dice each, and they represent the complete record of data for this experiment as contributed by the two subjects, H.H. and G.B. There can be no question of any improper option having been exercised in choosing the stopping-point, since the latter was determined by the termination of the school year and the graduation of the two experimenters. As a matter of fact, there was no total average of the results in terms of the criteria used in this report until many years later. We may, then, accept the results as calling for some explanation other than that of chance. The further presentation of experimental data will bear primarily upon other issues.

Question of Dice Imperfections

The next question in order of importance is that of the adequacy of the dice. Obviously, if the dice are faulty enough, they will themselves produce extrachance results if thrown for the faces that are

favored; in this instance, for the higher faces, since high dice were the objective. There are a number of ways of checking on this hypothesis. The most direct is that of the control described above, the results of which are given in Table 1. The fact that this control was limited to the medium-sized dice and that no similar comparable control series was carried out for the large and small size leaves something to be desired; but the lack of control series on the other dice is compensated for in various ways: First, as will be seen below in Table 4, the medium dice gave almost the same average as the total experiment, namely, 5.24 as compared to 5.27. Second, the fact that the medium dice were thrown for nearly as many runs as both the other pairs together is further reassurance. Third, there is no known reason to suspect the small and large dice of being in any way different from the medium; their manufacture and appearance were similar. Finally, the medium dice alone give a CR of 2.11 which is distinctly suggestive.

Also there are further controls bearing on the question of dice defects. One of these is incidentally provided by the results of throwing one die at a time. As mentioned previously, 67 runs of the Main Series were done in this manner. These results are given in Table 2.

Table 2

Comparison of Numbers of Dice per Throw

Size of Dice	One-Die			Two-Dice		
	Runs	Deviation	Average Score	Runs	Deviation	Average Score
Small	20	− 7	4.65	137	+ 63	5.46
Medium	25	+ 5	5.20	200	+ 49	5.25
Large	22	+11	5.50	80	+ 9	5.11
Total	67	+ 9	5.13	417	+121	5.29
Grand Total Including Misc. Series	74	+ 1	5.01	498	+112	5.22

Note especially that the small dice dropped below expectation in the one-die tests although they gave the highest average in the two-dice tests. All three pairs of dice were represented, though with a smaller number of runs than were made in the two-dice tests. The results of

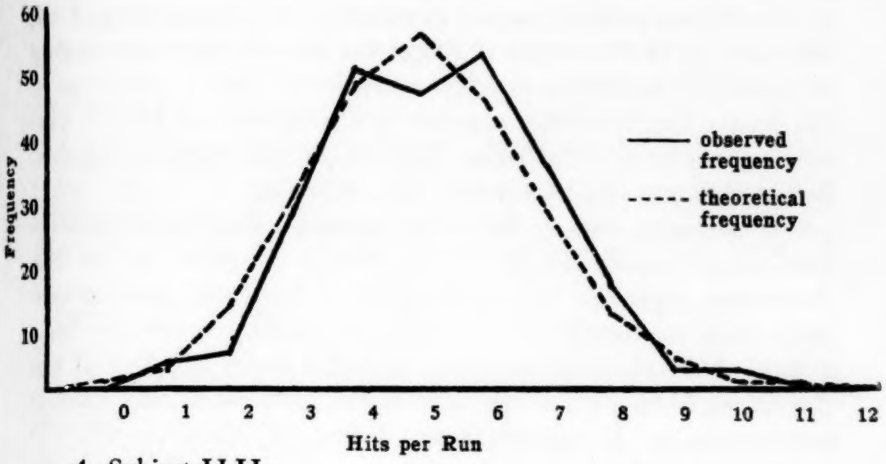
all the one-die tests conducted approximate very closely to expectation from chance if we include the one-die data from the Miscellaneous Series.[1] The latter add seven runs to the 67 shown in Table 2, giving an average of only 5.01 which suggests that, for some reason, the PK effect which functioned on the two-dice tests did not work on the one-die. (This is generally true of one-die throwing done in comparison with throws of more dice.[2]) But there is no reason whatever for supposing that if the dice were *faulty to a degree favoring high dice,* they would not have displayed this fault with one die per throw as well as with two dice per throw. Psychological differences can readily be supposed to exist, but not physical ones.

Other controls on the trueness of the dice are available or are revealed in the analyses of the Miscellaneous Series. The total of 88 runs made by the three subjects, W.S., J.A., and J.C.L., gave a negative deviation of 17. All three sizes of dice were involved; the medium dice were used in all but 16 runs, but both the large and small dice gave negative deviations (large, 14 runs, −16; small, 2 runs, −2). These data carry considerable weight against the hypothesis that faulty dice produced the results. And there are other such controls, as for example, the work of one of these three subjects, W.S., which consisted of 20 runs with all three sizes of dice represented and gave a *significantly negative deviation* (CR = 2.88). The same dice could not, by reason of being faulty, give both positive and negative scoring.

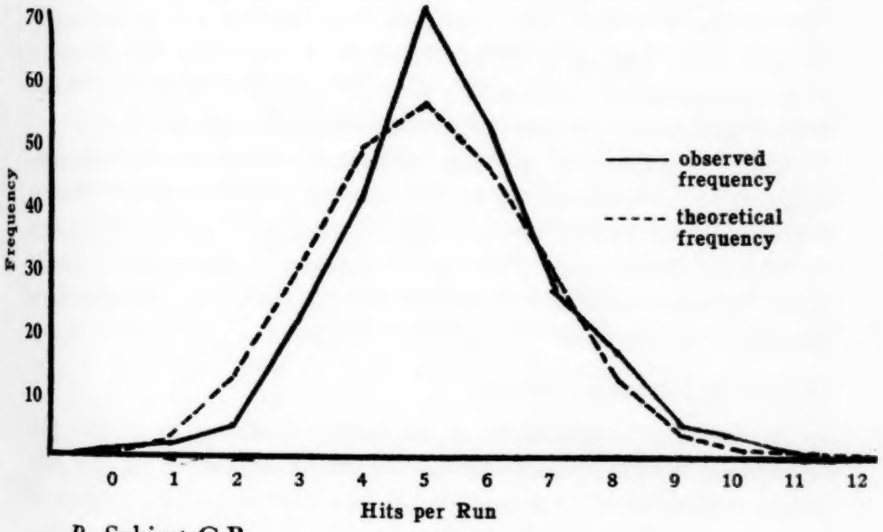
Two further types of evidence against the faulty dice interpretation are worth stressing: First, there is one type, dealt with under "Declines" below—especially that part devoted to the analysis of H.H.'s data—which will be found to constitute strong evidence against this counter-hypothesis. That discussion need not be anticipated here. Second, there is the evidence offered by the frequency of scores, when plotted, as in the curve of Figure 1, which shows for H.H. a lower frequency of fives than is expected by chance, with a

[1] This inclusion of the one-die results from the Miscellaneous Series is quite permissible since the conditions were essentially the same. The addition of the two-dice tests of that series to the two-dice section of the Main Series gives a CR of the whole of 2.94.

[2] In the paper on PK by J. B. Rhine in this issue (4), a similar drop to expectation occurred on the one-die tests. The explanation might be that one-die testing is slow and tedious in comparison with the two-dice.



A. Subject H.H.



B. Subject G.B.

FIG. 1. Score frequency distributions of subjects H.H. and G.B., showing both observed and theoretical distributions. Graph A represents H.H.'s work. (Compare it with Graph A, Fig. 1, of Hilton and Rhine report in this issue.) Graph B represents G.B.'s work.

greater number of both fours and sixes. No dice would be expected to give this two-peak or bimodal distribution. Most interesting of all, this curve of H.H.'s scores is almost the same as that representing his results in the Hilton and Rhine article (2) located elsewhere in this issue. This similarity suggests that some trait of H.H.'s personality may be involved rather than any physical defect in the dice. In fact, different dice were used in the two series.

At the same time, G.B.'s score frequency distribution shows a much larger number of fives, thus offering a control on the dice themselves, especially the combinations of the three pairs of dice from which the distributions of scores are pooled. Against the background of the observed frequency of G.B.'s scores and that of the theoretical, H.H.'s dip in the curve at the five-point in both series is most interesting. It suggests, though it does not prove, consistency in high scoring tendencies at times, with reversal to depressed tendencies at other times.

The score frequency distributions are given in the appendix to this article for the convenience of those who may wish to go over them. The results of certain other statistical tests likewise are presented in the appendix. They give further information regarding the character of the experimental results which may interest some readers, but they are not essential to the general conclusions of the report.

Taken together, the various checks and numerous controls constitute strong reassurance as to the adequate trueness of the dice—that is, sufficient for the requirements of the tests. As has often been stated it is obvious that there are no "perfect" dice. But it could hardly be maintained in the face of all these facts that imperfections produced the significant scoring reported here.

Manner of Throwing the Dice

A third counter-hypothesis which might be offered to explain the extrachance results is the possibility of skilled throwing by the subjects, a possibility which is ruled out by a comparison of the success of the two methods by which the dice were thrown: the hand-throwing and the mechanical-release. Table 3 summarizes the complete results for the two subjects who used both methods. It shows that the mechanical-release method is somewhat the better of the two in terms of score average: 5.36 as compared to 5.24. The results of the

Table 3

COMPARISON OF HAND-THROWING AND MECHANICAL-RELEASE PK TESTS

Subject	Hand-Throwing						Mechanical-Release					
	Runs	Av. Score	Total Hits	Dev.	SD	CR	Runs	Av. Score	Total Hits	Dev.	SD	CR
H. H....	180	5.25	945	+45	±22.95	1.96	64	5.14	329	+ 9	±13.68	0.66
G. B....	176	5.22	919	+39	±22.69	1.72	64	5.58	357	+37	±13.68	2.70
Total...	356	5.24	1,864	+84	±32.27	2.60	128	5.36	686	+46	±19.34	2.38
Grand Total.	H. H. and G. B. 484 runs: +130±37.62 CR=3.46											

mechanical-release tests, totaling 128 runs, have a CR that is nearly significant (2.38).

Here, as in the work of Rhine and Rhine, when the possibility of manually influencing the dice is entirely removed, the score average rises appreciably. The deviations of the average scores from expectation are .36 and .24, respectively, for mechanical-release and hand-throwing. In the earlier work, the corresponding figures were .65 and .51.

As Related to Conditions Affecting PK

The foregoing considerations clear the way for a discussion of some of the internal relations of the results on the assumption of the PK hypothesis. If chance, faulty dice, or manner of throwing afford no explanation, we must proceed with the assumption that either there is some defect in the experiment of which we are not aware or else the PK hypothesis is required to account for the results. We shall proceed, then, on the assumption that the PK effect is established and that it was operative in the production of the results of this investigation. This may be taken by the reader as merely a tentative position for purposes of discussion, for in dealing with a phenomenon so radically novel to the present scientific world-view and particularly to the psychological sciences, the student would do well to reserve judgment until more of the available research has been reported and brought into the perspective of critical consideration.

The three sizes of dice gave results which are in opposite order to size, that is, the smaller the dice, the higher the score average. As Table 4 below indicates, these averages are, for the small, medium,

Table 4

COMPARISON OF RESULTS WITH THREE SIZES OF DICE

Subject	Small		Medium		Large		Pooled		
	Runs	Aver.	Runs	Aver.	Runs	Aver.	Runs	Aver.*	CR†
H. H.....	80	5.34	109	5.10	55	5.29	244	5.22	2.02
G. B.....	77	5.38	116	5.37	47	5.09	240	5.32	2.87
Total...	157	5.36	225	5.24	102	5.20	484	5.27	3.46

*Expectation is 5.00.

†The total deviation from expectation, divided by the SD. In this case, the SD is found by the following process: $\sqrt{R \times 35/12}$ where R = number of runs. This is the familiar $\sqrt{npq}$ formula.

and large dice, respectively, 5.36, 5.24, and 5.20. Here, for the first time thus far in the PK literature, appears a set of experimental results which seem to offer some little conformity to expectation from mechanical principles. It is true that the proportion is far from corresponding (inversely) to the mass of the three sets of dice. The deviations of the average scores for the three sizes are .36, .24, and .20 and are of the following ratio, 9:6:5. This is quite out of line with the volume ratio of the dice, which is approximately 1:4:12.

A more definite disagreement with the mechanical analogy is found in the individual score averages, as may be seen in the table referred to. For example, subject H.H., while he did best on the small dice, did second best, and nearly as well, on the large dice, producing a much lower result on the medium-sized. Again, subject G.B. produced almost the same average score on the medium-sized dice as on the smallest. These individual peculiarities, although not significant, are sufficiently important to suggest that psychological factors rather than the mass of the dice are determinative. However, more experimental work is needed.

There is another aspect of the experiment which bears on the question of the applicability of mechanical principles to the results. It concerns the comparison of different numbers of dice thrown at a time. There was mentioned above the fact that a portion of the tests were conducted with one die thrown at a time. The objective was to determine whether the PK effect, if it appeared, would be more potent in dealing with a single die than with the pair. Physically regarded, this would be expected. However, the results actually show a much higher average for the two-dice than for the one-die

tests, as may be seen in Table 2. Their average on the two-dice tests was 5.29, and, for the one-die, 5.13. Whatever the determining force may be, it manifestly worked more effectively in the two-dice than in the one-die tests.

The mere fact that the small dice, which gave the highest positive deviation in the two-dice tests, gave the largest *negative* deviation in the one-die tests suggests (as far as the data are adequate) that the difference is due to some such psychological factor as the tediousness of throwing one little die at a time after the faster-moving tests with the pair.

Discussion

For the most part, the results speak for themselves. The general average of 5.27 is somewhat lower than the averages reported for the high-dice studies in the first two reports on the PK effect. But it is statistically significant for the relatively long Main Series of 484 runs. It is obvious from the comparison of the mechanical-release and the hand-throwing scores that the average would have been improved by introducing a larger portion of mechanical-release tests. Also, it seems probable that too great a burden of testing was placed upon the small number of subjects. Too many runs were made at a session, whereas in the earlier report referred to, the subject was, in the majority of sessions, stopped after two or three runs.

The role of individual difference is well brought out in the fact that while subject G.B. did nearly three times as well in terms of average deviation on the mechanical-release as he did on the hand-throwing, subject H.H. did considerably better on the hand-throwing, giving nearly twice as large an average deviation. Similar differences, too, appear in preferences for the different sizes of dice, as was pointed out above.

Declines

In the first article on the PK effect by Rhine and Rhine, there was reported a decline of score level in the three-run sets that were commonly conducted at a session. In the present investigation, the experiments were not done in short series but were often conducted through a long sequence of as many as 64 runs in a day. There was, however, an arrangement in the recording which placed four runs on

each page of the record book. It occurred to us in the analysis to inquire whether the four-run page showed a decline in score level across the page from left to right. The total pooling of pages on this plan did not, however, reveal any such decline but produced the following summary of average scores per run on the page: 5.13, 5.42, 5.29, 5.25. These results may not be strictly comparable to the case reported in the earlier paper, although the conditions were such that some degree of decline was anticipated when the analysis was begun. This is very exceptional in the PK reports, published and unpublished. For this reason, the decline of the last three of the four runs (figures just given) is at least worth noting. It is as if some interference or counter-effect entered into the first run. But this is, of course, a speculative matter.

However, another type of decline has been found since the appearance of the first PK report. It will be recalled that in that report, as in this one, there was a 12-trial run, each trial consisting of the throw of a pair of dice. It is important to mention, too, that the lines on the record page *divided each run into four segments.* When, recently, the data of the Main Series of that report were examined for vertical decline of scoring in the run, they showed an obviously *double* decline; that is, there was a drop in scoring from the first to the second segment of the run, then a rise in the third, and a drop again in the fourth. Immediately after this double decline was noted, the large Unwitnessed Series of that report was also examined and a similar double decline was found.[3]

[3] The data from this still unpublished analysis provide the following deviations, according to segments of the run:

Main Series

Segment of the Run	*1st*	*2nd*	*3rd*	*4th*
Deviations Pooled	+ 115.5	+ 51.5	+ 86.5	+ 41.5

Unwitnessed Series

Segment of the Run	*1st*	*2nd*	*3rd*	*4th*
Deviations Pooled	+ 49	+ 11	+ 48	+ 21

These results give differences that are significant when the two series are pooled. For instance, the first and second segments are significantly different, as are the first and fourth; while the third and fourth are almost so. But the way of estimating differences that has now been adopted as the standard for double declines is to test the difference between the sum of the first and third segments, on the one hand, and the sum of the second and fourth, on the other. By this test, the very significant CR of 3.44 is given for the whole experiment. These results, obtained since publication of the original report, will appear in fuller detail in a later survey of these and other double declines.

The ground for expecting a double decline in the data of the present report lies in the fact that, as stated, the run was segmented by lines on the record sheet.[4] What these lines do, merely in subdividing the run, to produce this type of decline is not yet known. It would seem that they may accomplish their effect by reminding the subject when he is midway. Then some subjects may reorient themselves and take a fresh start. Thus it may be that we get the effect of two beginnings and, hence, of the normal vertical decline repeated.

In view of these double declines, a search was made for other records from which there was reason to expect the same effect (records with 12 entries divided into four segments) and the present report was among these. Including the Miscellaneous Series, the deviations by segments are as follows:

Segment of the Run	*1st*	*2nd*	*3rd*	*4th*
Deviations Pooled	+ 23	− 4	+ 55	+ 39

The first and third quarters are to the second and fourth as 78:35. The difference is not a significant one, but with an approximately two-to-one ratio, it is sufficiently marked to be of interest.

This double decline is contributed entirely by H.H.; in fact, his deviations per quarter in the run are:

+ 21 − 6 + 25 + 14

This gives a ratio of first and third to second and fourth segments of 46:8. The principal advantage in calling attention to this distribution is to link it up with the quite similar distribution found in the analysis of H.H.'s work reported in another paper, also appearing in this number, by Hilton and Rhine. There, H.H.'s work, and again his alone, shows the double decline, the segments in the run giving the following deviations:

+ 28 + 10 + 60 + 32

The ratio here is 88:42.

It is interesting to obtain the distribution of hits by quarters of the record page or other unit of data, as was done in the Rhine and

[4] Of the PK work reported to date, the Rhine and Rhine report, the Hilton and Rhine (in this issue), and this one are the only ones having the segmented runs on the record sheet. The Reeves and Rhine report (3) has a 12-trial run but it is unsegmented and shows no double decline. Instead, it shows a simple vertical decline with a moderate turning-up at the end that provides a U-curve of hit distribution.

the Gibson and Rhine experiments. However, in view of the double decline found in H.H.'s work, the *half* page should be taken as a unit here rather than the whole page. Further details of the analysis are given in the Hilton and Rhine paper. The deviations are found for each quarter of the half page (upper and lower); these half-page quarter distributions are then pooled so as to give the total distribution of four quarters based on the half page as a unit or set. The final distribution by quarters for the two series by H. H. stated separately, as well as for the two series pooled, is as follows:

This Report	
+ 25	+ 21
+ 2	+ 6

Hilton and Rhine Report	
+ 57	+ 31
+ 45	− 3

H.H.'s Work, Both Series Pooled	
+ 82	+ 52
+ 47	+ 3

There is a *significant difference* (with a CR of 2.57) in the pooled work of H.H. between the two quarters which would be most expected to show such a difference, the upper left and lower right. This significant diagonal decline compares favorably with the earlier quarter declines mentioned.

The significant differences of H.H.'s work are helpful, too, in clarifying the interpretation of the results with respect to the faulty dice question. Biased dice would not explain this result.

G.B.'s results do not show a decline; rather, they show an *incline.* The horizontal distribution on the page is very slightly upward: + 37 to + 39; and the vertical more so: + 26 to + 50. The diagonal, upper left to lower right, of the quarter-page distribution is 12 to 25. Even if we arrange a distribution of the half page by quarters, as was done for H.H. because of his double declines, the diagonal (upper left to lower right) for G.B. is still an incline: + 18 to + 24. Finally, his rate of success for the entire experimental series was on the incline, the deviation of the first half to that of the second being + 33 to + 43.

The distributions of deviation for H.H. and G.B. are thus apparently individual. At any rate, they are opposite in trend and rather consistently so. H.H.'s distribution resembles that of his other work, not only in giving a double decline, but also in the quarter de-

cline on the half-page basis. The fact, too, of the similarity of his bimodal score frequency distribution for the two series combines to bring a total picture of considerable carry-over of pattern by H.H. from one series to the other. All this fits with the PK hypothesis, but hardly with any other.

Miscellaneous Series

As a feature of some interest, we mention an item from the work of minor subjects referred to as the Miscellaneous Series. The total of 88 runs of this work gave a negative deviation of 17. The negative effect was due to one subject who was so consistently negative as to constitute a phenomenon in himself. His 20 runs gave − 22, which has a CR of 2.88 and thus is significant. His scores were negative on all three dice, and for one-die as well as two-dice tests, making five negative subdivisions altogether. When the distribution of hits in the run was plotted in four segments (of three throws each), his deviations were: − 12, − 8, − 1, − 1 for the four subdivisions. Quite aside from its bearing on the question of the trueness of the dice, these results are of interest in considering the mode of functioning of the PK effect. Like ESP (and any other mental ability), it can be consistently misdirected, as well as consistently directed, at the target or objective.

Number of Dice per Throw

The general results with different numbers of dice per throw, which showed much better scoring with two dice than with one, are quite in line with those reported in the work by Rhine and by Gibson and Rhine showing a quite different relation than would be expected from a consideration of strictly mechanical principles. But if we remember that there is no reason whatsoever known to psychology to expect volitional attitudes to follow mechanical principles and that the appliction of physical analogy to the workings of the mind has been a purely hypothetical matter (even a relatively unfruitful endeavor), we find ourselves under much less conflict over the acceptance of these findings for what they are worth.

Size of Dice

The results from the comparison of sizes of dice were less clear-cut, favoring an inverse relation between size and success, but not

doing so to a degree that would be expected from known physical law. The relation is all out of proportion to that which would be required on the mechanical analogy. There is some conflict here with the results from the comparison of different numbers of dice, but the bearing of the latter is the more definite of the two. Again, however, it is appropriate to say that this work is a part of a sequence. It is not the first and it shall not be the last report bearing upon the relation of the PK effect to physical law.

It is much too early, then, to generalize regarding the possible significance of a nonmechanical system of causation identifiable with the psychical determinants in human personality. It is, however, fully in order to invite the attention to this inquiry of those students of human nature who have been hampered in their thinking by the defects of fitting all mental life into a system of laws and properties derived from the inorganic world.

Concluding Remarks

The work reported above has shown significant evidence of the PK effect in high dice tests. It confirms in this respect the five other reports which have appeared already or which accompany this one. Like the Rhine and Rhine series, it gives higher score averages with mechanical release of the dice than with hand- or cup-throwing.

It also offers considerable evidence of more than one kind against the faulty dice hypothesis. Tricky throwing as an explanation, too, may be dismissed. Also, the results from two dice per throw are superior to the one-die throwing just as in the work of the Gibson and Rhine report in which six dice per throw did as well as two per throw. The slight advantage in score average of the smaller dice more sharply focuses the question of size of dice as related to success, and further relevant work will be reported.

A significant "diagonal difference" appears in the quarter decline shown by H.H.'s work combined with that done by him in the other experimental series reported in this issue under his name. This report introduces the double decline in the run.

APPENDIX

Table 1

SCORE FREQUENCY DISTRIBUTIONS FOR THE EXPERIMENT

Hits	H. H.		G. B.		Miscellaneous Subjects		Grand Total		Control Series	
	Expected Frequency	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency	Observed Frequency
0.....	.38	0	.37	1	.14	0	.89	1	.20	0
1.....	3.25	4	3.19	2	1.17	0	7.61	6	1.70	4
2.....	12.75	7	12.55	5	4.60	6	29.90	18	6.69	5
3.....	30.37	28	29.87	21	10.95	13	71.19	62	15.93	18
4.....	48.81	50	48.01	40	17.60	20	114.42	110	25.60	20
5.....	55.78	46	54.87	70	20.12	22	130.77	138	29.26	33
6.....	46.48	52	45.72	51	16.76	12	108.96	115	24.38	20
7.....	28.46	35	27.99	26	10.26	10	66.71	71	14.93	18
8.....	12.71	16	12.50	16	4.58	5	29.79	37	6.66	7
9.....	4.03	3	3.97	5	1.45	0	9.45	8	2.12	2
10.....	.86	3	.85	3	.31	0	2.02	6	.45	0
11.....	.11	0	.11	0	.04	0	.26	0	.06	1
12.....	.01	0	.006	0	.002	0	.044	0	.003	0
Total...	244.00	244	240.006	240	87.982	88	572.014	572	127.983	128

Table 2

SUMMARY OF CHI SQUARE GOODNESS OF FIT TESTS

Grouping	$\Sigma\chi^2$	d.f.	P
Experimental			
H. H.	6.91	7	.43
G. B.	15.84	7	.03
Miscellaneous	2.44	5	.73
Total	17.32	9	.04
Control	.21	4	.98

Table 3

SUMMARY OF TESTS OF VARIANCE

Grouping	s^2	s	$\chi^2=\frac{Ns^2}{\sigma^2}$	d.f.	2P
Experimental					
H. H.	3.03	1.74	253.19	243	.64
G. B.	2.70	1.64	221.92	239	.44
Total	2.82	1.68	552.41	571	.58
Control	3.33	1.82	145.97	127	.24

References

1. Gibson, Edmond P., and Rhine, J. B. The PK effect: III. Some introductory series. *J. Parapsychol.*, 1943, **7**, 118-134.
2. Hilton, H., Jr., and Rhine, J. B. A second comparison of three sizes of dice in PK tests. *J. Parapsychol.*, [this number].
3. Reeves, Margaret P., and Rhine, J. B. The PK effect: II. A study in declines. *J. Parapsychol.*, 1943, **7**, 76-93.
4. Rhine, J. B. Dice thrown by cup and machine in PK tests. *J. Parapsychol.*, [this number].
5. Rhine, L. E., and Rhine, J. B. The psychokinetic effect: I. The first experiment. *J. Parapsychol.*, 1943, **7**, 20-43.

A SECOND COMPARISON OF THREE SIZES OF DICE IN PK TESTS

By HOMER HILTON, JR., and J. B. RHINE

ABSTRACT: On the completion of the research reported elsewhere in this issue and bearing his name, Mr. Hilton undertook a still more lengthy series of PK tests with himself and two members of his family as subjects. It is one of the earliest of the confirmatory PK researches. Eight hundred and twenty-four runs were made, in each of which a pair of dice were thrown 12 times for high dice. The scoring was highly significant. The dice were shaken and (in all but a few specified runs) thrown with the use of a cup so that the possibility of skilled throwing was relatively well eliminated. The dice were not specifically tested for bias; but the analysis of the records shows, at least in the case of subject H.H., that imperfections in the cubes could not have accounted for the extrachance results. This conclusion is strengthened particularly by the evidence offered by declines. Other findings in the experiment consist of a confirmation of the double decline found in H.H.'s earlier series and of his peculiar score frequency curve, which shows a twin peak or bimodal effect. Three different sizes of dice were used. The largest scored well below chance, the medium gave the highest results, with the small somewhat lower. Again, two dice per throw scored better than one die.—Ed.

INTRODUCTION

THE RESEARCH with which this report deals is a natural sequel to the earlier study entitled, "A Comparison of Three Sizes of Dice in PK Tests," by Hilton, Baer, and Rhine (2), which appears in this number. While in most respects the conditions and procedure are the same as in the earlier paper, there are certain distinctions to be made. In the present work, all but a specified few of the tests were conducted with cup-throwing, whereas in the earlier work the dice were mostly thrown from the hand with a considerable portion mechanically released on an inclined plane. For the most part, in this experiment as in the other, two dice were thrown at once, with only a small section of the work done with one die.

Only three subjects took part in the work reported here, namely, H.H., his sister, M.C., and her husband, S.T.C. The investigations were made in the summer of 1934, immediately following the period

in which the earlier investigation was made. They were conducted at the home of M.C. and S.T.C. in New York.

Procedure

Three pairs of dice were used which will be called "large," "medium," and "small." (These were not the dice used in the earlier work.) They measured, respectively, 1, 11/16, and 7/16 in. on the edge and were of plastic material, red, green, and amber in color. Most of the throwing was done from a leather dice cup, but a few runs were thrown by hand by M.C. and S.T.C. The dice were shaken and tossed into a tray set on a card-table, the tray having side walls to allow the subject to "bump" the dice about and to prevent the dice from falling off.

H.H. recorded for the other two subjects and M.C. recorded for H.H. The recording was done in sets of four runs to the page in bound composition books, one book for each subject.

The subject was given full freedom of choice as to which pair of dice was to be used for a given set or page of four runs, as well as when he wished to begin or stop the procedure, except that a session was always stopped with a completed page. As much spontaneity as possible was allowed. The few runs of hand-throwing were for purposes of comparison; the two subjects who did these thought they might do better thus.

The test conditions were maintained free from distractions as far as was possible in a normal household; but there were interruptions. It is especially noteworthy that two pages, or eight runs, done by H.H. while his little niece hovered about the low table trying to reach the dice, gave him his lowest scoring.

A control series on the trueness of the dice was projected and was to be done with naive subjects who were ignorant of the PK hypothesis and the general experimentation. The conditions suited to this control did not arise during the time available. However, a number of controls on the dice have emerged from the analyses, and they warrant, in our judgment, the consideration of the dice as adequate to the demands of the tests here described. On this point, the reader may himself judge, since this issue will be fully discussed below after the results of the experiment and of the analyses have been presented.

A word on subject attitudes may be interesting, though necessarily difficult to appraise. Perhaps these attitudes can best be described by quoting from a letter written by H.H. to J.B.R. on October 7, 1934, shortly after the summer's work: "I feel that I can say that I still have an open mind in regard to this work in dice throwing. Personally, I will neither accept it, nor reject it, and I'm sure that I would be unwilling to commit myself upon it until I had thrown a good many thousand more throws, or seen a cumulative mass of evidence that I could not disregard without being unreasonable. My sister and brother-in-law feel much the same as I do about it, and we are willing to do some more work at a later date in order to give you more material with which to work on in this field.

"As for the personalities of my two willing aides, I am at a loss to comment. My brother-in-law is the type who would become very angry at the dice when they failed to roll out with numbers above eight as he thought they should. He is a strenuous competitor in sports, never caring to participate unless he can do better than the average. His temper did not seem to affect his throwing adversely, in fact quite the contrary might be said. My sister was given to no displays or expressions of displeasure if success was not her lot."

Results

The Chance Hypothesis

The results will be presented in three general series, each associated with the subject who produced it. The total number of runs of 24 single die-throws (or 12 throws of a pair of dice) is 824. The average score for the entire research is 5.29 per run where 5.00 is expectancy (see Table 1). The total number of trials giving high

Table 1

Total Results of PK Tests Covered by the Report

Subjects	Runs	Average Score	Deviation	SD	CR
H.H.	404	5.32	+130	±34.37	3.78
S.T.C.	208	5.35	+ 73	±24.64	2.96
M.C.	212	5.19	+ 40	±24.88	1.61
Total	824	5.29	+243	±49.07	4.95

dice (faces totaling eight or above) is 4,363, or 243 above the chance expectation of 4,120. This is a highly significant deviation. The SD for 824 runs is ±49.07 and the CR, therefore, is 4.95. The itemized figures of Table 1 show that both H.H. and S.T.C. have significant CR's while that of M.C. is in the suggestive range.

There is, then, no need to consider further the question of whether these results are attributable to chance; the CR of 4.95 warrants the dismissal of the chance hypothesis without reservation. Either some physical factor or set of factors is responsible, or it is likely that the psychical principle being tested is at work. We must next distinguish between these two interpretations.

As stated above, the dice were thrown almost entirely from a cup. However, there were 52 runs thrown by hand in the sections by M.C. and S.T.C. But since these give an average score of 5.23 while the rest of their work gives 5.28, this small group may be ignored. Obviously, hand-throwing was not advantageous.

Unlike the preceding work in which H.H. participated, these experiments do not have a specific control series that was carried out to test the trueness of the dice. In our view, however, it will be possible to obtain ample light on the question of the applicability of the faulty dice hypothesis from the analyses of the results, and we will therefore keep primarily in mind, as the presentation of the results continues, the question of whether the dice were faulty.

Analysis for Declines

In the judgment of a number of investigators, the most important feature of the PK work, at least at this stage of the research, is the large amount of influence which mere position of a trial in the run or set of runs seems to exert on the rate of scoring. This effect is usually one of decline in scoring rate, both horizontally across the page and vertically in the run (hence, also, diagonally from upper left to lower right).

Each page of the record books contains four runs recorded in vertical columns so that from left to right there are four natural subdivisions to the page. In view of the considerable number of researches in which decline effects have been found, it is regarded as important to analyze the results of this investigation from that point of view. Accordingly, the pages were pooled for the entire work of

each subject so that a comparison could be made of the relative success on the four runs of the page. It was found that a marked general decline occurred from left to right, that is, as the subject proceeded with the test across the record page. Section A of Table 2 summarizes the results in terms of total deviation from expectation. The total number of runs is given and the work of all three subjects pooled. This gives an especially marked decline with the exception of the fourth run, on which there is a slight upturn. The upturn might in itself represent a degree of terminal salience (see glossary) attributable to the isolation of the end position, and it should be recalled that Rhine and Rhine also found such an upturn in the fourth run after a decline in the first three.[1] The consistency of the horizontal decline is remarkable when the various subdivisions in Table 2 are examined.

Table 2

SCORE DECLINE ACROSS THE PAGE BY RUNS

A. By Subjects

Subjects	Total Runs	Deviation by Order of Runs			
		1	2	3	4
H.H.	404	+49	+53	+ 8	+20
S.T.C.	208	+26	+ 1	+22	+24
M.C.	212	+16	+10	+ 8	+ 6
Total	824	+91	+64	+38	+50

B. By Number of Dice

Number of Dice per Throw	Total Runs	Deviation by Order of Runs			
		1	2	3	4
1	56	+10	+ 5	− 9	+ 4
2	768	+81	+59	+47	+46

C. By Size of Dice Used

Size of Dice	Total Runs	Deviations by Order of Runs			
		1	2	3	4
Large	88	+ 7	− 4	−16	− 9
Medium	412	+50	+37	+35	+33
Small	324	+34	+31	+19	+26

[1] The same thing happened in the Reeves data, but since that work was done entirely in three-run sets, the fourth run was the first of the next set on the page.

When the results are divided on the basis of how many dice are thrown at a time, we find that the one-die work, as well as the two-dice, shows a similar trend (see Section B of Table 2). When the distribution is obtained for the three sizes of dice, the decline is again evident in each instance (see Section C of Table 2). But while this decline is confirmatory in trend, it is not sufficient for a significant difference.

These declines are of considerable interest, even though they are not "significant" and regardless of how completely it can be established that the dice were adequately free from imperfections. When we compare the deviations of the total scores on the basis of half pages, left and right, as has been done in the Gibson and Rhine work (1), a marked drop in average score from left to right (see Table 3)

Table 3

COMPARISON OF DEVIATIONS OF SCORE AVERAGES, LEFT AND RIGHT HALF PAGE

A. As to Number of Dice Per Throw

Number of Dice Per Throw	Runs	Left Half Page Deviation of Average Score	Right Half Page Deviation of Average Score	$CR_{diff.}$
1	56	+.54	−.18	....
2	768	+.36	+.24	....
Total	824	+.38	+.21	1.37

B. As to Size of Dice

Size of Dice	Runs	Left Half Page Deviation of Average Score	Right Half Page Deviation of Average Score	$CR_{diff.}$
Large	88	+.07	−.57	....
Medium	412	+.42	+.33	....
Small	324	+.40	+.28	....
Total	824	+.38	+.21	1.37

is encountered. This appears in both the one-die and two-dice throwing and in all three sizes of dice. It is thus quite consistent and is in line with the instances of horizontal decline effects hitherto reported, with the exception of the Hilton, Baer, and Rhine work.

Vertical Declines. The vertical decline, or the decline of scoring in the course of the run, in this series shows the same kind of double

decline which appeared in the Hilton, Baer, and Rhine investigation and also the same as that appearing in the two series of the Rhine and Rhine work (4). Scoring fell off from the first segment (comprising three trials) to the second, rose again at the midpoint, and fell off from the third to the fourth. (See deviations of these segments of the run in the column marked "Total This Series," Table 4.) In the earlier series in which H.H. participated, he was the contributor of the double decline effect, and in this one likewise. (See Table 4 for a comparison of the double decline of H.H. in the two

Table 4

VERTICAL DECLINES BY SEGMENTS IN THE RUN

Segments of the Run	H.H. This Series	H.H. Earlier Series	S.T.C.	M.C.	Total This Series
1.....	+28	+21	+21	+22	+ 71
2.....	+10	− 6	+27	+14	+ 51
3.....	+60	+25	+18	−12	+ 66
4.....	+32	+14	+ 7	+16	+ 55
Total.....	+130	+54	+73	+40	+243

series.) The other two subjects, S.T.C. and M.C., show relatively normal declines. This difference between them and H.H. raises an interesting question and suggests, for what it is worth, a speculative answer. Is it not reasonable to suppose that H.H., who was the experimenter and who designed the recording procedure, was more conscious of the structure of the run as it was segmented on the record page than were the two who were subjects only? They on their part were presumably little concerned about the recording and methodological aspects of the test. Unfortunately, the question cannot be answered from these data, but it could be put to test.

Quarter Declines. There is a further analysis for position effects that has now become a customary step with all records to which it can be applied. This is a logical consequence of the finding of the horizontal and vertical declines. It is called the quarter decline analysis. This treatment consists in dividing the page (or section) into quarters and checking the hit distribution therein. The normal anticipation would be that if both horizontal and vertical declines occur, the upper left quarters would be highest in scoring, the lower right, lowest,

with the other two in between. Of course with a subject showing a double decline (vertically) as H.H. did, the upper and lower half pages should be quartered separately.

The distribution of hits by quarters for subjects M.C. and S.T.C. shown below, however, is given for the entire page since they did not show double declines.

DEVIATIONS BY QUARTER PAGE FOR S.T.C. AND M.C.

S.T.C.

+22	+26
+ 5	+20

M.C.

+18	+18
+ 8	− 4

These are not the characteristic quarter declines that have been given in the earlier report by Gibson and Rhine and the accompanying paper by Rhine on cup- and machine-thrown dice (3), but they are a part of the record. That of M.C. is close enough to the type to warrant interest.

Since H.H.'s vertical decline was "double," as it had been in the earlier series, it was decided here also to obtain the order distribution by quarters for each of the two halves of the page. These half-page distributions in terms of deviation per quarter are shown below. They are followed by a pooling of the quarter declines of the two halves of all the pages which H.H. did in this series. Finally, the quarter decline of H.H.'s earlier work is combined with that of the present experiment.

H.H. This Series, Top Half

+29.5	− 1.5
+21.5	−11.5

H.H. This Series, Bottom Half

+27.5	+32.5
+23.5	+ 8.5

Both Halves Pooled, This Series

+57 (1)	+31 (3)
+45 (2)	− 3 (4)

$CR_{diff.\ 1\text{-}4} = 2.47$

This and Earlier Series Pooled

+82 (1)	+52 (3)
+47 (2)	+ 3 (4)

$CR_{diff.\ 1\text{-}4} = 2.57$

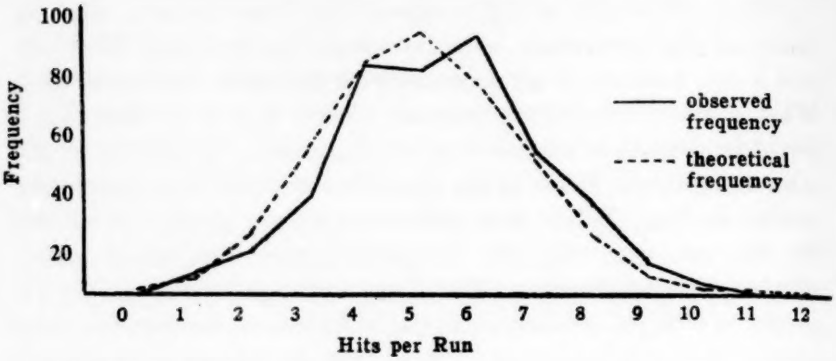
It will be seen from these data that there is a diagonal decline, a significant drop, that is, from the upper left to the lower right quarters, not only in the pooled work of H.H. for this report, but also when his work as a whole for both reports is combined.

These differences strongly support the other evidence that the results of this experiment were extrachance in character. That support is not, however, at all necessary with the CR in Table 1 of 4.95. What is much more to the point of present need is the fact that it would be difficult to suppose that the inequalities of the dice or any other recognizable factor in the experimental procedure caused these quarter declines. While these differences are not absolute proof that the dice were true, they offer a significant effect that cannot reasonably be ascribed to the dice. That alone is enough for the case for PK insofar as a single research goes; for H.H.'s work represents a large sample, and there is no need to draw upon the support of M.C.'s and S.T.C.'s results. For that matter, the same dice were used by all; H.H.'s score average was about the same as that of the combined work of the other two subjects; and, accordingly, it is not unreasonable to suppose that their results, too, were to some degree at least attributable to the PK effect. But of this we need not make an issue. H.H.'s work alone sufficiently establishes the need for a psychological explanation which, under the experimental conditions, could hardly be other than the PK hypothesis.

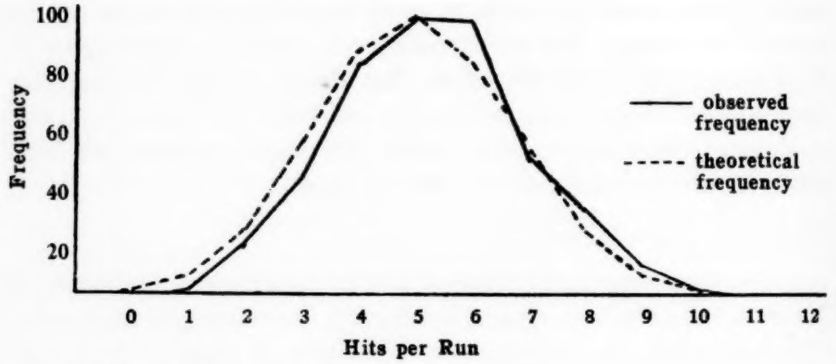
Score Frequency Distribution

The results, especially those of H.H., contain a strong indication, if not more definite evidence, against the hypothesis of faulty dice when the data are analyzed from the point of view of score frequency. The curve of distribution of score frequency is shown in Figure 1 for the different subjects, and along with these, the curves of expected frequency. It will be seen from these graphs that H.H. has produced in this series a two-peak or bimodal curve of score distribution as he did in the earlier series. (See Figure 1 of the Hilton, Baer, and Rhine paper.) The observed frequency of his scores of five is much below theoretical expectation. There is no kind of dice imperfection that would tend to produce curves such as those given by H.H.'s distribution in this series and his earlier work.

It might conceivably be argued, nevertheless, that the peculiar curve which H.H. obtained could be due to dice imperfections only if each of the three different pairs of dice used were loaded or biased differently. It will be noticed, however, that S.T.C. and M.C. do not show the bimodal curve; and since they used the same dice under



A. Subject H.H.



B. Subjects S.T.C. and M.C.

FIG. 1. Score frequency distributions of subjects, showing both observed and theoretical distributions. Graph A represents H.H.'s work. (Compare it with Graph A, Fig. 1, of Hilton, Baer, and Rhine report in this issue.) Graph B represents the work of S.T.C. and M.C.

the same conditions as H.H., this fact strengthens the interpretation just given to H.H.'s score distribution: that it argues against the faulty dice hypothesis.

The score frequency distributions and the tests of variance and chi square godness of fit are given in the appendix to this article for the convenience of those who may wish to go over them. These tests give further information regarding the character of the experimental results which may interest certain readers but they are not essential to the general conclusions of the report.

Bearing of the Results with Three Sizes of Dice

All three pairs of dice were of the same general manufacture and structure. None were supposed to be biased, though of course none were perfect, since perfection is only a hypothetical state. Further reassurance, then, that psychological rather than physical factors are playing a part in the dice-throwing tests may be obtained from an analysis of the results with respect to the size of dice as produced by the different subjects. (See Table 5 below for the comparative results in terms of deviations of average scores.) Here there are some im-

Table 5

RESULTS WITH THREE SIZES OF DICE, EXPRESSED AS DEVIATIONS OF AVERAGE SCORES

Subject	Large Dice		Medium Dice		Small Dice	
	Runs	Deviation of Average Score	Runs	Deviation of Average Score	Runs	Deviation of Average Score
H.H.	40	−.22	200	+.42	164	+.34
S.T.C.	36	−.14	92	+.39	80	+.53
M.C.	12	−.67	120	+.29	80	+.16
Total	88	−.25	412	+.38	324	+.34

portant differences. M.C. obtained a deviation on the large dice almost five times that of S.T.C., and S.T.C.'s deviation on the small dice is over three times as great as M.C.'s. All three subjects scored well below expectation on the large dice, and yet they did not score their best on the small dice, as one might be led to expect on the basis of physical analogies. On the contrary, the best scoring average was on the medium dice. If physical factors, such as excavations from the face of the dice produced in marking the spots, caused the favoring of high dice on the small dice, then we should have expected positive results on the large dice too. But they are negative for all subjects.

On the other hand, if we may be permitted a semi-speculative paragraph, the results are much easier to account for from the psychological point of view. The medium dice are about the right size for the "taste" of the average person. It is common observation that few people like to play with small dice for a very long time; they become tedious. This is almost invariable if there are medium

dice for comparison. Yet, if the dice are too large, they are less popular, too, than medium dice, and a certain amount of dissatisfaction accumulates with their continued use. The big dice were one inch on the edge, which is quite large for shaking in the ordinary leather dice cup and for throwing in the small confines of the tray. The size of the dice to be used was optional and the large ones were chosen for only 88 of the total of 824 runs. The medium dice were the favorites; the fact that they were thrown for 412 runs is some evidence of their greater popularity. This could be due also to their better scoring; but as a matter of fact, S.T.C. did better with the small dice, yet he threw more often with the medium. In the earlier Hilton, Baer, and Rhine work, likewise, the medium-sized dice were used the most frequently, even though on the whole they did not give the highest average score. In that work, too, the large dice, thrown by hand and with ample space in which to throw them, gave almost as high a score average as the medium dice (5.20 compared to 5.24). They did *not* go negative! On the whole, the low scoring with the large dice in this series seems therefore most reasonably interpreted as an effect of the degree of frustration that was experienced in trying to throw these large cubes from a small cup in a space too confined to give them a fling commensurate with their mass.

There is some statistical basis, too, for this psychical *vs.* physical interpretation of the negative deviation of the large dice. S.T.C.'s deviations per run on the page for the large dice are:

Total Runs	*1st Run*	*2nd Run*	*3rd Run*	*4th Run*
36	+11	+2	−11	−7

This is a marked horizontal decline, and it gives significant differences, first between the left half (1st + 2nd = + 13) and the right half (3rd + 4th = − 18). Then, too, the first and third have a significant difference and the first and fourth likewise. These results, even when corrected for the isolation of the case of a single subject out of three, argue strongly against the hypothesis of bias in the large dice. The significant differences are indicative of mental rather than of physical factors that entered into the course of the experiment.

Distraction

Finally, another instance of the role of psychological factors in the production of these results may be worth mentioning even though

it cannot be taken wtih full seriousness because it was not a deliberately planned aspect of the experiment. As mentioned earlier in the paper, there was some difficulty at times in obtaining freedom from distraction under the normal conditions of the household in which the tests were conducted. A major distraction was experienced in the course of eight runs which were carried out with H.H. as subject while his nineteen-months-old niece was present, hovering about the low table and trying to reach the dice. These eight runs were the only dice thrown under this condition. They gave, as compared to the 40 hits expected by chance, a total score of 23, or *17 hits below expectation.* This is, of course, a selected block of data, but such a result would not occur once in five thousand times. Tests thereafter were not made until after the little girl was in bed.

Discussion

When we consider the various cumulative evidences of psychical, as against physical, factors, it seems apparent that H.H.'s work, at least, stands out clearly as being reasonably interpretable in terms of the PK hypothesis. His data show, first, a score frequency which is clearly extranormal and which could not have been produced by a chance distribution of physically distorted dice. Second, his quarter decline in the present work, as well as in the combination of the present work and his earlier series, is significant and follows the direction to be expected from quarter declines found in other experiments. Third, the incident of the child's distracting H.H. for eight runs, which gave a very significant deviation on the negative side, again suggests that psychical factors are dominant.

In addition to H.H.'s work, there is a fourth piece of evidence in the negative scores produced by all subjects with the large dice, and especially by S.T.C. who showed a marked drop in scoring as he changed from one size of dice to another. Fifth, the consistency of the horizontal declines for all subjects in the various subdivisions of Tables 2 and 3 indicates a factor at work that is behavioral rather than inherent in the dice or attributable to chance. Finally, the overall view of the experiment reduces to a conclusion favoring psychical factors in explaining the scores. Since all possibility of skilled manipulation is ruled out and the presence of imperfections in the dice cannot account for the results, there is left only the alternative

hypothesis that the influence which produced the extrachance scores must have been a mental one.

We recognize that, in terms of controls, this research is less complete than is commonly the case. More analysis and discussion have been needed to interpret the results than in earlier reports; but the work has points worth bringing out, once the interpretation is clear. The superiority here again of the two-dice-per-throw tests over the one-per-throw is in line with all other comparable findings. The size-of-dice comparison, however, fails to offer a conclusive answer to the question of the relation between PK and physical law. Size is even less clearly correlated with success than in the earlier Hilton, Baer, and Rhine report. Moreover, the declines are in themselves a contribution, whether or not the differences involved are significant. They add to the growing collection of comparable examples of position effects that are emerging from the analyses of the PK results. The similarity of H.H.'s two series and of his double declines and quarter declines are the most valuable features of the study. The similarity, too, of his bimodal score frequency distributions, as shown in the graphs, makes this present report a link in the development of knowledge as to the factors at work.

Concluding Remarks

This experimental test of the PK hypothesis is one of the earliest investigations on the subject, but it represents a modest contribution to the evidence in favor of that hypothesis. It gave very significant results; accordingly, chance is not the explanation. The dice were almost always cup-thrown. In the work of only one of the three subjects, H.H., are the results clearly interpretable as not due to faulty dice; but this subject did nearly half of the total test runs. All used the same dice. His work shows a significant drop (diagonal decline) from the upper left to the lower right quarter of the half page distribution of deviations, the half page being used as a basis because of the double declines shown in this and H.H.'s earlier work.

The report contains several other pieces of evidence against the faulty dice interpretation. It was found that two dice per throw gave better scores than one. The scores with the large dice were very poor, those with the small dice fair, and those with the medium dice were the best of all.

The research raises several questions. The lowered frequency of five-scores in both series of H.H.'s work in itself poses a problem for inquiry, as does his double decline, and there is, in addition, the question raised by the very negative results with the large dice.

APPENDIX

Table 1

SCORE FREQUENCY DISTRIBUTIONS FOR THE EXPERIMENT

Hits	H.H.		S.T.C.		M.C.		TOTAL	
	Expected Frequency	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency	Observed Frequency
0.....	.63	0	.32	0	.33	0	1.28	0
1.....	5.38	7	2.77	0	2.82	2	10.96	9
2.....	21.12	16	10.87	8	11.08	10	43.07	34
3.....	50.28	34	25.89	23	26.39	16	102.56	73
4.....	80.81	80	41.61	36	42.41	43	164.82	159
5.....	92.36	79	47.55	46	48.46	51	188.37	176
6.....	76.96	89	39.63	45	40.39	50	156.98	184
7.....	47.12	51	24.26	25	24.73	23	96.11	99
8.....	21.04	33	10.83	17	11.04	12	42.91	62
9.....	6.68	11	3.44	7	3.50	4	13.62	22
10.....	1.43	4	.74	1	.75	1	2.92	6
11.....	.19	0	.10	0	.10	0	.38	0
12.....	.01	0	.01	0	.01	0	.02	0
	404.01	404	208.02	208	212.01	212	824.00	824

Table 2

SUMMARY OF CHI SQUARE GOODNESS OF FIT TESTS

Grouping	$\Sigma\chi^2$	d. f.	P
H.H.....	23.00	8	.0034
S.T.C.....	8.45	7	.29
M.C.....	7.08	7	.429
Total.....	32.86	9	.0004

Table 3

SUMMARY OF TESTS OF VARIANCE

Grouping	s^2	s	$\chi^2=\frac{Ns^2}{\sigma^2}$	d. f.	2P
H.H.....	2.59	1.61	358.34	403	.11
S.T.C.....	2.95	1.72	210.14	207	.86
M.C.....	2.63	1.62	190.95	211	.33
Total.....	3.05	1.75	860.68	823	.36

References

1. Gibson, Edmond P., and Rhine, J. B. The PK effect: III. Some introductory series. *J. Parapsychol.*, 1943, **7**, 118-134.
2. Hilton, H., Jr., Baer, G., and Rhine, J. B. A comparison of three sizes of dice in PK tests. *J. Parapsychol.*, [this number].
3. Rhine, J. B. Dice thrown by cup and machine in PK tests. *J. Parapsychol.*, [this number].
4. Rhine, L. E., and Rhine, J. B. The psychokinetic effect: I. The first experiment. *J. Parapsychol.*, 1943, **7**, 20-43.

DICE THROWN BY CUP AND MACHINE IN PK TESTS

By J. B. RHINE
Parapsychology Laboratory, Duke University

ABSTRACT: In this experiment conducted to test the PK hypothesis, the procedure, as usual, consisted of dice-throwing, with the subjects attempting to cause the dice, by direct volitional action, to fall with a specified face upturned. In one section, a pair of dice was thrown from a large cardboard cup; in another, the dice were automatically thrown by an electrically-driven, rotating wire cage. The objective of the experiment reported here was to compare the effects upon the PK process of the two methods of throwing.

The results of the experiment were significantly above chance expectation. Of the two main conditions, cup- and machine-throwing, those of the mechanical-throwing tests gave the higher score average. They were even independently significant.

No question of tricks in throwing the dice is relevant under the completely mechanical conditions of throwing. Likewise, there are two reasons why possible defects in the dice could not have produced the results obtained: First, there is the work mentioned below in which the same dice produced results below expectation with the same machine but with a different mental condition in the subject. Second, a significant decline effect was found between the top and bottom halves of the run. Such a decline could not be attributed to any dice defect. It is itself evidence of PK and an extrachance feature of prime importance also to the psychological study of the PK effect.—Ed.

INTRODUCTION

THE PK HYPOTHESIS, holding that a person is able to exert a statistically significant effect on the random fall of dice without contact of any recognized nature, has already been the subject of three reports in this *Journal*. Two more appear concurrently with this one in the present issue.

For appraising a PK experiment, there has already been established a certain order of importance of the various issues that must be considered. The primary consideration is whether the results are extrachance in character. Next, it is essential to show that the dice are not producing the statistical results by reason of imperfections, such as imbalance or favoring of the target face (i.e., the face which

the subject tries to make turn up in the act of throwing the dice). Then, if chance and biased dice do not afford an explanation, the next problem is the elimination of the possibility of tricky dice-throwing or of any influence on the test score due to the manner of handling the dice. This third question is the special point of focus of this paper, which, in this respect, follows logically upon the first report on the PK effect by Rhine and Rhine (5). In that research, there were two methods of throwing the dice: hand-throwing (including some cup-throwing) and the mechanical release of the dice by means of an inclined board down which they tumbled on being released.

The increased success which obtained in the earlier work, when the subjects were transferred from the hand-throwing to the mechanical-release method, encouraged further attempts in the same direction. In that experiment,[1] the hand-throwing averaged 5.51 hits per run (where 5.00 was expected) and the mechanical-release averaged 5.65. Accordingly, we were reasonably confident that a completely mechanized method of handling the dice could be substituted for the semi-mechanical procedure of the first report. In the latter procedure, the subject was allowed to pick up the dice and place them on the ledge from which they dropped when released. The next step, as taken in the present study, was to have them thrown in a motor-driven, rotating cage with no manual contact whatever by the subject.

Description of the Experiments

On the twenty-eighth and twenty-ninth of October, 1936, I undertook, with the help of Mr. J. L. Woodruff,[2] then a graduate student in psychology and my research assistant, and of Mr. A. J. Linzmayer, then secretary to the Parapsychology Laboratory, to make a comparison of cup-throwing and machine-throwing, the three of us acting as subjects. Woodruff was in charge of the experiment.

For the machine-throwing, we used a newly constructed, electrically-driven machine which had a rectangular cage, 25 inches in length and four by four inches at the ends, made of quarter-inch wire mesh. This cage rotated on an axis through the middle, throwing the dice

[1] A similar comparison occurs in the PK article by Hilton, Baer, and Rhine (2) in this issue wherein the mechanical-release method again gives the higher score average.

[2] Later, Dr. Woodruff. He is now serving with the armed forces of the United States.

from one end to the other over a rough course which made them bound about considerably. The roughness was accentuated by the addition of various baffles and bumping devices.

The motor was regulated by rheostat, and the cage was rotated at the rate preferred by the subject, usually so as to make a half revolution in from four to five seconds. The dice were read at the lower end of the cage on each half revolution while the cage was in motion, and an immediate record was made on prepared record sheets by J.L.W.—or by another of us, if he was the subject.

For the cup-throwing, a special dice table was used, with padded top and side-boards. The dice bounced freely on it and against its walls. For a cup, a heavy cylindrical cardboard container was found very suitable. It was better than a commercial dice cup since it was larger (of about one quart capacity) and allowed more room for shaking. The interior was not at that time roughened by rubber lining as it was later, but the subject was required either to shake the dice before throwing them, or else bump them against the side-boards of the table.

The subjects were allowed to select the face for which they wished to throw, and they chose the six-face throughout this experimental series. When the throw was made, they tried to exert a direct, determinative mental influence on the dice to make the six-face come up. Strong concentration of effort and attention was evident in manner, in posture, and through introspective reports. Sometimes the subject verbally expressed his attitude in words playfully directed toward the dice. Each subject had his own degree of expressiveness and his own manner of throwing.

The run record comprised 24 entries in a single column, each column consisting of 12 throws of a pair of dice. The tests were made in blocks of four runs to a set or sequence with alternating cup- and machine-throwing. Eleven sets were done by machine and nine by cup. Thirty additional runs made with the machine on the same days, though incidental to the main comparison, are included to make the report complete.

The run was checked and scored at once, and at the end of the two-day series, a summary of all the data was made by J.L.W. and filed with me. Thus there is every assurance that the record is com-

plete in spite of the lapse of more than six years' time between the experiment and this reporting of it. The double observation during the recording, the checking, and scoring, as well as the complete rechecking of the records, renders the likelihood of clerical errors comparatively small.

The pair of dice used were common white cubes of plastic material measuring 11/16 of an inch on the edge and marked with black paint. They were not inlaid. A control series was made to test the dice for serious inequalities; but before it became adequately extensive,[3] better controls appeared in the form of the significant difference obtained with the same dice under different conditions. And later on, still better controls emerged from the statistical analyses (see section on the "Decline Effect"). Discussion of the question of the adequacy of the dice will be given below.

Approximately six years after these experiments had been completed, the rechecking and analyses of the results were undertaken under the direction of Miss Betty M. Humphrey of the Parapsychology Laboratory at Duke. Outstanding among these analyses are the studies of hit distributions in the run and in the group of runs known as the "set." The results are presented under the heading of the "Decline Effect."

General Results

There were 110 runs made by the three subjects, or a total of 2,640 die-throws. A total of 74 runs were made with the machine and 36 by the cup method. Table 1 shows that the total of 110 runs gave 504 hits (for the face aimed at), which is 64 above expectation.

Table 1

Comparison of Cup- and Machine-Throwing

How Thrown	Runs	Average Score*	Total Hits	Deviation	SD	CR
Cup.....	36	4.53	163	+19	±10.95	1.74
Machine.....	74	4.61	341	+45	±15.69	2.87
Total.....	110	4.58	504	+64	±19.14	3.34

*Expectation here is 4.00.

[3] As far as it went, it gave almost exactly the theoretically expected score average: 279 trials with 46 hits; 46.5 are expected.

In this instance, expectation is 4.00 per run, or 440 altogether. With an SD of ± 19.14, this gives a CR of 3.34, which is quite significant. By accepted statistical standards, chance is thus ruled off the list of possible counter-hypotheses.

Special Results and Discussion

Before we go further, the biased dice hypothesis must be considered. This is the hypothesis that the extrachance score total was due to faulty dice. It can be refuted here on two counts: First, there are the results of a series of tests made with the same cage and dice but under different psychological conditions, with J.L.W. as subject and Miss Margaret M. Price as observer. Although a detailed report of that experiment will be published later, this much may be stated now: The social situation of the experiment was expressly intended to interfere with the subject's performance and cause him to go low in his scores. In the 80 runs of that experimental series, the average score was 3.84, while 4.00 is expectation. All throws were for the six-face, as in the present paper. Obviously, the hypothesis of dice defects cannot be used to explain both sets of results, those *above* expectation with a CR of 3.34, and those, as just cited, which were *below*. The fact that the two series are significantly different (CR of diff. = 2.76) strengthens this conclusion. The "Student's" *t*-test of the difference between these two series gives a *t* of 2.72 which, with 188 degrees of freedom, has a *P* of less than .01.

Possibly a still better treatment of the question of dice imperfections is available in the further analyses to which the results were subjected in 1942, after the decline effect was discovered in the first PK research by Rhine and Rhine. The data of this decline analysis will be presented and discussed below, where it will be seen that the significant difference obtained as a result of the decline could not reasonably be explained on the basis of imperfections in the dice. This represents an approach quite independent of the treatment in the preceding paragraph and one that is conclusive in itself.

With chance and biased dice off the slate, we can now compare the results with cup and machine. The figures of Table 1, as stated above, are not quite comparable for the two methods. The 74 runs of the machine have in them 20 runs of a special side-experiment in competition between A.J.L. and J.L.W., using one die at a time in-

stead of a pair as was usual. The results of these 20 runs were poor, giving a deviation of only + 1 for the whole. For several reasons, these should not properly be included in a comparison of the cup- and machine-throwing. No comparable tests were made with cup-thrown dice. Moreover, a group or team test was made on the machine, all three subjects "ganging up" on the dice to see if there was any combined effect. The ten runs gave a deviation of + 6 (average score of 4.60), which is not as good as we averaged on the machine when each was working alone. Again, there was nothing in the cup series comparable to this team test on the machine. As stated previously, these runs are reported simply to afford a complete record.

This leaves 44 runs with a positive deviation of 38 as the machine work which is strictly comparable to the cup data. The average score is 4.86, which is quite a bit above the 4.53 of the cup method. The SD for 44 runs is 12.10 and the CR of the positive deviation of 38 is 3.14. (See Table 2.) From this more correct basis of comparison, it

Table 2

A More Detailed Summary of Results from Comparison of Cup- and Machine-Throwing

Series	How Thrown	Runs	Average Score	Total Hits	Deviation	SD	CR
Main Series.....	Cup	36	4.53	163	+19	±10.95	1.74
Main Series.....	Machine	44	4.86	214	+38	±12.10	3.14
Special Tests.....	Machine	30	4.23	127	+ 7	±10.00	.70
Total.....	.....	110	4.58	504	+64	±19.14	3.34

is clear that the completely mechanized system of agitating the dice is certainly under no disadvantage when compared to cup-throwing. The average score of 4.86 represents a deviation of 22 percent above expectation, and this percentage is higher than that of the average score of 5.65 (where 5.00 was expectation) in the mechanical-release work of the first PK report. The latter gave a deviation 13 percent above expectation. However, it is not possible to make accurate comparisons between two such widely different situations. It is enough to say that the PK effect on the cage technique worked, and worked well—better than it did on the cup-thrown dice.

General Discussion

The lower scoring obtained with the throwing of a single die (20 runs with average of 4.05) is something to compare with the one-die tests of the other PK articles in this issue. A single die should, on physical grounds, be more readily influenced than a pair.

The brief series made to test group effort in PK, like that series devoted to competition, is not large enough to allow the drawing of any conclusions but is suggestive of further profitable lines of investigation, namely, the effects of various social conditions in the test environment.

The superior scoring of the mechanical-throwing test over cup-throwing (as suggested by the averages of 4.86 and 4.53, respectively), confirms the two earlier comparisons (**5, 2**), and incidentally it raises a question as to the psychological factors controlling success. It may well be that the cup- and hand-throwing involve operations which, while they do not affect the dice by PK, *do* demand a certain amount of attention and effort and, hence, distract the subject's real PK effort. While using the machine, he is more detached, less attentive to his own motor activity (to the way in which he throws the dice), and is therefore more "whole-souled" in the essential operation of the test. This view suggests a parallel pointed out (**4**) in the ESP work in which somewhat higher scoring was obtained with the test cards *completely* removed from the subject's sensory field—e.g., by removal to another room—than was given with the cards in view of the subject—i.e., with backs visible. In short, with ESP, *complete* elimination of sensory contact with the experimental object seemed to aid in the extrasensory perception. Now we see that, similarly in PK, *complete* elimination of motor contact with the experimental objects seems to aid in the *extra* motor causation or influence. Both in the ESP and in the PK researches, this relation remains as only a suggested one, but it may lead to further discoveries of conditions affecting performance.

The Decline Effect

After the above report had been written, and many years after the work had been completed, it occurred to us to look into these results for declines similar to those reported in the preceding papers on the PK effect. It was, as it turned out, a very fruitful inquiry, for

it yielded an independently significant relationship that would alone support the case for the PK effect. Moreover, coming after the experimental work was done, this finding has a certain control value that is unique, for the evidence it furnishes is not subject to the general run of counter-hypotheses.

Vertical Declines

To make the vertical decline analyses, the hits were counted separately for the top and the bottom halves of all the columns of the 110 runs. The record page was thus divided at the middle into an upper and lower section which gave the equivalent of 55 runs for each half. The top gave 59 hits above average expectation and the bottom gave only 5 above. The difference of 54 is significant, 2.82 times the SD of the difference (± 19.14). This, then, may be taken as an extrachance effect, and one that is not possibly explainable in terms of faulty dice or tricky throwing, for such would not be expected to cause a decline. Several other counter-hypotheses are similarly excluded. Here is a clear-cut verification of the vertical decline of the Reeves and Rhine report (3).

Horizontal Decline

Since *all* the runs had been made in four-run sequences or sets except the 20 one-die tests (for competition), it was possible to analyze the remaining 90 runs in the search for position effects. For instance, the comparison of success over the four-run set was made. It gave the following average scores for the four runs in the set:

4.74 5.00 4.73 4.32

These, too, show a general decline trend from left to right, the left half score average being to that of the right half in the ratio of 4.87: 4.52. They do not show the immediate drop from the first to the second which the Rhine and Rhine and the Reeves and Rhine reports showed, but the drop of .87 to .52 in the deviations of the score averages is sufficiently in line with earlier reports to warrant mention.

Decline by Quarters

Then the distribution of deviation was found by quarters of the four-run set. It will be recalled that in the Gibson and Rhine report

(1), a similar analysis gave a quarter-set decline pattern that was highest in the upper left quarter and lowest in the lower right, with the other two quarters in between. This same trend is found in the present study. The average scores per quarter are of course strictly comparable (4.00 is expectation):

	Left	*Right*
Upper.....	5.43	5.05
Lower.....	4.30	4.00

There is a significant difference between the upper left (5.43) and lower right (4.00) corners, with a CR of the difference of 2.63. This is the relation most to be expected from earlier work and needs no correction for selection.

The earlier (Gibson and Rhine) set was recorded on the record page so as to throw its long axis horizontally across the page with record entries six in the column, 20 columns to the set; while ours is vertical, with 24 entries in the column, four columns to the set. In view of this difference between the two researches under comparison, the similarity of decline pattern by quarters shown in the figures given below is all the more remarkable. Strictly speaking, in order to make the two distributions closely comparable, the + 22 of the Gibson and Rhine data should be compared with the + 23 of the present report. The two + 7's are comparable. Thus, the first part of the long axis of the set has, in both cases, the highest deviations, the higher of the two in each case being in the quarter on which the subject began the set. Likewise, in both reports, the second half of the long axis has the lowest deviations, the lower of the two being in the quarter on which the subject ended. The comparison is all the more interesting because the number of runs represented by the two researches is exactly the same: 90 runs in each. In the Gibson and Rhine series, they are evenly represented in the quarters of the set. In the present work, the two quarters on the left have 23 runs each; and those on the right, 22.

Gibson and Rhine Report

Long axis→

+ 36	+ 7
+ 22	+ 1

Present Report

←Long axis

+ 33	+ 23
+ 7	+ 0

On the other hand, the circumstances under which the two experiments were conducted were far from similar, at least in external conditions and procedure. In the one case, the dice were thrown from a cup and six were thrown at a time. In the other, two dice were thrown at a time and the throwing was divided between cup and machine. The subjects were different; the experimenters likewise. It is true that the general effort to influence the dice was perhaps as closely similar as personal motivation in two individuals is likely to be.

The emphasis upon this similarity can of course be overdone. As a matter of fact, it seems very likely that so striking an approximation of two wholly independent series of PK tests will not, in the very nature of human variability, be found again. On the other hand, the significant differences that obtain in both of these quarter declines between the upper left and lower right quarters permit us to regard them as not being wholly accidental occurrences. This appears to be evidence of the very best type in favor of the existence of some factor beyond the range of chance or experimental error. These results are evidently the effects of the lawful operation of a psychical principle.

This is the kind of verification that makes a research its own reward! It comes not of planning, nor yet of accident, but clearly as an evidence of a general relation running through these several investigations. What but psychical determinants could have given these results? And what could those determinants be?

Concluding Remarks

This is the sixth report of research contributing to the evidence for the PK effect that has appeared in publication. It adds further evidence for the decline effect as well. And finally, it shows that as good, if not better, results can be produced with mechanically-thrown dice as with cup-thrown.

At the outset, it was stated that no single study should or need be taken as a sole and sufficient basis for the establishment of the PK hypothesis and that a long series of reports was ready for the printer, all of which would be of value in any final judgment regarding the occurrence of the PK effect. This position should be re-emphasized, not out of any known weakness of the research as it has been conducted and reported, but rather because of the extremely revolutionary character of the PK hypothesis.

If one feature were to be singled out for special emphasis, it would be the quarter-set decline pattern. It is so similar to the Gibson and Rhine pattern, with both showing significant differences from top left to bottom right, that both look as if molded by the same forces. Other examples will be reported in due time.

References

1. Gibson, Edmond P., and Rhine, J. B. The PK effect: III. Some introductory series. *J. Parapsychol.*, 1943, **7**, 118-134.
2. Hilton, H., Jr., Baer, G., and Rhine, J. B. A comparison of three sizes of dice in PK tests. *J. Parapsychol.*, [this number].
3. Reeves, Margaret P., and Rhine, J. B. The PK effect: II. A study in declines. *J. Parapsychol.*, 1943, **7**, 76-93.
4. Rhine, J. B. The effect of distance in ESP tests. *J. Parapsychol.*, 1937, **1**, 172-184.
5. Rhine, L. E., and Rhine, J. B. The psychokinetic effect: I. The first experiment. *J. Parapsychol.*, 1943, **7**, 20-43.

GLOSSARY

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

EXPECTATION; see CHANCE.

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

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

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

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

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

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

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

MEAN CHANCE EXPECTATION; see CHANCE.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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