A CONTROVERSY OVER CHARGES OF FRAUD IN ESP

EDITORIAL INTRODUCTION

THE FOLLOWING four articles constitute a controversy which has recently arisen concerning the possibility of fraud in two ESP experiments of the thirties. The critic who makes the accusation is Mr. C. E. M. Hansel, Lecturer in Psychology at the University of Manchester in England, and the experiments against which he makes this charge are known as the Pearce-Pratt series, first published in 1937, and the Pratt-Woodruff series, published in 1939. The experimenters who were involved in those researches were Dr. J. B. Rhine, Director of the Parapsychology Laboratory of Duke University; Dr. J. G. Pratt, Assistant Director and one of the editors of this *Journal;* and Dr. J. L. Woodruff, Associate Professor of Psychology, City College of New York, and Secretary and Member of the Board of Trustees of the American Society for Psychical Research.

It has always been the editorial policy of the *Journal of Parapsychology* to welcome critical articles, although the editors have preferred, when possible, to clear up disagreements out of court, so to speak. This could have been done in the present instance if Mr. Hansel had so desired, for he was at the Parapsychology Laboratory for a number of weeks, having been invited there to facilitate his examination of the literature of the field. But he preferred to submit his papers for publication after his departure, thus precluding any other means of carrying on the discussion. On this account, his charges are being published here in full, together with the replies of Drs. Rhine, Pratt, and Woodruff.

A CRITICAL ANALYSIS OF THE PEARCE-PRATT EXPERIMENT

By C. E. M. Hansel

In the Pearce-Pratt experiment[1] the subject obtained high scores when guessing cards whose identities were unknown to any other person.

The subject was a student named Hubert Pearce. He obtained high scores under two conditions:

(A) When the cards were in a room on the top floor of the Social Science Building of Duke West Campus, approximately 100 yards away from him;

(B) When the cards were in a room on the top floor of the Medical Building approximately 250 yards away from him.

Procedure

The general procedure adopted in the experiments is described as follows in the report:—

> At the time agreed upon, H.E.P. visited J.G.P. in his research room on the top floor of what is now the Social Science Building on the main Duke campus. The two men synchronized their watches and set an exact time for starting the test, allowing enough time for H.E.P. to cross the quadrangle to the Duke Library where he occupied a cubicle in the stacks at the back of the building. From his window J.G.P. could see H.E.P. enter the Library.
>
> J.G.P. then selected a pack of ESP cards from several packs always available in the room. He gave this pack of cards a number of dovetail shuffles and a final cut, keeping them face-down throughout. He then placed the pack on the right-hand side of the table at which he was sitting. In the center of the table was a closed book on which it had been agreed with H.E.P. that the card for each trial would be placed. At the minute set for starting the test, J.G.P. lifted the top card from the inverted deck, placed it face-down on the book, and allowed it to remain there for approximately a full minute. At the beginning of the next minute this card was picked up with the left hand and laid, still face-down, on the left-hand side of the table, while with the right hand J.G.P. picked up the next card and put it on the

[1] Rhine, J. B., and Pratt, J. G. A review of the Pearce-Pratt distance series of ESP tests. *J. Parapsychol.*, 1954, **18**, 165-78.

book. At the end of the second minute, this card was placed on top of the one on the left and the next one was put on the book. In this way, at the rate of one card per minute, the entire pack of 25 cards went through the process of being isolated, one card at a time, on the book in the center of the table, where it was the target or stimulus object for that ESP trial.

In his cubicle in the Library, H.E.P. attempted to identify the target cards, minute by minute, and recorded his responses in pencil. At the end of the run, there was on most test days a rest period of five minutes before a second run followed in exactly the same way. H.E.P. made a duplicate of his call record, signed one copy, and sealed it in an envelope for J.B.R. Over in his room J.G.P. recorded the card order for the two decks used in the test as soon as the second run was finished. This record, too, was in duplicate, one copy of which was signed and sealed in an envelope for J.B.R. The two sealed records were delivered personally to J.B.R., most of the time before J.G.P. and H.E.P. compared their records and scored the number of successes. On the few occasions when J.G.P. and H.E.P. met and compared their unsealed duplicates before both of them had delivered their sealed records to J.B.R., the data could not have been changed without collusion, as J.G.P. kept the results from the unsealed records and any discrepancy between them and J.B.R.'s results would have been noticed. In Subseries D, J.B.R. was on hand to receive the duplicates as the two other men met immediately after each session for the checkup.

Uncontrolled Features of the Experiment

The results of the four subseries were all such as to make it certain that, in the absence of ESP, there must have been some premeditated plan to obtain high scores by normal means. If H.E.P. had wished to carry out a trick, it is clear that he had every opportunity to do so. From the time he entered the library, he was unobserved. He had a complete time schedule so that he knew to the minute the stage which the experiment had reached, and he also knew that during this time J.G.P. was occupied with the cards. He had 50 minutes to walk back to the Social Science Building or Medical Building to obtain sight of the cards. Thus we have to investigate the possibility of his being able to obtain information in this way. For this purpose it is necessary to know in detail the arrangements in each of the rooms used by J.G.P. We also note that the procedure adopted by J.G.P. would have made it virtually impossible for any sight to have been got of the cards except when he was recording them at the end of the experiment or when he left the room after the experiment.

While at the Parapsychology Laboratory I asked Dr. Pratt to demonstrate the procedure he adopted, and it was clear that the cards were made clearly visible during the recording of the targets at the end of the runs. I also ascertained that he did not lock his door, that the packs of cards were left unshuffled after the experiment, and that he made his record on notepaper. Dr. Pratt told me that he took his copy of the targets to Dr. Rhine but that he did not see H.E.P. until some time (unspecified) after each sitting. The following possibilities therefore present themselves at this point.

(1) That H.E.P. in some way got sight of the cards as they were turned up for recording.

(2) That H.E.P. waited until J.G.P. left the room after the experiment to take his copy of the target record to Dr. Rhine, and then entered the room and recorded the targets from the packs.

(3) That H.E.P. saw the second copy of the target lists, assuming that J.G.P. had left this on his desk, or obtained an imprint of

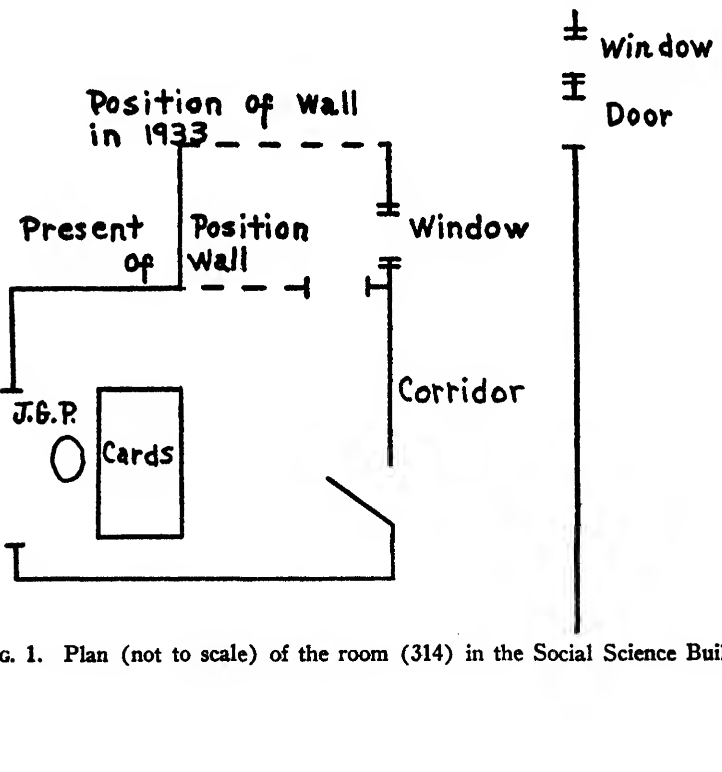

FIG. 1. Plan (not to scale) of the room (314) in the Social Science Building.

it from the next sheet on a pad of notepaper on which J.G.P. made his record.

(4) That H.E.P. had an assistant who obtained a record of the cards in one of the above ways.

The Room in the Social Science Building

In this room Dr. Pratt sat with his back to the window (Fig. 1). There was a pane of clear glass above the door, but it would have been a difficult matter for a short person like H.E.P. to stand on a chair and look through. I understood from J.G.P. that the wall on one side of the room had been moved since the experiment. In its previous position a clear glass window about 2′ square was available for anyone to look through from the corridor. The bottom of this window was about 5′ 10″ from the floor, and only a tall man in the corridor could conceivably have seen through it without standing on a chair. There is, however, a clear pane of glass above the door in a room on the opposite side further down the corridor through which sight might have been got of the cards by anyone standing on a chair inside the room with the door shut. This possibility was difficult to test owing to the structural alterations which had been made.

The Room in the Medical Building

The room I saw has structural alterations which have been made since 1933. There was a window into the corridor but this was made of ripple glass. The most obvious way in which sight might have been got of the cards was through a trap door in the ceiling. This trap door was about 4′ by 1½′ and was directly over the table at which J.G.P. sat. Although the room was on the top floor, the main staircase carries on up another flight into a large attic which extends over the lower rooms. It is thus possible that if this trap door was present in 1933, an observer could get into position over the trap door and, provided there was a hole in the cover, obtain sight of the cards. When I saw it, there was a circular hole in the cover which looked as if it had been added recently, but there was also a small piece of metal which appeared to cover a smaller hole. It would also have been a simple matter to drill a small hole and then fill it at a later date.

When the experiments were transferred to the room in the Medical Building, the scores dropped for the first five sittings to the chance level. If a trick was employed, this might have been due to difficulty in getting the relevant information under the new conditions. Scores also sometimes dropped for a complete sitting. Thus they were well above chance for the two sittings in which runs 37, 38, and runs 43, 44 were completed, but they were below the chance level for the two sittings in which runs 39, 40 and 41, 42 were completed. This effect might have arisen if on some days there was difficulty in getting into a position where sight of the cards or the records could be obtained.

Thus as the experimental conditions were such that above-chance scores could have been obtained by means of a trick, this experiment cannot be cited as providing evidence for ESP.

I also understand that statements have been made that H.E.P. was "seen somewhere where he should not have been." If by this is meant that he was seen wandering around when he should have been in the cubicle, or if he was seen leaving either of the rooms used by J.G.P. after the experiments, such statements should be brought into the light and carefully examined.

I would like to thank Dr. J. G. Pratt for giving me every assistance in obtaining information about the experiments which was not available in the experimental report.

Department of Psychology
Manchester University
Manchester, England

A REPLY TO THE HANSEL CRITIQUE OF THE PEARCE-PRATT SERIES

By J. B. Rhine and J. G. Pratt

In spite of his unconcealed eagerness to give the *coup de grâce* to a piece of ESP research, Mr. Hansel is entitled to our appreciation for bringing our experiment of twenty-seven years ago momentarily back into the limelight. This acknowledgment may be linked with the information that, although Hansel's negative approach to parapsychology was a matter of public knowledge, he was invited and given a travel grant by the Duke Laboratory to make the visit which resulted in his paper. There are still other points of value arising from his critique, but they can best be left to follow our evaluation of his remarks about the experiment.

We cannot, of course, condone Hansel's methods; instead of directly endeavoring to clear up his differences with the authors (which might have needed a few minutes' discussion when he was here), he has gone off into print without asking the reaction of the authors. Whether he really has the exceedingly strong case this course of action might suggest or is only exceedingly motivated to damage the evidence for ESP on the grounds of his own prejudgment, the reader will be able to determine for himself.

Comment might first be made on the technique of Hansel's climactic revelation that he has heard a rumor that H.E.P. was "seen somewhere where he should not have been." A rumor is hardly a proper basis for a scientific criticism; but even if such a rumor had any support (and we know of none), our comment would still be that if the conditions of the experiment were adequate (as we shall show they were), it should have been impossible for the rumored circumstance to have produced the results. In other words, the proper focus is so obviously on the adequacy of the methods that the introduction of the rumor is not in order.

Adequacy of Test Conditions

Were the test conditions adequate? On this issue, there is one very essential point to get straight at the outset, and Hansel has

not got it straight. Any piece of research under criticism obviously stands or falls on its strongest, best-controlled section. Anyone confining his attack to any other section is either misled or misleading or both. And that is precisely what Hansel does. He avoids mention of the most advanced section of the experiment, Series D, and confines his attention entirely to the section comprising Series A, B, and C. He is attacking a non-vital organ.

We ourselves, of course, did not stop with Series A, B, and C. If there were any reason to pause and analyze Hansel's remarks about the first three series in terms of the actual situation, it could be shown that, at that time and stage, the procedure represented a definite advance over previous work. The strained alternative hypotheses Hansel suggests are unrealistic to those acquainted with the situation. H.E.P. had no knowledge he was not being trailed, and any spying or collusion on his part would have been most obvious in a department (of psychology) in which there were skeptics ready to suspect him of trickery. The devices proposed by Hansel would have been conspicuous and clumsy in crowded corridors. But for general reasons of precaution, it was recognized at the time that the conditions were such that the validity of the results depended entirely on the experimenter, J.G.P. In order to broaden the base of responsibility and protect both the experimenter and the results from possible charges of fraud, such as those now leveled, the conditions were strengthened in the last sub-series of the experiment reported. It is to this improved stage, Section D, that attention must be given for a proper judgment.

We are now ready to look at that series. As a matter of fact, it is not easily overlooked and would be, for most readers, quite obviously the climax series in the paper. First of all, it is in its own right statistically significant, and its scoring average is above that of the paper as a whole. It can well bear the burden of the conclusion by itself. The next question, then, is how Hansel's criticisms apply to it and its conditions.

In this series, J.B.R., who had remained in the background previously, came into the test room with J.G.P. and sat through a series of six runs through the test pack (150 trials) for the purpose of scrutinizing the entire procedure from that point of vantage, to ensure that it was faithfully executed. He, like J.G.P., *could see* *the subject from the window as the latter entered the library* (and, of course, could see him exit as well). He was in the experimental room at the end of each session to receive the independent records from both J.G.P. and H.E.P. immediately on the arrival of the latter at the close of the session. Thus the subject was obviously allowed no opportunity to enter the room alone and copy the order of the cards or the impressions left on the record pad. Even with the somewhat imaginative supposition that H.E.P. had a collaborator, there was no time for the latter, even if he had (unnoticed by J.B.R.) observed the card-turning and recording by J.G.P., to have communicated the knowledge of card order thus gained to H.E.P. as he arrived in the building for the check-up. H.E.P. had to have his duplicate record in his own handwriting, with one copy sealed in an envelope, ready to hand to J.B.R. on entering the room. J.G.P. had to do the recording of the last run of each session after the test was over and H.E.P. was already on his way to the test room. Yet these final runs of the session were, in themselves, independently significant statistically.

It is clear, then, not only that Hansel's counterhypothesis does not apply to Series D, but also that he did not intend it to do so. One can only wonder why he did not deal with this series and, if he is interested in the case for psi, why he did not follow on up the trail of methodological advances into other researches in parapsychology over the intervening decades. He would have found that the question of honesty which (in this and other papers) preoccupies him with such peculiar personal absorption was practically banished as the research entered the investigation of precognition, in which, of course, the target series is not in existence for some time after the subject has committed his responses to paper.

But the motives of the critic are only of secondary importance here; of much more concern to us is the fact that many students of the field are properly seeking an objective evaluation of the case for parapsychological abilities. The body of fact in parapsychology is like a many-celled organism. Its strength is that of a growth-relationship, consisting not only of the compounding of one cell with another, but also of the many lawful interrelations that emerge in the growing structure. Going back as Hansel has done, with a one-cell perspective, to fix attention on some incomplete stage of development within a single experimental research is hard to understand in terms of healthy scientific motivation.

Positive Values of the Critique

So much for the lack of intrinsic worth in Hansel's paper. One can, however, as we have already indicated, find some incidental values coming from an attack such as this. First of all, unusual precautions are generally recognized as necessary in this branch of research. This is due partly to the incredibility and revolutionary character of the results and partly to their exceptional importance. The higher the value, the heavier the guard! With no formal schooling for psi research as such and with little organization in the field as yet, there is naturally a problem of maintaining the highest standards of precaution reached in some of the experiments. There may, of course, be better ways of alerting research workers to the continuing need for the exceptional research standards which our investigations require (that is, better ways than inviting the efforts of such critics as Hansel). But at least something can be said for the "bogeyman value" of his type of activity in parapsychology.

A less obvious service, however, may be credited to Hansel's critique—one that may have an important bearing on parapsychology at this stage. We may well surmise that there are many others than Hansel who feel suspicious of the psi investigations—even though perhaps more judiciously so than he—who suspect that a combination of loose test conditions and moral weaknesses on the part of subjects and even of experimenters could probably account for these rather undigestible results. We submit that it may be a good thing for such people to have an occasional spokesman. Some years ago Dr. George Price (after acknowledging that only this alternative remained) boldly expressed his grave doubts about the honesty of parapsychologists in general.[1] By the time the discussion had died down, it was considered by many parapsychologists that a large amount of much-needed education had been accomplished by the exchanges. Giving the silent skeptics a voice may serve for many of them to bring the issues out of the clouds of vague uncertainty and down to a more solid level of appraisal and judgment that they can appreciate.

[1] Price, G.R. Science and the supernatural. *Science,* Aug. 26, 1955.

But a more subtle value can be claimed for Hansel's sort of criticism. It concerns the question of whether or not parapsychology is really yet a science and, if it is, how much of a unit among the sciences it has become. Criteria of judgment on such a large question are extremely vague and ill-defined. Probably as decisive a factor as any affecting the general status a new field eventually reaches is the outcome of the various challenges presented to its methodology. Most critical students suspend judgment to see how well the new claims stand up under critical attack upon the methods that produced them. We in parapsychology can well afford to welcome these attempts to hack down the structure of our evidence, even if the criticisms are not always in keeping with the best scientific standards (and they seldom are). Radically new researches must be ever ready to stand trial with the expectation that the court is adversely biased and that the quality and quantity of the evidence must therefore be exceptionally strong. But our point here is that contests such as the one Hansel's critique has initiated must be recognized as part of the clumsy and wasteful way by which the test of survival is administered to an emerging radical discovery.

It may be of some advantage to see such a paper as Hansel's against the perspective of history. Some of us in the field today can recall the time when psychical research was a shadowy area in which the serious investigator felt the need of a good detective and when some workers thought that they must be trained in the arts of conjuring. Psychical research was at that stage something like the process of winnowing a few grains of probable fact from great masses of the chaff of human trickery and gullibility. That was a measure of its remoteness from an adequate scientific methodology. Over the years, however, the detective and the magician have left the scene completely, and no one (unless Hansel) would any longer miss them. The methods themselves, after the manner of science, have taken over the problems the detectives were to have solved. It was even in one of the *earliest* volumes of the *Journal of Parapsychology* (Vol. 2, p. 151) that the statement was made that the good faith and morality of the subject in an experiment were no longer a proper concern of the research worker; that if the methods were not adequate to deal with the subject who might be *assumed* to be dishonest, they were not adequate for the standards of a proper scientific field. Today it is important to hold to the high level of standards that have been achieved and to continue to do so until the understanding of the phenomena brings them more readily under predictable control, until the development of test apparatus takes the burden of safeguarding off the experimenter completely, and progress in the findings reduces the incredibility of psi phenomena to the level of science as a whole. Only at such a time, however distant it may be in the future, can the standards of parapsychology be lowered to those of the other psychological sciences (in which the question of the *honesty* of those involved is not raised).

The possibility of fraud is, of course, by no means confined to parapsychology, and it would even be difficult to say that it is greater or more likely to occur in this branch than in other branches. The challenge of psi and the importance of finding any possible alternative explanation have so exaggerated this hypothesis that it is difficult now to see it with adequate detachment.

It so happens, however, that the topic has been treated recently in a much more general context. Discussing the subject of morality in science in an address given at the 1960 A.A.A.S. annual meeting, Sir Charles P. Snow said, "We have all heard of perhaps half a dozen open and notorious ones [cases of fraud] which are on the record for anyone to read—ranging from the 'discovery' of the L radiation to the singular episode of the Piltdown man." Later he adds, "But the total number of all these men is vanishingly small by the side of the total number of scientists. . . . Science is a self-correcting system. That is, no fraud (or honest mistake) is going to stay undetected for long. There is no need for an extrinsic scientific criticism, because criticism is inherent in the process itself. So that all a fraud can do is waste the time of the scientist who has to clear it up."

Here Sir Charles is speaking of the value to the scientific process itself of "extrinsic criticism." With that we can heartily agree. Indeed, so far as parapsychology is concerned—and it has probably been concerned with a great deal more criticism proportionately than has any other branch of inquiry—it owes nothing that we can discover of its scientific advance to the extrinsic critic. It owes a great deal to the kind of critic who proceeds to do (or to help to do) a better experiment than the one in which he observed a flaw.

But in the present commentary on critics such as Hansel, we have in mind still other values beyond those of the improvement of research methods and the advance of discovery itself. Science-in-the-large is more than its research procedures alone. However adequate its methods may be, the progress of a branch of science may be brought to a veritable standstill if a sufficiently adverse climate of opinion prevails. If parapsychology is to continue to advance, it will have to deal competently not only with its problems of research methods but with its difficulties arising out of the currently prevailing idolatry of mechanism as well—difficulties of support, personnel, acceptance, and the interpretation of its larger meaning.

Parapsychology Laboratory
Duke University
Durham, North Carolina

A CRITICAL ANALYSIS OF THE PRATT-WOODRUFF EXPERIMENT

By C. E. M. Hansel

Introduction

If the design of an experiment is such that the result could have arisen by normal means, the experiment, by itself, cannot provide conclusive evidence for ESP. The result of the Pratt-Woodruff experiment[1] could quite easily have been brought about if the two experimenters had indulged in a trick. Thus, in itself, the experiment cannot prove the existence of ESP. The same can be said of any single experiment which purports to prove the existence of any new process. If the findings of this experiment had been confirmed by all other pairs of experimenters who cared to repeat it, the hypothesis of trickery on the part of the experimenters would itself have become unlikely, and at the same time any skeptic could have repeated the experiment and satisfied himself of the result so that eventually criticism would have tended to disappear. In the absence of confirmation of results, as is normally required where findings clash with existing knowledge, it becomes difficult to design an experiment which can conclusively demonstrate the existence of ESP.

It is to some extent the duty of the critic to repeat an experiment to which he raises these objections, if it has not been repeated by anyone else. But before going to the trouble of doing so he may wish to satisfy himself that he is not wasting his time. If he finds, on examining the experiment, evidence to support an alternative hypothesis to ESP, he would not regard the experiment as worth repeating. Thus we may postulate that an experimental result might have arisen in a particular manner which does not involve ESP, but as a second step we may investigate the experiment to see whether there is evidence that the results were, in fact, brought about in the manner postulated. If we find clear evidence that a particular form of error—or trickery—was responsible for the result there is little point in repeating the experiment.

[1] Pratt, J. G. and Woodruff, J. L. Size of stimulus symbols in extrasensory perception. *J. Parapsychol.*, 1939, **3**, 121-59.

In the case of the Pratt-Woodruff experiment, I would not consider the fact that the result could have been brought about by means of a trick involving both experimenters as sufficient in itself for rejecting it. If, on the other hand, trickery on the part of only one of the experimenters was sufficient, this constitutes a more serious criticism, as the experiment was designed to exclude such a possibility.

I have, in fact, found trickery on the part of one person to be sufficient to account for the result and the experiment could be discarded for that reason alone; but the possibility of such a trick having been used is further supported by characteristics found to be present in the score sheets which are difficult to explain in terms of any other hypothesis including that of ESP.

The Procedure Used in the Experiment

The essential features of the experiment which are pertinent to the present discussion are as follows: The subject and experimenter sat at opposite ends of the table. A screen was placed between them. It was 18" in height and 24" in breadth. There was a gap 20" in breadth and 5" in height at the bottom of this screen. Five cards (called "key" cards) bearing different symbols were placed in a row on pegs above the gap so that they were visible to the subject but not visible to the experimenter. Five blank cards were placed on the table below the gap so that they were visible to both the subject and experimenter. Each blank card was beneath one of the key cards. A second low screen on the experimenter's side of the gap was positioned so as to enable him to see the five blank cards, but to preclude the subject seeing the experimenter or the cards he was handling. The pack of ESP cards was held face downwards by the experimenter, and during each run he distributed them into five piles opposite the five blank cards in accordance with the subject's guesses. The subject denoted his guess by touching with a pencil the blank card which was situated beneath the relevant key card.

The testing procedure is described as follows (p. 127):

> One experimenter, Woodruff, and the subject sat facing each other across a table . . . The second experimenter, Pratt, sat about six feet from and almost directly behind the subject. The screen was placed in position. While Woodruff shuffled and cut the pack of cards to be used,

Pratt took the key cards from the pegs and handed them to the subject who changed their order and replaced them without giving Woodruff any indication of the new arrangement. In the last sub-series, Pratt re-arranged the key cards and put them on the pegs himself; during that period the experimenters were careful that the shuffling and cutting by Woodruff were not completed until Pratt had returned to his usual position, so that there could be no possibility of his seeing any of the cards held by Woodruff after they were shuffled. Woodruff then gave the signal to start and the subject proceeded to indicate his "guesses" by pointing to the blank cards in the opening under the screen. Woodruff distributed the cards following the subject's pointer, but he was in complete ignorance throughout the run of the symbol designation intended by the subject.

At the end of the run, the screen was left in position on the table while Woodruff recorded the actual distribution of the 25 cards on the appropriate record sheets and while Pratt recorded the order of the key cards on his record sheet bearing the same number. [Record sheets had been registered and serially numbered before the experiment.] The order of the key cards was recorded by Pratt in reverse order so as to make them correspond with Woodruff's record when the two sheets were juxtaposed later for checking. Pratt in addition recorded the name of the subject, the type of test, the date, and the initials of the experimenters. This recording was done without any communication between the experimenters or from the subject.

When Pratt finished his record, he carried it to the experimental table. Woodruff had usually finished his recording by this time. In case he had not, Pratt was careful to keep his record out of Woodruff's visual field until the other record was completed. Woodruff then clipped together the two independent records with the common serial number and deposited them without further marks or observation of the sheets themselves in a special locked box provided by the secretary for the purpose.

The screen with the key cards still on the pegs was then laid on its side, either by the subject or by Pratt, so that both the key cards and the 25 cards as distributed were visible to all three persons. Pratt then proceeded to sort out the hits from each pile, laying them nearer the key cards and counting aloud the number of hits for the run. This process was observed by Woodruff and the subject. The hits as segregated were then re-examined and re-counted. The score for the run as thus determined at the time from the cards themselves was recorded immediately by each experimenter in his personal record book.

To continue the test, the screen was again raised to its vertical position, the key cards were rearranged upon the pegs, Pratt returned to his seat behind the desk, and Woodruff, having shuffled and cut the pack of cards, gave the subject the signal to begin.

Other precautions were taken during the experiments which are not discussed here as they are irrelevant to the present argument. These may be seen listed in the original report.

Results

Although the total score for the group of 32 subjects was significantly above chance, only five of the individual subjects obtained scores which were themselves significant at the 0.05 level. One subject, however, obtained a result which yielded a probability of chance occurrence of 4.6×10^{-8}. The mean score for the group was 5.21 hits per 25 trials as compared with the chance expectation of 5.0. The probability of this result arising by chance (allowing for optional stopping) is given as 5×10^{-6}.

Analysis of the Experimental Conditions

When examining an experiment of this nature in order to see whether it is foolproof, we first assume that ESP is impossible and then seek some other cause of the high scores. Wherever a result has a very low probability of chance occurrence we may be reasonably certain that something has caused it to arise. The most important results to consider in this experiment are: (1) the over-all score obtained in the experiment ($P = 5 \times 10^{-6}$); (2) the score achieved by one subject, P.M., ($P = 4.6 \times 10^{-8}$).

If any form of trick was used we should expect its effect to be most easily detected in the case of the high-scoring subject, but the fact that four other subjects obtained scores significantly above the chance level ($P < 0.05$) makes it likely that a trick, if used by the subject, was used by more than one subject. An investigation of the experimental conditions, however, makes it difficult to see how a trick could have been used by any subject provided the experimenters were carrying out their duties effectively. In addition, the high scores of the one subject at some sittings could only have arisen through a trick which enabled two or more extra cards to be known in each run. Thus, if, for example, the cut before each run tended to bring certain symbols to the top of the pack and other symbols to the bottom, such an effect, combined with knowledge of it consciously utilized by the subject or combined with calling tendencies on the part of the subject, could not have produced sufficient hits to account for the result.

If the experimenter who handled the pack of target cards carried out his duties efficiently, it is difficult to see how high scores could have arisen through deliberate cheating by any other person involved in the experiments. If on the other hand this experimenter had any knowledge of the positions—or likely positions—of the key cards, he was clearly in a position to affect the result, since he was screened from sight of other persons and had ample opportunity, either while distributing the cards or while recording them, of changing the distributions of symbols in the five piles so as to bring about high scores.

An examination of the orders of the key cards on successive runs showed that they were by no means randomized. Thus when the position occupied by each key symbol in a run is tabulated against its position in the previous run, the contingency table shown in Table 1 is obtained for subject P.M. Thus the experimenter would see, for example, which symbol occupied position 5 at the end of a run and might know that it would most likely be in position 1 during the next run. Such a procedure is, however, difficult to put into operation. Moreover it would be difficult to account for scores

Table 1

CONTINGENCY TABLE SHOWING POSITION OF EACH KEY CARD AGAINST ITS POSITION IN THE PREVIOUS RUN FOR SUBJECT P.M.

Position in Previous Run	Position in Run: 1	2	3	4	5	Total
1	27	42	37	33	16	155
2	24	33	19	43	36	155
3	33	29	32	27	34	155
4	21	26	37	25	46	155
5	50	25	30	27	23	155
Total	155	155	155	155	155	775

Taking the expected frequencies in each cell as 31
$\chi^2 = 54.9$; $n = 16$; $P = < 0.00001$

of the size achieved by P.M. at some sittings. Also, as each subject replaced the cards himself on the pegs, the form of the contingency table would probably change markedly with different subjects.

When testing out this possibility, I found it far simpler to obtain high scores when acting as experimenter by noting the symbol occupying one of the end positions and detecting the position at which it was replaced on the pegs. I instructed an assistant to take the key cards from the pegs in order from left to right and then to replace them in different positions on the pegs. It was then quite easy to note the position at which the cards were replaced. This was done by listening to the sound of the cards being replaced and by observing the shadow of his arm on the table under the slit at the bottom of the screen. The positions of the first and last cards replaced were easiest to identify in this way.

If the experimenter had wished to influence the results in this manner he would have adopted the following procedure:

1. During the scoring at the end of a run he would have noted the symbol occupying position 1 (or position 5) in the row of key cards.

2. When the key cards were removed from the pegs by Dr. Pratt he would have noted whether they were removed in order from left to right or right to left.

3. When the subject "changed their order and replaced them" he would assume that the last card replaced would correspond with the card which occupied positions 1 or 5 in the last series of key cards (depending on whether they had been removed from the pegs from left to right or from right to left. This would arise if the subject took the key cards from Dr. Pratt and replaced them in different positions on the pegs without shuffling them.

4. He would have determined the position at which the first or last card was replaced.

5. While recording the cards in the piles at the end of the run, he would have exchanged cards so that extra hits were obtained in the pile for which he knew the identity of the key card.

Any effects of such a procedure should manifest themselves with the best possibility of being detected in the records of the high-

Table 2

POSITIONS WHICH HIGH-SCORING SYMBOLS OCCUPIED IN THE ROW OF KEY CARDS USED FOR THE PREVIOUS RUN IN THE DATA OF SUBJECT P.M.
(S = STAR, R = RECTANGLE, P = PLUS, W = WAVES, C = CIRCLE)

Date and Type of Sitting	Run	Score on Run	Highest Scoring Symbol	Hits on Symbol	Position of Symbol in this Run	Position of Symbol in Last Run	Critical Ratio for Sitting
Nov. 21, 1938	3	6	S	2	3	1	
	5	6	P	2	4	1	
	6	6	R	3	4	5	0.0
STM	12	10	W	4	2	5	
	14	8	S	3	3	5	
Nov. 28, 1938	2	7	P	3	1	5	
	3	8	W	3	1	5	
	4	6	C	2	1	5	
	5	6	R	3	1	5	
	6	6	C	3	4	2	
STM	8	9	C	4	3	3	4.85
	9	9	S	4	3	5	
	10	9	W	3	1	5	
	11	7	W	3	4	1	
	12	11	R	5	4	5	
	14	11	P	4	2	1	
	15	6	P	3	1	2	
Dec. 12, 1938	6	10	S	4	3	5	
	7	7	R	3	1	5	
STM	8	11	W	4	4	5	3.83
	10	6	C	3	5	5	
	12	10	C	4	1	5	
	15	8	R	4	1	5	
Jan. 9, 1939	2	6	C	3	3	5	
	3	7	P	3	1	1	
	6	10	C	4	5	1	
	8	7	C	3	4	4	
STM	12	6	S	3	5	1	2.13
	18	6	C	3	4	1	
	19	10	W	4	1	1	
	20	7	C	4	1	5	
	22	9	R	4	2	1	
Jan. 31, 1939	3	10	W	4	1	5	
	4	6	C	2	2	3	
	5	7	S	3	1	5	
STM	7	8	P	3	4	1	0.96
	11	7	S	3	2	1	
	13	6	W	3	4	3	
	15	6	R	2	3	1	
Feb. 3, 1939	2	6	C	3	4	1	
	7	6	P	2	5	2	1.18
STM	10	7	W	4	3	1	
	15	7	R	3	3	1	

Date and Type of Sitting	Run	Score on Run	Highest Scoring Symbol	Hits on Symbol	Position of Symbol in this Run	Position of Symbol in Last Run	Critical Ratio for Sitting
Feb. 10, 1939	2	7	S	3	3	4	
	15	7	P	3	3	2	
	16	9	S	3	1	2	
BSTM	17	7	R	3	2	3	0.57
	22	8	R	4	1	3	
	24	7	S	3	4	5	
	27	10	S	3	2	4	
Feb. 17, 1939	8	8	R	4	3	2	
	9	6	C	2	5	2	
BSTM	11	12	P	4	2	1	1.27
	16	7	C	3	1	5	
	25	7	W	3	3	4	

scoring subject P.M. I have therefore analyzed the record sheets of this subject in the following manner:

1. We take all runs made by this subject in which a score of above 5 was obtained.
2. We note the symbol which secured the maximum number of hits in each such run. Where two or more symbols secured an equally high number of hits we reject that run for the purposes of the present analysis.
3. Having identified the symbol which secured the maximum number of hits in a run, we check its position among the key cards (1-5) used for the previous run.
4. On the null hypothesis we should expect the highest score to arise on a symbol which previously occupied positions 1 or 5 on 40% of occasions, and for it to arise on a symbol which previously occupied positions 2, 3 or 4 on 60% of occasions.

The data from which this analysis is made are tabulated in Table 2. This contains all cases where the subject P.M. obtained above-chance scores and where a particular symbol obtained a maximum of hits.

In Table 3 it will be seen that out of 55 cases considered, there are 39 cases where the high-scoring symbol previously occupied positions 1 or 5 in the key card order and 16 cases where it occupied positions 2, 3 or 4. The following contingency table is obtained:

Table 3

POSITIONS OCCUPIED DURING PREVIOUS RUN, BY SYMBOLS SECURING MAXIMUM NUMBER OF HITS IN PRESENT RUN

Previous Position of Symbol	Cases Observed	Cases Expected
1 or 5	39	22
2, 3, or 4	16	33
Total	55	55

$\chi^2 = 21.89$; $n = 1$; $P < 10^{-5}$.

The three sittings at which the result was significantly above chance at the 0.05 probability level may be compared with the remaining five sittings. The result is shown in Table 4.

Table 4

COMPARISON OF SITTINGS AT WHICH RESULT IS SIGNIFICANTLY ABOVE CHANCE LEVEL ($P = 0.05$) AND REMAINING SITTINGS

Previous Position of Symbol	CR > 2	CR < 2	Totals
1 or 5	23	16	39
2, 3, or 4	4	12	16
Total	27	28	55

$\chi^2 = 3.98$; $n = 1$; $P < 0.05$ (with Yates correction)

Thus while the tendency to obtain high scores on symbols previously occupying positions 1 or 5 is present in both cases, it is present to an increased extent at those sittings in which the higher scores were obtained.

There were, in all, five subjects whose scores were significantly above the chance level ($P < 0.05$). Table 5 shows the number of cases in which the high-scoring symbol occupied each of the five positions in the key card order for each of these subjects when the STM procedure was used.

Taking the combined results for subjects other than P.M, it is found that the effect is mainly confined to position 1. If the data are tabulated so that each sitting for each subject is shown against

Table 5

POSITIONS OCCUPIED DURING PREVIOUS RUN BY HIGH-SCORING SYMBOLS (HIGH-SCORING SUBJECTS, STM PROCEDURE)

Previous Position of Symbol	Subject					Total
	C.C.	D.A.	D.L.	H.G.	P.M.	
1	13	2	2	13	16	46
2	3	1	4	5	3	16
3	6	2	2	5	3	18
4	3	0	3	6	1	13
5	5	0	5	3	20	33
Total	30	5	16	32	43	126

Taking the expected frequencies as 25.2 for each position in the total column, $\chi^2=34.8$; $n=4$; $P<10^{-6}$.

the date on which the sitting took place, the following table is obtained:—

Table 6

POSITIONS IN PREVIOUS TRIAL OCCUPIED BY HIGH-SCORING SYMBOLS (SITTINGS FOR EACH MONTH SHOWN SEPARATELY)

Month and Year	Position Occupied by Symbol in Previous Run					Total
	1	2	3	4	5	
Oct., 1938	7	3	3	2	2	17
Nov., 1938	6	5	3	1	10	25
Dec., 1938	1	1	3	0	11	16
Jan., 1939	26	6	9	10	10	61
Feb., 1939	6	1	0	0	0	7
Total	46	16	18	13	33	126

From this it appears that extra hits were obtained mainly on the symbol which previously occupied position 1, except from Nov., 1938, to the last sitting held in December, when hits were obtained on the symbol which previously occupied position 5.

Results for 1938 and 1939 are compared in Table 7.

THE BSTM SERIES

The BSTM runs were commenced in February after the STM series was completed. At these sittings the system for changing

Table 7

POSITION OCCUPIED BY HIGH-SCORING SYMBOL IN PREVIOUS RUN FOR SITTINGS HELD IN 1938 COMPARED WITH THOSE HELD IN 1939 (HIGH-SCORING SUBJECTS, STM PROCEDURE)

Observed Cases

Position in Previous Trial	1938	1939	Total
1	14	32	46
2	9	7	16
3	9	9	18
4	3	10	13
5	23	10	33
Total	58	68	126

$\chi^2 = 10.7$; $n = 2$; $P < 0.01$ (grouping positions 2, 3, 4)

the positions of the key cards was changed. In the report we read that during the STM procedure "Pratt took the key cards from the pegs and handed them to the subject who changed their order and replaced them." This statement is made when describing how the sitting started. Later when describing the procedure before subsequent runs were commenced we read that "the key cards were rearranged upon the pegs." It is not fully clear whether they were removed from the pegs by Dr. Pratt, as before the first run, or merely changed around on the pegs.

In the case of the BSTM procedure we read, in connection with the first run, that "in the last sub-series, Pratt rearranged the key cards and put them on the pegs himself." We are not told exactly what he did nor whether he adopted the same procedure for all the subsequent runs. If on only a few occasions he took the cards off the pegs from left to right (or right to left) and replaced them in different positions, or if he merely moved the cards around on the pegs, extra hits could have been arranged by the experimenter. Thus, if an attempt is made to obtain two extra hits in every run by exchanging two cards between the piles, and the cards are shuffled before being replaced on the pegs on seven occasions out of the eight, a score similar to that obtained in the experiment may be brought about.[2] When the cards are shuffled scores will be, on the average, at the chance level, and extra hits will be obtained when the shuffle is not made.

It will be seen in Table 2 that in the case of subject P.M. there is no tendency for the high-scoring symbol to occupy positions 1 or 5 on the previous run, during the BSTM sittings. When the results for all the high-scoring subjects under BSTM conditions are combined the following contingency table is obtained:

Table 8
PREVIOUS POSITION OF HIGH-SCORING SYMBOL FOR BSTM CONDITIONS
(14 SUBJECTS)

Position	Observed Number of Cases
1	8
2	14
3	5
4	22
5	21
Total	70

The distribution is by no means random: $\chi^2 = 16.4$; $n = 4$; $P < 0.01$

As the BSTM series commenced after the STM series was completed, there is no reason why a further modification should not have been made so that extra hits were obtained on key cards occupying other positions than 1 and 5 among the key cards.

RANDOMIZATION OF THE KEY CARDS

The possibility of the experimenter obtaining knowledge of the positions of one or more of the key cards obviously depends on whether the key cards were shuffled before being replaced on the pegs. Examination of Table 1 shows that randomization was certainly incomplete in the case of the sittings carried out with subject P.M. The values shown in Table 1 might well arise if the order of the key cards was changed by the subject when he replaced the cards. It may be argued that such values might arise through shuffling habits; but further data make it appear very unlikely that the

[2] Chance expectation of hits where n cards are displaced in a run of 25 trials $= 5.0 \times 0.85n$.

Table 9

ARRANGEMENTS OF KEY CARDS IN SUCCESSIVE RUNS WITH SUBJECT D.A.

Jan. 11, 1939						Feb. 1, 1939						Feb. 22, 1939					
Run	Position of Symbol					Run	Position of Symbol					Run	Position of Symbol				
	1	2	3	4	5		1	2	3	4	5		1	2	3	4	5
1	P	S	C	R	W	1	S	P	W	R	C	1	R	C	W	S	P
2	W	P	C	S	R	2	C	R	W	P	S	2	R	P	C	W	S
3	P	S	C	R	W	3	C	R	W	P	S	3	W	S	R	C	P
4	W	C	R	P	S	4	S	W	C	R	P	4	S	P	C	R	W
5	W	C	R	P	S	5	S	C	R	P	W	5	C	W	P	R	S
6	S	P	R	C	W	6	C	R	S	P	W	6	P	R	C	S	W
7	C	R	W	P	S	7	C	R	P	W	S	7	S	W	R	P	C
8	S	P	W	R	C	8	C	R	P	W	S	8	S	C	P	R	W
9	S	P	W	R	C	9	W	P	R	C	S	9	P	S	C	W	R
10	S	P	W	R	C	10	C	R	P	S	W	10	W	P	S	R	C
11	S	P	R	W	C	11	C	R	P	S	W	11	W	P	S	C	R
12	S	P	W	R	C	12	C	R	P	S	W	12	C	R	W	S	P
13	S	P	W	R	C	13	C	R	P	S	W	13	C	W	P	S	R
14	S	P	W	R	C	14	C	R	P	W	S	14	S	R	C	W	P
						15	P	R	S	W	C	15	C	P	S	R	W
						16	S	P	R	C	W	16	C	W	R	P	S
						17	S	P	R	C	W	17	C	W	S	R	P
						18	R	P	S	C	W	18	C	S	P	W	R

cards were shuffled on all occasions. The orders of the key cards used in all the runs made by the second highest scoring subject (D.A). are shown in Table 9.

There are, in all, 12 occasions when the order of the key cards remained unchanged from the previous run in a total of 50 runs made by this subject. The probability of this number of cases arising by chance is obviously minute. In fact, the key cards remained in the same positions for the last three runs of a sitting held on January 11 and then were in the same positions again for the first run of the next sitting held more than a fortnight later. For the following two runs, they were in the same order but reversed from left to right. Thus, at these sittings, it would appear likely that Dr. Pratt removed the cards from the pegs and handed them to the subject who replaced them, on some occasions, straight back in order from right to left (or left to right if they were removed from right to left). Had the cards been shuffled, or even cut, before being handed to the subject, it is difficult to see how the runs of similar orders among the key cards could have arisen.

Dr. Pratt has told me that he shuffled the cards before handing them to the subject. He may have done so on most occasions, but it appears unlikely that he did so during the sittings made with D.A. There is no mention in the report of the cards being shuffled, and failure to do so on even a small proportion of runs would be sufficient to explain the results of the experiment.

Conclusions

1. The design of the Pratt-Woodruff experiment permits the main experimenter to create above-chance scores by obtaining knowledge of the positions of one or more of the key cards and changing the distribution of cards in the five piles.

2. Characteristics are present in the record sheets of the high-scoring subject P.M. which would be expected to arise if the main experimenter was influencing the experiment in this way.

3. Similar characteristics, but to a less marked extent, are present in the record sheets of the other four subjects whose results were significantly above the chance level.

4. Under STM conditions high scores tended to arise on symbols which occupied positions 1 and 5 in the key-card order for the previous run. They arose mainly on symbols which occupied position 5 during sittings made in November and December of 1938 and position 1 in the remaining sittings.

5. Under BSTM conditions high scores tended to arise on symbols which previously occupied positions 4 and 5 in the previous run.

6. In the case of most of the subjects (other than P.M.) their results could have been achieved by using such a trick during a small proportion of the runs made.

7. There is evidence in the records of the subject D.A. that the key card randomization was far from satisfactory, and the characteristics present in these records would arise if the order of the key cards was being changed by placing them back in different positions on the board, or if the key cards were merely moved to different positions on the pegs.

Discussion

It is possible to obtain high scores under the conditions of the Pratt-Woodruff experiment by means of a trick. It is clear that such a trick, if used with care, could be extremely difficult to detect. Thus, if extra hits were arranged on symbols occupying the five positions in the key card series approximately equal numbers of times, its detection using the analysis described above would be impossible. In that case, failing detection by a more elaborate analysis, it could only be said that, as the experimental conditions were not fool-proof, the experiment could not provide evidence for ESP.

The analyses made in this paper are by no means exhaustive, nor as complete as is desirable. They were limited by the time available with the records.

Taking the results obtained with the high-scoring subject P.M. it is difficult to see how an alternative explanation to that of trickery on the part of the main experimenter can be provided. It is possible, however, that there is an alternative explanation which has eluded the writer.

I would like to thank Dr. J. B. Rhine for arranging my visit to the Parapsychology Laboratory and Dr. J. G. Pratt who let me see the record sheets and who has given me every assistance in carrying out the investigation.

University of Manchester
Manchester, England
Department of Psychology

REFUTATION OF HANSEL'S ALLEGATION CONCERNING THE PRATT-WOODRUFF SERIES

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

Introduction

The preceding paper by Mr. Hansel may be unique in scientific literature. It claims to offer evidence on a statistical basis that an investigator cheated in his own experiment. However, his "case" is characterized by logical weakness and by inconsistencies. We list below the main points that we shall develop more fully later on.

I. Hansel quotes our report out of context to create the impression of a weakness in the experimental procedure which did not exist.

II. He makes an issue of the lack of randomness of the key cards in spite of the fact that complete randomness of these permutations was neither claimed by us nor required for the adequacy of our experimental procedure and even though he admits that his experimenter-fraud hypothesis could not be based upon this factor.

III. Hansel's method of analysis is fallacious. It fails to take account of different factors which may explain the statistical significance in his findings, and therefore his interpretation of experimenter fraud is without support.

IV. Hansel claims that his "effect" exists generally in the statistically significant sections of our series, and he argues that Woodruff, as a constant factor in these runs involving a number of subjects, was fraudulently responsible for the extra-chance scores. We shall describe a more appropriate method of analysis for finding out whether the rate of scoring among the five symbols is related to the preceding positions of the key cards. Our analysis shows that a run-to-run effect is *not* found in all of the significant sections of the experiment, but is present only in the work of one high-scoring subject. Thus Hansel's attempt to use the data of the other high-scoring subjects from our series to "confirm" his hypothesis of a trick by Woodruff is a failure.

V. We shall point out a number of respects in which our critic shifts his argument or selects his facts, apparently seeking to give his "evidence" an illusory appearance of strength or plausibility. Some of these unwarranted departures from objectivity should have been apparent to Hansel from a reading of his own paper as it stands; others could hardly escape the notice of anyone with access to the data of our experiment, as Hansel himself had.

One would suppose that any fellow scientist invited to the Parapsychology Laboratory for the purpose of studying the evidence for ESP would have attempted preliminary discussion in private of any questions raised in his mind about our data before submitting a paper for publication. Hansel, however, has chosen to present his case without such discussion. Moreover, he has, without waiting for publication, meanwhile been giving his manuscript wide private circulation. His actions appear to represent a deliberate attempt to discredit parapsychology by any means. Under the circumstances we have no choice but to point out his mistakes in print.

I. Our Procedure for Re-arranging the Key Cards

Near the beginning of his paper, Hansel quotes several sentences from our report. He presents these as if they were complete, original paragraphs cited without omissions. These sentences deal mainly with the method followed in changing the key cards between runs.

Actually, the beginning of the quotation given by Hansel does not form the opening of a new paragraph in our report. He has incorrectly given the impression that he has quoted the full paragraph, whereas he has in fact omitted the first, topic sentence. This sentence in our report reads: "The actual testing procedure for each run may be described as follows." To take the place of our topic sentence, Hansel has substituted his own which is similar to ours, except for the fact that the three key words "for each run" are left out.

Within the quotation given by Hansel, the significant sentence about the method of changing the key cards is: "While Woodruff shuffled and cut the pack of cards to be used, Pratt took the key cards from the pegs and handed them to the subject who changed their order and replaced them without giving Woodruff any indication of the new arrangement." Hansel recognized that it would not have been possible for Woodruff to use his hypothetical trick when this procedure was followed. To get around this difficulty in the way of his experimenter-fraud hypothesis, he conveniently but erroneously has implied that the description given in our report applied only to the *first run of the session.*

Our report clearly shows that the general description of the procedure applied to each run of the series, and not merely to the first run of each session. Hansel could scarcely have overlooked this fact, since it was the essential part of the topic sentence of the first paragraph he quoted from our report. The omission of that sentence enabled him to misrepresent the conditions so as to lend a superficial plausibility to his hypothesis that Woodruff resorted to a trick. The omission and the misrepresentation could hardly be a mere coincidence.

The extent to which his criticism depends upon this misinterpretation of the procedure is amply apparent from the following statements in his paper:

> ...In the [Pratt-Woodruff] report we read that during the STM procedure "Pratt took the key cards from the pegs and handed them to the subject who changed their order and replaced them." This statement is made when describing how the sitting started. Later when describing the procedure before subsequent runs were commenced we read that "the key cards were rearranged upon the pegs." It is not fully clear whether they were removed from the pegs by Dr. Pratt as before the first run or merely changed around on the pegs.
>
> In the case of the BSTM procedure we read in connection with the first run that "in the last sub-series Pratt rearranged the key cards and put them on the pegs himself." We are not told exactly what he did nor whether he adopted the same procedure for all the subsequent runs.

There was no change in the procedure after the first run of the session. Hansel misread our twenty-year-old report on a simple point of procedure—a point that has never been questioned before.

II. Was the Re-arrangement of the Key Cards Adequate?

Nowhere in our report did we speak of the re-arrangement of the key cards as providing a random sequence of permutations. We deliberately spoke not of *randomness* but of *re-arrangement* (e.g., the subject "changed their order and replaced them without giving Woodruff any indication. . . . In the last sub-series, Pratt re-arranged the key cards and put them on the pegs himself. . . ."). The purpose of having the key cards on the pegs outside of Woodruff's sight was to keep him from knowing their positions during the run. This is all that was claimed. To achieve this purpose it was not necessary that the relationship between the order of key cards for one run and that for the next should be statistically random.

Having raised the randomness question, Hansel concedes that it is not relevant to his fraud hypothesis. Nevertheless, he continues to confuse the issue with it. For example, Table 1 in his paper is an analysis of the run-to-run sequence of key cards in the data of the highest-scoring subject, P.M. In discussing this table, Hansel admits that the results of the experiment could not be accounted for on the assumptions that Woodruff had learned the subject's habits of replacing the key cards and had fraudulently misplaced cards to increase the *probability* of "hits." Having rejected this particular hypothesis, Hansel introduces another type of analysis which does not depend upon the nonrandomness of the key cards. However, he leaves the reader with the false impression that the randomness question is pertinent to his basic thesis.

In fact, a later section of his paper is devoted entirely to this irrelevant topic. There he interprets observed nonrandomness in the key cards for subject D.A. as evidence that they were not "shuffled" before being replaced on the pegs. There is another quite obvious interpretation—one that is entirely consistent with the procedure described in the report. The instruction given the subjects was that they should change the order of the key cards and then place them on the pegs in a different arrangement from the one used in the preceding run. The subjects in the STM series, *after having changed the order of the cards in their hands,* placed them on the pegs *while looking at the symbols on the faces of the cards.* The effort to achieve an order unlike that of the preceding run or an order that Woodruff did not know is quite sufficient to account for a nonrandom relationship in the order from one run to the next.

In the case of subject D.A., there occurred, as shown in Hansel's Table 9, an unusually high number of instances in which the order of the key cards was repeated from one run to the next. Hansel interprets this as evidence that the key cards were taken from the pegs without shuffling and were handed to the subject, who placed them straight back on without changing the order. This seems an untenable hypothesis. Subject D.A. was (and is) a member of the faculty of the Duke University Psychology Department. The most likely explanation is that he decided to find out for himself the effect of putting the key cards in a certain order, and thus the same order appeared from run to run. If this is the case, he said nothing about it to the experimenters. The experimental method involved locking up the written record of the cards as soon as it was made before the score was checked. This procedure made it unlikely that the investigators would notice that the subject was "trying his own experiment" by repeating certain permutations. But regardless of whether our explanation is the correct one, the fact of the matter is that we did not notice the repetition at that time—and this fact we now find neither surprising nor disturbing.

In Hansel's paper, however, the matter is presented as something sinister. The implication is that, when the key cards were not changed, Woodruff might have been aware of their positions and thus he could have placed cards erroneously to increase the number of hits. Indeed, if Woodruff had been trying to make hits in this manner, he had his best opportunity to do so when D.A. put the cards back in the same order (if he had noticed it was being done).

The repeated key cards in the D.A. data thus offer a method of testing the Hansel hypothesis of experimenter trickery. This test is to compare the rates of scoring in those runs (a) when the order was changed and (b) when it was not changed. This is an obvious thing to do. Why didn't Hansel do it? If he had, he would have found that (a) the 12 runs with unchanged key cards scored low, with only 7 more hits than mean chance expectation; (b) the 21 runs with the changed key cards gave a positive deviation of 21 hits, a higher average. Thus he would have known that the repeated orders were not related to the scores.

Hansel even selects facts out of context to favor his interpretation. He shows that the key cards were in the same position for the last runs of D.A. on January 11 and again for the first run of February 1. He then offers this fact as evidence that the key cards remained unused on the screen throughout this period. Yet the records show that other sessions comprising a total of 540 runs intervened between these dates and the key cards were re-arranged this number of times. Thus the use of a preferred order by D.A. in his first run of February 1 could not possibly mean that the cards remained unchanged since his last session. Therefore, instead of indicating a flaw in the procedure, this observation suggests that our own interpretation is probably the correct one: D.A. preferred certain arrangements of the key cards; and without telling the experimenters, he sometimes placed the cards back in a chosen order after mixing them in his hands. This did not conflict with the experimental conditions described in our report.

III. The Fallacy of Hansel's Method of Analysis

The question which Hansel has raised may be restated as follows: Is the distribution of the score among the five key cards in a given run dependent upon the position of these symbols in the preceding run? In attempting to answer this question, he relies upon a simple count of the hits in each of the five positions in those runs which gave a total score of more than five.

Hansel's method has flaws which render the results ambiguous and therefore uninterpretable. (1) Counting the *number* of hits in each position fails to take account of any tendency on the part of the subject to have more cards placed in one or more of the piles than in the others. (2) His method overlooks the habits of the subjects in replacing the key cards. (3) His method fails to recognize that subjects might have a tendency to score at a higher rate *by ESP* in one or more of the five positions than in the others, or on certain symbols more than on others. (4) These factors might work in combination to produce an "effect" by his method.

Usually, subjects show a tendency to respond more often to the symbols occupying the three inner key-card positions. If there exists at the same time a tendency for subjects to place the key cards that were in the end positions back on the three inner pegs, these two concomitant habits would account for finding *more hits* on the symbols that had been on the ends. Thus this hypothesis might explain Hansel's results without the necessity of bringing in either his trickery interpretation or any variation of the ESP hypothesis.

The point we are making here is that one simply cannot tell what his results mean. A further study of the data is needed before one can say whether his figures mean anything at all in regard to a run-to-run effect on the scoring. Only after such an effect has been found to exist on the basis of an adequate method of analysis does one properly consider its interpretation. Hansel appears to have been partially aware of the weakness of his case when he wrote: "The analyses made in this paper are by no means exhaustive, nor as complete as is desirable." Whether or not he knew how truthfully he spoke in those lines, they form an astonishing confession with which to end a critique containing so serious a charge.

What the matter boils down to is this: The statistical significance of the results of our paper is not at issue. Hansel's argument against our findings as evidence of ESP is based upon a misinterpretation of our conditions. In the effort to support the trickery hypothesis which he chooses to offer, he has carried out analyses of the data which yield uninterpretable results. If his criticism has *any* valid bearing on our experiment, we have been unable to find what it is.

We made analyses of the data of the four highest-scoring subjects to find out to what extent the main complicating factors mentioned above actually exist in the records. Table 1 shows the frequencies of *responses* over the five positions and the chi-square evaluation of the distribution for each subject. Only one subject, D.A., shows an approximately even distribution of trials. The other three all show a tendency to point more often to the second, third, and fourth positions and to neglect the two end ones. This tendency only approaches a level of statistical significance in P.M., but it is highly significant in H.G. and C.C.

A second factor to be considered is the distribution of the success rate (percentage of hits) over the five positions. Table 2 shows the results of the analysis for this factor. It is apparent that there is a considerable range of variation in each of the four subjects.

Table 1

Distribution of Responses Over the Five Positions for the Four Highest-Scoring Subjects

Subject	Trials in Each Position					χ^2_4	P
	1	2	3	4	5		
P.M.	756	843	862	803	786	9.05	.06
D.A.	251	249	251	238	261	1.07	.9
H.G.	842	1006	1044	924	859	33.66	< .000005
C.C.	672	1074	1148	1128	828	182.22	< .0000001

Note: The subscript numerals given with χ^2 indicate degrees of freedom. Thus each χ^2 value in this table has 4 d.f.

While the differences in scoring rate are not statistically significant, they do exist; and they would therefore have to be taken into account in any method for studying a run-to-run relation of key card positions and scoring.

We have not analyzed the two subjects, H.G. and C.C., for tendencies shown in the replacement of the key cards. As we have said, we recognized that the procedure might not generate a random order of permutations, and the evidence presented by Hansel in support of this anticipation from the data of P.M. and D.A. is sufficient to show that this is a real factor that any method would have to take into account.

The amount of effort that Hansel devotes to applying his method of analysis to the other high-scoring subjects after his discovery of the "effect" in P.M. indicates that he felt (though he did not explicitly state) the need for confirmation of the findings. We agree that this need for confirmation is one that must be met before any

Table 2

Percentage of Hits to Trials in Each Position (Chance Expectation: 20%)

Subject	Position				
	1	2	3	4	5
P.M.	26.46	22.42	23.55	23.04	21.50
D.A.	25.10	25.70	22.31	18.91	24.90
H.G.	20.78	21.87	20.59	21.43	23.63
C.C.	20.98	23.56	20.03	21.81	21.14

kind of conclusion can properly be reached regarding an *a posteriori* result such as the one he found in P.M.'s data. But his claims that the data of the other high-scoring STM subjects (his Table 5) and of the BSTM series (his Table 8) confirm the "effect" are without any foundation. In the first place, he has not treated the additional data separately in these analyses, but has always loaded the results in favor of his prediction by re-using the P.M. data. In the second place, the two subjects, H.G. and C.C., who superficially gave a suggestion of a run-to-run effect (his Table 5), showed, as we have indicated, a strong favoring of the three inner positions, and therefore they would be expected to get a larger number of hits on them. If Hansel's Table 5 shows anything at all for these two subjects, it may only be that they tended to place the key card that had been in the first position back on one of these three inner pegs. (But we shall show in the next section that these data, when properly analyzed, show no Hansel effect of any statistical significance.)

IV. A More Appropriate Statistical Analysis

An analysis that more adequately meets the requirements for a test of a dependence of scoring upon the key card positions of the preceding run is the following: First, tabulate the hits and misses in five groups defined by the position of the key card in the preceding run. Then test by chi-square the 2×5 matrix of hits and misses to see whether there is any significant variation in scoring rate. We have done this kind of analysis for different groupings of the data with which Hansel was concerned, and the results are as follows:

(1) *P.M. Data for the STM Series, Runs with Scores of Six and Above.* With this subject, there is an effect of such a high significance level ($\chi_4^2 = 31.67$, $P = .000005$) that one is justified in forming hypotheses about it, but not in attempting to reach a conclusion without confirmation or without considering alternative possibilities of interpretation. The contributions to chi-square of the "hit" cells for those two groups having the key cards that were formerly in positions one and five are based on positive deviations, and the other three hit cells all have negative deviations. This analysis, based on the data shown in our Table 3, shows that the run-

Table 3

P. M.'s Results in STM Runs With Scores of Six or More in Relation to Previous Key Card Positions

Key Card Position in Previous Run	Hits	Misses	Total
1	108	194	302
2	66	199	265
3	61	207	268
4	77	205	282
5	125	183	308
Total	437	988	1425

$\chi^2 = 31.67$ (4 d.f.)
P = .000005

to-run key card effect in subject P.M. is indeed a striking one that deserves further attention. We shall return to it later.

(2) *The High-scoring STM Runs for C.C. and H.G.* These two subjects, according to Hansel, came closest to providing a "confirmation" of his effect in the P.M. data. By our analysis neither subject alone showed any tendency for the scoring rate to depend upon the previous key card positions: For H. G., $\chi_4^2 = 2.81$ with $P = .6$; for C.C., $\chi_4^2 = 4.30$ with $P = .3$. Combining the analysis

Table 4

Results of H. G. and C. C. in STM Runs With Scores of Six or More in Relation to Previous Key Card Positions

Key Card Position in Previous Run	Hits	Misses	Total
1	228	482	710
2	190	514	704
3	192	486	678
4	198	478	676
5	199	508	707
Total	1007	2468	3475

$\chi^2 = 5.15$ (4 d.f.)
P = .28

for these two subjects (see our Table 4) yields results that are still below the level of significance ($\chi_4^2 = 5.15$ with $P = .28$).

(3) *The BSTM Data.* Table 8 in Hansel's paper is especially revealing on the question of the unreliability of his method of analysis. He finds there for the BSTM data a value of χ_4^2 with $P < .01$. Our analysis (our Table 5) gives a chi-square that is *not* significant ($\chi_4^2 = 6.04$, $P = .2$). But even if the BSTM data had supported Hansel's "effect," this finding would have presented strong evidence against his "interpretation." The reason is that the method of rearranging the key cards in this sub-series, when Pratt changed the order and replaced the cards face-inward on the pegs, completely eliminates the question of Woodruff's having been able to keep track of one or more symbols. Yet the rate of hitting in these BSTM records was the same as in the series as a whole, and this fact points to something in the subjects' responses as the explanation. The conditions of the experiment left only the ESP interpretation.

Hansel began his investigation with the data before him and with the conviction that there must be some explanation besides ESP for the results. Searching through the work of the highest scoring subject, he came upon something which he could interpret as evidence of fraud by Woodruff. Because of this groping, "after-the-fact" approach, this initial finding could not be conclusive even if the method of analysis had been adequate. Confirmation of the

Table 5

Results of BSTM Runs With Scores of Six or More in Relation to Previous Key Card Positions

Key Card Position in Previous Run	Hits	Misses	Total
1	216	561	777
2	219	561	780
3	215	556	771
4	253	525	778
5	234	560	794
Total	1137	2763	3900

$\chi^2 = 6.04$ (4 d.f.)
$P = .2$

effect in the data of other high-scoring subjects was therefore of paramount importance. Hansel's efforts to achieve this objective show that he recognized this need. These efforts failed—as we have shown—in spite of his claims to the contrary. This failure was all the more dismal since he was not consistent in applying the test he used on P.M.'s data, but shifted the basis of analysis in the "confirmatory" applications. This fact, coupled with his repeated use of the P.M. data in the later tests of significance, makes much of his paper downright misleading.

Hansel deals in his analysis with only a small portion of data selected from a large body. Out of P.M.'s series of 4,050 trials, he selects a relationship which depends upon only 55 observations. This creates the likelihood that the criteria for data selection were established after the large body of data had been examined and its internal characteristics noted. As can be easily demonstrated, when such procedures of selection are used, even "random" data can be made to yield highly significant results.[1]

V. Some General Observations and Concluding Remarks

In his "Introduction" Hansel seeks to convey the impression that attempts have been made to repeat the Pratt-Woodruff exper-

[1] To illustrate this point, one of us examined the first block of 1,000 numbers in the Kendall and Smith, *Table of Random Sampling Numbers* in search of samples of 55 consecutive items which would show significant departures from chance and would therefore provide "evidence" that someone used a "trick" in making up the table. Here is some of the "evidence" which was easily found in support of this "hypothesis." (1) In columns 2 and 3 and the first four numbers of column 4, there are 41 odd digits and 14 even digits ($\chi_1{}^2 = 13.25$, $P = .0003$). (2) Again the sample of 55 digits made up of columns 29 and 30 and the bottom five digits of column 31 contain 15 odd numbers and 40 even numbers ($\chi_1{}^2 = 11.36$, $P = .0008$). (3) The 40 items in the twenty-fifth row plus the first 15 items of the twenty-fourth row contain 27 of the digits 2, 3, and 4 combined, and only 28 of the remaining seven digits ($\chi_1{}^2 = 9.54$, $P = .002$). (4) The first two blocks of digits spanning columns 21–24 and the first 15 items in block 3 in these same columns contain 14 fives where only 5.5 are expected and 41 of the other nine digits where 49.5 are expected ($\chi_1{}^2 = 14.60$, $P = .00014$). (5) Extending the above sample through the fifth block of columns 21–24 shows that the excess number of fives increased still further (for this 100-item sample, $\chi_1{}^2 = 16.00$, $P = .00006$). Findings of this kind can be obtained at will if one takes almost any body of data and makes selections from it upon the basis of certain criteria which one establishes by inspection and then applies rigorously in selecting a smaller sample. Such a procedure proves nothing except that, when selected in this fashion, even "random" data may have extreme idiosyncrasies. It would have been surprising indeed if a person with the degree of ingenuity demonstrated by Hansel had not been able to obtain some "suspicious" results.

iment without success. In actuality, insofar as efforts to replicate our procedure are concerned, the score stands: Attempts = 0, Successes = 0, Failures = 0! Confirmation is, of course, a different matter. The numerous positive results obtained in ESP research over the 21-year span since the publication of our series constitute, in a very precise and real sense, confirmation of our ESP evidence.

Hansel employs certain "verbalisms" as if they were established facts. Thus, in accounting for the fact that the BSTM data do not reflect the same effect as that found in P.M., he states that "... there is no reason why a further modification should not have been made ..." He fails to specify: (a) why a modification should have been made if the earlier and "easiest" form of "trickery" which he has already offered was so effective; (b) what the "modification" is and how it was accomplished; and (c) what the evidence is to support the existence of such a "modification." Are we to infer that mere supposition is adequate evidence for him when the reputation of a fellow-psychologist is at stake?

Now let us deal with the question: How are the P.M. results to be explained? The existence of an effect in her STM data, even when the matter is examined by a proper method, still leaves something of a question for which an answer is in order and which may be considered even though a final answer is not now within reach. Hansel has too eagerly offered only the hypothesis of trickery. Other hypotheses which have been or can be considered are:

(1) Actually, one can offer a consistent and reasonable ESP hypothesis, as follows: For the subject P.M., the run began, in the psychological sense, when she re-arranged and placed the target cards. The ESP task being a difficult one, she dealt with it by a "narrowing of attention" procedure. For her the task became one of attempting to identify only *some* of the cards in the deck: those with the particular symbols which had become salient because of their prominent, end positions in the preceding run.

(2) There may be an alternative ESP interpretation, such as a differential rate of scoring on the five symbols coupled with some habitual tendency in the placement of the symbols on the pegs.

(3) Finally, as stated above, this may be a selected, meaningless, statistical effect, for statistical oddities are a dime a dozen. To take

one seriously it is necessary to confirm it. The data of other subjects in our series fail to support this oddity, whereas they do support the significant scoring level of the experiment. Therefore the Hansel effect is still unconfirmed and unexplained, and it certainly could not explain the Pratt-Woodruff results.

It must, of course, be recognized that conjectures of possible ESP functioning to account for this new effect suffer the same shortcoming as does Hansel's "trickery" hypothesis: they are offered in relation to highly selected data, and the same data which led to the formulation of a hypothesis cannot be used to verify it. The advantage of the ESP hypothesis lies in the fact that it allows for differential results for different subjects while Hansel's, involving alleged trickery by an experimenter who was always present, requires consistency for its verification. Thus, for verification of his hypothesis, as has been pointed out, (a) all high-scoring *subjects* should show the Hansel effect, which they do not; and (b) all high-scoring *series* should show the Hansel effect, which they do not. (At the point where Hansel says, regarding the supposed method of trickery, that "there is no reason why a further modification should not have been made," he was attempting to justify his hypothesis as the explanation for a different kind of effect in the BSTM data. His effort was uncalled for, however, because there was no statistically significant "effect" in the BSTM data to be explained).

We who have worked in parapsychology have learned not only to operate with special safeguards, but to expect a degree of suspicion. However, we are entitled to *responsible* questioning and to *mature* consideration of the evidence. What experience has Hansel had with his professional fraternity in psychology that makes it almost (with him) a foregone conclusion that Woodruff was a conscious cheat? If Hansel is a sincere student of our branch of science, he must know that since our paper was published in 1939 Woodruff has been either author or co-author of eight published psi research reports. Of the eight, only one reported total results which deviated significantly from chance. Three reported no significant results of any kind. Simple failure to publish these three non-supportive (for the ESP hypothesis) papers would have been a less blatant form of bias than that which Hansel has alleged. Other specific details of these pieces of research and about the published articles could be cited as evidence of objectivity. Our critic may, of course, attempt the task of furnishing a similar dossier of his objectivity.

In the last analysis, it becomes apparent that the existence of psi phenomena does not depend upon what the parapsychologist believes or what the critic believes. It is also apparent that this kind of controversy settles nothing insofar as the basic question is concerned. We find ourselves in substantial agreement with Hansel's contention that it is most difficult to arrange a test of ESP which will rule out the possibility of some form of fraud[2], so charges of fraud have been and will continue to be leveled against investigators who report positive results. Parapsychologists must accept this fact and only those who are temperamentally suited to handle this kind of charge should attempt parapsychological research and publication.

But does the critic not also have an obligation? Criticism comes easily; research is more difficult. Let the critic repeat the research he criticizes. Let him introduce the modifications he feels are needed. The danger for him lies in the possibility that he will find supportive evidence for ESP. Then he too will face a test of his objectivity in whether he publishes or not. If he passes this test, he may expect to be faced with the same allegations of fraud that he has imposed on others.

As a rule, one scientist does not question another's good faith. Research workers in parapsychology have taken more notice of the possibility of fraud in the investigator than is the case in any other branch of science. The reason is not that fraud has been found more often in this branch, but rather that the revolutionary impact of the findings upon current scientific theories has for some diehard skeptics made even this extreme alternative attractive as a possible escape from accepting the conclusions of the parapsychologists. Thus the raising of the question of fraud by the experimenter is only the ultimate stage of criticism: the final, reluctant recognition of the challenge of the field and of the strength of the evidence. Parapsychologists can therefore accept the need to be on guard both

[2] This is not a novel position for us, as we discussed the question in this vein in our original report (p. 140).

against fraud itself and against unwarranted charges of fraud as normal professional hazards on their frontier of science. The final answer to allegations such as the one Hansel has made is in the *number* of independent confirmations. In the two decades since our experiment was reported, the structure of the evidence for ESP and for psi in general has improved in both quality and quantity. And in the large, well-buttressed edifice that the total evidence now forms, we think the section of the foundation identified as the Pratt-Woodruff series is still firm.

Parapsychology Laboratory
Duke University
Durham, North Carolina

Department of Psychology
City College of New York
New York, N. Y.