The Evidence for ESP: A Critique

A basic requirement of science is that a demonstration of some new process should rarely fail when other investigators try to verify it. This requirement has not been satisfied in parapsychology. Hence some conclusions can be drawn.

C. E. M. Hansel

WHEN CONSIDERING the subject of the evidence for parapsychology, I assumed that this *evidence* comprised what has been published in the form of experimental results, consisting essentially of scores obtained during guessing experiments. It is then necessary to consider whether this evidence can provide proof for the existence of the various hypothetical processes postulated by parapsychologists.

Both the experimental results and the supposed phenomena have changed in nature over the years. Back in 1882, a committee set up by the Society for Psychical Research (SPR) was interested in thought transference from one person to another. Since then the range of abilities has increased until amost any conceivable process that appears to be inconsistent with contemporary scientific theory is likely to be included in the repertoire. At the same time, the range of organisms claimed to possess these abilities has increased to include not only humans but horses, cats, mice, cockroaches, chicken eggs, and plants.

A fundamental feature when considering any ability is its distribution in the population. When John Dalton reported red-green color-blindness in 1794, he gave figures for its incidence in his students that are remarkably

C. E. M. Hansel is emeritus professor of psychology at University College of Swansea, Wales. He is the author of ESP and Parapsychology: A Critical Re-evaluation.

consistent with values accepted today. But the distribution claimed for ESP after a hundred years of investigation varies with each investigator. In the thirties, J. B. Rhine, after more than five years of research, came to the conclusion that some ability was present in at least a fifth of the population of Durham, North Carolina. In addition he found it to be present in almost every psychology student that he tested. Investigators in five psychology departments, who used Rhine's tests with very large samples following the publication of his *Extra-Sensory Perception* in 1934, found no ability to score at above the chance expectation level.

In England S. G. Soal, repeating Rhine's work between 1934 and 1939, found scoring at only chance levels after testing 160 subjects. These included students, professional clairvoyants, and mediums. He also tested Eileen Garrett, a medium who had produced remarkably high scores when tested by Rhine three years earlier, and got only chance results. Thus experiments at that time appeared to indicate that subjects could not score above the chance level unless they were tested in Rhine's laboratory, and they confirmed contemporary ideas in psychology that sensory information is necessary for perception.

Following a general tightening up of experimental procedures at Duke, high-scoring subjects became difficult to find. By 1938, the Pratt-Woodruff experiment, proclaimed by Rhine to be the most carefully controlled experiment in the history of science, produced only one high-scoring subject among the 32 tested. The group as a whole scored only one extra hit in each 200 trials.

In the 1940s, following the disappearance of high-scoring subjects, the emphasis was on group experiments. But even these presented difficulty until a new type of experiment providing a new source of experimental error appeared on the scene. Here a large sample of subjects was divided into two categories, called "sheep" and "goats," and a comparison was made of the mean scores from the two subgroups. Often a difference was found, but the overall score after combining scores from the subgroups—rather in the manner of Burt and Thouless—was at only chance level.

In more recent times a massive survey of the population in Britain carried out by the SPR failed to find a single individual who could consistently score above the chance level. But in the United States, after throwing experimental precautions to the wind and returning to a single experimenter plan, Helmut Schmidt appears to have no more difficulty finding subjects than did Rhine in his heyday. Other investigators, after dropping the use of Rhine's five symbols and returning to techniques reminiscent of the early days of the SPR, appear to have no difficulty finding suitable subjects.

Some years back Rhine suggested that the whole mass of experiments considered in toto indicate the existence of ESP. While disagreeing completely, I would suggest that facts can be established by looking at the

mass of evidence, but these facts are more likely to indicate the absence of any paranormal process than to demonstrate the existence of any.

Most people who have studied the experiments would probably agree on the following facts, which are well documented in the literature: (1) High scores have been obtained in some investigations that make it almost certain that something is causing them to arise. (2) There is a high incidence of trickery on the part of those taking part in the experiments. (3) Some experiments for which the most ambitious claims have been made contain elementary faults in design, in execution, and in the experimental report.

Consider two important features of experimental work in science: (1) *repeatability;* (2) the use of *control series.* These assume particular importance where the conclusion drawn from an experiment arises from a statistical analysis of the data.

The conditions under which it becomes possible to demonstrate the existence of a phenomenon have been stated by R. L. Fisher as follows: "We may say that a phenomenon is experimentally demonstrable when we know how to conduct an experiment which will rarely fail to give a statistically significant result."

When an experiment is repeated in order to check its findings, the new investigators are likely to tighten up the experimental conditions if they consider any possible source of error to be present. In these circumstances, if a phenomenon is real it rapidly becomes established. But, if the result is not confirmed, it will rapidly be rejected. The original result may then be attributed to a weakness in the design that was eliminated in the second experiment. Thus failure to confirm the original result may provide evidence to indicate its cause.

Alternatively, the effect of a new precaution can be determined by carrying out an experiment in which the new feature is present in the experimental series and absent in a control series, the two series being similar in all other respects.

At Stanford University in 1915, J. E. Coover carried out an extensive investigation of telepathy in which 105 guessers and 97 senders took part. Ten thousand trials were made in which the guessers attempted to guess the identity of items in a pack of 40 playing cards, the picture cards having been removed. Before each run of 40 guesses a die was thrown to decide whether the sender should look at the cards during the run—making telepathy possible—or merely put the card aside without looking at it, thus making telepathy impossible. These latter runs constituted a control series for comparison with the experimental series. Coover found no appreciable difference between scores in the control and experimental series and concluded that no evidence had been provided for telepathy.

It was later claimed by Burt and Thouless that the total combined score of the experimental and control series had odds of about 200 to one against arising by chance. For Coover, the use of a control series ensured

that the effects of small sources of error common to the two series would be canceled out. Burt and Thouless were claiming that clairvoyance rather than error was operating in the control series and responsible for any above-chance score.

Rhine later claimed that subjects in his laboratory scored as easily under clairvoyance conditions as under telepathy conditions, thus making a transmitter redundant and a control series, as used by Coover, inappropriate.

With the emergence of clairvoyance, a control series in which the possibility of ESP was eliminated might have been possible by shuffling the pack of targets after the guessing had been completed; but Rhine found that high scores were still obtained and attributed this to precognition.

It is of interest that one investigator more recently dispersed with both the sender and the receiver. He merely compared two sets of tables of "random numbers" and claimed to get significantly above-chance scores. Koestler, Hardy, and Harvie attributed this effect to some new principle of "synchronicity" operating in nature. Others might deduce that it indicated the possibility of error when a single person has to check through two lists of numbers each containing 60,000 items.

The main use of a control series is to determine the effect on the result of some possible factor. The parapsychologist seeks to design his experiment so that all known factors that could cause high scores to arise are removed. If any known process can account for the result, it is unnecessary to invoke ESP. To incorporate a control series in which the possibilities of the experimenter being lax or cheating are removed may be achieved by tightening up the conditions; but this is unlikely to be done if the experimenter is a fraud or even anxious to keep on getting results. If, however, a further experiment is conducted by another investigator, he may attribute any failure to confirm the original result to removal of a source of error, and he may deduce the nature of that error from the difference in the experimental conditions between the control and experimental series.

Some interesting features arise when looking at the whole mass of ESP experiments and noting the effects of changes in procedure that may be regarded as acting to produce a control series in this manner. In some cases the data from a single investigation are sufficient. Such an example arises in the case of the Turner-Owmbey experiment, at one time considered by Rhine and his collaborators as one of the eight crucial experiments providing evidence for ESP.

Ms. Turner was guessing symbols generated by the experimenter, Ms. Owmbey, who was 250 miles away. There were five different possibilities at each guess; and each day, according to the report, independent records of 25 targets and guesses were sent to Rhine by the two women.

For the first three days scores were 19, 16, 16, giving odds against chance occurrence that were enormous. It is clear that something was responsible other than chance. Then, on the following five days, scores dropped to around the chance level (7, 7, 8, 6, 2). But it had been disclosed by Rhine when the experiment was in progress that during the first three days the list of guesses had been sent by Turner to Owmbey instead of directly to Rhine, and that Owmbey had the chance to amend her list of targets before passing it on.

Further similar tests with another subject carried out by Owmbey produced only chance scores throughout. This second subject had obtained consistently high scores in other ESP tests with the same experimenter when he was situated a few feet away.

Thus the presence of a loophole that permitted cheating to take place in the first experimental series gave high scores that were not present when the source of error was removed in the control series. It will also be observed that the second subject, who could obtain high scores under Rhine's usual lax conditions, was unable to do so when situated 250 miles away and when the possibility of error was reduced.

A further case arises from an experiment reported from Maimonides Dream Research Laboratory. Here an attempt made to identify targets from the reports of a subject following dreaming episodes was stated to give odds around 10,000 to 1 against chance occurrence. Two repeats of the experiment, in which a further outside investigator took part who imposed his own safeguards on the experimental procedure, failed to give a result significantly above chance. If these two repeat experiments are regarded as a control series in which a possible loophole was removed, and the original experiment is regarded as the experimental series in which the loophole was present, then it might be deduced that the result of the first experiment was due to the loophole in the original conditions.

An interesting example arises in the experiment on Uri Geller carried out by Targ and Puthoff and reported in *Nature,* where Geller made 10 attempts to guess the face-up on a die after it had been shaken inside a box. In 10 attempts he passed twice and obtained 8 hits in the remaining eight trials. The fact that Geller passed on two occasions is of interest, since parapsychologists in general have assumed—without seeking much experimental evidence—that their subjects are unable to say whether they are right or wrong when making their guesses.

Although these tests were made in the early part of Geller's stay at the Stanford Research Institute, no further tests were made to confirm the result. Since it is difficult to believe that Targ and Puthoff could have terminated their tests after making ten trials with a result that surpassed anything achieved since the heyday at Duke, it would seem likely that Geller would no longer take part following the imposition of some extra experimental precaution. Thus it would seem likely that the investigators

were more interested in publishing positive results than in finding out what was going on in their experiment.

In remote-viewing experiments with Geller reported in *Nature,* the results of 13 tests were reported. In all but three of the tests (Nos. 5, 6, 7), the targets were known to the people taking part in the experiment and present during the tests. In experiments 5, 6, and 7, however, an outside person prepared the drawing and it was revealed to no one else until Geller completed his drawing, Geller "passed" on all three of these tests. Thus it seems that withholding knowledge of the target's identity from persons who could have passed on the information to Geller removed his ability to perform. In three other tests the drawings were prepared by outside persons, but they were revealed to the investigators during the course of the experiment. In these three tests Geller was "successful."

Around 1938, following discussions between parapsychologists and psychologists concerning conditions required in ESP experiments, Rhine started using two experimenters who each ensured—or were meant to ensure—that the other experimenter could not affect the result. While this is desirable, it is not sufficient in itself to prevent fraud. In the United Kingdom, Soal quite clearly was able to cheat in spite of the presence of a second experimenter—Mrs. Goldney. In fact on the one occasion when he was seen to be cheating by an observer, this was kept quiet by the second experimenter and reported only following some pressure to do so twenty years later.

In the Pratt-Woodruff experiment, Rhine's "stringent" experimental conditions were insufficient to stop one of the investigators using trickery had he wished to do so. Similarly, in the experiments on mice culminating in the dismissal of the main investigator from his post, the presence of one cheat among a group of investigators escaped notice until the results of 12 experiments had been published. Thus an essential feature of a repeat experiment is that new investigators and none of the previous investigators take part.

Today card-guessing as employed by Rhine has been largely dropped. At the same time experiments have become more complex, less well controlled, and in some cases conducted by a single investigator. When an experimenter carries out a series of tests until he has consistently obtained significant results with very high anti-chance odds, the position is clear-cut: If, as he claims, his experiment is foolproof, then either he has demonstrated some new process that others should be able to confirm or else it is extremely likely that he is either incompetent or fraudulent.

A difficulty arises when considering Rhine's claim that only experimenters who believe ESP to be possible obtain results and that failure to obtain results by the critical investigators is not due to their care in carrying out the experiments but to some inhibitory effect that they have on the process in question. If this is so, the experimenter becomes part of

the experiment and he has to be controlled by other investigators. With present-day computer-controlled techniques, there is no difficulty in doing this and at the same time keeping the critical investigators well away from the scene of the experiment; but no attempt appears to have been made to do this.

One further investigation that should be mentioned is the Veritac series, carried out by five investigators for the U.S. Air Force. Here care was paid to the design of the experiments. The procedures of generating targets and recording subjects' guesses were automated. Tests were carried out for telepathy, clairvoyance, and precognition with groups of subjects who were classified as "sheep" or "goats." No evidence was found in the groups as a whole or in any individual subjects for any of the paranormal processes. The investigators state that at the start they were inclined to believe in ESP and were mainly interested in finding out more about it.

The failure to provide a satisfactory demonstration of any parapsychological process confirms the long-established belief that in the absence of information via the senses perception does not take place. Were it not for the misinformation provided by the media for the general public, there would be little point in paying attention to the claims of parapsychologists until they produce a repeatable demonstration. •