

## An Assessment of the Evidence for Psychic Functioning <sup>1</sup>

Jessica Utts

University of California, Davis

[In this issue of the *Journal*, we are pleased to publish evaluations of government-sponsored research of psychic functioning, predominantly remote viewing, conducted at SRI International and Science Applications International Corporation (SAIC) during the past three decades. These evaluations were commissioned by the American Institutes for Research (AIR) at the request of Congress and the Central Intelligence Agency. The purpose was to assess the validity of psychic functioning and its potential applications as a means of determining whether the research should receive continued government funding. The AIR, in turn, formed a panel consisting primarily of Dr. Jessica Utts (a leading proponent of psi research) and Dr. Ray Hyman (a leading critic of parapsychology) to undertake the evaluations...  
–[Original] *Editor's Note*]

**ABSTRACT:** Research on psychic functioning, conducted over a two-decade period, is examined to determine whether the phenomenon has been scientifically established. A secondary question is whether it is useful for government purposes. The primary work examined in this report was government-sponsored research conducted at Stanford Research Institute (later known as SRI International) and at Science Applications International Corporation (SAIC).

Using the standards applied to any other area of science, it is concluded that psychic functioning has been well established. The statistical results of the studies examined are far beyond what is expected by chance. Arguments that these results could be due to methodological flaws in the experiments are soundly refuted. Effects of a magnitude similar to those found in government-sponsored research at SRI and SAIC have been replicated at a number of laboratories around the world. Such consistency cannot be readily explained by claims of flaws or fraud.

The magnitude of psychic functioning exhibited appears to be in the range between what social scientists call a small and a medium effect. It is thus reliable enough to be replicated in properly conducted experiments, with sufficient trials to achieve the long-run statistical results needed for replicability.

A number of other patterns have been found, suggestive of how to conduct more productive experiments and to produce applied psychic functioning. For instance, it does not appear that a sender is needed. Precognition, in which the relevant information is known to no one until a future time, appears to work quite well. Recent experiments suggest that, if there is a psychic sense, it works much as our other five senses do, by detecting change. Physicists are currently grappling with an understanding of time, and it may be that a psychic sense scans the future for major change, much as our eyes scan the environment for visual change or our ears allow us to respond to sudden changes in sound.

---

<sup>1</sup> [Originally published in *JP*, 1995, 59, 289-320.]

**The author recommends that future experiments focus on understanding how psychic functioning works and on how to make it as useful as possible. There is little benefit in continuing experiments designed to offer proof, since there is little more to be offered to anyone who does not accept the current collection of data.**

## 1. INTRODUCTION

The purpose of this report is to examine a body of evidence collected over the past few decades in an attempt to determine whether psychic functioning is possible. Secondary questions include whether such functioning can be used productively for government purposes and whether the research to date provides any explanation for how it works.

There is no reason to treat this area differently from any other area of science that relies on statistical methods. Any discussion based on belief should be limited to questions that are not data driven, such as whether there are any methodological problems that could substantially alter the results. Too often people on both sides of the question debate the existence of psychic functioning on the basis of their personal belief systems rather than on an examination of the scientific data.

One objective of this report is to provide a brief overview of recent data as well as the scientific tools necessary for careful readers to reach their own conclusions based on those data. The tools consist of a rudimentary summary of how statistical evidence is typically evaluated and a listing of methodological concerns particular to experiments of this type.

Government-sponsored research in psychic functioning dates back to the early 1970s, when a program was initiated at what was then the Stanford Research Institute (now called SRI International). That program existed until 1989. The following year, government sponsorship moved to a program at Science Applications International Corporation (SAIC) under the direction of Dr. Edwin May, who had been employed in the SRI program since the mid-1970s and had been Project Director from 1986 until the close of the program.

This report will focus most closely on the most recent work, that done at SAIC. Section 2 describes the basic statistical and methodological issues required to understand this work. Section 3 discusses the program at SRI. Section 4 covers the SAIC work (with some of the details in Appendix B). Section 5 deals with the issue of external validation by exploring related results from other laboratories. Section 6 includes a discussion of the usefulness of this capability for government purposes. Section 7 provides conclusions and recommendations.

## 2. SCIENCE NOTES

### 2.1 *Definitions and Research Procedures*

There are two basic types of functioning that are generally considered under the broad heading of psychic or paranormal abilities. These are classically known as *extrasensory perception (ESP)*, in which one acquires information through unexplainable means, and *psychokinesis*, in which one physically ma-

nipulates the environment through unknown means. The SAIC laboratory uses more neutral terminology for these abilities; they refer to ESP as *anomalous cognition* (AC) and to psychokinesis as *anomalous perturbation* (AP). The vast majority of work at both SRI and SAIC investigated anomalous cognition rather than anomalous perturbation, although there was some work done on the latter.

Anomalous cognition is further divided into categories based on the apparent source of the information. If it appears to come from another person, the ability is called *telepathy*; if it appears to come in real time but not from another person, it is called *clairvoyance*, and if the information could have been obtained only by knowledge of the future, it is called *precognition*.

It is possible to identify apparent precognition by asking someone to describe something for which the correct answer will not be known until later in time. It is difficult to rule out precognition in experiments attempting to test telepathy or clairvoyance, since it is almost impossible to be sure that subjects in such experiments never learn the correct answer at some point in the future. These distinctions are important in the quest to identify an explanation for anomalous cognition, but they do not bear on the issue of its existence.

The vast majority of anomalous cognition experiments at both SRI and SAIC used a technique known as *remote viewing*. In these experiments, a *viewer* attempts to draw or describe (or both) a *target* location, photograph, object, or short video segment. All known channels for receiving the information are blocked. Sometimes the viewer is assisted by a *monitor* who asks the viewer questions; of course, in such cases the monitor is blind to the answer as well. Sometimes a *sender* is looking at the target during the session, but sometimes there is no sender. In most cases the viewer eventually receives *feedback* in which he or she learns the correct answer, thus making it difficult to rule out precognition as the explanation for positive results, whether or not there was a sender.

Most anomalous cognition experiments at SRI and SAIC were of the *free-response* type, in which viewers were asked simply to describe the target. In contrast, *forced-choice* experiments are ones in which there are a small number of known choices from which the viewer must choose. The latter may be easier to evaluate statistically, but they have been traditionally less successful than free-response experiments. Some of the work at SAIC addressed potential explanations for why that might be the case.

## 2.2 Statistical Issues and Definitions

Few human capabilities are perfectly replicable on demand. For example, even the best hitters in the major baseball leagues cannot hit on demand. Nor can we predict when they will hit or when they will score a home run. In fact, we cannot even predict whether or not a home run will occur in a particular game. That does not mean that home runs do not exist.

Scientific evidence in the statistical realm is based on replication of the same average performance or relationship *over the long run*. We would not expect a fair coin to result in 5 heads and 5 tails over each set of 10 tosses, but we can expect the proportion of heads and tails to settle down to about one half over a very long series of tosses. Similarly, a good baseball hitter will not hit the ball exactly the same proportion of times in each game but should be relatively consistent over the long run.

The same should be true of psychic functioning. Even if there truly is an effect, it may never be replicable on demand in the short run even if we understand how it works. However, over the long run in well-controlled laboratory experiments we should see a consistent level of functioning, above that expected by chance. The anticipated level of functioning may vary based on the individual players and the conditions, just as it does in baseball; but if players of similar ability are tested under similar conditions, the results should be replicable over the long run. In this report we will show that replicability in that sense has been achieved.

*2.2.1 P values and comparison with chance.* In any area of science, evidence based on statistics comes from comparing what actually happened to what should have happened by chance. For instance, without any special interventions, about 51% of births in the United States result in boys. Suppose someone claimed to have a method that enabled one to increase the chances of having a baby of the desired sex. We could study their method by comparing how often births resulted in a boy when that was the intended outcome. If that percentage was higher than the chance percentage of 51% *over the long run*, the claim would have been supported by statistical evidence.

Statisticians have developed numerical methods for comparing the results of an experiment to what is expected by chance. The *p* value is the answer to the following question: *If chance alone is responsible for the results, how likely would we be to observe results this strong or stronger?* If the answer to that question (i.e., the *p* value) is very small, then most researchers are willing to rule out chance as an explanation. In fact, it is commonly accepted practice to say that if the *p* value is 5% (.05) or less, we can rule out chance as an explanation. In such cases, the results are said to be *statistically significant*. Obviously, the smaller the *p* value, the more convincingly chance can be ruled out.

Notice that when chance alone is at work, we *erroneously* find a statistically significant result about 5% of the time. For this reason and others, most reasonable scientists require replication of nonchance results before they are convinced that chance can be ruled out.

*2.2.2 Replication and effect sizes.* In the past few decades scientists have realized that true replication of experimental results should focus on the *magnitude* of the effect, or the *effect size*, rather than on replication of the *p* value. This is because the latter is heavily dependent on the size of the study. In a very large study, it will take only a small magnitude effect to rule out chance convincingly. In a very small study, it would take a huge effect to rule out chance convincingly.

In our hypothetical sex-determination experiment, suppose 70 out of 100 births intended to be boys actually resulted in boys, giving a rate of 70% instead of the 51% expected by chance. The experiment would have a *p* value of .0001, quite convincingly ruling out chance. Now suppose someone attempted to replicate the experiment with only 10 births and found 7 boys, that is, also 70%. The smaller experiment would have a *p* value of .19 and would not be statistically significant. If we were simply to focus on that issue, the result would appear to be a failure to replicate the original result, even though it achieved exactly the same 70% boys! In only 10 births it would require 90% of them to be boys before chance could be ruled out. Yet the 70% rate is a more exact replication of the result than the 90%.

Therefore, although  $p$  values should be used to assess the overall evidence for a phenomenon, they should not be used to define whether or not a replication of an experimental result was “successful.” Instead, a successful replication should be one that achieves an effect that is within expected statistical variability of the original result or that achieves an even stronger effect for explainable reasons.

A number of different *effect size* measures are in use in the social sciences, but in this report we will focus on the one used most often in remote viewing at SRI and SAIC. Because the definition is somewhat technical, it is given in Appendix A. An intuitive explanation will be given in the next subsection. Here, we note that an effect size of 0 is consistent with chance, and social scientists have, by convention, declared an effect size of 0.2 as small, 0.5 as medium, and 0.8 as large. A medium effect size is supposed to be visible to the naked eye of a careful observer, whereas a large effect size is supposed to be evident to any observer.

*2.2.3 Randomness and rank-order judging.* At the heart of any statistical method is a definition of what should happen “randomly” or “by chance.” Without a random mechanism, there can be no statistical evaluation.

There is nothing random about the responses generated in anomalous cognition experiments; in other words, there is no way to define what they would look like “by chance.” Therefore, the random mechanism in these experiments must be in the choice of the target. In that way, we can compare the response to the target and answer the question: “If chance alone is at work, what is the probability that a *target* would be chosen that matches this *response* as well as or better than does the actual target?”

To accomplish this purpose, a properly conducted experiment uses a set of targets defined in advance. The target for each remote viewing is then selected randomly, in such a way that the probability of getting each possible target is known.

The SAIC remote-viewing experiments and all but the early ones at SRI used a statistical evaluation method known as *rank-order judging*. After the completion of a remote viewing, a judge who is blind to the true target (called a *blind judge*) is shown the response and five potential targets, one of which is the correct answer and the other four of which are “decoys.” Before the experiment is conducted, each of those five choices must have had an equal chance of being selected as the actual target. The judge is asked to assign a rank to each of the possible targets, where a rank of 1 means it matches the response most closely, and a rank of 5 means it matches the least.

The rank of the correct target is the numerical score for that remote viewing. By chance alone the actual target would receive each of the five ranks with equal likelihood, since, despite what the response said, the target matching it best would have the same chance of selection as the one matching it second best and so on. The average rank by chance would be 3. Evidence for anomalous cognition occurs when the average rank over a series of trials is significantly lower than 3. (Notice that a rank of 1 is the best possible score for each viewing.)

This scoring method is conservative in the sense that it gives no extra credit for an excellent match. A response that describes the target almost perfectly will achieve the same rank of 1 as a response that

contains only enough information to pick the target as the best choice out of the five possible choices. One advantage of this method is that it is still valid even if the viewer knows the set of possible targets. The probability of a first place match by chance would still be only one in five. This is important because the later SRI and many of the SAIC experiments used the same large set of photographs as targets. Therefore, the experienced viewers would eventually become familiar with the range of possibilities since they were usually shown the answer at the end of each remote-viewing session.

For technical reasons explained in Appendix A, the effect size for a series of remote viewings using rank-order judging with five choices is  $(3.0 - \text{average rank}) / \sqrt{2}$ . Therefore, small, medium, and large effect sizes (0.2, 0.5 and 0.8) correspond to average ranks of 2.72, 2.29 and 1.87, respectively. Notice that the largest effect size possible using this method is 1.4, which would result if every remote viewing achieved a first place ranking.

### *2.3 Methodological Issues*

One of the challenges in designing a good experiment in any area of science is to close the loopholes that would allow explanations other than the intended one to account for the results.

There are a number of places in a remote-viewing experiment where information could be conveyed by normal means if proper precautions are not taken. The early SRI experiments suffered from some of those problems, but the later SRI experiments and the SAIC work were done with reasonable methodological rigor, with some exceptions noted in the detailed descriptions of the SAIC experiments in Appendix B.

The following list of methodological issues shows the variety of concerns that must be addressed. It should be obvious that a well-designed experiment requires careful thought and planning:

1. No one who has knowledge of the specific target should have any contact with the viewer until after the response has been safely secured.
2. No one who has knowledge of the specific target or even of whether or not the session was successful should have any contact with the judge until after the judging has been completed.
3. No one who has knowledge of the specific target should have access to the response until after the judging has been completed.
4. Targets and decoys used in judging should be selected using a well-tested randomization device.
5. Duplicate sets of target photographs should be used, one during the experiment and one during the judging, so that no cues (such as fingerprints) can be inserted onto the target that would help the judge recognize it.
6. The criterion for stopping an experiment should be defined in advance so that it is not called to a halt when the results just happen to be favorable. Generally, that means specifying the number of trials in advance, but some statistical procedures require or allow other stopping rules. The important point is that the rule be defined in advance in such a way that there is no ambiguity about when to stop.
7. Reasons, if any, for excluding data must be defined in advance and followed consistently, and the reasons should not be dependent on the data. For example, a rule specifying that a trial could be

aborted if the viewer felt ill would be legitimate, but only if the trial was aborted before anyone involved in that decision knew the correct target.

8. Statistical analyses to be used must be planned before collecting the data so that a method most favorable to the data is not selected post hoc. If multiple methods of analysis are used, the corresponding conclusions must recognize that fact.

#### 2.4 *Prima Facie Evidence*

According to *Webster's Dictionary*, in law prima facie evidence is "evidence having such a degree of probability that it must prevail unless the contrary be proved." There are a few examples of applied, nonlaboratory remote viewings provided to the review team that would seem to meet that criterion for evidence. These are examples in which the sponsor or another government client asked for a single remote viewing of a site, known to the requestor in real time or in the future, and the viewer provided details far beyond what could be taken as a reasonable guess. Two such examples are given by May (1995) in which it appears that the results were so striking that they far exceed the phenomenon as observed in the laboratory. Using a post hoc analysis May concluded that in one of the cases the remote viewer was able to describe a microwave generator with 80% accuracy, and that of what he said almost 70% of it was reliable. Laboratory remote viewings rarely show that level of correspondence.

Notice that standard statistical methods cannot be used in these cases because there is no standard for probabilistic comparison. But evidence gained from applied remote viewing cannot be dismissed as inconsequential just because we cannot assign specific probabilities to the results. It is most important to ascertain whether or not the information was achievable in other standard ways. In Section 3 an example is given in which a remote viewer allegedly gave codewords from a secret facility that he should not have even known existed. Suppose the sponsors could be absolutely certain that the viewer could not have known about those codewords through normal means. Then even if we cannot assign an exact probability to the fact that he guessed them correctly, we can agree that it would be very small. That would seem to constitute prima facie evidence unless an alternative explanation could be found. Similarly, the viewer who described that microwave generator allegedly knew only that the target was a technical site in the United States. Yet he drew and described the microwave generator, including its function, its approximate size, how it was housed, and the fact that it had "a beam divergence angle of 30 degrees" (May, 1995, p. 15).

Anecdotal reports of psychic functioning suffer from a similar problem in terms of their usefulness as proof. They have the additional difficulty that the "response" isn't even well defined in advance, unlike applied remote viewing, where the viewer provides a fixed set of information on request. For instance, if a few people each night happen to dream of plane crashes, then some will obviously do so on the night before a major plane crash. Those individuals may interpret the coincidental timing as meaningful. This is undoubtedly the reason many people think the reality of psychic functioning is a matter of belief rather than science, since they are more familiar with the provocative anecdotes than with the laboratory evidence.

### 3. THE SRI ERA

#### 3.1 *Early Operational Successes and Evaluation*

According to Puthoff and Targ (1975), the scientific research endeavor at SRI may never have been supported had it not been for three apparent operational successes in the early days of the program. These are detailed by Puthoff and Targ (1975), although the level of the matches is not clearly delineated.

In one of the apparent successes, concerning the “West Virginia Site,” two remote viewers purportedly identified an underground secret facility. One of them apparently named codewords and personnel in this facility accurately enough to set off a security investigation to determine how that information could have been leaked. Given only the coordinates of the site, the viewer first described the above-ground terrain, and he then proceeded to describe details of the hidden underground site.

The same viewer then claimed that he could describe a similar Communist Bloc site and proceeded to do so for a site in the Urals. According to Puthoff and Targ, “the two reports for the West Virginia Site, and the report for the Urals Site were verified by personnel in the sponsor organization as being substantially correct” (p. 8).

The third reported operational success concerned an accurate description of a large crane and other information at a site in Semipalatinsk, USSR. Again the viewer was provided with only the geographic coordinates of the site and was asked to describe what was there.

Although some of the information in these examples was verified to be highly accurate, the *evaluation* of operational work remains difficult, in part because there is no chance baseline for comparison (as there is in controlled experiments) and in part because of differing expectations of different evaluators. For example, a government official who reviewed the Semipalatinsk work concluded that there was no way the remote viewer could have drawn the large gantry crane unless “he actually saw it through remote viewing, or he was informed of what to draw by someone knowledgeable of [the site].” Yet that same analyst concluded that “the remote viewing of [the site] by subject S1 proved to be unsuccessful” because “the only positive evidence of the rail-mounted gantry crane was far outweighed by the large amount of negative evidence noted in the body of this analysis.” In other words, the analyst had the expectation that in order to be “successful” a remote viewing should contain accurate information only.

Another problem with evaluating this operational work is that there is no way to know with certainty that the subject did not speak with someone who had knowledge of the site, however unlikely that possibility may appear. Finally, we do not know to what degree the results in the reports were selectively chosen because they were correct. These problems can all be avoided with well-designed controlled experiments.

#### 3.2 *The Early Scientific Effort at SRI*

During 1974 and early 1975, a number of controlled experiments were conducted to see whether various types of target material could be successfully described with remote viewing. The results reported by Puthoff and Targ (1975) indicated success with a wide range of material, from “technical”

targets, such as a xerox machine, to natural settings, such as a swimming pool. These and some of the subsequent experiments, however, were criticized on statistical and methodological grounds. I briefly describe one series of experiments and criticisms of it to show the kinds of problems that existed in the early scientific effort.

The largest series during the 1973 to 1975 time period involved remote viewing of natural sites. Sites were randomly selected for each trial from a set of 100 possibilities. They were selected “without replacement,” meaning that sites were not reused once they had been selected. The series included eight viewers, including two supplied by the sponsor. Many of the descriptions showed a high degree of subjective correspondence, and the overall statistical results were quite striking for most of the viewers.

Critics attacked these experiments on a number of issues, including the selection of sites without replacement and the statistical scoring method used. The results were scored by having a blind judge attempt to match the target material with the transcripts of the responses. A large fraction of the matches were successful. But critics noted that some successful matching could be attained just from cues contained in the transcripts of the material, such as instances when a subject mentioned in one session what the target had been in the previous session. Because sites were selected without replacement, knowing what the answer was on one day would exclude that target site from being the answer on any other day. There was no way to determine the extent to which these problems influenced the results. The criticisms of these and subsequent experiments, although perhaps unwelcome at the time, have resulted in substantially improved methodology in these experiments.

### *3.3 An Overall Analysis of the SRI Experiments: 1973-1988*

In 1988 an analysis was made of all of the experiments conducted at SRI from 1973 until that time (May et al., 1988). The analysis was based on all 154 experiments conducted during that era, consisting of over 26,000 individual trials. Of those, almost 20,000 were of the forced-choice type, and just over 1000 were laboratory remote viewings. There was a total of 227 subjects in all experiments.

The statistical results were so overwhelming that results that extreme or more would occur only about once in every 1020 such instances if chance alone were the explanation (i.e. the  $p$  value was less than 10-20). Obviously some explanation other than chance must be found. Psychic functioning may not be the only possibility, especially since some of the earlier work contained methodological problems. However, the fact that the same level of functioning continued to hold in the later experiments, which did not contain those flaws, lends support to the idea that the methodological problems cannot account for the results. In fact, there was a talented group of subjects (labeled G1 in that report) for whom the effects were stronger than for the group at large. According to May, the majority of experiments with that group were conducted later in the program, when the methodology had been substantially improved.

In addition to the statistical results, a number of other questions and patterns were examined. A summary of the results revealed the following:

1. “Free-response” remote viewing, in which subjects describe a target, was much more successful than “forced-choice” experiments, in which subjects were asked to choose from a small set of possibilities.

2. There were six selected individuals whose performance far exceeded that of unselected subjects. The fact that these same selected individuals consistently performed better than others under a variety of protocols provides a type of replicability that helps substantiate the validity of the results. If methodological problems were responsible for the results, they should not have affected this group differently from others.
3. Mass-screening efforts found that about 1% of those who volunteered to be tested were consistently successful at remote viewing. This indicates that remote viewing is an ability that differs across individuals, much like athletic ability or musical talent. (Results of mass-screenings were not included in the formal analysis because the conditions were not well controlled, but the subsequent data from subjects found during mass screening were included.)
4. Neither practice nor a variety of training techniques consistently worked to improve remote-viewing ability. It appears that it is easier to find than to train good remote viewers.
5. It is not clear whether feedback (showing the subject the right answer) is necessary, but it does appear to provide a psychological boost that may increase performance.
6. Distance between the target and the subject does not seem to impact the quality of the remote viewing.
7. Electromagnetic shielding does not appear to inhibit performance.
8. There is compelling evidence that precognition, in which the target is selected after the subject has given the description, is also successful.
9. There is not evidence to support anomalous perturbation (psychokinesis)--that is, physical interaction with the environment by psychic means.

#### *3.4 Consistency with Other Laboratories in the Same Era*

One of the hallmarks of a real phenomenon is that its magnitude is replicable by various researchers working under similar conditions. The results of the overall SRI analysis are consistent with results of similar experiments in other laboratories. For instance, an overview of forced-choice precognition experiments (Honorton & Ferrari, 1989) found an average effect size per experimenter of 0.033, whereas all forced-choice experiments at SRI resulted in a similar effect size of 0.052. The comparison is not ideal because the SRI forced-choice experiments were not necessarily precognitive, and they used different types of target material than the standard card-guessing experiments.

Methodologically sound remote viewing has not been undertaken at other laboratories, but a similar regime called the ganzfeld (described in more detail in Section 5) has been similarly successful. The largest collection of ganzfeld experiments was conducted from 1983 to 1989 at the Psychophysical Research Laboratories in Princeton, NJ. Those experiments were also reported by separating novices from experienced subjects. The overall effect size for novice remote viewing at SRI was 0.164, and the effect size for novices in the ganzfeld at PRL was a very similar 0.17. For experienced remote viewers at SRI the overall effect size was 0.385; for experienced viewers in the ganzfeld experiments it was 0.35. These consistent results across laboratories help refute the idea that the successful experiments at any one lab are the result of fraud, sloppy protocols, or some methodological problem, and they also provide an indication of what can be expected in future experiments.

## 4. THE SAIC ERA

### 4.1 An Overview

The review team decided to focus more intensively on the experiments conducted at Science Applications International Corporation (SAIC), because they provide a manageable yet varied set to examine in detail. They were guided by a Scientific Oversight Committee consisting of experts in a variety of disciplines, including a winner of the Nobel Prize in Physics, internationally known professors of statistics, psychology, neuroscience and astronomy, and a medical doctor who is a retired U.S. Army Major General. Further, we have access to the details for the full set of SAIC experiments, but not for those conducted at SRI. Whatever detail may be missing from the written reports are obtainable from the principal investigator, Edwin May, to whom we have been given unlimited access.

In a memorandum dated July 25, 1995, May listed the set of experiments conducted by SAIC. There were 10 experiments, all designed to answer questions about psychic functioning raised by the work at SRI and other laboratories, rather than just to provide additional proof of its existence. Some of the experiments were of a format similar to the remote-viewing experiments conducted at SRI, and we can examine those to see whether they replicated the SRI results. We will also examine what new knowledge can be gained from the results of the SAIC work.

### 4.2 The 10 Experiments

Of the 10 experiments done at SAIC, six of them involved remote viewing and four did not. Rather than list the details in the body of this report, I give a brief description of the experiments in Appendix B. What follows is a discussion of the methodology and results for the experiments as a whole. Because of the fundamental differences between remote viewing and the other types of experiments, we discuss them separately.

In the memorandum of July 25, 1995, May provided the review team with details of the 10 experiments, including a short title, number of trials, effect size, and overall  $p$  value for each. His list was in chronological sequence. It is reproduced in Table 1, using his numbering system, with the experiments categorized by type and then sequentially within type. The effect size estimates are based on a limited number of trials, and thus they are augmented with an interval to show the probable range of the true effect (e.g.,  $0.124 \pm 0.071$  indicates a range from 0.053 to 0.195). Remember that an effect size of 0 represents chance, and a positive effect size indicates positive results.

### 4.3 Assessing the Remote-Viewing Experiments by Homogeneous Sets of Sessions

Although Table 1 provides an overall assessment of the results of each experiment, it does so at the expense of information about variability among viewers and types of targets. In terms of understanding the phenomenon, it is important to break the results down into units that are as homogeneous as possible in terms of procedure, individual viewer, and type of target. This is also important in order to assess the impact of any potential methodological problems. For example, in one pilot experiment (E6, AC in Lucid Dreams) viewers were permitted to take the targets home with them in sealed envelopes.

Table 2 presents the effect size results at the most homogeneous level possible based on the information provided. For descriptions of the experiments, refer to Appendix B. Overall effect sizes for each viewer and total effect sizes for each experiment are weighted according to the number of trials, so that each trial receives equal weight.

TABLE 1  
SAIC EXPERIMENTS LISTED BY DR. EDWIN MAY

Experiment	Title	Trials	Effect Size	p value
Remote-Viewing Experiments				
1	Target dependencies	200	0.124 ± 0.071	.040
4	AC with binary coding	40	-0.067 ± 0.158	.664
5	AC lucid dreams, base	24	0.088 ± 0.204	.333
6	AC lucid dreams, pilot	21	0.368 ± 0.218	.046
9	ERDa AC behavior	70	0.303 ± 0.120	.006
10	Entropy II	90	0.550 ± 0.105	9.1 × 10 <sup>-8</sup>
Other Experiments				
2	AC of binary targets	300	0.123 ± 0.058	.017
3	MEG replication	12,000sb	MCE	MCE
7	Remote observation	48	0.361 ± 0.144	.006
8	ERD EEG investigation	7,000s	MCE	MCE

aERD = event related desynchronizations.

bs = stimuli in physiological experiments.

#### 4.4 Consistency and Replicability of the Remote-Viewing Results

One of the most important hallmarks of science is replicability. A phenomenon with statistical variability, whether it is scoring home runs in baseball, curing a disease with chemotherapy, or observing psychic functioning, should exhibit about the same level of success in the long run, over repeated experiments of a similar nature. The remote-viewing experiments are no exception. Remember that such events should not replicate with any degree of precision in the short run because of statistical variability, just as we would not expect to always get five heads and five tails if we flip a coin 10 times, or to see the same batting averages in every game.

TABLE 2  
INDIVIDUAL EFFECT SIZES

Experiment	Expert remote viewers					Viewer unknown/ other	Total
	009	131	372	389	518		
Static targets (photographs)							
E1: Static	0.424	-0.071	0.424	0.177	0.283	n.a.	0.247
E9	0.432	n.a.	0.354	0.177	n.a.	n.a.	0.303
E10: Static	0.566	n.a.	0.801	-0.071	0.778	n.a.	0.550
E5a	n.a.	n.a.	n.a.	n.a.	n.a.	0.088	0.088
E6b	n.a.	n.a.	n.a.	n.a.	n.a.	0.370	0.370
E4c	-0.112	n.a.	0	0.112	n.a.	-0.559	-0.067
Dynamic targets (video film clips)							
E1: Dynamic	0	0.354	-0.283	0	-0.071	n.a.	0.000
E10: Dynamic	0.919	n.a.	0.754	0	0.424	n.a.	0.550
Overall	0.352	0.141	0.340	0.090	0.271	n.a.	

<sup>a</sup>Experiment 5 did not include any expert viewers.

<sup>b</sup>Experiment 6 included four expert viewers, but separate results were not provided.

<sup>c</sup>Experiment 4 used a specially designed target set and only four choices in judging.

The analysis of SRI experiments conducted in 1988 singled out the laboratory remote-viewing sessions performed by six “expert” remote viewers (numbers 002, 009, 131, 372, 414, and 504). These six individuals contributed 196 sessions. The resulting effect size was 0.385 (May et al., 1988, p. 13). The SRI analysis does not include information individually by viewer, nor does it include information about how many of the 196 sessions used static versus dynamic targets. One report provided to the review team (May, Lantz, & Piantineda, 1994) included an additional experiment conducted after the 1988 review was performed, in which viewer 009 participated with 40 sessions. The effect size for viewer 009 for those sessions was 0.363. None of the other five SRI experts were participants.

The same identifying numbers for subjects were used at SAIC, and so we can compare the performance for these individuals at SRI and SAIC. Of the six, three were specifically mentioned as participating in the SAIC remote-viewing experiments. As can be seen in Table 2, viewers 009, 131, and 372 all participated in Experiment 1, and viewers 009 and 372 participated in Experiments 4, 9, and 10 as well.

The overall effect sizes for two of the three viewers (009 and 372) were very close to the SRI effect size of 0.385 for these subjects, at 0.35 and 0.34, respectively, and the 0.35 effect size for viewer 009 was very similar to his 0.363 effect size in the report by May, Lantz, and Piantineda (1994). Therefore, we see a repeated and, more importantly, hopefully repeatable level of functioning above chance for these individuals. An effect of this size should be reliable enough to be sustained in any properly conducted experiment with enough trials to obtain the long run statistical replicability required to rule out chance.

It is also important to notice that viewers 009 and 372 did well on the same experiments and poorly on the same experiments. In fact, the correlation between their effect sizes across experiments is .901, which is very close to a perfect correlation of 1.0. This kind of consistency warrants investigation to determine whether it is the nature of the experiments, a statistical fluke, or some methodological problems that led these two individuals to perform so closely to one another. If methodological problems are responsible, then they must be subtle indeed because the methodology was similar for many of the experiments, yet the results were not. For instance, procedures for the sessions with static and dynamic targets in Experiment 1 were almost identical to each other, yet the dynamic targets did not produce evidence of psychic functioning ( $p = .50$ ) and the static targets did ( $p = .0073$ ). Therefore, a methodological problem would have had to affect results differentially for the two types of targets, even though the assignment of target type was random across sessions.

#### *4.5 Methodological Issues in the Remote-Viewing Experiments at SAIC*

As noted in Section 2.3, there are a number of methodological considerations needed to perform a careful remote-viewing experiment. Information necessary to determine how well each of these were addressed is generally available in the reports, but in some instances I consulted May for additional information. As an example of how the methodological issues in Section 2.3 were addressed, an explanation will be provided for Experiment 1.

In this experiment the viewers all worked from their homes (in New York, Kansas, California, and Virginia). Nevin Lantz, who resided in Pennsylvania, was the principal investigator. After each session, viewers faxed their response to Lantz and mailed the original to SAIC. Upon receipt of the fax, Lantz mailed the correct answer to the viewer. The viewers were supposed to mail their original responses to SAIC immediately, after faxing them to Lantz. According to May, the faxed versions were later compared with the originals to make sure the originals were sent without any changes. The other methodological issues in Section 2.3 were handled as follows:

1. *No one who has knowledge of the specific target should have any contact with the viewer until after the response has been safely secured.*

No one involved with the experiment had any contact with the viewers, since they were not in the vicinity of either SAIC or Lantz's home in Pennsylvania.

2. *No one who has knowledge of the specific target or even of whether or not the session was successful should have any contact with the judge until after that task has been completed.*

Lantz and the individual viewers were the only ones who knew the correct answers, but according to May, they did not have any contact with the judge during the period of this experiment.

3. *No one who has knowledge of the specific target should have access to the response until after the judging has been completed.*

Again, since only the viewers and Lantz knew the correct target, and since the responses were mailed to SAIC by the viewers before they received the answers, this condition appears to have been met.

4. *Targets and decoys used in judging should be selected using a well-tested randomization device.*

This has been standard practice at both SRI and SAIC.

5. *Duplicate sets of target photographs should be used, one during the experiment and one during the judging, so that no cues (such as fingerprints) can be inserted onto the target that would help the judge recognize it.*

This was done; Lantz maintained the set used during the experiment, and the set used for judging was kept at SAIC in California.

6. *The criterion for stopping an experiment should be defined in advance so that it is not called to a halt when the results just happen to be favorable. Generally, that means specifying the number of trials in advance, but some statistical procedures require other stopping rules. The important point is that the rule be defined in advance in such a way that there is no ambiguity about when to stop.*

In advance it was decided that each viewer would contribute 40 trials, 10 under each of four conditions (all combinations of sender/no sender and static/dynamic). All sessions were completed.

7. *Reasons, if any, for excluding data must be defined in advance and followed consistently, and the reasons should not be dependent on the data. For example, a rule specifying that a trial could be aborted if the viewer felt ill would be legitimate, but only if the trial was aborted before anyone involved in that decision knew the correct target.*

No such reasons were given, nor was there any mention of any sessions being aborted or discarded.

8. *Statistical analyses to be used must be planned before collecting the data so that a method most favorable to the data is not selected post hoc. If multiple methods of analysis are used, the corresponding conclusions must recognize that fact.*

The standard rank-order judging had been planned, with results reported separately for each of the four conditions in the experiment for each viewer. Thus 20 effect sizes were reported, four for each of the five viewers.

#### 4.6 Was Anything Learned at SAIC?

4.6.1 *Target selection.* In addition to the question of whether psychic functioning is possible, the experiments at SAIC were designed to explore a number of hypotheses. Experiments 1 and 10 were

both designed to see whether there is a relationship between the “change in visual entropy” in the targets and the remote-viewing performance.

Each of the five sense with which we are familiar is a change detector. Our vision is most readily drawn to something that is moving, and in fact if our eyes are kept completely still, we cease to see at all. Similarly, we hear because of moving air, and our attention is drawn to sudden changes in sound levels. Other senses behave similarly. Thus, it is reasonable that if there really is a “psychic sense,” then it would follow that same pattern.

Experiments 1 and 10 were designed to test whether remote-viewing performance would be related to a particular type of change in the target material, namely, the “change in visual entropy.” A target with a high degree of change would be one in which the colors changed considerably throughout the target. A detailed explanation can be found in the SAIC reports of this experiment or in May, Spottiswoode, and James (1994). There was indeed a correlation between the change in entropy in the target and the remote-viewing quality. This result was initially shown in Experiment 1 and replicated in Experiment 10. A simulation study matching randomly chosen targets to responses showed that this was unlikely to be an artifact of target complexity or other features.

It is worth speculating on what this might mean for determining how psychic functioning works. Physicists are currently grappling with the concept of time, and precognition may in fact be consistent with current understanding. Perhaps it is the case that we do have a psychic sense, much like our other senses, and that it works by scanning the future for possibilities of major change much as our eyes scan the environment for visual change and our ears are responsive to auditory change. That idea is consistent with anecdotal reports of precognition, which are generally concerned with events involving major life change. Laboratory remote viewing may in part work by someone directing the viewer to focus on a particular point in the future, that in which he or she receives the feedback from the experiment. It may also be the case that this same sense can scan the environment in actual time and detect change as well.

Another hypothesis put forth as SAIC was that laboratory remote-viewing experiments are most likely to be successful if the pool of potential targets is neither too narrow nor too wide in terms of the number of possible elements in the target. They called this feature the “target-pool bandwidth” and described it as the number of “differential cognitive elements.” They reasoned that if the possible target set was too small, the viewer would see the entire set and be unable to distinguish that information from the psychic information. If the set was too broad, the viewer would not have any means for editing an extensive imagination.

Combining these two results would indicate that a good target set would contain targets with high change in visual entropy, but that the set would contain a moderately sized set of possibilities. The set of 100 photographs used in the later days at SRI and at SAIC may have inadvertently displayed just those properties.

*4.6.2 Remote staring.* Experiment 7, described in Appendix B, provided results very different from the standard remote-viewing work. That experiment was designed to test claims, made in the

former Soviet Union and by some researchers in the United States, that individuals could influence the physiology of another individual from a remote location. The study was actually two separate replications of the same experiment, and both replications were successful from a traditional statistical perspective. In other words, it appeared that the physiology of one individual was activated when he or she was being watched by someone in a distant room. If these results are indeed sound, then they may substantiate the folklore indicating that people know when they are being observed from behind.

*4.6.3 Enhanced binary computer guessing.* Experiment 2 was also very different from the standard remote-viewing experiments, although it was still designed to test anomalous cognition. Three subjects attempted to use a statistical enhancement technique to increase the ability to guess forced-choice targets with two choices. This clever computer experiment showed that for one subject, guessing was indeed enhanced from a raw rate of just above chance (51.6% instead of 50%) to an enhanced rate of 76%. The method was extremely inefficient, and it is difficult to imagine practical uses for this ability, if indeed it exists.

## 5. EXTERNAL VALIDATION: REPLICATIONS OF OTHER EXPERIMENTS

### *5.1 Conceptual Similarity: Ganzfeld Experiments*

While remote viewing has been the primary activity at SRI and SAIC, other researchers have used a similar technique to test for anomalous cognition, called the ganzfeld. As noted in the SAIC Final Report of 29 September 1994, the ganzfeld experiments differ from remote viewing in three fundamental ways. First, a “mild altered state is used”; second, senders are usually used, so that telepathy is the primary mode; and third, the *receivers* (viewers) do their own judging just after the session, rather than having an independent judge.

The ganzfeld experiments conducted at the Psychophysical Research Laboratories (PRL) were mentioned in Section 3.4. Since the time those results were reported, other laboratories have also been conducting ganzfeld experiments. At the 1995 annual Meeting of the Parapsychological Association, three replications were reported, all published in the peer-reviewed *Proceedings* of the conference.

The ganzfeld experiments differ in the preferred method of analysis as well. Rather than using the sum of the ranks across sessions, a simple count is made of how many first place matches resulted from a series. Four rather than five choices are given, and so by chance there should be about 25% of the sessions resulting in first place matches.

### *5.2 Ganzfeld Results from Four Laboratories*

In publishing the ganzfeld results from the Psychophysical Research Laboratories, Bem and Horton (1994) excluded one of the studies from the general analysis for methodological reasons, and found that the remaining studies showed 106 hits out of 329 sessions, for a hit rate of 32.2% when 25% was expected by chance. The corresponding *p* value was .002. As mentioned earlier, the hallmark of science is replication. This result has now been replicated by three additional laboratories.

Bierman (1995) reported four series of experiments conducted at the University of Amsterdam. Overall, there were 124 sessions and 46 hits, for a hit rate of 37%. The hit rates for the four individual experiments were 34.3%, 37.5%, 40%, and 36.1%, and so the results are consistent across his four experiments.

Morris, Dalton, Delanoy, and Watt (1995) reported results of 97 sessions conducted at the University of Edinburgh in which there were 32 successes, for a hit rate of 33%. They conducted approximately equal numbers of sessions under each of three conditions. In one condition there was a known sender, and in the other two conditions it was randomly determined at the last minute (and unknown to the receiver) that there would either be a sender or not. Hit rates were 34% when there was a known sender and when there was no sender, and 28% when there was a sender but the receiver did not know whether there would be. They did discover post hoc that one experimenter was more successful than the other two at achieving successful sessions, but the result was not beyond what would be expected by chance as a post hoc observation.

Broughton and Alexander (1995) reported results from 100 sessions at the Institute for Parapsychology in North Carolina. They too found a similar hit rate, with 33 hits out of 100 sessions, or 33%.

Results from the original ganzfeld work and these three replications are summarized in Table 3, along with the SRI and SAIC remote-viewing results. The effect sizes for the ganzfeld replications are based on Cohen's  $h$ , which is similar in type to the effect size used for the remote-viewing data. Both effect sizes measure the number of standard deviations by which the results fall above chance, using the standard deviation for a single session.

TABLE 3  
REMOTE VIEWING AND GANZFELD REPLICATIONS

Laboratory	Sessions	Hit rate	Effect size
All remote viewing at SRI	770	n.a.	0.209
All remote viewing at SAIC	445	n.a.	0.230
PRL, Princeton, NJ	329	32%	0.167
University of Amsterdam, the Netherlands	124	37%	0.261
University of Edinburgh, Scotland	97	33%	0.177
Institute for Parapsychology, NC	100	33%	0.177

### 5.3 Conclusions about External Replication

The results shown in Table 3 show that remote viewing has been conceptually replicated across a number of laboratories, by various experimenters and in different cultures. This is a robust effect that,

were it not in such an unusual domain, would no longer be questioned by science as a real phenomenon. It is unlikely that methodological problems could account for the remarkable consistency of results shown in Table 3.

## 6. IS REMOTE VIEWING USEFUL?

Even if we were all to agree that anomalous cognition is possible, there remains the question of whether it would have any practical use for government purposes. The answer to that question is beyond the scope of this report, but some speculations can be made about how to increase the usefulness.

First, it appears that anomalous cognition is to some extent possible in the general population. None of the ganzfeld experiments used exclusively selected subjects. However, it also appears that certain individuals possess more talent than others and that it is easier to find those individuals than to train people. It also appears to be the case that certain individuals are better at some tasks than others. For instance, viewer 372 at SAIC appears to have a facility with describing technical sites.

Second, if remote viewing is to be useful, the end users must be trained in what it can do and what it cannot. At our current level of understanding, remote viewing is rarely 100% accurate, and there is no reliable way to learn what is accurate and what is not. The same is probably true of most sources of intelligence data.

Third, what is useful for one purpose may not be useful for another. For instance, suppose a remote viewer could describe the setting in which a hostage is being held. That information may not be any use at all to those unfamiliar with the territory, but could be useful to those familiar with it.

## 7. CONCLUSIONS AND RECOMMENDATIONS

It is clear to this author that anomalous cognition is possible and has been demonstrated. This conclusion is not based on belief, but rather on commonly accepted scientific criteria. The phenomenon has been replicated in a number of forms across laboratories and cultures. The various experiments in which it has been observed have been different enough that if some subtle methodological problems can explain the results, then there would have to be a different explanation for each type of experiment, yet the impact would have to be similar across experiments and laboratories. If fraud were responsible, similarly, it would require an equivalent amount of fraud on the part of a large number of experimenters or an even larger number of subjects.

What is not so clear is that we have progressed very far in understanding the mechanism for anomalous cognition. Senders do not appear to be necessary at all; and feedback of the correct answer may or may not be necessary. Distance in time and space do not seem to be an impediment. Beyond those conclusions, we know very little.

I believe that it would be wasteful of valuable resources to continue to look for proof. No one who has examined all of the data across laboratories, taken as a collective whole, has been able to suggest methodological or statistical problems to explain the ever-increasing and consistent results to date.

Resources should be directed to the pertinent questions about how this ability works. I am confident that the questions are no more elusive than any other questions in science dealing with small to medium sized effects, and that if appropriate resources are targeted to appropriate questions, we can have answers within the next decade.

## **APPENDIX A**

### **EFFECT SIZE MEASURE USED WITH RANK-ORDER JUDGING**

In general, effect sizes measure the number of standard deviations the true population value of interest falls from the value that would be true if chance alone were at work. The standard deviation used is for one subject, trial, and so forth, rather than being the standard error of the sample statistic used in the hypothesis test.

In rank-order judging, let  $R$  be the rank for one trial. If the number of possible choices is  $N$ , then we find:

$$E(R) = (N + 1)/2$$

and

$$\text{Var}(R) = (N^2 - 1)/12.$$

Therefore, when  $N = 5$ , we find  $E(R_i) = 3$  and  $\text{Var}(R_i) = 2$ . The effect size is therefore:

$$\text{Effect Size} = (3.0 - \text{Average Rank})/\sqrt{2}.$$

## **APPENDIX B**

### **A BRIEF DESCRIPTION OF THE SAIC EXPERIMENTS**

#### *Experiments Involving Remote Viewing*

There were six experiments involving remote viewing, done for a variety of purposes.

#### *Experiment 1: Target and Sender Dependencies*

*Purpose.* This experiment was designed to test whether a sender is necessary for successful remote viewing and whether dynamic targets, consisting of short video clips, would result in more successful remote viewing than the standard photographs used in most of the SRI experiments.

*Method:* Five experienced remote viewers participated, three of whom (numbers 009, 131, and 372) were included in the experienced group at SRI; their identification numbers were carried over to the SAIC experiments. Each viewer worked from his or her home and faxed the results of the sessions to the principal investigator, Nevin Lantz, located in Pennsylvania. Whether the target was static or dynamic and whether there was a sender was randomly determined and unknown to the viewer. Upon receiving the fax of the response, Lantz mailed the correct answer to the viewer. The original response was sent to SAIC in California, where the results were judged by an analyst blind to the correct target. Standard rank-order judging was used.

Since it is not explicitly stated, I asked May what measures were taken to make sure the viewer actually mailed the original response to SAIC *before* receiving the correct answer in the mail. He said that the original faxed responses were compared with the responses received by SAIC to make sure they were the same, and they all were.

*Results.* Each viewer contributed 10 trials under each of the four possible conditions (sender/no sender and static/dynamic target), for a total of 40 trials per viewer. There was a moderate difference (effect size = 0.121,  $p = .08$ ) between the static and dynamic targets, with the traditional photographs faring better than the dynamic video clips. There was no noticeable difference based on whether a sender was involved, supporting the same conclusion reached in the overall analysis of the SRI work. Combined over all conditions and all viewers, the effect size was 0.124 ( $p = .04$ ); for the static targets alone it was 0.248 (exact  $p = .0073$ ), and for the dynamic targets it was 0.00 ( $p = .50$ ).

*Discussion:* The SAIC staff speculated that the dynamic targets were not successful because the possibilities were too broad. They chose a new set of dynamic targets to be more similar to the static targets and performed another experiment the following year to compare the static targets with the more similar set of dynamic ones. That experiment is described below (Experiment 10).

#### *Experiment 4: Enhancing Detection of AC with Binary Coding*

*Purpose:* This experiment was designed to see whether remote viewing could be used to develop a message-sending capability by focusing on the presence or absence of five specific features of a target. The target set was constructed in packets of four, with possible combinations of the absence (0) or presence (1) of each of the five features chosen to correspond to the numbers 00000, 01110, 10101, and 11011. This is standard practice in information theory when trying to send a two-digit number (00, 01, 10, or 11); the remaining three bits are used for "error correction." Different sets of five features were used for each of 10 target packs.

*Method:* Five viewers each contributed eight trials, but the same eight targets were used for all five viewers. There was no sender, and viewers were told that each target would be in a fixed location for one week. They were to spend 15 minutes trying to draw the target, then fax their responses to SAIC in California. The results were blind-judged, and the binary features were coded both by the viewers and by an independent analyst.

*Results:* The results were unsuccessful in showing any evidence of psychic functioning. Neither standard rank-order judging nor analysis based on the binary guesses showed any promise that this method works to send messages.

#### *Experiment 5: AC in Lucid Dreams (Baseline)*

*Purpose:* Despite its name, this experiment did not involve lucid dreaming. Instead, it was used to test three novice remote viewers who were to participate in an experiment involving remote viewing while dreaming. This baseline experiment was designed to see whether these individuals would be successful at standard laboratory remote viewing.

*Method:* For this baseline experiment, each of the three viewers contributed eight trials using a standard protocol common in the SRI era. For each trial, a target was randomly chosen from the set of 100 photograph targets used at SRI and SAIC. The target was placed on a table (no sender was used) while the viewer, in another room, was asked to provide a description. The response was later blind-judged by comparing it to the target and four decoys and providing a rank ordering of the five choices.

*Results:* Of the three novice viewers, one obtained a promising effect size of 0.265, although the result was not statistically significant due to the small number of trials (eight). Individual results were not provided for the other two viewers, but the overall effect size was reported as 0.088 for the three viewers.

#### *Experiment 6: AC in Lucid Dreams (Pilot)*

*Purpose:* A lucid dream is a dream in which one becomes aware that he or she is dreaming and can control subsequent events in the dream. This ability has apparently been successfully trained by Stephen LaBerge of the Lucidity Institute. He was the principal investigator for this experiment. The experiment was designed to see whether remote viewing could be successfully employed while the viewer was having a lucid dream.

*Method:* Seven remote viewers were used; four were experienced SAIC remote viewers, and three were experienced lucid dreamers from the Lucidity Institute. The latter three were the novice viewers used in Experiment 5. The experienced SAIC remote viewers were given training in lucid dreaming. The number of trials contributed by each viewer could not be fixed in advance because of the difficulty of attaining the lucid dream state. A total of 21 trials were conducted, with the seven viewers contributing anywhere from one to seven trials each. The report did not mention whether the stopping criterion was fixed in advance, but according to Dr. May the experiment was designed to proceed for a fixed time period and to include all sessions attained during that time period.

In contrast to well-controlled protocols, the viewers were allowed to take the target material home with them. The targets, selected from the standard pool of photographs, were sealed in opaque envelopes with covert threads to detect possible tampering (there were no indications of such tampering). Viewers were instructed to place the targets at bedside and to attempt a lucid dream in which the envelope was opened and the target viewed. Drawings and descriptions were then to be produced upon awakening.

*Results:* The results were blind-judged using the standard sum of ranks. Since the majority of viewers contributed only one or two trials, analysis by individual viewer would be meaningless. For the 21 trials combined, the effect size was 0.368 ( $p = .046$ ). Information was not provided to differentiate the novice remote viewers from the experienced ones.

#### *Experiment 9: ERD (Event Related Desynchronization) AC Behavior*

*Purpose:* The remote viewing in this experiment was conducted in conjunction with measurement of brain waves using an EEG. The purpose of the experiment was to see whether EEG activity would change when the target the person was attempting to describe was briefly displayed on a computer monitor in a distant room. Details of the EEG portion will be explained as Experiment 8. Here, we summarize the remote-viewing part of the study.

*Method:* Three experienced remote viewers (numbers 009, 372, and 389) participated. Because of the pilot nature of the experiment, the number of trials differed for each viewer based on availability, with viewers 009, 372, and 389 contributing 18, 24, and 28 trials, respectively. Although it is not good protocol to allow an unspecified number of trials, it does not appear that this problem can explain the results of this experiment.

*Results.* Responses were blind-judged using standard rank-order analysis. The effect sizes for viewers 009, 372, and 389 were 0.432 ( $p = .033$ ), 0.354 ( $p = .042$ ) and 0.177 ( $p = .175$ ), respectively. The overall effect size was .0303 ( $p = .006$ ).

#### *Experiment 10: Entropy II*

*Purpose:* This experiment was designed as an improved version of Experiment 1. After the unsuccessful showing for the dynamic targets in Experiment 1, the SAIC team speculated that the “target-pool band-width”—defined as the number of “cognitively differentiable elements” in the target pool—might be an important factor. If the possible target material was extremely broad, viewers might have trouble filtering out extraneous noise. If the set of possibilities was too small, as in forced-choice experiments, the viewer would see all choices at once and would have trouble filtering out that knowledge. An intermediate range of possibilities, too large to be considered all at once, was predicted to be ideal. The standard photograph pool seemed to fit that range. For this experiment, a pool of dynamic targets was created with a similar “band-width.” In both experiments (1 and 10) the researchers predicted that remote-viewing success would correlate with the change in visual entropy of the target, as explained earlier.

TABLE 4  
RESULTS OF EXPERIMENT 10

Viewer	Rank	Static Targets		Dynamic Targets		
		Effect size	$p$	Rank	Effect size	$p$
009	2.20	0.565	.037	1.70	0.919	$1.8 \times 10^{-3}$
372	1.87	0.801	$9.7 \times 10^{-4}$	1.93	0.754	$1.8 \times 10^{-3}$
389	3.10	-0.071	.589	3.00	0.000	.500
518	1.90	0.778	$7.2 \times 10^{-3}$	2.40	0.424	.091
Total	2.22	0.550	$1.1 \times 10^{-5}$	2.22	0.550	$1.1 \times 10^{-5}$

*Method:* Four of the five viewers from Experiment 1 were used (numbers 009, 372, 389, and 518). They each contributed equal numbers of sessions with static and dynamic targets, with the viewers blind to which trials had which type. Senders were not used, and all sessions were conducted

at SAIC in California (unlike Experiment 1, in which the viewers worked at home). Viewer 372 contributed 15 of each type, and the others each contributed 10 of each type. Standard rank-order judging was used.

*Results:* Table 4 shows the results for this experiment. Unlike in Experiment 1, the static and dynamic targets produced identical effect sizes, with both types producing very successful results. The combined effect size for all trials is 0.55, resulting in a z score of 5.22.

#### *The Other Experiments at SAIC*

There were four additional experiments at SAIC, not involving remote viewing. Two of them (Experiments 3 and 8) involved trying to measure brain activity related to psychic functioning and will be described briefly. Experiment 3 used a magnetoencephalograph (MEG) to attempt to detect anomalous signals in the brain when a remote stimulus was present. Due to the background noise in the brain measurements and the expected strength of the signal, the experimenters realized too late that they would not be able to detect a signal even if it existed. Experiment 8 utilized an EEG to try to detect the interruption of alpha waves when a remote-viewing target was briefly displayed on a computer monitor in another room. The area of the brain tested was that corresponding to visual stimuli. No significant change in alpha was seen.

The remaining two experiments were replications of previous work measuring psychic functioning in areas other than remote viewing. They will be described in detail.

#### *Experiment 2: AC of Binary Targets*

*Purpose:* This experiment attempted to replicate and enhance random number generator experiments conducted at SRI. In these types of experiments a computer randomly selects one of two choices to be the target, denoted as 0 or 1. The internal workings of the computer then rapidly oscillate between 0 and 1, and the subject pushes a mouse button when he or she thinks the internal choice matches the target choice. This process is repeated over many trials. The computer tabulates the results, and the experiment is a success if the subject guesses the correct answer more often than would be expected by chance. The purpose is to see whether humans can correctly guess the computer-selected binary targets and, hopefully by extension, correctly solve binary choice problems in real situations. If that were to be the case, then real problems could be posed as binary ones (e.g., is the lost child still in this city or not?) to narrow down possibilities.

*Method:* This SAIC experiment was designed to enhance the accuracy of binary guessing by using a statistical technique called sequential analysis. Rather than just give one guess for each decision, the subject continues to guess until the computer ascertains that a decision has been reached. The computer keeps track of the number of times 0 and 1 have each been guessed and announces a decision when one of the choices has clearly won out over the other, or when it is clear that it is essentially an ongoing tie. In the latter case, no decision is recorded. Three subjects participated (numbers 007, 083, and 531) in this experiment. Subject 531 had been successful in similar experiments at SRI.

*Results:* Using this method for enhancing the accuracy of the guesses, subject 531, who had been successful in previous similar experiments, was able to achieve 76 correct answers out of 100 tries. This remarkable level of scoring for this type of experiment resulted in an effect size of 0.520 and a z score of 5.20. The other two subjects did not differ from chance results, with 44 and 49 correct decisions out of 100 or 101. (One subject accidentally contributed an additional trial.)

Although the result for subject 531 is remarkably successful, it does not represent a very efficient method of obtaining the decision. To reach the 100 decisions required a total of 21,337 guesses, that is, over 200 guesses for each decision. Of the individual guesses, only 51.6% were correct, for an effect size of 0.032, similar to other forced-choice experiments. Due to the large number of guesses, the corresponding z score was 4.65. Combined over all three subjects, 56% of the 301 decisions were correct, and the effect size was 0.123. The combined results were still statistically significant, with  $p = .017$ , as shown in Table 1.

#### *Experiment 7: Remote Observation*

*Purpose:* It is often reported anecdotally that people know when they are being watched. Two experiments were conducted at SAIC to determine whether these anecdotes could be supported by a change in physiology when someone is being observed from a distance. The experimental design was essentially the same for the two experiments. This work was a conceptual replication of results reported by researchers in the former Soviet Union, the United States, and Scotland. The experiments in the former Soviet Union were interpreted to mean that the physiology of the recipient was being manipulated by the sender, an effect that if real could have frightening consequences.

*Method:* The “observee” was seated in a room with a video camera focused on him or her and with galvanic skin response measurements being recorded. In a distant room the “observer” attempted to influence the physiology of the observee at randomly spaced time intervals. During those time intervals, an image of the observee appeared on a computer monitor watched by the observer. During “control” periods, the video camera remained focused on the observee, but the computer monitor did not display his or her image to the observer. There were 16 “influence” periods randomly interspersed with 16 “control” periods, each of 30 seconds, with blank periods of 0 to 5 seconds inserted to rule out patterns in physiology.

*Result:* To determine whether the galvanic skin response of the observees was activated while they were being watched, the response during the control periods was compared with the response during the “influence” periods for each subject. The results were then averaged across subjects. In both experiments, there was greater activity during the periods of being watched than there was during the control periods. The results were statistically significant in each case ( $p = .036$  and  $.014$ ), and the effect sizes were similar, at 0.39 and 0.49. As preplanned, the results were combined, yielding an effect size of 0.39 ( $p = .005$ ). As an interesting post hoc observation, it was noted that the effect was substantially stronger when the observer and observee were of opposite sexes than when they were of the same sex.

*Discussion:* This experiment differs from the others conducted at SAIC since it involves interaction between two people rather than one person ascertaining information about the environment or the

future. It raises substantially more questions than it answers, since the mechanism for the shift in physiology is unknown. Possibilities range from the idea that the observee was able to know when the computer in the distant room was displaying his or her image (not unlike remote viewing) to the possibility that the observer actually did influence the physiology of the observee. Further experimentation as well as a review of similar past experiments may be able to shed light on this important question.

## References

- BEM, D. J., & HONORTON, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin*, **115**, 4-18.
- BIERMAN, D. J. (1995). The Amsterdam Ganzfeld Series III & IV: Target clip emotionality, effect sizes and openness. *Proceedings of Presented Papers: The Parapsychological Association 38th Annual Convention*, 27-37.
- BROUGHTON, R. & ALEXANDER, C. (1995). Autoganzfeld II: The first 100 sessions. *Proceedings of Presented Papers: The Parapsychological Association 38th Annual Convention*, 53-61.
- HONORTON, C., & FERRARI, D. C. (1989). "Future telling": A meta-analysis of forced-choice precognition experiments, 1935-1987. *Journal of Parapsychology*, **53**, 281-308.
- MAY, E. C. (1995, May). *AC technical trials: Inspiration for the target entropy concept*. SAIC Technical Report, Palo Alto, CA.
- MAY, E. C., LANTZ, N. D., & PIANTINEDA, T. (1994, November). *Feedback considerations in anomalous cognition experiments*. SAIC Technical Report, Palo Alto, CA.
- MAY, E. C., UTTS, J. M., TRASK, V. V., LUKE, W. W., FRIVOLD, T. J., & HUMPHREY, B. S. (1988). *Review of the psychoenergetic research conducted at SRI International (1973-1988)*. SRI International Technical Report, Menlo Park, CA.
- MORRIS, R. L., DALTON, K., DELANOY, D., & WATT, C. (1995). Comparison of the sender/no sender condition in the ganzfeld. *Proceedings of Presented Papers: The Parapsychological Association 38th Annual Convention*, 244-259.
- PUTHOFF, H. E., & TARG, R. (1975, December). *Perceptual augmentation techniques: Part two—Research report*. Stanford Research Institute Final Report, Menlo Park, CA.

## Une Évaluation des Preuves du Fonctionnement Psi

[Dans ce numéro du journal, nous sommes heureux de publier des évaluations des recherches sur le fonctionnement psychique sponsorisées par le gouvernement, en particulier la vision à distance, conduite au SRI International and Science Applications International Corporation (SAIC) au cours des trois dernières décennies. Ces évaluations ont été diligentées par l'American Institutes for Research (AIR) à la requête du Congrès et de la Central Intelligence Agency. Le but était d'évaluer la validité du fonctionnement psychique et ses potentielles applications afin de déterminer si la recherche pourrait continuer à recevoir des financements gouvernementaux. L'AIR, à son tour, a formé un panel consistant d'abord du Dr. Jessica Utts (un chercheur de pointe dans la recherche psi) et du Dr. Ray Hyman (un critique de pointe sur la parapsychologie) pour réaliser ces évaluations. Ce sont leurs rapports, en plus d'une brève réponse d'Utts à Hyman, que nous publions ici. Le lecteur devra garder à l'esprit que ces évaluations n'étaient pas conçues pour rendre compte de l'ensemble des recherches psi, et les auteurs étaient limités par les instructions reçues de la part de l'AIR. Nous publions ces évaluations telles que soumises à l'AIR, avec un minimum de travail éditorial. —Ed.]

RESUME : Une recherche sur le fonctionnement psi, conduite sur une période de deux décennies, est examinée pour déterminer si le phénomène a été scientifiquement établi. Une question secondaire est celle de son utilité pour des objectifs gouvernementaux. Le premier travail examiné dans ce rapport porte sur la recherche sponsorisée par le gouvernement conduite au Stanford Research Institute (dit ensuite SRI International) et au Science Applications International Corporation (SAIC).

En utilisant les standards appliqués dans d'autres zones de la science, nous concluons que le fonctionnement psi a bien été établi. Les résultats statistiques des études examinées sont bien au-delà de ce que nous pourrions attendre du hasard seul. Les arguments selon lesquels ces résultats pourraient être dus à des biais méthodologiques dans les expérimentations sont clairement réfutés. Des effets sur la magnitude similaires à ceux trouvés dans les recherches sponsorisées par le gouvernement au SRI et au SAIC ont été répliqués dans plusieurs laboratoires autour du monde. Une telle consistance ne peut pas être expliquée simplement par des affirmations non prouvées de biais ou de fraude.

La magnitude du fonctionnement psi manifestée apparaît dans la portion que les chercheurs en sciences sociales appellent un effet faible ou médium. Il est donc suffisamment fiable pour être répliqué dans des expérimentations correctement conduites, avec suffisamment d'essais pour obtenir les résultats statistiques accumulés nécessaires pour la réplification.

Nombre d'autres patterns ont été trouvés, suggérant des façons de conduire des expérimentations plus productives et de produire des applications du fonctionnement psi. Par exemple, il ne semble pas nécessaire d'inclure un émetteur. La précognition, dans laquelle l'information pertinente n'est connue par personne avant une période future, semble très bien fonctionner. Des expérimentations récentes suggèrent que, s'il y a un sens psi, il fonctionne plutôt comme nos cinq autres sens, en détectant des changements. Les physiciens sont actuellement en train de s'échiner à comprendre le temps, et il se pourrait bien qu'existe un sens psi capable de scanner le futur pour prévenir des changements majeurs, à la manière de nos yeux scannant l'environnement pour détecter des changements visuels ou de nos oreilles pour réagir des changements soudains de son.

L'auteur recommande que des expérimentations futures se focalisent sur la compréhension du fonctionnement psi et la façon de le rendre aussi utile que possible. Il y a peu de bénéfices à poursuivre des expérimentations orientées vers la preuve, puisqu'il sera difficile d'offrir mieux à ceux qui n'acceptent pas déjà les données actuellement collectées.

### **Eine Bewertung des Beweismaterials Für Parapsychisches Funktionieren**

[Wir freuen uns, in dieser Ausgabe des *Journal* Auswertungen der von der Regierung geförderten Forschungen zu parapsychischen Funktionsweisen, überwiegend Remote Viewing, zu veröffentlichen, die in den letzten drei Jahrzehnten von der SRI International and Science Applications International Corporation (SAIC) durchgeführt worden waren. Diese Auswertungen wurden von den American Institutes for Research (AIR) auf Wunsch des Kongresses und der Central Intelligence Agency in Auftrag gegeben. Ziel war es, die Validität der parapsychischen Funktionsweisen und ihrer möglichen Anwendungen zu bewerten, um festzustellen, ob die Forschung weiterhin staatliche Mittel erhalten sollte. Die AIR wiederum bildete ein Panel, das hauptsächlich aus Dr. Jessica Utts (einer führenden Vertreterin der Psi-Forschung) und Dr. Ray Hyman (einem führenden Kritiker der Parapsychologie) bestand, um diese Bewertungen durchzuführen. Ihre Berichte, zusammen mit einer kurzen Antwort von Utts an Hyman,

veröffentlichen wir hier. Der Leser sollte sich darüber im Klaren sein, dass diese Bewertungen nicht als umfassende Übersichtsarbeiten zur Psi-Forschung gedacht und die Autoren durch die Vorgaben von AIR eingeschränkt waren. Wir veröffentlichen die Bewertungen, wie sie bei AIR eingereicht wurden, mit minimaler Bearbeitung.- Anm. d. Red.]

**ZUSAMMENFASSUNG:** Die Forschung über parapsychische Funktionsweisen, die über einen Zeitraum von zwei Jahrzehnten durchgeführt wurde, wird bewertet, um festzustellen, ob das Phänomen wissenschaftlich belegt ist. Eine zweite Frage betrifft die Nützlichkeit für staatliche Zwecke. Im Mittelpunkt dieses Berichts steht die von der Regierung geförderte Forschungsarbeit am Stanford Research Institute (später bekannt als SRI International) und an der Science Applications International Corporation (SAIC). Unter Verwendung der für alle anderen Wissenschaftsbereiche geltenden Maßstäbe wird der Schluss gezogen, dass parapsychische Funktionsweisen gut etabliert sind. Die statistischen Ergebnisse der untersuchten Studien liegen weit über dem, was durch Zufall erwartet wird. Argumente, dass diese Ergebnisse auf methodische Fehler in den Experimenten zurückzuführen sein könnten, werden fundiert widerlegt. Effekte von ähnlicher Größenordnung wie bei der von der Regierung geförderten Forschung bei SRI und SAIC wurden in einer Reihe von Laboratorien auf der ganzen Welt reproduziert. Diese Konsistenz lässt sich nicht ohne weiteres aufgrund von Fehlern oder Betrug erklären.

Die Größenordnung der hier gezeigten parapsychischen Funktionsweisen scheint im Bereich dessen zu liegen, was in den Sozialwissenschaften als schwacher und mittlerer Effekt bezeichnet wird. Diese ist somit zuverlässig genug, um in ordnungsgemäß durchgeführten Experimenten repliziert zu werden, mit einer ausreichenden Anzahl von Versuchen, um die für die Replizierbarkeit erforderlichen langfristigen statistischen Ergebnisse zu erzielen.

Eine Reihe weiterer Muster wurden gefunden, die darauf hinweisen, wie man produktivere Experimente durchführen und anwendbare parapsychische Funktionsweisen erzeugen kann. So scheint beispielsweise kein Sender benötigt zu werden. Präkognition, bei der relevante Information jemandem erst in der Zukunft bekannt wird, scheint recht gut zu funktionieren. Neuere Experimente deuten darauf hin, dass ein etwaiger parapsychischer Sinn wie unsere anderen fünf Sinne funktioniert, indem er auf Veränderungen reagiert. Physiker bemühen sich derzeit um ein Verständnis von Zeit, und es kann sein, dass ein parapsychischer Sinn die Zukunft nach bedeutsamen Veränderungen durchsucht, so wie unsere Augen die Umgebung nach visuellen Veränderungen absuchen oder unsere Ohren es uns erlauben, auf plötzliche Klangveränderungen zu reagieren.

Der Autor empfiehlt, dass sich zukünftige Experimente darauf konzentrieren sollten, zu verstehen, wie parapsychische Funktionsweisen zustande kommen und wie man diese so nützlich wie möglich anwenden kann. Beweisorientierte Experimente lediglich fortzusetzen, bringt kaum noch einen Nutzen, da demjenigen, der den vorliegenden Bestand an Daten nicht akzeptiert, kaum mehr geboten werden kann.

## **Una Evaluación de la Evidencia de Funcionamiento Psíquico**

[En este número de la revista, nos complace publicar evaluaciones de investigaciones patrocinadas por el gobierno sobre el funcionamiento psíquico, en su mayoría de visión remota (remote viewing), realizadas en SRI International y Science Applications International Corporation (SAIC) durante las úl-

timas tres décadas. Estas evaluaciones fueron encargadas por el American Institutes for Research (AIR) a solicitud del Congreso y la Agencia Central de Inteligencia Norteamericanos. El propósito fue evaluar la validez del funcionamiento psíquico y sus posibles aplicaciones como un medio para determinar si la investigación debe recibir financiamiento gubernamental continuo. El AIR, a su vez, formó un panel compuesto principalmente por la Dra. Jessica Utts (una de las principales defensoras de la investigación psi) y el Dr. Ray Hyman (un destacado crítico de la parapsicología) para llevar a cabo las evaluaciones... El lector debe tener en cuenta que estas evaluaciones no pretendían ser revisiones exhaustivas de la investigación psi y que los autores estaban limitados por las instrucciones que recibieron de AIR. Publicamos las evaluaciones tal como se enviaron a AIR, con una edición mínima. -Nota del editor]

RESUMEN: Se examina la investigación sobre el funcionamiento psíquico, realizada durante un período de dos décadas, para determinar si el fenómeno se ha establecido científicamente. Una pregunta secundaria es si es útil para propósitos gubernamentales. El trabajo principal examinado en este informe fue una investigación patrocinada por el gobierno realizada en el Stanford Research Institute (más tarde conocido como SRI International) y la Science Applications International Corporation (SAIC).

Usando los estándares aplicados a cualquier otra área de la ciencia, se concluye que el funcionamiento psíquico ha sido bien establecido. Los resultados estadísticos de los estudios examinados van mucho más allá de lo que se esperaría por casualidad. Los argumentos de que estos resultados podrían deberse a fallas metodológicas en los experimentos son ampliamente refutados. Los efectos de una magnitud similar a los encontrados en la investigación patrocinada por el gobierno en SRI y SAIC se han replicado en varios laboratorios de todo el mundo. Dicha consistencia no puede explicarse fácilmente como fallas o fraude.

La magnitud del funcionamiento psíquico exhibido parece estar en el rango de lo que los científicos sociales llaman efectos pequeño a mediano. Por lo tanto, es lo suficientemente confiable como para ser replicado en experimentos conducidos adecuadamente, con datos suficientes para lograr los resultados estadísticos a largo plazo necesarios para la replicabilidad.

Se han encontrado varios otros patrones que sugieren cómo realizar experimentos más productivos y producir un funcionamiento psíquico aplicado. Por ejemplo, no parece que se necesite alguien que envíe la información. La precognición, en la cual nadie conoce la información relevante hasta un tiempo futuro, parece funcionar bastante bien. Experimentos recientes sugieren que, si existe un sentido psíquico, funciona de manera similar a como lo hacen nuestros otros cinco sentidos, al detectar cambios. Los físicos actualmente están tratando de entender al tiempo, y puede ser que un sentido psíquico explore el futuro en busca de un cambio importante, así como nuestros ojos escanean el entorno en busca de cambios visuales o nuestros oídos nos permiten responder a cambios repentinos en el sonido. El autor recomienda que los experimentos futuros se centren en comprender cómo trabaja el funcionamiento psíquico y en cómo hacerlo lo más útil posible. Hay pocos beneficios en los experimentos continuos diseñados para obtener confirmación, ya que hay poco más que ofrecer a quien no quiera aceptar los datos ya recopilados.