An Evaluation of the Remote Vewing Program: Research and Operational Applications

DRAFT REPORT

Prepared by:
The American Institutes for Research

September 22, 1995

Note: The page sequence in this document has been corrected from the original record.

Approved For Release 2002/05/22 : CIA-RDP96-00791R000200180005-5

Executive Summary

Studies of paranormal phenomena have nearly always been associated with controversy. Despite the controversy concerning their nature and existence, many individuals and organizations continue to be avidly interested in these phenomena. The intelligence community is no exception: beginning in the 1970s, it has conducted a program intended to investigate the application of one paranormal phenomenon — ***remote viewing,*** or the ability to describe locations one has not visited. of which one has no prior knowledge?

Conceptually, remote viewing would seem to have tremendous potential utility for the intelligence community. Accordingly, a three-component program involving basic research, operations, and foreign assessment has been in place for some time. Prior to transferring this program to a new sponsoring organization within the intelligence community, a thorough program review was initiated.

The part of the program review conducted by the American Institutes for Research (AIR), a nonprofit, private research organization, consisted of two main components. The first component was a review of the research program. The second component was a review of the operational application of the remote viewing phenomenon in intelligence gathering. Evaluation of the foreign assessment component of the program was not within the scope of the present effort.

Research Evaluation

To evaluate the research program, a "blue ribbon" panel was assembled. The panel included two noted experts in the area of parapsychology: *Dr. Jessica Utts*, a Professor of Statistics at the University of California/Davis, and *Dr. Raymond Hyman*, a Professor of Psychology at the University of Oregon. In addition to their extensive credentials, they were selected to represent both sides of the paranormal controversy: Dr. Utts has published articles that view paranormal interpretations positively, while Dr. Hyman was selected to represent a more skeptical position. Both, however, are viewed as fair and open-minded scientists. In addition to these experts, this panel included two Senior Scientists from AIR; both have recognized methodological expertise, and both had no prior background in parapsychological research. They were included in the review panel to provide an unbiased methodological

perspective. In addition, *Dr. Lincoln Moses*, an Emeritus Professor at Stanford University, provided statistical advice, while *Dr. David A. Goslin*, President of AIR, served as coordinator of the research effort.

Panel members were asked to review all laboratory experiments and meta-analytic reviews conducted as part of the research program; this consisted of approximately 80 separate publications, many of which are summary reports of multiple experiments. In the course of this review, special attention was given to those studies that (a) provided the strongest evidence for the remote viewing phenomenon, and (b) represented new experiments controlling for methodological artifacts identified in earlier reviews. Separate written reviews were prepared by Dr. Utts and Dr. Hyman. They exchanged reviews with other panel members who then tried to reach a consensus.

In the typical remote viewing experiment in the laboratory, a remote viewer is asked to visualize a place, location, or object being viewed by a "beacon" or sender. A judge then examines the viewer's report and determines if this report matches the target or, alternatively, a set of decoys. In most recent laboratory experiments reviewed for the present evaluation, National Geographic photographs provided the target pool. If the viewer's reports match the target, as opposed to the decoys, a hit is said to have occurred. Alternatively, accuracy of a set of remote viewing reports is assessed by rank-ordering the similarity of each remote viewing report to each photograph in the target set (usually five photographs). A better-than-chance score is presumed to represent the occurrence of the paranormal phenomenon of remote viewing, since the remote viewers had not seen the photographs they had described (or did not know which photographs had been randomly selected for a particular remote viewing trial).

In evaluating the various laboratory studies conducted to date, the reviewers reached the following conclusions:

- A statistically significant laboratory effort has been demonstrated in the sense that hits occur more often than chance.

- It is unclear whether the observed effects can unambiguously be attributed to the paranormal ability of the remote viewers as opposed to characteristics of the

judges or of the target or some other characteristic of the methods used. Use of the same remote viewers, the same judge, and the same target photographs makes it impossible to identify their independent effects.

- Evidence has not been provided that clearly demonstrates that the *causes* of hits are due to the operation of paranormal phenomena; the laboratory experiments have not identified the sources or origins of the remote viewing phenomenon.

Operational Evaluation

The second component of the program involved the use of remote viewing in gathering intelligence information. Here, representatives of various intelligence groups—"end users" of intelligence information — presented targets to remote viewers, who were asked to describe the target. Typically, the remote viewers described the results of their experiences in written reports, which were forwarded to the end users for evaluation and, if warranted, action.

To assess the operational value of remote viewing in intelligence gathering, a multifaceted evaluation strategy was employed. First, the relevant research literature was reviewed to identify whether the conditions applying during intelligence gathering would reasonably permit application of the remote viewing paradigm. Second, members of three groups involved in the program were interviewed: (1) end users of the information; (2) the remote viewers providing the reports, and (3) the program manager. Third, feedback information obtained from end user judgments of the accuracy and value of the remote viewing reports were assessed.

This multifaceted evaluation effort led to the following conclusions:

- The conditions under which the remote viewing phenomenon is observed in laboratory settings do not apply in intelligence gathering situations. For example, viewers cannot be provided with feedback and targets may not display the characteristics needed to produce hits.

- The end users indicating that, although some accuracy was observed with regard to broad background characteristics, the remote viewing reports failed to produce the concrete, specific information valued in intelligence gathering.

- The information provided was inconsistent, inaccurate with regard to specifics, and required substantial subjective interpretation.

- In no case had the information provided ever been used to guide intelligence operations. Thus, Remote viewing failed to produce actionable intelligence.

Conclusions

The foregoing observations provide a compelling argument against continuation of the research program within the intelligence community. Even though a statistically significant effect has been observed in the laboratory, it remains unclear whether the existence of a paranormal phenomenon, remote viewing, has been demonstrated. The laboratory studies do not provide evidence regarding the sources or origins of the phenomenon, nor do they address an important methodological issue of inter-judge reliability.

Further, even if it could be demonstrated unequivocally that a paranormal phenomenon occurs under the conditions present in the laboratory paradigm, these conditions have limited applicability and utility for intelligence gathering operations. For example, the nature of the remote viewing targets are vastly dissimilar, as are the specific tasks required of the remote viewers. Most importantly, the information provided by remote viewing is vague and ambiguous, making it difficult, if not impossible, for the technique to yield information of sufficient quality and accuracy ~~of information~~ for actionable intelligence. Thus, we conclude that continued use of remote viewing in intelligence gathering operations is not warranted.

1. Background and History

In their continuing quest to improve effectiveness, many organizations have sought techniques that might be used to enhance performance. For the most part, the candidate techniques come from rather traditional lines of inquiry stressing interventions such as selection, training, and performance appraisal. However, some other, more controversial performance enhancement techniques have also been suggested. These techniques range from implicit learning and mental rehearsal to the enhancement of paranormal abilities.

In the mid-1980s, at the request of the Army Research Institute, the National Research Council of the National Academy of Sciences established a blue ribbon panel charged with evaluating the evidence bearing on the effectiveness of a wide variety of techniques for enhancing human performance. This review was conducted under the overall direction of David A. Goslin, then Executive Director of the Commission on Behavioral and Social Sciences and Education (CBASSE), and now President of the American Institutes for Research (AIR). The review panel's report, *Enhancing Human Performance: Issues, Theories, and Techniques,* was published by the National Academy Press in 1988 and summarized by Swets and Bjork (1990). They noted that although the panel found some support for certain alternative performance enhancement techniques — for example, guided imagery — little or no support was found for the usefulness of many other techniques, such as learning during sleep and remote viewing.

Although the findings of the National Research Council (NRC) were predominantly negative with regard to a range of paranormal phenomena, work on remote viewing has continued under the auspices of various government programs. Since 1986, perhaps 50 to 100 additional studies of remote viewing have been conducted. At least some of these studies represent significant attempts to address the methodological problems noted in the review conducted by the NRC panel.

At the request of Congress, the Central Intelligence Agency (CIA) is considering assuming responsibility for this the remote viewing program. As part of its decision-making process, the CIA was asked to evaluate the research conducted since the NRC report. This evaluation was intended to determine: (a) whether this research has any long-term practical

value for the intelligence community, and (b) if it does, what changes should be made in methods and approach to enhance the value of remote viewing research. To achieve these goals, the CIA contracted with the American Institutes for Research to supervise and conduct the evaluation. This report contains the results of our evaluation.

Before presenting our results, we begin by presenting a brief overview of the remote viewing phenomenon and a short history of the applied program that involves remote viewing.

Remote Viewing

Although parapsychological research has a long history, studies of "remote viewing" — also referred to as a form of "anomalous cognition" — as a unique manifestation of psychic functioning began in the 1970s. In its simplest form, a typical remote viewing study during this early period of investigation consisted of the following: A person, referred to as a "beacon" or "sender," travels to a series of remote sites. The remote viewer, a person who putatively has the parapsychological ability, is asked to describe the locations of the beacon. Typically, these location descriptions include drawings and a verbal description of the location. Subsequently, a judge evaluates this description by rank ordering the set of locations against the descriptions. If the judge finds that the viewer's description most closely matched the actual location of the sender, a hit is said to have occurred. If hits occur more often than chance, or if the assigned ranks are more accurate than a random assignment, one might argue that a psychic phenomenon has been observed: the viewer has described a location not visited during the session. This phenomenon has been studied by various investigators throughout the intervening period, using several variants of this basic paradigm.

If certain people (or all people to a greater or lesser extent, as has been proposed by some investigators) possess the ability to see and describe target locations they have not visited, this ability might prove of great value to the intelligence community. As an adjunct method to gathering intelligence, people who possess this ability could be asked to describe various intelligence targets. This information, especially if considered credible and reliable, could supplement and enhance more time-consuming and perhaps dangerous methods for collecting data. Although certain (perhaps unwarranted) assumptions, such as the availability of a sender, are implicit in this argument, the possibility of gathering intelligence through this mechanism has provided the major impetus for government interest in remote viewing.

Remote viewing was and continues to be a controversial phenomenon. Early research on remote viewing was plagued by a number of statistical and methodological flaws.[1] One statistical flaw found in early studies of remote viewing, for example, was due to failure to control for the elimination of locations already judged. For example, if there were five targets in the set, judges might lower their rankings for a viewing already judged as a "hit" or ranked first. In other words, all targets did not have an equal probability of being assigned all ranks. Another commonly noted methodological flaw was that cues in the remote viewing paradigm, such as the time needed to drive to various locations, may have allowed viewers to produce hits without using any parapsychological ability.

More recent research has attempted to control for many of these problems. New paradigms have been developed where, for example, viewers — in double-blind conditions — are asked to visualize pictures drawn from a target pool consisting of National Geographic photographs. In addition to this experimental work, an applied program of intelligence operations actually using remote viewers has been developed. In the following section, we describe the history of the government's remote viewing program.

Program History

"Star Gate" is a Defense Intelligence Agency (DIA) program which involved the use of paranormal phenomena, primarily "remote viewing," for intelligence collection. During Star Gate's history, DIA pursued three basic program objectives: "Operations," using remote viewing to collect intelligence against foreign targets; "Research and Development," using laboratory studies to find new ways to improve remote viewing for use in the intelligence world; and "Foreign Assessment," the analysis of foreign activities to develop or exploit the paranormal for any uses which might affect our national security.

Prior to the advent of Star Gate in the early 1990s, the DIA, the Central Intelligence Agency (CIA), and other government organizations conducted various other programs pursuing some or all of these objectives. CIA's program began in 1972, but was discontinued in 1977. DIA's direct involvement began about 1985 and has continued up to the time of this

[1]Many of these problems are described in the National Research Council Report.

review. During the last twenty years, all government programs involving parapsychology have been viewed as highly controversial, high-risk, and have been subjected to various reviews.

In 1995, the CIA declassified its past parapsychology program efforts in order to facilitate a new, external review. In addition, CIA worked with DIA to continue declassification of Star Gate program documents, a process which had already begun at DIA. All relevant CIA an DIA program documents were collected and inventoried. In June of 1995, CIA's Office of Research and Development (ORD) then contracted with AIR for this external review, based on our long-standing expertise in carrying out studies relating to behavioral science issues and our neutrality with respect to the subject matter.

Evaluation Objectives

The CIA asked AIR to address a number of key objectives during the technical review of Star Gate. These included:

- a comprehensive evaluation of the research and development in this area, with a focus on the validity of the technical approach(es) according to acceptable scientific standards.

- an evaluation of the overall program utility or usefulness to the government.

The CIA believes that the controversial nature of past parapsychology programs within the intelligence community, and the scientific controversy clouding general acceptance of the validity of paranormal phenomena, demand that these two issues of utility and scientific validity be addressed separately.

- consideration of whether any changes in the operational or research and development activities of the program might bring about improved results if the results were not already optimum.

- develop recommendations for the CIA as to appropriate strategies for program activity in the future.

We were directed to base our findings on the data and information provided as a result of DIA and CIA program efforts, since it was neither possible nor intended that we review the entire field of parapsychological research and its applications. Also, we would not review or evaluate the "Foreign Assessment" component of the program.

In the next chapter, we present our methodology for conducting the evaluation. A major component of the evaluation was to commission two nationally-regarded experts to review the program's relevant research studies; their findings are presented in Chapter 3, along with our analysis of areas of agreement and disagreement. In Chapter 4, we present our findings concerning the operational component of the program. Finally, in Chapter 5 we present our conclusions and recommendations.

2. Evaluation Plan

The broad goal of the present effort was to provide a thorough and objective evaluation of the remote viewing program. Because of the multiple components of the program, a multifaceted evaluation plan was devised. As mentioned previously, only the research and intelligence gathering components of the program were considered here. In this section, we describe the general approach used in evaluating these two components of the program, beginning with the research program.

Remote Viewing Research

The Research Program. The government-sponsored research program had three broad objectives. The first and primary objective was to provide scientifically compelling evidence for the existence of the remote viewing phenomenon. It could be argued that if unambiguous evidence for the existence of the phenomenon cannot be provided, then there is little reason to be concerned with its potential applications.

The second objective of the research program was to identify causal mechanisms that might account for or explain the observed (or inferred) phenomenon. This objective of the program is of some importance; an understanding of the origins of a phenomenon provides a basis for developing potential applications. Further, it provides more compelling evidence for the existence of the phenomenon (Cook & Campbell, 1979; James, Muliak, & Brett, 1982). Thus, in conducting a thorough review, an attempt must be made to assess the success of the program in developing an adequate explanation of the phenomenon.

The third objective of the research program was to identify techniques or procedures that might enhance the utility of the information provided by remote viewings. For example, how might more specific information be obtained from viewers and what conditions set boundaries on the accuracy of viewings? Research along those lines is of interest primarily because it provides the background necessary for operational applications of the phenomenon.

The NRC provided a thorough review of the unclassified remote viewing research through 1986. In this review (summarized in Swets & Bjork, 1990), the nature of the research methods led the reviewers to question whether there was indeed any effect that could

clearly be attributed to the operation of paranormal phenomena. Since then, the Principal Investigator, Dr. Edwin May, under formerly classified government contracts, has conducted a number of other studies not previously reviewed. These studies were expressly intended to address many of the criticisms raised in the initial NRC report. Because these studies might provide new evidence for the existence of the remote viewing phenomenon, its causal mechanisms, and its boundary conditions, a new review seemed called for.

The Review Panel. With these issues in mind, a blue-ribbon review panel was commissioned, with the intent of ensuring a balanced and objective appraisal of the research. Two of the reviewers were scientists noted for their interest, expertise, and experience in parapsychological research. The first of these two expert reviewers, Dr. Jessica Utts, a Professor of Statistics at the University of California-Davis, is a nationally-recognized scholar who has made major contributions to the development and application of new statistical methods and techniques. Among many other positions and awards, Dr. Utts is an Associate Editor of the *Journal of the American Statistical Association* (Theory and Methods) and the Statistical Editor of the *Journal of the American Society for Psychical Research.* She has published several articles on the application of statistical methods to parapsychological research and has direct experience with the remote viewing research program.

The second expert reviewer, Dr. Raymond Hyman, is a Professor of Psychology at the University of Oregon. Dr. Hyman has published over 200 articles in professional journals on perception, pattern recognition, creativity, problem solving, and critiques of the paranormal. He served on the original NRC Committee on Techniques for the Enhancement of Human Performance. Dr. Hyman serves as a resource to the media on topics related to the paranormal, and has testified as an expert witness in court cases involving paranormal claims. He is recognized as one of the most important and fair-minded skeptics working in this area. Curriculum Vitae for Dr. Utts and Dr. Hyman are included in Appendix A.

In addition to these two experts, four other scientists were involved in the work of the review panel. Two senior behavioral scientists and experts in research methods at the American Institutes for Research, ***Dr. Michael Mumford*** and ***Dr. Andrew Rose,*** served both as members of and staff to the panel. Dr. Mumford holds a Ph.D. in Industrial/Organizational Psychology from the University of Georgia. He is a Fellow of the American Psychological Association's Division 5, Measurement, Evaluation, and Statistics. Dr. Rose is a cognitive

psychologist with a Ph.D. from the University of Michigan. He has over 22 years of experience in designing and conducting basic and applied behavioral science research. Dr. Rose is Chief Scientist of the Washington Office of AIR. They were to bring to the panel a methodological perspective unbiased by prior work in the area of parapsychology. The third participant was ***Dr. Lincoln Moses,*** an Emeritus Professor of Statistics at Stanford University, who participated in the review as a resource with regard to various statistical issues. Finally, ***Dr. David A. Goslin,*** President of AIR, participated as both a reviewer and coordinator for the review panel.

Research Content. Prior to convening the first meeting of the review panel, the CIA transferred to AIR all reports and documents relevant to the review. We organized and copied these documents. In addition, the Principal Investigator for the program, Dr. Edwin May, was asked to provide two other pieces of information for the panel. First, he was asked to list those studies which he believes provide the strongest evidence bearing on the nature and significance of the remote viewing phenomenon. Second, he was asked to identify all unique studies conducted since the initial NRC report that provide evidence bearing on the nature and significance of the phenomenon. Additionally, he was asked to participate in an interview with members of the review panel following its first meeting to clarify any ambiguities about these studies. The complete list of documents, including notations of the "strongest evidence" set and the "unique" set, is included in Appendix B.[1]

Review Procedures. Remote viewing, like virtually all other parapsychological phenomena, represents one of the most controversial research areas in the social sciences (e.g., Bem & Honorton, 1994; Hyman, 1994). Therefore, any adequate review of the research program must take this controversy into account in such a way that the review procedures are likely to result in a fair and unbiased assessment of the research.

To ensure a fair and comprehensive review, Drs. Utts and Hyman agreed to examine all program documents. In the course of this review it was agreed that all members of the review panel would carefully consider:

[1]One document pertaining to the program remained classified during the period of this review. One of the review panel (Dr. Mumford) examined this document and provided an unclassified synopsis to the review panel.

- those studies recommended by the Principal Investigator as providing compelling evidence for the phenomenon, and

- those empirical studies conducted since the NRC review that might provide new evidence about the existence and nature of the phenomenon.

The members of the review panel convened at the Palo Alto office of AIR to structure exactly how the review process would be carried out. To ensure that different perspectives on paranormal phenomena would be adequately represented, Drs. Utts and Hyman were asked to prepare independent reports based on their review. In this review, they were to cover four general topics:

- Was there a statistically significant effect?

- Could the observed effect, if any, be attributed to a paranormal phenomenon?

- What mechanisms, if any, might plausibly be used to account for any significant effects and what boundary conditions influence these effects?

- What would the findings obtained in these studies indicate about the characteristics and potential applications of information obtained through the remote viewing process?

After they had each completed their reports, they presented the reports to other members of the panel. After studying these reports, all members of the review panel (except Dr. Moses) participated in a series of conference calls. The primary purpose of these exchanges was to identify the conclusions on which the experts agreed and disagreed. Next, in areas where they disagreed, Drs. Utts and Hyman were asked to discuss the nature of the disagreements, determine why they disagreed, and if possible, attempt to resolve the disagreements. Both the initial reports and the dialogue associated with discussion of any disagreements were made a part of the written record. In fact, Dr. Hyman's opinions on areas of agreement and disagreement are included in his report; in addition to her initial report, Dr. Utts prepared a reply to Dr. Hyman's opinions of agreement and disagreement. This reply, in addition to their original reports, are included in Section 3 below.

If disagreements could not be resolved through this dialogue, then the other members of the review panel were to consider the remaining issues from a general methodological perspective. Subsequently, they were to provide an addendum to the dialogue indicating which of the two positions being presented seemed to be on firmer ground both substantively and methodologically. This addendum concludes Section 3 below.

Intelligence Gathering: The Operational Program

The Program. In addition to the research component, the program included two operational components. One of those components was "foreign assessment," or analysis of the paranormal research being conducted by other countries. This issue, however, is beyond the scope of the present review. The other component involved the use of remote viewing as a technique for gathering intelligence information.

In the early 1970s, the CIA experimented with applications of remote viewing in intelligence gathering. Later in the decade, they abandoned the program. However, other government agencies, including the Department of Defense, used remote viewers to obtain intelligence information. The viewers were tasked with providing answers to questions posed by various intelligence agencies. These operations continued until the Spring of 1995, when the program was suspended.

Although procedures varied somewhat during the history of the program, viewers typically were presented with a request for information about a target of interest to a particular agency. Multiple viewings were then obtained for the target. The results of the viewings then were summarized in a three- or four-page report and sent to the agency that had posed the original question. Starting in 1994, members of the agencies receiving the viewing reports were formally asked to evaluate their accuracy and value.

Any comprehensive evaluation of the remote viewing program must consider how viewings were used by the intelligence community. One might demonstrate the existence of a statistically significant paranormal phenomenon in experiments conducted in the laboratory; however, the phenomenon could prove to be of limited operational value either because it does not occur consistently outside the laboratory setting or because the kind of information

provided is of limited value to the intelligence community.

General Evaluation Procedures. No one piece of evidence provides unequivocal support for the usefulness of a program. Instead, a more accurate and comprehensive picture can be obtained by considering multiple sources of evidence (Messick, 1989). Three basic sources of information were used in evaluation of the intelligence gathering component:.

- Prior research studies
- Interviews with program participants, and
- Analyses of user feedback.

Prior Research Studies. As noted above, one aspect of the laboratory research program was to identify those conditions that set bounds on the accuracy and success of the remote viewing process. Thus, one way to analytically evaluate potential applications in intelligence gathering is to enumerate the conditions under which viewers were assigned tasks and then examine the characteristics of the remote viewing paradigm as studied through experimentation in the laboratory. The conditions under which operational tasks occur — that is, the requirements imposed by intelligence gathering — could then provide an assessment of the applicability of the remote viewing process.

Interviews. As part of the Star Gate program, the services of remote viewers were used to support operational activities in the intelligence community. This operational history provides an additional basis for evaluating the Star Gate program; ultimately, if the program is to be of any real value, it must be capable of serving the needs of the intelligence community. By examining how the remote viewing services have been used, it becomes possible to draw some initial, tentative conclusions about the potential value of the Star Gate program. Below, we describe how information bearing on intelligence applications of the remote viewing phenomenon was gathered. Later, in Section 4, we describe the results of this information-gathering activity and draw some conclusions from the information we obtained.

Although a variety of techniques might be used to accrue retrospective information (questionnaires, interviews, diaries, etc.), the project team decided that structured interviews

examining issues relevant to the various participants would provide the most appropriate strategy. Accordingly, structured interviews were developed for three participant groups in intelligence operations:

- end-users: representatives from agencies requesting information from remote viewers,
- the Program Manager, and
- the remote viewers.

Another key issue to be considered in an interview procedure is the nature of the people to be interviewed. Although end-users, program managers, and viewers represent the major participants, many different individuals have been involved in intelligence applications of remote viewing over the course of the last twenty years. Nevertheless, it was decided to interview only those persons who were involved in the program at the time of its suspension in the Spring of 1995. This decision was based on the need for accurate, current information that had not been distorted by time and could be corroborated by existing documentation and follow-up interviews.

Information about operational applications was gathered in a series of interviews conducted during July and August of 1995. We interviewed seven representatives of end-user groups, three remote viewers, and the incumbent Program Manager. With regard to the data collection procedures that we employed, a number of points should be borne in mind. First, members of the groups we interviewed could only speak to recent operations. Although it would have been desirable to interview people involved in earlier operations, for example during the 1970s, the problems associated with the passage of time, including forgetting and the difficulties involved in verifying information, effectively precluded this approach. Accordingly, the interviews focused on current operations.

Second, it should be noted that the end-user representatives represented a range of current concerns in the intelligence community. The relevant user groups were involved in operations ranging from counterintelligence and drug interdiction to search and rescue operations. This diversity permitted operational merits to be assessed for a number of different contexts.

The interviews were conducted by one of the two panel members from AIR. A retired intelligence officer took notes during the interviews. A representative of the CIA attended interviews as necessary to describe the reasons the interviews were being conducted and to address any security concerns.

Each interview was conducted using a standard protocol. Different protocols were developed for members of the three groups because they had somewhat different perspectives on current operations. Appendix C presents the instructions given to the interviewer. This Appendix also lists the interview questions presented to users, viewers, and the program manager. User interviews were conducted in the offices of the client organization; interviews with the program manager and the viewers were conducted at the Washington Office of AIR. The interviews were one to two hours long. A total of 12 to 16 questions were asked in the interviews.

We developed the questions presented in each interview as follows: Initially, the literature on remote viewing and available information bearing on operations within the intelligence community were reviewed by AIR scientists. This review was used to formulate an initial set of interview questions. Subsequently, these candidate questions were presented to a panel of three psychologists at AIR. In addition, review panel members were asked to review these candidate questions to insure they were not leading and covered the issues that were relevant to the particular group under consideration.

With regard to operational users, four types of questions were asked. These four types of questions examined the background and nature of the tasks presented to the remote viewers, the nature and accuracy of the information resulting from the viewings, operational use of this information, and the utility of the resulting information.

The remote viewers were asked a somewhat different set of questions. The four types of questions presented to them examined recruitment, selection, and development; the procedures used to generate viewings; the conditions that influenced the nature and success of viewings; and the organizational factors that influenced program operations.

The Program Manager was not asked about the viewing process. Instead, questions presented to the program manager primarily focused on broader organizational issues. The

four types of managerial questions focused on the manager's background, client recruitment, factors influencing successes and failures, and needs for effective program management.

The interview questions presented in each protocol were asked in order, as specified in Appendix D. Typically these interviews began by asking for objective background information. Questions examining broader evaluative issues were asked at the end of the interview. The AIR scientist conducting the interviews produced reports for each individual interview. They also are contained in Appendix C.

Analyses of User Feedback. In addition to the qualitative data provided by the interviews, some quantitative information was available. For all of the operational tasks conducted during 1994, representatives from the requesting agencies were asked to provide two summary judgments: one with respect to the accuracy of the remote viewing, and the second of the actual or potential value of the information provided. These data — the accuracy and value evaluations obtained for viewings as program feedback from the users — were analyzed and summarized in a report prepared prior to the current evaluation. A copy of this report is provided in Appendix D. Although these judgments have been routinely collected for only a relatively short period of time, they provided an important additional source of evaluative information. This information was of some value as a supplement to interviews in part because it was collected prior to the start of the current review, and in part because it reflects user assessments of the resulting information.

We present the findings flowing from this multifaceted evaluation of the operational component of the program in Section 4 of this report. In that section, we first present the findings emerging from prior research and the interviews and then consider the results obtained from the more quantitative evaluations. Prior to turning to this evaluation of operations, however, we first present the findings from review of the basic research, examining evidence for the existence and nature of the remote viewing phenomenon.

3. Research Reviews

In this section, we present the conclusions drawn by the two experts after reviewing the research studies bearing on remote viewing. We begin by presenting the review of Dr. Jessica Utts. Subsequently, a rejoinder is provided by Dr. Raymond Hyman. Finally, Dr. Utts presents a reply to Dr. Hyman. The major points of agreement and disagreement are noted in the final section, along with our conclusions.

In conducting their reviews, both Dr. Hyman and Dr. Utts focused on the remote viewing research. However, additional material is provided as indicated by the need to clarify certain points being made. Furthermore, both reviewers provided unusually comprehensive reviews considering not only classified program research, but also a number of earlier studies having direct bearing on the nature and significance of the phenomenon.

An Assessment of the Evidence for Psychic Functioning

Dr. Jessica Utts
Division of Statistics, University of California, Davis
September 1, 1995

ABSTRACT

Research on psychic functioning, conducted over a two decade period, is examined to determine whether or not the phenomenon has been scientifically established. A secondary question is whether or not 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, known as 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 similar magnitude to those found in government-sponsored research at SRI and SAIC have been replicated at a number of laboratories across 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 medium effect. That means that it is 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 applied psychic functioning. For instance, it doesn't appear that a sender is needed. Precognition, in which the answer is known to no one until a future time, appears to work quite well. Recent experiments suggest that if there is a psychic sense then it works much like our other five senses, by detecting change. Given that physicists are currently grappling with an understanding of time, it may be that a psychic sense exists that scans the future for major change, much as our eyes scan the environment for visual change or out ears allow us to respond to sudden changes in sound.

It is recommended that future experiments focus on understanding how this phenomenon works, and on how to make it as useful as possible. There is little benefit to 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 or not psychic functioning is possible. Secondary questions include whether or not such functioning can be used productively for government purposes, and whether or not 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 or not there are any methodological problems that could substantially alter the results. It is too often the case that 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 a careful reader to reach his or her own conclusions based on that data. The tools consist of a rudimentary overview 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 was in existence 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, done by 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 an Appendix); Section 5 is concerned with external validation by exploring related results from other laboratories; Section 6 includes a discussion of the usefulness of this capability for government purposes and 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 manipulates 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 only been obtained 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 isn't known until later in time. It is more 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 see the correct answer at some point in the future. These distinctions are important in the quest to identify an explanation for anomalous cognition, but do not bear on the existence issue.

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 simply asked to describe the target. In contrast, a **forced-choice** experiment is one 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 done at SAIC addresses 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

someone will hit or when they will score a home run. In fact, we cannot even predict whether or not a horse run will occur in a particular game. That does **not** mean that home runs don't 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 five heads and five tails over each set of ten 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 given players of similar ability 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 percent 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 *percent over the long run,* then the claim would have been supported by statistical evidence.

Statisticians have developed numerical methods for comparing results to what is expected by chance. Upon observing the results of an experiment, 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 percent (0.05) or less, then 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 percent of the time. For this reason and others, most reasonable scientists require replication of non-chance 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 convincingly rule out chance. In a very small study, it would take a huge effect to convincingly rule out chance.

In our hypothetical sex-determination experiment, suppose 70 out of 100 births designed to be boys actually resulted in boys, for a rate of 70 percent instead of the 51 percent expected by chance. The experiment would have a *p-value of* 0.0001, quite convincingly ruling out chance. Now suppose someone attempted to replicate the experiment with only ten births and found 7 boys, i.e also 70 percent. The smaller experiment would have a *p-value of* 0. 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 percent boys! In only ten births it would require 90 percent of them to be boys before chance could be ruled out. Yet the 70 percent rate is a more exact replication of the result than the 90 percent.

Therefore, while *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 1. 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, while 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?"

In order 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 one means it matches the response most closely, and a rank of five 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 three. Evidence for anomalous cognition occurs when the average rank over a series of trials is significantly lower than three. (Notice that a rank of one 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 one 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 *National Geographic* 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 1, the effect size for a series of remote viewings using rank-order judging with five choices is (3.0 - average rank). Therefore, small, medium and large effect sizes (0. 1, 0. 5 and 0. 8) correspond to average ranks of 2.86, 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 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 2.

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:

- 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 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.

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

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

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

- 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.

- Reasons, if any, for excluding data must be defined in advance and followed consistently, and 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.

- Statistical analyses to be used must be planned in advance of collecting the data so that a method most favorable to the data isn't 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, non-laboratory 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 requester 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, Dr. May concluded that in one of the cases the remote viewer was able to describe a microwave generator with 80 percent accuracy, and that of what he said almost 70 percent 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 can't 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 the 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 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 in 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.

3.1 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.

One of the apparent successes concerned the "West Virginia Site" in which two remote viewers purportedly identified an underground secret facility. One of them apparently named codewords and personnel in this facility accurately enough that it set off a security investigation to determine how that information could have been leaked. Based only on the coordinates of the site, the viewer first described the above ground terrain, 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 if 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 like a xerox machine to natural settings, like a swimming pool. But these and some of the subsequent experiments were criticized on statistical and methodological grounds; we briefly describe one of the experiments and criticisms of it to show the kinds of problems that existed in the early scientific effort.

The largest series during in 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, like 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 influence the results. The criticisms of these and subsequent experiments, while 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 a thousand were laboratory remote viewings. There were a total of 227 subjects in all experiments.

The statistical results were so overwhelming that results that extreme or more so would occur only about once in every 10^{20} such instances if chance alone is the explanation (i.e., the p-*value* was less than 10^{-}'). 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 Dr. 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 was a group of 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 one percent 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 or not 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 no evidence to support anomalous perturbation (psychokinesis), i.e. 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 and 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 .052. The comparison is not ideal since 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 shown to be 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, while 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 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, unlike for the set conducted at SRI. Whatever details may be missing from the written reports are obtainable from the principal investigator, Dr. Edwin May, to whom we have been given unlimited access.

In a memorandum dated July 25, 1995, Dr. Edwin May listed the set of experiments conducted by SAIC. There were ten 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 similar format to the remote viewing experiments conducted at SRI and we can examine those to see whether or not 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 TEN EXPERIMENTS

Of the ten 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, Appendix 2 gives a brief description of the experiments. 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 25 July 1995, Dr. May provided the review team with details of the ten experiments, including a short title, number of trials, effect size and overall p-value for each one. His list was in time sequence. It is reproduced in Table 1, using his numbering system, with the experiments categorized by type, then sequentially within type. The effect

size estimates are based on a limited number of trials, so they are augmented with an interval to show the probable range of the true effect (e.g. .124+.071 indicates a range from .053 to .195). Remember that an effect size of 0 represents chance, while a positive effect size indicates positive results.

TABLE 1: SAIC EXPERIMENTS LISTED BY DR. EDWIN MAY				
Expr	Title	Trials	Effect Size	*p-value*
Remote Viewing Experiments				
1	Target dependencies	200	.124±.071	0.040
4	AC with binary coding	40	-.067±.158	0.664
5	AC lucid dreams, base	24	.088±.204	0.333
6	AC lucid dreams, pilot	21	.368±.218	0.046
9	ERD AC Behavior	70	.303±.120	0.006
10	Entropy II	90	.550±.105	$9.1X10^{-8}$
Other Experiments				
2	AC of binary targets	300	.123±.058	0.017
3	MEG Replication	12,000s	MCE	MCE
7	Remote observation	48	.361±.144	0.006
8	ERD EEG investigation	7,000s	MCE	MCE

4.3 ASSESSING THE REMOTE VIEWING EXPERIMENTS BY HOMOGENEOUS SETS OF SESSIONS

While Table I 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 2. Overall effect sizes for each viewer and total effect sizes for each experiment are weighted according to the number of trials, so each trial receives equal weight.

TABLE 2: INDIVIDUAL EFFECT SIZES

Experiment	Experiment Remote Viewers					Viewer	
	009	131	372	389	518	Unknown /Other	Total
	Static Targets (*National Geographic*)						
E1: Static	.424	-.071	.424	.177	.283	n.a.	.247
E9	.432	n.a.	.354	.177	n.a.	n.a.	.303
E10: Static	.566	n.a.	.801	-.071	.778	n.a.	.550
E5 (Note 1)	n.a.	n.a.	n.a.	n.a.	n.a.	.088	.088
E6 (Note 2)	n.a.	n.a.	n.a.	n.a.	n.a.	.370	.370
E4 (Note 3)	-.112	n.a.	0	n.a.	n.a.	-.559	-.067
	Dynamic Targets (Video Film Clips)						
E1: Dynamic	0	.354	-.283	0	-.071	n.a.	.000
E10: Dynamic	.919	n.a.	.754	0	.424	n.a.	.550
Overall	.352	.141	.340	.090	.271	n.a.	

1. Experiment 5 did not include any expert viewers.
2. Experiment 6 included 4 expert viewers but separate results were not provided.
3. Experiment 4 used a specially designed target set and only 4 choices in judging.

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 ten times, or see the same batting averages in every game.

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 and Piantineda) 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 .363. None of the other six SRI experts were participants.

The same subject identifying numbers were used at SAIC, 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 Experiments I 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 .35 and .34, respectively, and the .35 effect size for Viewer 009 was very similar to his .363 effect size in the report by May, Lantz and Piantineda (1994). Therefore, we see a repeated and, more importantly, hopefully a 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-value* = .50) and the static targets did *(P-value* = .0073). Therefore, a methodological problem would have had to differentially effect results 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 Dr. 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). Dr. Nevin Lantz, who resided in Pennsylvania, was the principal investigator. After each session, viewers faxed their response to Dr. Lantz and mailed the original to SAIC. Upon receipt of the fax, Dr. Lantz mailed the correct answer to the viewer. The viewers were supposed to mail their original responses to SAIC immediately, after taxing them to Dr. Lantz. According to Dr. May, the taxed versions were later compared with the originals to make sure the originals were sent without any changes. Here are how the other methodological issues in Section 2.3 were handled:

- *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 Dr. Lantz's home in Pennsylvania.

- *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.*

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

- *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 Dr. 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.

- *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.

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

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

- *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, ten under each of four conditions (all combinations of sender/no sender and static/dynamic). All sessions were completed.

Reasons, if any, for excluding data must be defined in advance and followed consistently, and 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.

- *Statistical analyses to be used must be planned in advance of collecting the data so that a method most favorable to the data isn't 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 or not psychic functioning is possible, the experiments at SAIC were designed to explore a number of hypotheses. Experiments I and 10 were both designed to see if there is a relationship between the "change in visual entropy" in the targets and the remote viewing performance.

Each of the five senses 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 or not 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 the article "Shannon Entropy: A Possible Intrinsic Target Property" by May, Spottiswoode and James, in the *Journal of Parapsychology,* December 1994. It was indeed found that there was a correlation between the change in entropy in the target and the remote viewing quality. This result was initially shown in Experiment I and replicated in Experiment 10. A simulation study matching randomly chosen targets to response 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 -cannot rule out precognition as being 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 at 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 "differentiable 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 National Geographic* 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 2, 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 percent. 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 Sept. 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 Psychophysical Research Laboratories (PRL) were already 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 places matches resulted from a series. Four rather than five choices are given, 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 PRL, Bem and Honorton (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 percent when 25 percent 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 percent. The hit rates for the four individual experiments were 34.3 percent, 37.5 percent, 40 percent and 36.1 percent, 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 percent. 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 percent when there was a known sender and when there was no sender, and 28 percent when there was a sender but the receiver did not know whether or not 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 percent hits.

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 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	.209
All Remote Viewing at SAIC	455	N/A	.230
PRL, Princeton, NJ	329	32 percent	.167
University of Amsterdam, Netherlands	124	37 percent	.261
University of Edinburgh, Scotland	97	33 percent	.177
Institute of Parapsychology, NC	100	33 percent	.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 or not 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. Given our current level of understanding, it is rarely 100 percent 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; 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.

8. REFERENCES

Bem, Daryl J. and Charles Honorton (1994). "Does psi exist? Replicable evidence for an anomalous process of information transfer," *Psychological Bulletin,* 115, 4-18.

Bierman, Dick J. (1995). "The Amsterdam Ganzfeld Series III & IV: Target clip emotionality, effect sizes and openness," *Proceedings of the 38th Annual Parapsychological Association Convention, 27-37.*

Broughton, Richard and Cheryl Alexander (1995). "Autoganzfeld II: The first 100 sessions," *Proceedings of the 38th Annual Parapsychological Association Convention, 53-61.*

May, Edwin C. (1995). "AC Technical trials: Inspiration for the target entropy concept," May 26, 1995, SAIC Technical Report.

May, Edwin C., Nevin D. Lantz and Tom Piantineda (1994). "Feedback Considerations in Anomalous Cognition Experiments," Technical Report, 29 November 1994.

May, Edwin C., J.M. Utts, V.V. Trask, W.W. Luke, T.J. Frivold and B.S. Humphrey (1988). "Review of the psychoenergetic research conducted at SRI International (1973-1988)"
SRI International Technical Report, March 1989.

Morris, Robert L., Kathy Dalton, Deborah Delanoy and Caroline Watt (1995). "Comparison of the sender/no sender condition in the Ganzfeld, " *Proceedings of the 38th Annual Parapsychological Association Convention, 244-259.*

Puthoff, Harold E. and Russell Targ (1975). "Perceptual Augmentation Techniques: Part Two--Research Report," Stanford Research Institute Final Report, Dec. 1, 1975.

APPENDIX 1

EFFECT SIZE MEASURE USED WITH RANK ORDER JUDGING

In general, effect sizes measure the number of standard deviation 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, etc., rather than being the standard error of the sample statistic used in the hypothesis test.

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

$$E(R_{ii}) = (N + 1)/2$$

and

$$Var(R_1) = \frac{N^2 - 1}{12}$$

Therefore, when N = 5, we find *E(Rd* 3 *and Var(Rd* = 2. The effect size is therefore:

$$EffectSize \quad \frac{(3.0 - Average\ Rank)}{F2}$$

APPENDIX 2

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 or not a sender is necessary for successful remote viewing and whether or not dynamic targets, consisting of short video clips, would result in more successful remote viewing than the standard *National Geographic* photographs used in most of the SRI experiments.

METHOD: Five experienced remote viewers participated, three of whom (#s 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 or not there was a sender was randomly determined and unknown to the viewer. Upon receiving the fax of the response, Dr. 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 Dr. 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 ten 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 = .121, p = .08) between the static and dynamic targets, with the traditional *National Geographic* photographs faring better than the dynamic video clips. There was no noticeable difference based on whether or not 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 .248 (exact p = .0073) while 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 if 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, 01 1 10, 10101, and I 101 1. 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 corrections." Different sets of five features were used for each of ten target packs.

METHOD: Five viewers each contributed eight trials, but the same eight targets were used for all five viewers. There was no sender used, 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 by both the viewers and 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 if 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 National Geographic* targets used at SRI and SAIC. The target was placed on a table (so 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 .265, although the result was not statistically significant due to the small number of trials (8). 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 Dr. Stephen LaBerge of the Lucidity Institute. He was the Principal Investigator for this experiment. The experiment was designed to see if 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 or not 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.

Unlike with standard well-controlled protocols, the viewers were allowed to take the target material home with them. The targets, selected from the standard *National Geographic pool,* 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 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 or not 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 (#s 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 the 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 0.303 (p = 0.006).

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 bandwidth" 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 *National Geographic* 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 in Section 4.6.1.

METHOD: Four of the five viewers from Experiment 1 were used (#s 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 I in which the viewers worked at home. Viewer #372 contributed 15 of each type while 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 .55, resulting in a z-score of 5.22.

TABLE 4: RESULTS FOR EXPERIMENT 10						
	Static Targets			Dynamic Targets		
Viewer	Rank	ES	p	Rank	ES	p
009	2.20	.565	.037	1.70	.919	$1.8X10^{-3}$
372	1.87	.801	$9.7X10^{-4}$	1.93	.754	$1.8X10^{-3}$
389	3.10	-.071	.589	3.0	.000	.500
518	1.90	.778	$7.2X10^{-3}$	2.4	.424	.091
Total	2.22	.550	$1.1X10^{-5}$	2.22	.550	$1.1X10^{-5}$

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 magnetoenchephalograph (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 I 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 if humans can correctly guess 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 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 zero and one 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 (#s 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 .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, i.e. over 200 guesses for each decision. Of the individual guesses, only 51.6 percent were correct, for an effect size of .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 percent of the 301 decisions were correct and the effect size was 0. 123. The combined results were still statistically significant, with p = 0. 0 17, 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 or not 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 (FSU), the United States and Scotland. The experiments in the FSU 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.

RESULTS: To determine whether or not 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 .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.

Approved For Release 2002/05/22 : CIA-RDP96-00791R000200180005-5

Evaluation of Program on "Anomalous Mental Phenomena"

Ray Hyman
University of Oregon
Department of Psychology
Eugene, Oregon
September 11, 1995

INTRODUCTION

Professor Jessica Utts and I were given the task of evaluating the program on "Anomalous Mental Phenomena" carried out at SRI International (formerly the Stanford Research Institute) from 1973 through 1989 and continued at SAIC (Science Applications International Corporation) from 1992 through 1994. We were asked to evaluate this research in terms of its scientific value. We were also asked to comment on its potential utility for intelligence applications.

The investigators use the term "Anomalous Mental Phenomena" to refer to what the parapsychologists label *as psi. Psi* includes both extrasensory perception (called *Anomalous Cognition* by the present investigators) and psychokinesis (called *Anomalous Perturbation* by the present investigators). The experimenters claim that their results support the existence of Anomalous Cognition--especially clairvoyance (information transmission from a target without the intervention of a human sender) and precognition. They found no evidence for the existence of Anomalous Perturbation.

Our evaluation will focus on the 10 experiments conducted at SAIC. These are the most recent in the program as well as the only ones for which we have adequate documentation. The earlier SRI research on remote viewing suffered from methodological inadequacies. Another reason for concentrating upon this more recent set of experiments is the limited time frame allotted for this evaluation.

I will not ignore entirely the earlier SRI research. I will also consider some of the contemporary research in parapsychology at other laboratories. This is because a proper scientific evaluation of any research program has to place it in the context of the broader scientific community. In addition, some of this contemporary research was subcontracted by the SAIC investigators.

Professor Utts has provided an historical overview of the SRI and SAIC programs as well as descriptions of the experiments under consideration. I will not duplicate what she has written on these topics. Instead, I will focus on her conclusions that:

> *Using the standards applied to any other area of science, it is concluded that psychic functioning has been well established [Utts, Sept 1995, p I]*

> *Arguments that these results could be due to methodological flaws in the experiments are soundly refuted Effects of similar magnitude to those found in government-sponsored research at SRI and SAIC have been replicated at a number of laboratories across .he world. Such consistency cannot be readily explained by claims of flaws or fraud* [Utts, Sept 1995, p I]

Because my report will emphasize points of disagreement between Professor Utts and me, I want to state that we agree on many other points. We both agree that the SAIC experiments were free of the methodological weaknesses that plagued the early SRI research. We also agree that the SAIC experiments appear to be free of the more obvious and better known flaws that can invalidate the results of parapsychological investigations. We agree that the effect sizes reported in the SAIC experiments are too large and consistent to be dismissed as statistical flukes.

I also believe that Jessica Utts and I agree on what the next steps should be.

We disagree on key questions such as:

1. Do these apparently non-chance effects justify concluding that the existence of anomalous cognition has been established?

2. Has the possibility of methodological flaws been completely eliminated?

3. Are the SAIC results consistent with the contemporary findings in other parapsychological laboratories on remote viewing and the ganzfeld phenomenon?

The remainder of this report will try to justify why I believe the answer to these three questions is "no".

SCIENTIFIC STATUS OF THE PROGRAM

Science is basically a *communal* activity. For any developed field of inquiry, a community of experts exist. This community provides the *disciplinary matrix* which determines what questions are worth asking, which issues are relevant, what variables matter and which can be safely ignored, and the criteria for judging the adequacy of observational data. The community provides checks and balances through the referee system, open criticism, and independent replications. Only those relationships that are reasonably lawful and replicable across independent laboratories become part of the shared scientific store of "knowledge."

An individual investigator or laboratory can contribute to this store. However, by itself, the output of a single investigator or laboratory does not constitute science. No matter how careful and competent the research, the findings of a single laboratory count for nothing unless they can be reliably replicated in other laboratories. This rule is true of ordinary claims. It holds true especially for claims that add something new or novel to the existing database. When an investigator, for example, announces the discovery of a new element, the claim is not accepted until the finding has been successfully replicated by several independent laboratories. Of course, this rule is enforced even more when the claim has revolutionary implications that challenge the fundamental principles underlying most sciences.

GENERAL SCIENTIFIC HANDICAPS OF THE SAIC PROGRAM

The brief characterization of scientific inquiry in the preceding section alerts us to serious problems in trying to assess the scientific status of the SAIC research. The secrecy under which the SRI and SAIC programs was conducted necessarily cut them off from the communal aspects of scientific inquiry. The checks and balances that come from being an open part of the disciplinary matrix were absent. With the exception of the past year or so, none of the reports went through the all-important peer-review system. Worse, promising findings did not have the opportunity of being replicated in other laboratories.

The commendable improvements in protocols, methodology, and data-gathering have not profited from the general shake-down and debugging that comes mainly from other laboratories trying to use the same improvements. Although the research program that started in 1973 continued for over twenty years, the secrecy and other constraints have produced only ten adequate experiments for consideration. Unfortunately, ten experiments--especially from one laboratory (considering the SAIC program as a continuation of the SRI program)--is far too few to establish reliable relationships in almost any area of inquiry. In the traditionally elusive quest for psi, ten experiments from one laboratory promise very little in the way of useful conclusions.

The ten SAIC experiments suffer another handicap in their quest for scientific status. The principal investigator was not free to run the program to maximize scientific payoff. Instead, he had to do experiments and add variables to suit the desires of his sponsors. The result was an attempt to explore too many questions with too few resources. In other words, the scientific inquiry was spread too thin. The 10 experiments were asked to provide too many sorts of information.

For these reasons, even before we get to the details (and remember the devil is usually in the details), the scientific contribution of this set of studies will necessarily be limited.

PARAPSYCHOLOGY'S STATUS AS A SCIENCE

Parapsychology began its quest for scientific status in the mid-1800s. At that time it was known as Psychical research. The Society for Psychical Research was founded in London in 1882. Since that time, many investigators--including at least four Nobel laureates--have tried to establish parapsychology as a legitimate science. Beginning in the early 1930s, J.B. Rhine initiated an impressive program to distance parapsychology from its tainted beginnings in spiritualistic seances and turn it into an experimental science. He pulled together various ideas of his predecessors in an attempt to make the study of ESP and PK a rigorous discipline based on careful controls and statistical analysis.

His first major publication caught the attention of the scientific community. Many were impressed with this display of a huge database, gathered under controlled conditions, and analyzed with the most modem statistical tools. Critics quickly attacked the statistical basis of the research. However, Burton Camp, the president of the Institute of Mathematical Statistics, came to the parapsychologists' defense in 1937. He issued a statement that if the critics were going to fault parapsychological research they could not do so on statistical grounds. The critics then turned their attention to methodological weaknesses. Here they had more success.

What really turned scientists against parapsychological claims, however, was the fact that several scientists failed to replicate Rhine's results. This problem of replicability has plagued parapsychology ever since. The few, but well-publicized, cheating scandals that were uncovered also worked against parapsychology's acceptance into the general scientific community.

Parapsychology shares with other sciences a number of features. The database comes from experiments using controlled procedures, double-blind techniques where applicable, the latest and most sophisticated apparatus, and sophisticated statistical analysis. In addition, the findings are reported at annual meetings and in refereed journals.

Unfortunately, as I have pointed out elsewhere, parapsychology has other characteristics that make its status as a normal science problematic. Here I will list only a few. These are worth mentioning because they impinge upon the assessment of the scientific status of the SAIC program. Probably the most frequently discussed problem is the issue of replicability. Both critics and parapsychologists have agreed that the lack of consistently replicable results has

been a major reason for parapsychology's failure to achieve acceptance by the scientific establishment.

Some parapsychologists have urged their colleagues to refrain from demanding such acceptance until they can put examples of replicable experiments before the scientific community. The late parapsychologist J.G.Pratt went further and argued that parapsychology would never develop a replicable experiment. He argued that psi was real but would forever elude deliberate control. More recently, the late Honorton claimed that the ganzfeld experiments had, indeed, achieved the status of a replicable paradigm. The title of the landmark paper in the January 1994 issue of the *Psychological Bulletin* by Bem and Honorton is "Does psi exist? Replicable evidence for an anomalous process of information transfer.?' In her position paper "Replication and meta-analysis in parapsychology" *(Statistical Science, 1991, 6,* pp. 363-403), Jessica Utts reviews the evidence from meat-analyses of parapsychological research to argue that replication has been demonstrated and "that the overall evidence indicates that there is an anomalous effect in need of explanation."

In evaluating the SAIC research, Utts points to the consistency of effect sizes produced by the expert viewers across experiments as well as the apparent consistency of average effect sizes of the SRI and SAIC experiments with those from other parapsychological laboratories. These consistencies in effect sizes across experiments and laboratories, in her opinion, justify the claim that anomalous mental phenomena can be reliably replicated with appropriately designed experiments. This is an important breakthrough for parapsychology, if it is true. However, to anticipate some of my later commentary, I wish to emphasize that simply replicating effect size is not the same thing as showing the repeated occurrence of anomalous mental phenomena. Effect size is nothing more than a standardized difference between an observed and an expected outcome hypothesized on the basis of an idealized probability model. An indefinite number of factors can cause departures from the idealized probability model. An investigator needs to go well beyond the mere demonstration that effect sizes are the same before he/she can legitimately claim that they are caused by the same underlying phenomenon.

In my opinion, a more serious challenge to parapsychology's quest for scientific status is the lack of cumulativeness in its database. Only parapsychology, among the fields of inquiry

claiming scientific status, lacks a cumulative database. Physics has changed dramatically since Newton conducted his famous experiment using prisms to show that white light contained all the colors of the spectrum. Yet, Newton's experiment is still valid and still yields the same results. Psychology has changed its ideas about the nature of memory since Ebbinghaus conducted his famous experiments on the curve of forgetting in the 1880s. We believe that memory is more dynamic and complicated than can be captured by Ebbinghaus' ideas about a passive, rote memory system. Nevertheless, his findings still can be replicated and they form an important part of our database on memory.

Parapsychology, unlike the other sciences, has a shifting database. Experimental data that one generation puts forth as rock-solid evidence for psi is discarded by later generations in favor of new data. When the Society for Psychical Research was founded in 1882, its first president Henry Sidgwick, pointed to the experiments with the Creery sisters as the evidence that should convince even the most hardened skeptic of the reality of psi. Soon, he and the other members of the Society argued that the data from Smith-Blackburn experiments provided the fraud-proof case for the reality of telepathy. The next generation of Psychical researchers, however, cast aside these cases as defective and we no longer hear about them. Instead, they turned to new data to argue their case.

During the 1930s and 1940s, the results of Rhine's card guessing experiments were offered as the solid evidence for the reality of psi. The next generation dropped Rhine's data as being flawed and difficult to replicate and it hailed the Soal-Goldney experiments as the replicable and rock-solid basis for the existence of telepathy. Next came the Sheep-Goat experiments. Today, the Rhine data, the Sheep-Goats experiments, and the Soal-Goldney experiments no longer are used to argue the case for psi. Contemporary parapsychologists, instead, point to the ganzfeld experiments, the random-number generator experiments, and--with the declassifying of the SAIC experiments--the remote viewing experiments as their basis for insisting that psi exists.

Professor Utts uses the ganzfeld data and the SAIC remote viewing results to assert that the existence of anomalous cognition has been proven. She does not completely discard earlier data. She cites meat-analyses of some of the earlier parapsychology experiments. Still, the cumulative database for anomalous mental phenomena does not exist. Most of the data accumulated by previous investigators has been discarded. In most cases the data have been

discarded for good reasons. They were subsequently discovered to be seriously flawed in one or more ways that was not recognized by the original investigators. Yet, at the time they were part of the database, the parapsychologists were certain that they offered incontestable evidence for the reality of psi.

How does this discussion relate to our present concerns with the scientific status of the SAIC program? This consideration of the shifting database of parapsychology offers a cautionary note to the use of contemporary research on the ganzfeld and remote viewing as solid evidence for anomalous mental phenomena. More than a century of parapsychological research teaches us that each generation of investigators was sure that it had found the 'Holy Grail'--the indisputable evidence for psychic functioning. Each subsequent generation has abandoned their predecessors' evidence as defective in one way or another. Instead, the new generation had its own version of the holy grail.

Today, the parapsychologists offer us the ganzfeld experiments and, along with Jessica Utts, will presumably will include the SAIC remote viewing experiments as today's reasons for concluding that anomalous cognition has been demonstrated. Maybe this generation is correct. Maybe, this time the "indisputable" evidence will remain indisputable for subsequent generations. However, it is too soon to tell. Only history will reveal the answer. As E.G.Boring once wrote, when writing about the Soal-Goldney experiments, you cannot hurry history.

Meanwhile, as I will point out later in this report, there are hints and suggestions that history may repeat itself Where Utts sees consistency and incontestable proof, I see inconsistency and hints that all is not as rock-solid as she implies.

I can list other reasons to suggest that parapsychology's status as a science is shaky, at best. Some of these reasons will emerge as I discuss specific aspects of the SAIC results and their relation to other contemporary parapsychological research.

THE CLAIM THAT ANOMALOUS COGNITION EXISTS

Professor Utts concludes "that psychic functioning has been well established." She bases this conclusion on three other claims: 1) the statistical results of the SAIC and other parapsychological experiments "are far beyond what is expected by chance", 2) "arguments that these results could be due to methodological flaws are soundly refuted"; and 3) "Effects of similar magnitude to those found in government-sponsored research at SRI and SAIC have been replicated at a number of laboratories across the world.

Later, in this report, I will raise questions about her major conclusion and the three supporting claims. In this section, I want to unpack just what these claims entail. I will start with the statistical findings. Parapsychological is unique among the sciences in relying solely on significant departures from a chance baseline to establish the presence of its alleged phenomenon. In the other sciences the defining phenomena can be reliably observed and do not require indirect statistical measures to justify their existence. Indeed, each branch of science began with phenomena that could be observed directly. Gilbert began the study of magnetism by systematically studying a phenomenon that had been observed and was known to the ancients as well as his contemporaries. Modern physics began by becoming more systematic about moving objects and falling bodies. Psychology became a systematic science by looking for lawful relationships among sensory discriminations. Another starting point was the discovery of lawful relationships in the remembering and forgetting of verbal materials. Note that in none of these cases was the existence of the defining phenomena in question. No one required statistical tests and effect sizes to decide if magnetism was present or if a body had fallen. Psychophysicists did not need to reject a null hypothesis to decide if sensory processes were operating and memory researchers did not have to rely on reaching accepted levels of significance to know if recall or forgetting had occurred.

Each of the major sciences began with phenomena whose presence was not in question. The existence of the primary phenomena was never in question. Each science began by finding systematic *relationships among variations* in the magnitudes of attributes of the central phenomena and the attributes of independent variables such as time, location, etc. The questions for the investigation of memory had to do with how best to describe the forgetting curve and what factors affected its parameters. No statistical tests or determination of effect sizes were required to decide if, in fact, forgetting was or was not present on any particular occasion.

Only parapsychology claims to be a science on the basis of phenomena (or a phenomenon) whose presence can be detected only by rejecting a null hypothesis. To be fair, parapsychologists also talk about doing process research where the emphasis is on finding systematic relationships between attributes of psi and variations in some independent variable. One conclusion from the SRI/SAIC project, for example, is that there is no relationship between the distance of the target from the viewer and the magnitude of the effect size for anomalous cognition. However, it is still the case that the effect size, and even the question of whether anomalous cognition was present in any experiment, is still a matter of deciding if a departure from a chance base line is non-accidental.

At this point I think it is worth emphasizing that the use of statistical inference to draw conclusions about the null hypothesis assumes that the underlying probability model adequately represents the distributions and variations in the real world situation. The underlying probability model is an *idealization* of the empirical situation for which it is being used. Whether or not the model is appropriate for any given application is an empirical matter and the adequacy of the model has to be justified for each new application. Empirical studies have shown that statistical models fit real world situations only approximately. The tails of real-world distributions, for example, almost always contain more cases than the standard statistics based on the normal curve assume. These departures from the idealized model do not have much practical import in many typical statistical applications because the statistical tests are *robust*. That is, the departures of the actual situation from the assumed probability model typically do not distort the outcome of the statistical test.

However, when statistical tests are used in situations beyond their ordinary application, they can result in rejections of the null hypothesis for reasons other than a presumed departure from the expected chance value. Parapsychologists often complain that their results fail to replicate because of inadequate power. However, because the underlying probability models are only approximations, *too much power* can lead to rejections of the null hypothesis simply because the real world and the idealized statistical model are not exact matches. This discussion emphasizes that significant findings can arise for many reasons--including the simple fact that statistical inference is based on idealized models that mirror the real world only approximately.

I agree with Jessica Utts that the effect sizes reported in the SAIC experiments and in the

recent ganzfeld studies probably cannot be dismissed as due to chance. Nor do they appear to be accounted for by multiple testing, file-drawer distortions, inappropriate statistical testing or other misuse of statistical inference. I do not rule out the possibility that some of this apparent departure from the null hypothesis might simply reflect the failure of the underlying model to be a truly adequate model of the experimental situation. However, I am willing to assume that the effect sizes represent true effects beyond inadequacies in the underlying model. Statistical effects, by themselves, do not justify claiming that anomalous cognition has been demonstrated--or, for that matter, that an anomaly of any kind has occurred.

So, I accept Professor Utts' assertion that the statistical results of the SAIC and other parapsychological experiments "are far beyond what is expected by chance." Parapsychologists, of course, realize that the truth of this claim does not constitute proof of anomalous cognition. Numerous factors can produce significant statistical results. Operationally, the presence of anomalous cognition is detected by the elimination of all other possibilities. This reliance on a *negative definition of its central phenomenon* is another liability that parapsychology brings with its attempt to become a recognized science. Essentially, anomalous cognition is claimed to be present whenever statistically significant departures from the null hypothesis are observed under conditions that preclude the operation of all mundane causes of these departures. As Boring once observed, every success in parapsychological research is a failure. By this he meant that when the investigator or the critics succeed in finding a scientifically acceptable explanation for the significant effect the claim for ESP or anomalous cognition has failed.

Having accepted the existence of non-chance effects, the focus now is upon whether these effects have normal causes. Since the beginning of Psychical research, each claim that psychic functioning had been demonstrated was countered by critics who suggested other reasons for the observed effects, Typical alternatives that have been suggested to account for the effects have been fraud, statistical errors, and methodological artifacts. In the present discussion I am not considering fraud or statistical errors. This leaves only methodological oversight as the source for a plausible alternative to psychic functioning. Utts has concluded that "arguments that these results could be due to methodological flaws are soundly refuted.@' If she is correct, then I would have to agree with her bottom line "that psychic functioning has been well established.??

Obviously I do not agree that all possibilities for alternative explanations of the non-chance results have been eliminated. The SAIC experiments are well-designed and the investigators have taken pains to eliminate the known weaknesses in previous parapsychological research. In addition, I cannot provide suitable candidates for what flaws, if any, might be present. Just the same, it is impossible in principle to say that any particular experiment or experimental series is completely free from possible flaws. An experimenter cannot control for every possibility--especially for potential flaws that have not yet been discovered.

At this point, a parapsychologist might protest that such "in principle?? arguments can always be raised against any findings, no matter how well conceived was the study from which they emerged. Such a response is understandable, but I believe my caution is reasonable in this particular case. Historically, many cases of evidence for psi were proffered on the grounds that they came from experiments of impeccable methodological design. Only subsequently, sometimes by fortunate accident, did the possibility of a serious flaw or alternative explanation of the results become available. The founders of the Society for Psychical Research believed that the Smith-Blackburn experiments afforded no alternative to the conclusion that telepathy was involved. They could conceive of no mundane explanation. Then Blackburn confessed and explained in detail just how he and Smith had tricked the investigators.

The critics became suspicious of the Soal-Goldney findings not only because the results were too good, but also because Soal lost the original records under suspicious circumstances. Hansel, Scott, and Price each generated elaborate scenarios to explain how Soal might have cheated. Hansel and Scott reported finding peculiar patterns in the data. The scenarios, for accounting for these data, however, were extremely complicated and required the collusion of several individuals--some of whom were prominent statesmen and academics. The discovery of how Soal actually had cheated was made by the parapsychologist Betty Markwick. The finding came about through fortuitous circumstances. The method of cheating turned out to involve only one person and employed an ingenious, but simple, method that none of the critics had anticipated.

During the first four years of the original ganzfeld-psi experiments, the investigators asserted that their findings demonstrated psi because the experimental design precluded any normal alternative. Only after I and a couple of parapsychologists independently pointed out how the

use of a single set of targets could provide a mundane alternative to psychic communication did the ganzfeld experimenters realize the existence of this flaw. After careful and lengthy scrutiny of the ganzfeld database, I was able to generate a lengthy list of potential flaws.

Honorton and his colleagues devised the autoganzfeld experiments. These experiments were deliberately designed to preclude the flaws that I and others had eventually discovered in the original ganzfeld database. When the statistically significant results emerged from these latter experiments, they were proclaimed to be proof of anomalous communication because all alternative mundane explanations had been eliminated. When I was first confronted with these findings, I had to ad@t that the investigators had eliminated all but one of the flaws that I had listed for the original database. For some reason, Honorton and his colleagues did not seem to consider seriously the necessity of insuring that their randomization procedures were optimal. However, putting this one oversight aside, I could find no obvious loopholes in the experiments as reported.

When I was asked to comment on the paper that Daryl Bem and Charles Honorton wrote for the January 1994 issue of the *Psychological Bulletin, I* was able to get much of the raw data from Professor Bem. My analyses of that data revealed strong patterns that, to me, pointed to an artifact of some sort. One pattern, for example, was the finding that all the significant hitting above chance occurred only on the second or later occurrence of a target. All the first occurrences of a target yielded results consistent with chance. Although this was a post hoc finding, it was not the result of a fishing expedition. I deliberately looked for such a pattern as an indirect way of checking for the adequacy of the randomization procedures. The pattern was quite strong and persisted in every breakdown of the data that I tried--by separate investigator, by target type, by individual experiment, etc. The existence of this pattern by itself does not prove it is the result of an artifact. As expected, Professor Bem seized upon it as another peculiarity of psi. Subsequent to finding this pattern, I have learned about many other weaknesses in this experiment which could have compromised the results. Robert Morris and his colleagues at the University of Edinburgh took these flaws as well as some additional ones that they uncovered, into account when they designed the ganzfeld replication experiments.

The point of this discussion is that it takes some time before we fully recognize the potential flaws in a newly designed experimental protocol. In some cases, the discovery of a serious

flaw is the result of a fortuitous occurrence. In other cases, the uncovering of flaws came about only after the new protocol had been used for a while. Every new experimental design, as is the case for every new computer program, requires a shakedown period and debugging. The problems with any new method or design are not always apparent at first. Obvious flaws may be eliminated only to be replaced by more subtle ones.

How does this apply to the SAIC experiments? These experiments were designed to eliminate the obvious flaws of the previous remote viewing experiments at SRI. Inspection of the protocol indicates that they succeeded in this respect. The new design and methodology, however, has not had a chance to be used in other laboratories or to be properly debugged. Many of the features that could be considered an asset also have possible down sides. I will return to this later in the report when I discuss the use of the same viewers and the same judge across the different experiments. For now, I just want to suggest some general grounds for caution in accepting the claim that all possible methodological flaws have been eliminated.

The third warrant for Jessica Utts' conclusion that psi has been proven is that "Effects of similar magnitude to those found in government-sponsored research at SRI and SAIC have been replicated at a number of laboratories across the world." I will discuss this matter below. For now, I will point out that effects of similar magnitude can occur for several different reasons. Worse, the average effect size from different parapsychological research programs is typically a meaningless composite of arbitrary units. As such, these averages do not represent meaningful parameters in the real world. For example, Honorton claimed that the autoganzfeld experiments replicated the original ganzfeld experiments because the average effect size for both databases was approximately identical. This apparent similarity in average effect size is meaningless for many reasons. For one thing, the similarity in size depends upon which of many possible averages one considers. In the case under consideration the average effect size was obtained by adding up all the hits and trials for the 28 studies in the database. One experimenter contributed almost half to this total. Others contributed in greatly unequal numbers. The average will differ- if each experimenter's contribution is given equal weight.

In addition, the heterogeneity of effect sizes among separate investigators is huge. All the effect sizes, for example, of one the investigators were negative. Another investigator

contributed mostly moderately large effect sizes. If the first investigator had contributed more trials to the total, then the average would obviously have been lower. Similar problems exist for the average from the autoganzfeld experiments. In these latter experiments, the static targets--which most closely resembled the overwhelming majority of targets in the original database--yielded an effect size of zero. The dynamic targets yielded a highly significant and moderate effect size. Is the correct average effect size for these experiments based on a composite of the results of the static and dynamic targets or should it be based only the dynamic targets?

THE SAIC PROGRAM

As I have indicated, the SAIC experiments are an improvement on both the preceding SRI experiments as well as previous parapsychological investigations. The investigators seem to have taken pains to insure that randomization of targets for presentation and for judging was done properly. They have eliminated the major flaw in original SRI remote viewing experiments of non-independence in trials for a given viewer. Some of the other features can be considered as improvements but also as possible problems. in this category I would list the use of the same experienced viewers in many experiments and the use of the same target set across experiments. The major limitations that I see in these studies derive from their newness and their having been conducted in secrecy. The newness simply means that we have not had sufficient time to debug and to grasp fully both the strengths and weaknesses of this protocol. The secrecy aggravated this limitation by preventing other investigators from reviewing and criticizing the experiments from the beginning, and by making it impossible for independent laboratories to replicate the findings.'

The fact that these experiments were conducted in the same laboratory, with the same basic protocol, using the same viewers across experiments, the same targets across experiments, and the same investigators aggravates, rather than alleviates, the problem of independent replication. If subtle, as-yet-undetected bias and flaws exist is the protocol, the very consistency of elements such as targets, viewers, investigators, and procedures across experiments enhances the possibility that these flaws will be compounded.

Making matters even worse is the use of the same judge across all experiments. The judging of viewer responses is a critical factor in free-response remote viewing experiments. Ed May,

the principle investigator, as I understand it, has been the sole judge in all the free response experiments. May's rationale for this unusual procedure was that he is familiar with the response styles of the individual viewers. If a viewer, for example, talks about bridges, May--from his Familiarity with this viewer--might realize that this viewer uses bridges to refer to any object that is on water. He could then interpret the response accordingly to make the appropriate match to a target. The SAIC experiments did benefit from the input of a distinguished oversight committee. But this still falls far short of what could have taken place in an open forum. Whatever merit this rationale has, it results in a methodological feature that violates some key principles of scientific credibility. One might argue that the judge, for example, should be blind not only about the correct target but also about who the viewer is. More important, the scientific community at large will be reluctant to accept evidence that depends upon the ability of one specific individual. In this regard, the reliance on the same judge for all free-response experiments is like the experimenter effect. To the extent that the results depend upon a particular investigator the question of scientific objectivity arises. Scientific proof depends upon the ability to generate evidence that, in principle, any serious and competent investigator--regardless of his or her personality--can observe.

The use of the same judge across experiments further compounds the problem of non-independence of the experiments. Here, both Professor Utts and I agree. We believe it is important that the remote viewing results be obtainable with different judges. Again, the concern here is that the various factors that are similar across experiments, count against their separate findings as independent evidence for anomalous cognition.

HAS ANOMALOUS COGNITION BEEN PROVEN?

Obviously, I do not believe that the contemporary findings of parapsychology, including those from the SRI/SAIC program, justify concluding that anomalous mental phenomena have been proven. Professor Utts and some parapsychologists believe otherwise. I admit that the latest findings should make them optimistic. The case for psychic functioning seems better than it ever has been. The contemporary findings along with the output of the SRI/SAIC program do seem to indicate that something beyond odd statistical hiccups is taking place. I also have to admit that I do not have a ready explanation for these observed effects. Inexplicable statistical departures from chance, however, are a far cry from compelling evidence for

anomalous cognition.

So what would be compelling evidence for the reality of anomalous cognition? Let's assume that the experimental results from the SAIC remote viewing experiments continue to hold up. Further assume that along with continued statistical significance no flaws or mundane alternative possibilities come to light. We would then want to ensure that similar results will occur with new viewers, new target pools, and several independent judges. Finally, to satisfy the normal standards of science, we would need to have the findings successfully replicated in independent laboratories by other parapsychologists as well as non-parapsychologists.

If the parapsychologists could achieve this state of affairs, we are faced with a possible anomaly, but not necessarily anomalous cognition. As the parapsychologist John Palmer has recognized, parapsychologists will have to go beyond demonstrating the presence of a statistical anomaly before they can claim the presence of psychic functioning. This is because, among other things, the existence of a statistical anomaly is defined negatively. Something is occurring for which we have no obvious or ready explanation. This something may or may not turn out to be paranormal, According to Palmer, parapsychologists will have to devise *a positive theory of the paranormal* before they will be in a position to claim that the observed anomalies indicate paranormal functioning.

Without such a positive theory, we have no way of specifying the boundary conditions for anomalous mental phenomena. Without such a theory we have no way of specifying when psi is present and when it is absent. Because psi or anomalous cognition is currently detected only by departures from a null hypothesis all kinds of problems beset the quest for the claim and pursuit of psychic functioning. For example, the *decline effect,* which was investigated in one of the SAIC experiments, was once used as an important sign for the presence of psi. J.B.Rhine discovered this effect not only in some of his data but in his re-analyses of data collected by earlier investigators. He attached great importance to his effect because it existed in data whose investigators neither knew of its existence nor had they been seeking it. In addition, the decline effect helped Rhine to explain how seemingly null results really contained evidence for psi. This is because the decline effect often showed up as an excess of hitting in the early half of the experiment and as a deficit of hitting in the second half of the experiment. These two halves, when pooled together over the entire experiment, yielded an overall hit rate consistent with chance.

Although Rhine and other parapsychologists attached great importance to the decline effect as a reliable and often hidden sign of the presence of psychic functioning, the reliance on this indicator unwittingly emphasizes serious problems in the parapsychologist's quest. As the SAIC report on binary coding states, the decline effect is claimed for a bewildering variety of possibilities. Some investigators have found a decline effect going from the first quarter to the last quarter of each separate score sheet in their experiment. Other investigators have reported a decline effect as a decrease in hit rate from the first half to the second half of the total experiment. Still others find a decline effect across separate experiments. Indeed, almost any variation where the direction is from a higher hit rate to a lower hit rate has been offered as evidence for a decline effect. To confuse matters further, some investigators have claimed finding evidence for an *incline effect.*

If the decline effect is a token for the presence of psi, what should one conclude when the data, as was the case in the SAIC experiment on binary coding, show a significant departure from the null hypothesis *but no decline effect ?* We know what the parapsychologist's conclude. As long as they get a significant effect, they do not interpret the absence of the decline effect as the absence of psychic functioning. This state of affairs holds as well for several other effects that have been put forth as tokens or signs of anomalous mental functioning. Several such signs are listed in the *Handbook of parapsychology* [!977, B.B. Wolman, Editor].

Typically, such signs are sought when the attempt to reject the ordinary null hypothesis *fails. Displacement effects* are frequently invoked. When his attempts to replicate Rhine's results failed, Soal was persuaded to re-analyze his data in terms of displacement effects. His retrospective analysis uncovered two subjects whose guesses significantly correlated with the target one or two places ahead of the intended target. In his subsequent experiments with these two subjects, one kept hitting on the symbol that came after the intended target while the other produced significant outcomes only when her guesses were matched against the symbol that occurred just before the intended target. Negative hitting, increased variability, and other types of departures from the underlying theoretical probability model have all been used as hidden signs of the presence of psychic functioning.

What makes this search for hidden tokens of psi problematic is lack of constraints. Any time the original null hypothesis cannot be rejected, the eager investigator can search through the

data for one or more these markers. When one is found, the investigator has not hesitated in offering this as proof of the presence of psi. However, if the null hypothesis is rejected and none of these hidden signs of psi can be found in the data, the the investigator still claims the presence of psi. This creates the scientifically questionable situation where any significant departure from a probability model is used as proof of psi but the absence of these departures does not count as evidence against the presence of psi.

So, acceptable evidence for the presence of anomalous cognition must be based on a positive theory that tells us when psi should and *should not* be present. Until we have such a theory, the claim that anomalous cognition has been demonstrated is empty. Without such a theory, we might just as well argue that what has been demonstrated is a set of *effects--each one of which be the result of an entirely different cause.*

Professor Utts implicitly acknowledges some of the preceding argument by using consistency of findings with other laboratories as evidence that anomalous cognition has been demonstrated. I have already discussed why the apparent consistency in average effect size across experiments cannot be used as an argument for consistency of phenomena across these experiments. To be fair, parapsychologists who argue consistency of phenomena across experiments often go beyond simply pointing to consistency in effect sizes.

One example is the claim that certain personality correlates replicate across experiments. May and his colleagues correctly point out, however, that these correlations tend to be low and inconsistent. Recently, parapsychologists have claimed that extroversion correlates positively with successful performance on anomalous cognition tasks. This was especially claimed to be true of the ganzfeld experiments. However, the apparently successful replication of the autoganzfeld experiments by the Edinburgh group (under subcontract to the SAIC program) found that the introverts, if anything, scored higher than the extroverts.

The autoganzfeld experiments produced significant effects only for the dynamic targets. The static targets produced zero effect size. Yet the bulk of the targets in the original ganzfeld database were static and they produced an effect size that was significantly greater than the zero effect size of the autoganzfeld experiments (I was able to demonstrate that there was adequate power to detect an effect size of the appropriate magnitude for the static targets in the autoganzfeld experiments). Further indication of inconsistency is the SAIC experiment

which found that the only the static targets produced a significant effect size, whereas the dynamic targets yielded a zero effect size. May and his colleagues speculated that the failure of the dynamic targets was due to a 'bandwidth' that was too wide. When they apparently narrowed the bandwidth of the dynamic targets in a second experiment, both dynamic and static targets did equally well. It is unclear whether this should be taken as evidence for consistency or inconsistency. Note that the hypothesis and claim for the autoganzfeld experiments is that dynamic targets should be significantly better than static ones. As far as I can tell the original dynamic targets of the ganzfeld experiments are consistent with an unlimited bandwidth.

Other important inconsistencies exist among the contemporary databases. *The raison d'etre* for the ganzfeld experiments is the belief among some parapsychologists that an altered state facilitates picking up the psi signal because it lowers the noise-to-signal ratio from external sensory input. The touchstone of this protocol is the creation of an altered state in the receiver. This contrasts sharply with the remote viewing experiments in which the viewer is always in a normal state. More important is that the ganzfeld researchers believe that they get best results when each subject serves as his/her own judge. Those experiments in the ganzfeld database that employed both external judges and subjects as their own judges found that their results were more successful using subjects as their own judges. The reverse is true in the remote viewing experiments. The remote viewer experimenters believe that external judges provide much better hit rates than viewer-judges. This difference is even more extreme in the SAIC remote viewing where a single judge was used for all experiments. This judge, who was also the principal investigator, believed that he could achieve best results if he did the judging because of his familiarity with the response styles of the individual viewers.

So even if the ganzfeld and the SAIC remote viewing experiments have achieved significant effects and average effect sizes of approximately the same magnitude, there is no compelling reason to assume they are dealing with the same phenomena or phenomenon. To make such a claim entails showing that the alleged effect shows the same pattern of relationships in each protocol. Almost certainly, a positive theory of anomalous mental phenomena that predicts lawful relationships of a recognizable type will be necessary before a serious claim can be made that the same phenomenon is present across different research laboratories and experiments. Such a positive theory will be necessary also to tell us when we are and when

we are not in the presence of this alleged anomalous cognition.

WHAT NEEDS TO BE EXPLAINED?

Professor Utts and many parapsychologists argue that they have produced evidence of an anomaly that requires explanation. They assert that the statistical effects they have documented cannot be accounted for in terms of normal scientific principles or methodological artifact. After reviewing the results from the SAIC experiments in the context of other contemporary parapsychological research, Utts is confident that more than an anomaly has been demonstrated. She believes the evidence suffices to conclude that the anomaly establishes the existence of psychic functioning.

This evidence for anomalous cognition, according to Utts and the parapsychologists, meets the standards employed by the other sciences. By this, I think Professor Utts means that in many areas of scientific inquiry the decision that a real effect has occurred is based on rules of statistical inference. Only if the null hypothesis of no difference between two or more treatments is rejected can the investigator claim that the differences are real in the sense that they are greater than might be expected on the basis of some baseline variability. According to this standard, it seems that the SAIC experiments as well as the recent ganzfeld experiments have yielded effects that cannot be dismissed as the result of normal variability.

While the rejection of the null hypothesis is typically a *necessary* step for claiming that an hypothesized effect or relationship has occurred, it is never *sufficient.* Indeed, because the underlying probability model is only an approximation, everyone realizes that the null hypothesis is rarely, if ever, strictly true. In practice, the investigator hopes that the statistical test is sufficiently *robust* that it will reject the null hypothesis only for meaningful departures from the null hypothesis. With sufficient power, the null hypothesis will almost certainly be rejected in most realistic situations. This is because effect sizes will rarely be exactly zero. Even if the true effect size is zero in a particular instance, sufficient power can result in the rejection of the null hypothesis because the assumed statistical model will depart from the real-world situation in other ways. For most applications of statistical inference, then, *too much power* can result in mistaken inferences as well as *too little power.*

Here we encounter another way in which parapsychological inquiry differs from typical scientific inquiry. In those sciences that rely on statistical inference, they do so as an aid to weeding out effects that could be the result of chance variability. When effect sizes are very small or if the experimenter needs to use many more cases than is typical for the field to obtain significance, the conclusions are often suspect. This is because we know that with enough cases an investigator will get a significant result, regardless of whether it is meaningful or not. Parapsychologists are unique in postulating a null hypothesis that entails a true effect size of zero if psi is not operating. Any significant outcome, then, becomes evidence for psi. My concern here is that small effects and other departures from the statistical model can be expected to occur in the absence of psi. The statistical model is only an approximation. When power is sufficient and when the statistical test is pushed too far, rejections of the null hypothesis are bound to occur. This is another important reason why claiming the existence of an anomaly based solely on evidence from statistical inference is problematic.

This is one concern about claiming the existence of an anomaly on the basis of statistical evidence. In the context of this report, I see it as a minor concern. As I have indicated, I am willing to grant Professor Utts' claim that the rejection of the null hypothesis is probably warranted in connection with the SAIC and the ganzfeld databases. I have other concerns. Both have to do with the fact that no other science, so far as I know, would draw conclusions about the existence of phenomena solely on the basis of statistical findings. Although it is consistent with scientific practice to use statistical inference to reject the null hypothesis, it is not consistent with such practice to postulate the existence of phenomena on this basis alone. Much more is required. I will discuss at least two additional requirements.

Thomas Kuhn's classic characterization of normal and revolutionary science has served as the catalyst for many discussions about the nature of scientific inquiry. He popularized the idea that normal scientific inquiry is guided by what he called *a paradigm.* Later, in the face of criticisms, he admitted that he had used the *term paradigm* to cover several distinct and sometimes contradictory features of the scientific process. One of his key uses of the term *paradigm* was to refer to the store of *exemplars* or textbook cases of standard experiments that every field of scientific inquiry possesses. These exemplars are what enable members of a scientific community to quickly learn and share common principles, procedures, methods, and standards. These exemplars are also the basis for initiating new members into the

community. New research is conducted by adapting one or more of the patterns in existing exemplars as guidelines about what constitutes acceptable research in the field under consideration.

Every field of inquiry, including parapsychology, has its stock of exemplars. In parapsychology these would include the classic card guessing experiments of J.B. Rhine, the Sheep-Goat experiments, etc. What is critical here is the striking difference between the role of exemplars in parapsychology as contrasted with their role in all other fields of scientific inquiry. These exemplars not only serve as models of proper procedure, but they also are teaching tools. Students in a particular field of inquiry can be assigned the task of replicating some of these classic experiments. The instructor can make this assignment with the confident expectation that each student will obtain results consistent with the original findings. The physics instructor, for example, can ask novice students to try Newton's experiments with colors or Gilbert's experiments with magnets. The students who do so will get the expected results. The psychology instructor can ask novice students to repeat Ebbinghaus' experiments on forgetting or Peterson and Peterson's classic experiment on short-term memory and know that they will observe the same relationships as reported by the original experimenters.

Parapsychology is the only field of scientific inquiry that *does not have even one exemplar that can be assigned to students with the expectation that they will observe the original results!* In every domain of scientific inquiry, *with the exception of parapsychology,* many core exemplars or paradigms exist that will reliably produce the expected, lawful relationships. This is another way of saying that the other domains of inquiry are based upon robust, lawful phenomena whose conditions of occurrence can be specified in such a way that even novices will be able to observe and/or produce them. Parapsychologists do not possess even one exemplar for which they can confidently specify conditions that will enable anyone —let alone a novice—to reliably witness the phenomenon.

The situation is worse than I have so far described. The phenomena that can be observed with the standard exemplars do not require sensitive statistical rejections of the null hypothesis based on many trials to announce their presence. The exemplar in which the student uses a prism to break white light into its component colors requires no statistics or complicated inference at all. The forgetting curve in the Ebbinghaus experiment, requires

nothing more than plotting proportion recalled against trial number. Yet, to the extent that parapsychology is approaching the day when it will possess at least one exemplar of this sort, the "observation?F of the "phenomenon' will presumably depend upon the indirect use of statistical inference to document its presence.

In the standard domains of science, this problem of having not a single exemplar for reliably observing its alleged phenomenon, would be taken as a sign that the domain has *no central phenomena.* When Soviet scientists announced the discovery of mitogenetic radiation, some western scientists attempted to replicate the findings. Some reported success; others reported mixed results; and many failed entirely to observe the effect. Eventually scientists, including the Soviets, abandoned the quest for mitogenetic radiation. Because no one, including the original discover, could specify conditions under which the phenomenon--if there be one--could be observed, the scientific community decided that there was nothing to explain other than as-yet-undetected artifacts. The same story can be told about N-Rays, Polywater, and other candidate phenomena that could not be reliably observed or produced. We cannot explain something for which we do not have at least some conditions under which we can confidently say it occurs. Even this is not enough. The alleged phenomenon not only must reliably occur at least under some conditions but it also *must reliably vary in magnitude or other attributes as a function of other variables.* Without this minimal amount of lawfulness, the idea that there is something to explain is senseless. Yet, at best, parapsychology's current claim to having demonstrated a form of anomalous cognition rests on the possibility that it can generate significant differences from the null hypothesis under conditions that are still not reliably specified.

I will suggest one more reason for my belief that it is premature to try to account for what the SAIC and the ganzfeld experiments have so far put before us. On the basis of these experiments, contemporary parapsychologists claim that they have demonstrated the existence of an "anomaly". I will grant them that they have apparently demonstrated that the SAIC and the ganzfeld experiments have generated significant effect sizes beyond what we should expect from chance variations. I will further admit that, at this writing, I cannot suggest obvious methodological flaws to account for these significant effects. As I have previously mentioned, this admission does not mean that these experiments are free from subtle biases and potential bugs. The experimental paradigms are too recent and insufficiently evaluated to know for sure. I can point to departures from optimality that might harbor potential flaws--

such as the use of a single judge across the remote viewing experiments, the active coaching of viewers by the experimenter during judging procedures in the ganzfeld, my discovery of peculiar patterns of scoring in the ganzfeld experiments, etc. Having granted that significant effects do occur in these experiments, I hasten to add that without further evidence, I do not think we can conclude that these effects are all due to the same cause—let alone that they result from a single phenomenon that is paranormal in origin.

The additional reason for concern is the difference in the use of 'anomaly' ;n this context and how the term 'anomaly' is used in other sciences. In the present context, the parapsychologists are using the term 'anomaly' to refer to apparently inexplicable departures from the null hypothesis. These departures are considered inexplicable in the sense that apparently all normal reasons for such departures from the null hypothesis have been excluded. But these departures are not lawful in the sense that the effect sizes are consistent. The effect sizes differ among viewers and subjects; they also differ for different experimenters; they come and go in inexplicable ways within the same subject. Possibly some of these variations in effect size will be found to exhibit some lawfulness in the sense that they will correlate with other variables. The SAIC investigators, for example, hope they have found such correlates in the entropy and bandwidth of targets. At the moment this is just a hope.

The term "anomaly" is used in a much more restricted sense in the other sciences. Typically an anomaly refers to a lawful and precise departure from a theoretical baseline. As such it is something the requires explaining. Astronomers were faced with a possible anomaly when discrepancies from Newtonian theory were reported in the orbit of Uranus. In the middle 1800s, Urban Leverrier decided to investigate this problem. He reviewed all the data on previous sightings of Uranus--both before and after it had been discovered as new planet. On the basis of the previous sightings, he laboriously recalculated the orbital path based on Newtonian theory and the reported coordinates. Sure enough, he found errors in the original calculations. When he corrected for these errors, the apparent discrepancy in Uranus' orbit was much reduced. But the newly revised orbit was still discrepant from where it should be on Newtonian theory. With this careful work, Leverrier had transformed a potential anomaly into an actual anomaly. Anomaly in this sense meant a precise and lawful departure from a well-defined theory. It was only after the precise nature, direction, and magnitude of this discrepancy was carefully specified did Leverrier and the scientific community decide that

here was an anomaly that required explanation. What had to explain was quite precise. What was needed was an explanation that exactly accounted for this specific departure from the currently accepted theory.

Leverrier's solution was to postulate a new planet beyond the orbit of Uranus. This was no easy task because it involved the relatively unconstrained and difficult problem of inverse perturbations. Leverrier had to decide on a size, orbit, location, and other attributes of a hitherto unknown body whose characteristics would be just those to produce the observed effects on Uranus without affecting the known orbit of Saturn. Leverrier's calculations resulted in his predicting the location of this hitherto unknown planet and the astronomer Galle located this new planet, Neptune, close to where Leverrier had said it would be.

The point of this story is to emphasize the distinction between the parapsychologists' use of anomaly from that of other scientists'. Anomalies in most domains of scientific inquiry are carefully specified deviations from a formal theory. What needs to be explained or accounted for is precisely described. The anomalies that parapsychologists are currently talking about differ from this standard meaning in that the departures are from the general statistical model and are far from having the status of carefully specified and precise deviations from a theoretical baseline. In this latter case we do not know what it is that we are being asked to explain. Under what conditions can we reliably observe it? What theoretical baselines are the results a departure from? How much and in what direction and form do the departures exist? What specifically must our explanation account for?

Finally, I should add that some parapsychologists, at least in the recent past, have agreed with my position that parapsychological results are not yet ready to be placed before the scientific community. Parapsychologists such as Beloff, Martin Johnson, Gardner Murphy, J.G. Pratt and others have complained that parapsychological data are volatile and messy. Some of these investigators have urged their colleagues to first get their house in order before they ask the scientific community at large to take them seriously. Martin Johnson, especially, has urged his colleagues to refrain from asking the scientific community to accept their findings until they can tame them and produce lawful results under specified conditions. Clearly, parapsychology has still not reached this desired state. At best, the results of the SAIC experiments combined with other contemporary findings offer hope that the parapsychologists may be getting closer to the day when they can put something before the scientific

community and challenge it to provide an explanation.

POTENTIALS FOR OPERATIONAL APPLICATIONS

It may seem obvious that the utility of remote viewing for intelligence gathering should depend upon its scientific validity. If the scientific research cannot confirm the existence of a remote viewing ability, then it would seem to be pointless to try an use this non-existent ability for any practical application. However, the matter is not this simple. If the scientific research confirms the existence of anomalous cognition, this does not guarantee that this ability would have useful applications. Ed May, in his presentation to the evaluation panel, gave several reasons why remote viewing could be real and, yet, not helpful for intelligence gathering. In his opinion, approximately 20 percent of the information supplied by a viewer is accurate. Unfortunately, at the time the remote viewer is generating the information, we have no way of deciding which portion is likely to be the accurate one. Another problem is that the viewer's information could be accurate, yet not relevant for the intelligence analyst's purposes.

This question is related to the problem of boundary conditions which I discussed earlier in this report. From both a scientific and an operational viewpoint the claim that anomalous cognition exists is not very credible until we have ways to specify when and when it is not present. So far, parapsychology seems to have concentrated only in finding ways to document the existence of anomalous cognition. The result is a patchwork quilt of markers that, when present, are offered as evidence for the presence of psi. These markers or indicators include the decline effect, negative hitting as well as positive hitting, displacement hitting, the incline effect, increased variability, decreased variability and just about any other way a discrepancy from a probability model can occur. A cynic will note that the absence of any or most of these markers is not used as evidence for the *absence* of psi. This lack of way to distinguish between the presence and absence of anomalous cognition creates many challenges for parapsychology, some of which I have already discussed.

So, even if remote viewing is a real ability possessed by some individuals, its usefulness for intelligence gathering is questionable. If May is correct, then 80% of the all the information supplied by this talented viewer will be erroneous. Without any way to tell which statements

of the views are reliable and which are not, the use of this information may make matters worse rather than better.

Can remote viewing have utility for information gathering even if it cannot be scientifically validated? I can imagine some possibilities for remote viewing to be an asset to the intelligence analyst even when the viewer possesses no valid paranormal powers. The viewer might be a person of uncommonly good sense or have a background that enables him or her to provide helpful information even if it does not come from a paranormal source. Another possibility is that the viewer, even though lacking in any truly accurate intelligence information, might say things or open up new ways of dealing with the analyst's problem. In this latter scenario the remote viewer is a catalyst that may open up new ways of looking at an intelligence situation much like programs for problem solving and creative thinking stimulate new ways of looking at a situation. However, if the usefulness of the remote viewer reduces to a matter of injecting common sense or new perspectives into the situation, I believe that we can accomplish the same purpose in more efficient ways.

In considering potential utility, I am most concerned about separation of the operational program in remote viewing from the research and development phase. By default, the assessment of the usefulness of the remote viewing in the operational arena is decided entirely by subjective validation or what May and Utts *call prima facie* evidence. Granted it is difficult to assess adequately the effectiveness of remote viewing in the operational domain. Nevertheless, better ways can be devised than have apparently been used up to now. In our current attempt to get an initial idea about the effectiveness of the current operational use of remote viewing, we have simply been asking individuals and agencies who have used the services of the remote viewers, if the information they received was accurate and useful. Whatever information we get from this survey is extremely limited for the purposes of judging the utility of remote viewing in the operational domain.

Even psychologists who should know better underrate the power of subjective validation. Anyone who relies *on prima facie* evidence as a basis for affirming the validity of remote viewing should carefully read that portion of Marks and Kamman's *The psychology of the psychic* [1981] in which they discuss the SRI and their own experiments on remote viewing. In the early stages of their attempt to replicate the SRI remote viewing experiments, they were astonished at the high quality of their subject's protocols and the apparent accuracy of

the viewing. After each session, the experimenters and the subject (viewer) would visit the target site and compare the verbal protocol with the actual site. The specific details of the viewers' responses appeared to match specific objects in the target site with uncanny accuracy. When they gave the verbal protocols to the judge, a distinguished professor, to blindly match against the actual target sites, he was astonished at how well what he considered the closest matching protocol for each site matched actual details of the target. He had no doubt that the viewers had demonstrated strong remote viewing abilities.

So, both the viewers and the judge quickly became convinced of the reality of remote viewing on the basis of the uncanny matches between the verbal descriptions and the actual target sites. The experimenters received a rude awakening when they discovered that, despite the striking matches observed between target and verbal description, the judge had matched the verbal protocols to the wrong target sites. When all parties were given the results the subjects could not understand how the judge could have matched any but the actual target site to their descriptions. For them the match was so obvious that it would be impossible for the judge to have missed it. The judge, on the other hand, could not accept that any but the matches he made could be paired with the actual target sites.

This phenomenon of subjective validation is pervasive, compelling and powerful. Psychologists have demonstrated it in a variety of settings. I have demonstrated it and written about in the context of the psychic reading. In the present context, subjective validation comes about when a person evaluates the similarity between a relatively rich verbal description and an actual target or situation. Inevitably, many matches will be found. Once the verbal description has been judged to be a good match to a given target, the description gets locked in and it becomes virtually impossible for the judge to see the description as fitting any but the original target.

Unfortunately, all the so-called *prima facie* evidence put before us is tainted by subjective validation. We are told that the many details supplied by the viewers were indeed inaccurate. But some details were uncannily correct and even, in one case, hidden code words were correctly revealed. Such accounts do indeed seem compelling. They have to be put in the context, however, of all such operational attempts. We have to know the general background and expectations of the viewers, the questioners, etc. Obviously, the targets selected for the

viewers in the operational setting will have military and intelligence relevance. If the viewer [some of the viewers have intelligence backgrounds] suspects the general nature of the target, then previous background knowledge might very well make the presence, say of a gantry, highly likely. In addition, the interactions and questioning of the viewers in these settings appear to be highly suggestive and leading.

I can imagine that the preceding paragraph might strike a reader as being unreasonable. Even allowing for subjective validation, the possibility that a viewer might accurately come up with secret code words and a detailed description of particular gantry is quite remote on the basis of common sense and sophisticated guessing. I understand the complaint and I realize the reluctance to dismiss such evidence out of hand. However, I have had experience with similarly *compelling prima facie* evidence for more than a chance match between a description and a target. In the cases I have in mind, however, the double blind controls were used to pair descriptions with the true as well as with the wrong target sites. In all these test cases with which I am familiar, the unwitting subjects found the matches between their descriptions and the presumed target equally compelling regardless of whether the presumed target was the actual or the wrong one.

What this says about operational effectiveness, is that, for evaluation purposes, half of the time the viewers and the judges should be mislead about the what was the actual target. In these cases, both the interrogator and the viewer, as well as the judge, have to be blind to the actual targets. Under such conditions, if the judges and the others find the matches between the verbal descriptions and the actual targets consistently better than the matches between the verbal descriptions and the decoy targets, then this would constitute some evidence for the effectiveness of remote viewing. I can confidently predict, regardless of the outcome of such an evaluation, that many of the verbal descriptions when matched with decoy targets will be judged to be uncanny matches.

SUGGESTIONS: WHAT NEXT?

I have played the devil's advocate in this report. I have argued that the case for the existence of anomalous cognition is still shaky, at best. On the other hand, I want to state that I believe that the SAIC experiments as well as the contemporary ganzfeld experiments display methodological and statistical sophistication well above previous parapsychological research.

Despite better controls and careful use of statistical inference, the investigators seem to be getting significant results that do not appear to derive from the more obvious flaws of previous research. I have argued that this does not justify concluding that anomalous cognition has been demonstrated. However, it does suggest that it might be worthwhile to allocate some resources toward seeing whether these findings can be independently replicated. If so, then it will be time to reassess if it is worth pursuing the task of determining if these effects do indeed reflect the operation of anomalous cognition. This latter quest will involve finding lawful relationships between attributes of this hypothesized phenomenon and different independent variables. Both the scientific and operational value of such an alleged phenomenon will depend upon how well the conditions for its occurrence can be specified and how well its functioning can be brought under control.

Both Professor Utts and I agree that the very first consideration is to see if the SAIC remote viewing results will still be significant when independent judges are used. I understand Ed May's desire to use a judge who is very familiar with the response styles of the experienced viewers. However, if remote viewing is real, then conscientious judges, who are blind to the actual targets, should still be able to match the verbal descriptions to the actual targets better than chance. If this cannot be done, the viability of the case for remote viewing becomes problematical. On the other hand, assuming that independent judges can match the descriptions to the correct targets reasonably well, then it becomes worthwhile to try to independently replicate the SAIC experiments.

At this point we face some interesting questions. Should we try to replicate the remote viewing studies by using the same viewers, the same targets, and the same protocol? Perhaps change only the experimenters, the judge, and the laboratory? At some point we would also want to change the targets. For completeness, we would also want to search for new, talented viewers.

If independent replications confirm the SAIC findings, we still have a long way to go. However, at this stage in the proceedings, the scientific community at large might be willing to acknowledge that an anomaly of some sort has been demonstrated. Before the scientific community will go beyond this acknowledgment, the parapsychologists will have to devise a positive theory of anomalous communication from which they can make testable predictions about relationships between anomalous communication and other variables.

CONCLUSIONS

THE SCIENTIFIC STATUS OF THE SAIC RESEARCH PROGRAM

1. The SAIC experiments on anomalous mental phenomena are statistically and methodologically superior to the earlier SRI remote viewing research as well as to previous parapsychological studies. In particular, the experiments avoided the major flaw of non-independent trials for a given viewer. The investigators also made sure to avoid the problems of multiple stastical testing that was characteristic of much previous parapsychological research.

2. From a scientific viewpoint, the SAIC program was hampered by its secrecy and the multiple demands placed upon it. The secrecy kept the program from benefiting from the checks and balances that comes from doing research in a public forum. Scrutiny by peers and replication in other laboratories would accelerated the scientific contributions from the program. The multiple demands placed on the program meant that too many things were being investigated with too few resources. As a result, no particular finding was followed up in sufficient detail to pin it down scientifically. Ten experiments, no matter how well conducted, are insufficient to fully resolve one important question, let alone the several that were posed to the SAIC investigators.

3. Although, I cannot point to any obvious flaws in the experiments, the experimental program is too recent and insufficiently evaluated to be sure that flaws and biases have been eliminated. Historically, each new paradigm in parapsychology has appeared to its designers and contemporary critics as relatively flawless. Only subsequently did previously unrecognized drawbacks come to light. Just as new computer programs require a shakedown period before hidden bugs come to light, each new scientific program requires scrutiny over time in the public arena before its defects emerge. Some possible sources of problems for the SAIC program are its reliance on experienced viewers, and the use of the same judge--one who is familiar to the viewers, for all the remote viewing.

4. The statistical departures from chance appear to be too large and consistent to attribute to statistical flukes of any sort. Although I cannot dismiss the possibility that these

rejections of the null hypothesis might reflect limitations in the statistical model as an approximation of the experimental situation, I tend to agree with Professor Utts that real effects are occurring in these experiments. *Something* other than chance departures from the null hypothesis has occurred in these experiments.

5. However, the occurrence of statistical effects does not warrant the conclusion that psychic functioning has been demonstrated. Significant departures from the null hypothesis can occur for several reasons. Without a positive theory of anomalous cognition, we cannot say that these effects are due to a single cause, let alone claim they reflect anomalous cognition. We do not yet know how replicable these results will be, especially in terms of showing consistent relations to other variables. The investigators report findings that they believe show that the degree of anomalous cognition varies with target entropy and the 'bandwidth' of the target set. These findings are preliminary and only suggestive at this time. Parapsychologists, in the past, have reported finding other correlates of psychic functioning such as extroversion, sheep/goats, altered states only to find that later studies could not replicate them.

6. Professor Utts and the investigators point to what they see as consistencies between the outcome of contemporary ganzfeld experiments and the SAIC results. The major consistency is similarity of average effect sizes across experiments. Such consistency is problematical because these average effect sizes, in each case, are the result of arbitrary combinations from different investigators and conditions. None of these averages can be justified as estimating a meaningful parameter. Effect size, by itself, says nothing about its origin. Where parapsychologists see consistency, I see inconsistency. The ganzfeld studies are premised on the idea that viewers must be in altered state for successful results. The remote viewing studies use viewers in a normal state. The ganzfeld experimenters believe that the viewers should judge the match between their ideation and the target for best results; the remote viewers believe that independent judges provide better evidence for psi than viewers judging their own responses. The recent autoganzfeld studies found successful hitting only with dynamic targets and only chance results with static targets. The SAIC investigators, in one study, found hitting with static targets and not with dynamic ones. In a subsequent study they found hitting for both types of targets. They suggest that they may have

solution to this apparent inconsistency in terms of their concept of bandwidth. At this time, this is only suggestive.

7. The challenge to parapsychology, if it hopes to convincingly claim the discover, of anomalous cognition, is to go beyond the demonstration of significant effects. The parapsychologists need to achieve the ability to specify conditions under which one can reliably witness their alleged phenomenon. They have to show that they can generate lawful relationships between attributes of this alleged phenomenon and independent variables. They have to be able to specify boundary conditions that will enable us to detect when anomalous cognition is and *is not* present.

SUGGESTIONS FOR FUTURE RESEARCH

1. Both Professor Utts and I agree that the first step should be to have the SAIC protocols rejudged by independent judges who are blind to the actual target.

2. Assuming that such independent judging confirms the extra-chance matchings, the findings should be replicated in independent laboratories. Replication could take several forms. Some of the original viewers from the SAIC experiments could be used. However, it seems desirable to use a new target set and several independent judges.

OPERATIONAL IMPLICATIONS

1. Tthe current default assessment of the operational effectiveness of remote viewing is fraught with hazards. Subjective validation is well known to generate compelling, but false, convictions that a description matches a target in striking ways. Better, double blind, ways of assessing operational effectiveness can be used. I suggest at least one way in the report.

2. The ultimate assessment of the potential utility of remote viewing for intelligence gathering cannot be separated from the findings of laboratory research.

THE REPLY

Response to Ray Hyman's Report of September 11, 1995 "Evaluation of Program in Anomalous Mental Phenomena"

Jessica Utts
Division of Statistics
University of California, Davis

Ray Hyman's report of September 11, 1995, written partially in response to my written report of September 1, 1995 elucidates the issues on which he and I agree and disagree. I basically concur with his assessment, but there are three issues he raises with regard to the scientific status of parapsychology to which I would like to respond.

1. "Only parapsychology, among the fields of inquiry claiming scientific status, lacks a cumulative database (p.6)."

It is simply not true that parapsychology lacks a cumulative database. In fact, the accumulated database is truly impressive for a science that has had so few resources. While critics are fond of relating, as Professor Hyman does in his report, that there has been "more than a century of parapsychological research (p. 7)" psychologist Sybo Schouten (1993, p.316) has noted that the total human and financial resources devoted to parapsychology since 1882 is at best equivalent to the expenditures devoted to fewer than two months of research in conventional psychology in the United States.

On pages 4 and 5 of their September 29, 1994 SAIC final report, May, Luke and James summarize four reports that do precisely what Professor Hyman claims is not don in parapsychology; they put forth the accumulated evidence for anomalous cognition in a variety of formats. Rather than dismissing the former experiments, parapsychologists build on them. As in any area of science, it is of course the most recent experiments that receive the most attention, but that does not mean that the field would divorce itself from past work. Quite to the contrary, past experimental results and methodological weaknesses are used to design better and more efficient experiments.

As an example of the normal progress of inquiry expected in any area of science, the autoganzfeld experiments currently conducted by parapyscholgists did not simply spring out of thin air. The original ganzfeld experiments followed from Honorton's observation at Maimondies Medical Center, that anomalous cognition seemed to work well in dreams. He investigated ways in which a similar state could be achieved in normal waking hours, and found the ganzfeld regime in another area of psychology. The automated ganzfeld followed from a critical evaluation of the earlier ganzfeld experiments, and a set of conditions agreed upon by Honorton and Professor Hyman. The current use of dynamic targets in autoganzfeld experiments follows from the observation that they were more successful than static targets in the initial experiments. The investigation of entropy at SAIC follows from this observation as well. This is just one example of how current experiments are built from past results.

2. "Only parapsychology claims to be a science on the basis of phenomena (or a phenomenon) whose presence can be detected only by rejecting a null hypothesis (p.8)."

While it is true that parapsychology has not figured out all the answers, it does not differ from normal science in this regard. It is the norm of scientific progress to make observations first, and then to attempt to explain them. Before quantum mechanics was developed there were a number of anomalies observed in physics that could not be explained. There are many observations in physics and in the social and medical sciences that can be observed, either statistically or deterministically, but which cannot be explained.

As a more recent example, consider the impact of electromagnetic fields on health. An article in Science (Vol. 269, 18 August, p. 911) reported that "After spending nearly a decade reviewing the literature on electromagnetic fields (EMFs), a panel of the National Council on Radiation Protection and Measurements (NCRP) has produced a draft report concluding that some health effects linked to EMFs--such as cancer and immune deficiencies--appear real and warrant steps to reduce EMF exposure... Biologists have failed to pinpoint a convincing mechanism of action." In other words, a statistical effect has been convincingly established and it is not the responsibility of science to attempt to establish its mechanism, just as in parapsychology.

As yet another example, consider learning and memory, which have long been studied in psychology. We know they exist, but brain researchers are just beginning to understand how they work by using sophisticated brain imaging techniques. Psychologists do not understand these simple human capabilities, and they certainly do not understand other observable human phenomena such as what causes people to fall in love. Yet, no one would deny the existence of these phenomena just because we do not understand them.

In any area involving the natural inherent in humans, science progresses by first observing a statistical difference and then attempting to explain it. At this stage, I believe parapsychology has convincingly demonstrated that an effect is present, and future research attempts should be directed at finding an explanation. In this regard parapsychology in on par with scientific questions like the impact of electromagnetic fields on health, or the cross-cultural differences in memory that have been observed by psychologists.

3. "Parapsychology is the only field of scientific inquiry that does not have even one exemplar that can be assigned to students with the expectation that they will observe the original results (p. 18)."

I disagree with this statement for two reasons. First, I can name other phenomena for which students could not be expected to do a simple experiment and observe a result, such as the connection between taking aspirin and preventing heart attacks or the connection between smoking and getting lung cancer. What differentiates these phenomena from simple experiments like splitting light with a prism is that the effects are statistical in nature and are not expected to occur every single time. Not everyone who smokes gets lung cancer, but we can predict the proportion who will. Not everyone who attempts anomalous cognition will be successful, but I think we can predict the proportion of time success should be achieved.

Since I believe the probability of success has been established in the autoganzfeld experiments, I would offer them as the exemplar Professor Hyman requests. The problem is that to be relatively assured of a successful outcome requires several hundred trials, and no student has the resources to commit to this experiment. As I have repeatedly tried to explain to Professor Hyman and others, when dealing with a small to medium effect it takes hundreds or sometimes thousands of trials to establish "statistical significance." In fact, the Physicians Health Study that initially established the link between taking aspirin and reducing heart

attacks studies over 22,000 men. Had it been conducted on only 2,200 men with the same reduction in heart attacks, it would not have achieved statistical significance. Should students be required to recruit 22,000 participants and conduct such an experiment before we believe the connection between aspirin and heart attacks is real?

Despite Professor Hyman's continued protests about parapsychology lacking repeatability, I have never seen a skeptic attempt to perform an experiment with enough trials to even come close to insuring success. The parapsychologists who have recently been willing to take on this challenge have indeed found success in their experiments, as described in my original report.

REFERENCE:

Schouten, Sybo (1993). "Are we making progress?" In Psi Research Methodology: A Re-examination, Proceedings of an International Conference, Oct 29-30, 1988, edited by L. Coly and J. McMahon, NY: Parapsychology Foundation, Inc., pgs. 295-322.

Points of Agreement and Disagreement

The exchanges between Dr. Utts and Dr. Hyman has converged upon a number of issues bearing on the remote viewing research, and parapsychological research in general. In this section of the report, we summarize the major points of agreement and disagreement particularly as they pertain to the remote viewing laboratory experiments conducted as part of the current program. By adopting this more narrow focus, we are intentionally bypassing the evidence for remote viewing or other paranormal phenomena that have arguably been demonstrated in other paradigms (e.g., the autoganzfeld studies) or prior to the present set of laboratory studies. Our position is that the charter for the present evaluation is to examine the current program; evidence obtained by other researchers in other laboratories is not a direct component of this program.

Before beginning this discussion, we feel it important to recognize that, in our opinion, both review papers are of exceptional quality. One of Dr. Hyman's first comments about Dr. Utts' review was that he considered it perhaps the best defense of parapsychological research he had come across. We concur; likewise, we feel that Dr. Hyman's paper represents one of the clearest expressions of the skeptic position we have seen.

At the outset, it should be noted that the two reviewers agreed far more than they disagreed. One central point of agreement concerns the existence of a statistically significant effect: Both reviewers note that the evidence accrued to date in the experimental laboratory studies of remote viewing indicate that a statistically significant effect has been obtained. Likewise, they agree that the current (e.g., post-NRC review) experimental procedures contain significant improvements in methodology and experimental control.

The major remaining area of disagreement concerns attribution of causality. It is Dr. Hyman's position that existence of a statistically significant effect does not allow one to infer that these laboratory experiments have provided an unequivocal demonstration of remote viewing. A statistically significant effect might arise from many sources; more simply stated, the results could have occurred for many different reasons. Until the reasons for the results can be pinpointed, one can not say that fully adequate evidence has been obtained for the existence of any phenomenon, including paranormal phenomena such as remote viewing.

Typically, the process of identifying causes involves both eliminating competing explanations and developing and testing, in Dr. Hyman's terms, "positive" hypotheses. Dr. Hyman states that these competing explanations have not been adequately eliminated; Dr. Utts believes that they have.

Our conclusion is that, although the remote viewing research has made substantial methodological progress in recent years, a central problem remains in the laboratory experiments conducted as part of the present program. The problem is that most if not all significant findings have been obtained by using the same remote viewers, the same judge, the same target set, and the same scoring procedures. This characteristic of the remote viewing research in the program raises the possibility of what has been termed "monomethod bias" — a factor limiting the confidence we can have in drawing conclusions about the generality utility of the phenomenon. Basically, the concept here is that a specific method could have some very subtle influences on performance that only appear over many repetitions or trials; these influences could have nothing to do withh the phenomenon being investigated. Recognition of this point is why both reviewers stress the need for independent replication.

Another consequence of this experimental situation (the same remote viewers, etc.) is that it makes attribution of causation difficult. For example, one interpretation of the significant outcomes obtained is that the **judge** has the ability to influence the results of the remote viewing. As unlikely as this seems, this interpretation cannot be ruled out unless different judges are used.

Although this general problem was of some concern to the reviewers, both were more concerned about one specific aspect of the problem: The methods used in the current program to study remote viewing involve having a judge assess the degree of correspondence between the viewings and a set of targets. In the current studies, this judge has typically been the Principal Investigator, a person who has substantial familiarity with both the viewers and the research paradigm. In addition to the potential confounding noted above, this kind of procedure can potentially to result in "criterion contamination"— some unconscious, nondeliberate influence on the results in a particular direction. As a result, both reviewers stress the point that cross-judge agreement needs to be demonstrated: the studies need to be replicated using independent judges before strong statements can be made about the existence

of remote viewing.

Another issue where there seems to be some remaining disagreement is a variant of the causation problem. Methodological problems are, of course, not the only reason a statistically significant effect might occur but still fail to provide evidence for the existence of a phenomenon. It is quite possible that an effect occurred for reasons other than the proposed cause. In other words, normal psychological processes rather than paranormal processes might account for the effects obtained in the remote viewing studies. Recognition of this fact led both reviewers to stress the point that more attention needs to be given to the identification of underlying causes in remote viewing research.

The reviewers however, differ somewhat in the way they would approach this issue. Dr. Utts believes that the existence of the phenomenon has been adequately demonstrated; therefore, the focus of future research should be on causative mechanisms — how the ability works. Dr. Hyman takes the position that competing hypotheses have not yet been eliminated; thus, he stresses the continued need for replication in different laboratories, with different investigators, etc., so that causal mechanisms, specifically paranormal mechanisms, can be identified.

Our conclusion about this issue is direct: If laboratory research is to continue, this research must be conducted in such a way that causal mechanisms can be articulated, and that also serves to rule out competing explanations.

A final point of disagreement has to do with the nature of the remote viewing phenomenon itself. Dr. Hyman argues that the effect has not been readily replicated in other laboratory settings. Instead, it appears inconsistently, occurring only for certain people at certain times. He argues that the size of the observed effects are small, they do not consistently emerge under certain specified conditions, and that a coherent pattern of results has not been obtained in studies of remote viewing. At the risk of misrepresenting his position, the implication is that unless more is uncovered about the phenomenon, it is not likely to be a fruitful area for further research. Dr. Utts counters with a compelling argument, noting that many effects are statistical in nature, occurring only rarely and under certain conditions. This should not preclude further research into the causes and operation of the phenomenon.

We agree with Dr. Utts that the conditions under which a phenomenon occurs should not determine whether or not it is worth exploring and pursuing scientifically. On the other hand, Dr. Hyman's observations point to a broader problem. If weak, inconsistent effects are obtained in a laboratory-based research program, it is open to question whether the phenomenon under investigation has any practical value. The anomaly may occur from time to time but it occurs in such an erratic fashion that it cannot be used to guide ones' actions. In this regard, we find some agreement between the two reviewers, both of whom question whether the kind of strong consistent effects needed for practical applications have been demonstrated.

Conclusions from the Expert Reviews

In the preceding section we noted the points of agreement and disagreement among the reviewers. We tried furthermore to clarify and reconcile these points of agreement and disagreement. With this background in mind, we now return to the basic questions presented to the reviewers and attempt to draw some firm conclusions about the implications to be drawn from this research review.

The first question presented to the reviewers was whether the evidence indicated the presence of a statistically significant effect. This question was answered in a straightforward fashion: the reviewers agreed that, considered broadly, statistically significant effects have been obtained in these studies. It appears that viewers' descriptions produce hits more frequently than would be expected by chance.

The second question presented to the reviewers considers the nature of these effects. The question to be answered was whether the effects could be attributed to paranormal phenomena. In this regard, the reviewers disagreed, with Dr. Utts arguing positively and Dr. Hyman negatively. Our conclusion from the discussions is that ***direct evidence has not been provided indicating that this paranormal ability of the remote viewers is the source of these effects.*** Attribution in general is difficult to demonstrate; for the present set of laboratory experiments, a primary concern for us is that the same viewers, the same judge, the same target set, and the same scoring procedures were repetitively used. This makes it difficult or impossible to localize the source of the phenomenon.

The third question presented to the panel asked whether we have obtained an adequate understanding of the phenomenon. Do we know how the ability, if it exists, works? Here it is clear that the present research program has failed to identify mechanisms explaining the source of these effects.

The fourth and final question presented to the reviewers was whether the research provides support for intelligence gathering operations. Here the magnitude of the observed effects, their consistency and replicability, and the need for subjective interpretation all seem to argue against potential applications.

Taken as a whole, these answers lead to relatively straightforward general conclusions:

- The laboratory research conducted as part of the present program has identified a statistically significant "anomaly."
- However, the experiments have not provided a convincing demonstration that a paranormal ability is involved.
- The research studies have not identified the nature and source of the effect.
- There is no evidence that the phenomenon would prove useful in intelligence gathering.

4. Evaluating the Utility of Remote Viewing in Intelligence Operations

In the preceding section, we presented several conclusions bearing on evidence for the existence of remote viewing and on the nature of the phenomenon. In this section, we will, for the moment, assume that the phenomenon does indeed exist. This leads to questions of the utility of the phenomenon for intelligence applications. As noted earlier, our evaluation of remote viewing in terms of its value to the intelligence community had three components:

- assessment of task requirements in relation to the boundary conditions believed to influence application of the phenomenon,
- interviews with various participants in the program, and
- user assessments of the information provided by the remote viewing process.

In this section, we consider each of these components.

Research on Boundary Conditions

Historically, the literature on paranormal phenomena, and on remote viewing in particular, has paid relatively little attention to establishing boundary conditions — more specifically, the conditions under which effects are and are not observed. This point has been made in numerous earlier critiques of parapsychological research (Druckman & Swets, 1988; Hyman, 1994; Swets & Bjork, 1990). In response to these critiques, several investigators have initiated research intended to provide more clear-cut evidence bearing on the conditions influencing when effects are and are not observed in parapsychological studies (Bem & Honorton, 1994; Honorton, 1994). Research on remote viewing reflects this same general trend. In fact, many of the studies conducted in recent years have expressly examined this issue (May, Luke, & James, 1994; May, Luke & Lantz, 1993).

Senders and Distance. When one considers the remote viewing paradigm, one boundary condition immediately comes to the fore with regard to intelligence applications. During information gathering, it is unlikely that a "beacon" or sender will be at the site. In fact, the interviews with end users and the remote viewers indicate that only in the case of missing persons is a sender likely to be involved in information collection. If senders are necessary for the phenomenon to occur, its utility would be dramatically reduced.

The laboratory experiments conducted as part of the current program avoid this issue; since the targets are photographs kept in another part of the laboratory, the equivalent of a "sender" would be the laboratory assistant who retrieves the photograph (or the computer that selects it). If one considers a "sender" as either a transmitter of information or an active participant in the remote viewing, this paradigm essentially does not have one. However, a recent study by May, Spottswood, and James (1994) has specifically addressed this issue. Their findings indicate that significant remote reviewing effects are obtained even when senders are not available.

Another boundary condition that might influence the utility of remote viewing in intelligence collection is the distance of the target from the viewer. More specifically, for many of the tasks involved in intelligence applications, it may be impossible to place the viewer anywhere near the target. Thus, if the accuracy of remote viewing is limited by distance, this condition may, in turn, restrict its value. However, the work of May and his colleagues (May, Luke, & James, 1994; May, Luke, & Lantz, 1994) indicates that remote viewing may not be related to distance.

Targets. A set of studies conducted by Lantz, Luke, and May (1994) examined another boundary condition of particular importance when assessing potential applications of the remote viewing phenomenon. Most recent research on remote viewing has been based on a fixed set of "static" targets, drawn from National Geographic photographs. Lantz, Luke, and May varied target type by including a new set of targets covering a wider range of novel stimuli, referred to as "dynamic" targets. Perhaps another relevant aspect of these new targets is that the remote viewers in the standard paradigm are very familiar with the National Geographic photographs, but were unfamiliar with the dynamic targets. The researchers found that significant effects were obtained for static but not dynamic targets.

This finding is of some importance from an operational intelligence perspective. In Appendix C, the interview reports, end users and remote viewers describe the operational targets presented to remote viewers. These targets represent a diverse set of potentially novel stimuli ranging from the location of ships to the likely background characteristics of a person. The findings obtained in the Lantz, Luke, and May study suggest that accurate viewings are less likely to be obtained for these dynamic "real-world" targets, a finding which, if replicated, represents a severe constraint on the utility of remote viewing to the intelligence community. Even if, as May, Spottswood, and James (1994) point out, those effects can be explained in terms of target "bandwidth," there is no reason to assume that intelligence targets will conform to a limited range of predefined bandwidths.

Information Requirements and Specificity. Another set of boundary conditions examined in recent studies considers the nature of the targets and the information to be assessed by ***judges.*** Intelligence consumers place a premium on the availability of specific, concrete, potentially verifiable information. On the other hand, the information obtained from remote viewings tends to be stated in broad, vague terms; a critical role must be played by judges or intelligence analysts, who must interpret the remote viewers' reports.

Recognizing this problem, a series of studies (May, Luke, & Lantz, 1993) were initiated intended to provide more concrete judgments framed in terms of the presence or absence of specific categories of objects, a technique referred to as "binary coding." Results from these studies indicated that when relatively specific information was used in judging matches, weak or insignificant effects were obtained.

Specificity, as a boundary condition, can also be assessed in terms of targets. This issue was examined in a meta-analysis (May, Luke, & Lantz, 1993) contrasting the effects obtained for different types of targets. They found that effect sizes were related to target complexity, with more dynamic targets showing larger effect sizes. Incidentally, and seemingly in contradiction to the Lantz, Luke, and May (1994) study cited above, they found insignificant effects for static photograph targets. This pattern of effects, if intrinsic to the remote viewing process, would seem to limit applications to those involving more complex targets requiring less specific descriptions. Furthermore, it may also represent a significant limitation on applications of the remote viewing phenomenon where specific information is required.

Another aspect of specificity has to do with the degree of accuracy of the remote viewers' reports. Most remote viewing reports contain a large number of potentially interpretable components, an unknown percentage of which may not be related to the target. The end users cannot know which components are or are not related. This makes it difficult for end users, particularly if they are not highly trained, to separate valid information from irrelevant information, a problem that may prohibit effective operational applications of the results of the viewing process.

Feedback. Perhaps the most widely accepted boundary condition applying to remote viewing is the need for feedback. It is commonly held that the accuracy of viewings depends on the viewer receiving feedback following production of the viewing. In fact, one explanation of the remote viewing phenomenon is that viewings represent a form of precognition where the viewers identify future outcomes or events that they themselves will experience; in effect, they "communicate" with themselves across time (May, Lantz, & Piantineda, 1994). The important point here is that in "real-world" information gathering situations, it may be impossible for viewers to acquire the kind of feedback needed for accurate viewings.

Remote Viewer Training. A final boundary condition frequently noted in the literature on remote viewing pertains to the nature and development of the presumptive remote viewing ability. Typically, remote viewing is held to represent a relatively rare talent not widely distributed in the population. Recent studies have suggested that associative learning techniques may contribute to improved performance on remote viewing tasks (May, Luke & Lantz, 1993). However, little evidence is available indicating that the capacity for producing accurate viewings can be systematically developed. This has implications for information gathering operations, since the burden would fall on the ability of the sponsoring agency to recruit and/or select individuals who already possess remote viewing skills.

Other Boundary Conditions. In the preceding discussion, we have limited commentary to those boundary conditions directly addressed in the literature. However, a number of other boundary conditions exist that might limit potential applications of the remote viewing phenomenon. One such limitation on potential applications of the phenomenon involves the assessments of the information provided by the remote viewers. Typically, when judgmental data, such as target matches, are to be used in decision making,

evidence is required indicating that end-user analysts arrive at much the same decisions. If analysts (or trained judges) cannot arrive at similar decisions, it is impossible to know what kind of actions should be taken. Unfortunately, evidence is not available indicating the degree of interrater, or cross-judge, agreement. Accordingly, it is difficult to assess whether different analysts or judges will reach similar conclusions given the same data. These differences in assessments, if indeed they exist, may set another important limitation on potential applications of remote viewings.

Another boundary condition that has potential importance for remote viewing operational applications pertains to the degree of prior practice. In the remote viewing research studies conducted as part of the present program, both judges and the viewers have had years working together on a rather limited target set. Putting aside for the moment potential cueing and rule-based learning effects resulting from long periods of practice, in most field settings viewers and judges or analysts will have only limited opportunities to work together. As a result, it is open to question whether the findings obtained in these laboratory experiments can be extended to field settings.

Boundary Conditions: Conclusions

The issue of boundary conditions is part of a broader problem in generalizing research findings to operational settings. Typically, generalization from laboratory to operational settings is contingent on three requirements (Cook & Campbell, 1979). First, there is a need to demonstrate that an observed effect is sufficiently robust that it can be observed using different methods. Second, explicit verified causal explanations, reflecting an understanding of the source of observed effects, need to be provided. Third, alternative causal explanations, for example judges and viewers learning implicit rules, must be ruled out. The laboratory experiments in the current research program conducted to date have relied on one method, judgmental matching; plausible causal mechanisms has not been identified; and studies have not been conducted drawing out competing explanations. Lacking this evidence, drawing strong conclusions concerning the potential operational applications of remote viewing is at best tenuous, and at worst misleading.

Taken as a whole, prior laboratory experiments examining the boundary conditions related to remote viewing have clearly provided an important foundation for establishing the

nature of the phenomenon. On the other hand, however, the findings obtained in these studies actually argue ***against*** operational applications in the intelligence community. Broadly speaking, it appears that the conditions under which intelligence information is gathered, the nature of the targets, the unavailability of feedback, and the inconsistency with which accurate viewings are obtained may all limit the usefulness of the phenomenon in intelligence operations. Further, many significant boundary conditions have not been examined and the scientific basis for generalizing from laboratory to field settings has not been provided. Given these observations, it appears that the existing research does not justify operational applications and, in fact, vis a vis the known boundary conditions, argues against operational applications.

Interview Findings

The second component of the operational evaluation was the interviews conducted with the various parties who had direct involvement in the remote viewing program in various intelligence operations. A detailed description of the procedures used in interviewing program participants and the results obtained in individual interviews is presented in Appendix C; here, we briefly summarize the interview procedures before turning to the principal findings which they generated.

Structured interviews were conducted with the three principal types of program participants: recent users of the remote viewing service, the remote viewers themselves, and the Program Manager. Separate interview protocols were developed for each type of respondent to capture their unique perspectives. The interview questions were written to obtain objective information about characteristics of the program as opposed to more subjective evaluative assessments of parapsychological phenomena. All questions were reviewed by AIR psychologists.

Interview Results: End Users. The interviews conducted with representatives of end user groups were particularly noteworthy: these interviews directly examined the accuracy and utility of the information being provided by the remote viewers and provided user assessments of the operational value of the viewings.

As can be seen in the individual interview reports (Appendix C), initial interest in the potential uses of remote viewing was linked to contact with the Program Manager. Typically, users decided to try out the viewings on an exploratory basis. The primary reasons users were willing to explore the potential value of this technique were that it might (a) provide information otherwise difficult or impossible to obtain, and (b) provide information more rapidly and economically than other sources. All of the users indicated that viewings might be especially attractive when other, more traditional options had been exhausted.

The tasks assigned to viewers involved a number of different types of targets. Although all of these targets were relevant to operational intelligence issues, the targets were not the same as those reported in the laboratory experiments. The targets included people, their backgrounds, and their actions, as well as their locations. Likewise, the typical tasks asked of the remote viewers were dissimilar to laboratory conditions: they were asked to identify objects at specified locations and where certain objects might be found. When responding to these task demands, the remote viewers typically had at least some background knowledge; in some cases,this background knowledge was substantial. Typically, three independent viewings were obtained. The results of each viewing were summarized in a three or four page report which included drawings and verbal descriptions.

With regard to the information provided in each report, five general observations emerged across all interviews:

- The information provided in the reports was stated in broad, vague terms.
- The reports were most likely to prove accurate with regard to general stereotypical characteristics of the situation. Such results might be attributable to the background information available to the viewers.
- The reports were most likely to prove inaccurate with regard to concrete specifics of the task. For example, the reports often were not consistent with key known facts nor did they provide information about the unique features of a location.

- A large amount of irrelevant and often inaccurate information was contained in the reports, thus making them difficult to apply without substantial interpretive effort.

- The reports independently provided by different remote viewers displayed many inconsistencies.

Because the viewings were not consistent with each other, because inaccuracies were observed, and because the information lacked the specificity needed for intelligence operations, ***the viewings were never used as a primary source of evidence in making decisions.*** In fact, even the most favorable of the user groups found the information inadequate for operational decisions. Instead, it was used to fill in background information on people that could not be readily obtained through available assets. All of the users noted that viewings should only be considered as providing supplemental information, and should be judiciously interpreted if used at all.

The perceived utility of the information varied with the nature of the organization using the viewing services. Typically, viewings were seen as more valuable when the organization involved did not have adequate assets available and when there was no other convenient way of obtaining requisite information. When other information sources were available, the users found the viewings to be less valuable. In all cases, the users remarked that they would continue to consider working with the viewers on an exploratory basis. However, in most cases the users were not willing to invest their own operational funds to obtain viewings. In part this was an issue of cost; in part, however, the users felt that because the remote viewing process was controversial, official acknowledgement of its value was essential for routine use in intelligence gathering. This point was illustrated by the manager of the most favorable user group, who noted that many of the analysts were critical of the technique and unwilling to apply it unless specifically directed to do so.

Remote Viewers. The three remote viewers we interviewed were all working in the intelligence community when they were recruited to be viewers. Although one viewer was subjected to a formal screening process, all were primarily selected on the basis of their interest in parapsychology. The viewers all received formal training as described in Appendix C and continued to extend this initial training through active exploration of various

parapsychological techniques. The viewers used a variety of techniques to produce viewings including automatic writing, meditation, and channelling.

With regard to the procedures used in generating viewings, a number of factors were noted that influenced the nature and outcomes of the viewers' efforts. First, the viewers disagreed as to how useful it was to have background information concerning the nature of their tasking. Second, the viewers noted that the usefulness of viewings depended on having sophisticated, knowledgeable users who accepted the technique and were willing to actively work with what was necessarily somewhat ambiguous information. Third, they uniformly agreed that the types of information requested — such as activities of individuals and specific locations of specific people or objects — were not particularly compatible with their remote viewing experiences.

In addition to these observations, the viewers indicated a number of other considerations that might contribute to program outcomes. Some of these considerations, such as balancing work load and the repetitive nature of specific tasks, related to better management. Other considerations, such as management support and the availability of managers familiar with the process, related to the day-to-day management of the program.

From a parapsychological perspective, a potentially disturbing aspect of the interview was that the remote viewers commented that their written reports were occasionally inconsistent with their viewings. They themselves considered it appropriate to make the written reports consistent with known characteristics of the target, even if the report was different from their viewing. Similarly, the remote viewers commented that previous managers also changed viewers' reports to make them more consistent.

The Program Manager. Appendix C presents a summary of the interview with the most recent program manager. Although the manager did not have any prior background in parapsychology, he did bring to the program a background in operational intelligence. This background played an important role in allowing the manager to recruit clients. Most clients, however, had been interested in the program in the past — an interest maintained in part because the program provided a rapid turnaround of otherwise difficult-to-obtain information.

The manager noted that none of the user groups had used the viewings as a primary basis for operational decisions. Instead, they were using the service to explore potential applications of the technique or, alternatively, as a source of supplemental information. Roughly half of the groups contacted agreed to participate in the program on an exploratory basis.

During his interview, the manager stated that recruitment of clients was in large measure due to the Congressional mandate. He noted however, that acceptance of the program was limited by the controversy surrounding remote viewing. The research component was seen as potentially helpful in addressing this issue, although it did not contribute much to operational management. The manager felt that the program's long-term results, particularly in foreign assessment, would be enhanced by declassifying the research work and assigning responsibility for it to another government organization such as the Department of Justice or the National Science Foundation. Along similar lines, he suggested that remote viewing services be provided on a contract basis whenever the need for them arose.

Conclusions

Before turning to the broader conclusions flowing from these interviews, certain limitations inherent in this particular evaluative effort should be mentioned. First, the evaluation was limited to recent operations. Although this factor reduces and focuses the range of information considered, it does not help to ensure the accuracy of the information reported. Indeed, if previous managers or the viewers themselves changed viewer reports, then relying on this body of information could potentially be misleading with respect to the remote viewing phenomenon or to the operational use of remote viewing. Second, it should be recognized that interviews generate qualitative information that does not permit us to draw strong quantitative conclusions. Third, it is clear that this report necessarily focuses on actual operations and not on operations as they might occur under ideal conditions.

Even bearing these points in mind, we believe that the interviews have a number of significant implications. The most important implication pertains to the nature of the information provided by the remote viewing process and the potential applications of this information in intelligence operations. The information provided by viewings:

- is vague and general in nature;
- is not consistent across independent viewings;
- lacks specific content consistent with known facts of the case; and
- includes a large amount of irrelevant, often erroneous, information.

Intelligence operations are contingent on the availability of relatively specific information which is reliable, consistent, and potentially verifiable. The lack of specifics apparent in the viewings may well be an intrinsic characteristic of the process — a point noted by the remote viewers. Nonetheless, this vagueness and the limited agreement evident in viewers' reports make it impossible to apply this information in decision making. One potential strategy for addressing this issue would be to seek more specific information in viewings. However, existing laboratory research provides little support for the likely success of this approach. Alternatively, one might explicitly train users to work with this kind of broad, rather vague information in such a way that they would attend to both its strengths and its weaknesses. However, this kind of effort would require a strong commitment from sponsoring organizations and unambiguous, high-level support for the program.

In addition to the problems associated with the reliability and vagueness of the information provided, the viewings typically contain a large amount of irrelevant information and often fail to capture key aspects of the case. These characteristics of the viewings reduce user confidence in the information provided. Further, the nature of this information is such that the burden is effectively placed on the user in sorting out what is and is not relevant. Under these conditions it is quite possible that the information provided by viewings will lead users down a number of blind alleys, thereby resulting in misallocation of intelligence resources. Alternatively, it may do little more than reinforce existing preconceptions and stereotypes about the targets — a trend evident in the user interviews.

This observation brings us to the accuracy issue. Although the viewings tended not to be accurate with regard to concrete specifics, a characteristic which held true regardless of the type of tasking, the users felt that they were more accurate in describing broad background factors involved in the case. The problem here, of course, is that this outcome may simply

reflect the availability of background information and the logical use of relevant analytic cues. This problem was, in fact, noted by the viewers when they indicated that they modified reports to make them consistent with know background factors. This observation suggests that the viewers may not be providing any more information than could be obtained from perceptive analysis working without the aid of remote viewing techniques.

Still another limitation on the operational value of the information being provided pertains to its source. Remote viewing is a controversial phenomenon. Accordingly, routine operational use of the resulting information across multiple units in the intelligence community may well be contingent on a compelling demonstration of the phenomenon and broad acceptance of the resulting findings. Without this acceptance, users will discount the information provided and attempts to force analysts to use this information are likely to cause conflict. If, however, comparable information were provided by the viewings, it could be argued that the program was more cost-effective than conventional techniques.

The fact that the information being provided is too vague and unreliable to permit operational application suggests that the viewing program is of limited value to the intelligence community. This conclusion, however, pertains to the current program. It is, of course, always possible that a substantially restructured and redirected research effort along with a substantial additional investment of resources might yield better results. Given the costs involved, however, and the need for specific, reliable information, any investment along those lines should be contingent on successful demonstration of not only the existence of the phenomenon but also its ability to consistently produce specific accurate information.

Our foregoing observations of the interview results can be summarized as follows:

- The viewing process has not provided information of adequate specificity and reliability for use in intelligence gathering.
- The process is not widely accepted and is not seen as essential to intelligence gathering.
- The process is often used experimentally because the needed information is difficult to obtain.

- Accuracy is limited to broad superficial characteristics of the case and may simply reflect good logical analysis.

- The available research does not directly support operations and new types of research are needed to justify current operational use.

User Assessments

A third source of evidence used to evaluate operational utility consisted of analyses of user feedback. This feedback was in the form of assessments of product (i.e., the remote viewers' reports) accuracy and value collected from end users. This information was collected by a representative of the U.S. Government in the Spring and Summer of 1995. The results of these analyses have been compiled in a report; this report is included as Appendix D; we summarize the findings and implications below.

The program office asked each user submitting tasks to evaluate the accuracy and value of the viewings. Evaluations were available for forty viewings obtained in 1994 and 1995. Accuracy (i.e., "Is the information accurate?") was rated on a six-point scale with 1 indicating high degrees of accuracy. Value (i.e., "What is the value of the source's information?") was rated on a five-point scale where a 1 indicated major significance, while a 5 indicated no value. It should be noted that separate ratings were obtained for each viewer's report on each tasking requested by an organization.

The average accuracy rating was 3.0 across viewers and tasks, indicating that there might "possibly" be some accurate information provided. The average value rating was 3.5 (out of 5) indicating that the information was of relatively low value. This pattern of results is consistent with the interview findings. Remote viewings were better at capturing broad background information rather than the concrete, specific information needed for intelligence operations.

Bearing this general conclusion in mind, a number of other questions about the remote viewings can be addressed by these data. For example, one important question is whether any particular remote viewer did particularly well. The accuracy scores of all three viewers were within a tenth of a standard deviation from the mean, while value scores were within a

quarter of a standard deviation. Thus, it appears that no one viewer performed exceptionally well or poorly.

Another question is whether the remote viewers performed better on some tasks than others. Because tasks differed by the nature of the organization, this question can be addressed by contrasting the evaluations obtained from different organizations. Two organizations, one concerned with tracking people and one concerned with combat intelligence, indicated that the information provided was of substantially greater accuracy and value than was reported by the other user organizations. In the case of the two organizations providing positive evaluations, however, the viewers had substantial general background information concerning the targets, a finding suggesting that the ratings of outcomes of viewings may be dependent on the amount of pre-existing background information.

These findings are noteworthy because they confirm a conclusion derived from the interviews. More specifically, viewings apparently were of limited accuracy, with accuracy being linked to the availability of general background information. Such background information, however, was not believed to have much operational value.

Information Gathering Applications: Operational Conclusions

The findings obtained in this evaluation of the intelligence applications of remote viewing lead us to the conclusion that ***remote viewing*** as used in the present program has limited value for the intelligence community as an information gathering technique. The basic considerations that lead us to this conclusion are:

- Conditions under which significant effects are observed in experimental laboratory research are, for the most part, unlikely to occur in intelligence operations. Not only will feedback be unavailable, but the target pool will typically be unconstrained.

- Information provided by the remote viewing technique tends to be vague and ambiguous and it appears difficult, if not impossible, to consistently obtain accurate information from the remote viewers across a range of targets.

These problems with the consistency and accuracy of viewings were also apparent in the end user interviews. In these interviews, viewings were found to be too broad and vague; the viewings failed to provide the concrete, specific information needed for actionable intelligence. Further, there were indications that its potential usefulness was limited to information which could be acquired in other ways.

Thus, the evidence accrued from research, interviews, and user assessments all indicate that the remote viewing phenomenon has no real value for intelligence operations at present. In fact, given the findings obtained to date and the nature of intelligence operations, one must question whether any further applications can be justified without major theoretical and practical advances in our understanding of the phenomenon, assuming it exists at all.

5. Conclusions

In the preceding sections of this report, we have presented a variety of evaluative data bearing on the existence of the paranormal phenomenon known as remote viewing and its potential applications in intelligence gathering. This multifaceted evaluation effort was structured to ensure a fair, unbiased evaluation of both the research program and its intelligence applications. As is the case with any objective, relatively sophisticated program evaluation effort, many pieces of evidence bearing on different aspects of the program have been presented. In this section, we summarize the basic conclusions flowing from this evaluation effort.

Summary of Key Findings

Two expert reviewers, one known to be a sophisticated advocate of the study of paranormal phenomena and one viewed as a fair-minded skeptic, reviewed the laboratory experiments conducted as part of the current program that bear on the existence of the remote viewing phenomenon. They focused primarily on recent, better-controlled laboratory studies, drawing from other sources as needed to ensure a comprehensive evaluation of the research literature. Although the reviewers disagreed on some points, on many points they reached substantial agreement.

The first important point of agreement concerns the existence of a statistically significant effect, which leads to the following finding:

- ***A statistically significant effect has been observed in the recent laboratory experiments of remote viewing.***

However, the existence of a statistically significant effect did not lead both reviewers to the conclusion that this research program has provided an unequivocal demonstration that remote viewing exists. A statistically significant effect might result either from the existence of the phenomenon, or, alternatively, to methodological artifacts or other alternative explanations for the observed effects.

It is with regard to the explanation for these effects that the two reviewers differ most clearly. One reviewer argues that the procedures used in recent studies rule out many, but not all, methodological explanations. The other reviewer argues that the consistency of the results obtained across experiments strongly suggests the existence of the paranormal phenomenon. We concluded that

- ***The experimental research conducted as part of the current program does not unambiguously support the interpretation of the results in terms of a paranormal phenomenon.***

Both reviewers agreed that one important methodological problem has not yet been addressed. Specifically, only one judge — apparently the Principal Investigator — was used in assessing matches throughout these experimental studies. As a consequence, there is no evidence for agreement across independent judges as to the accuracy of the remote viewings. Failure to provide evidence that independent judges arrive at similar conclusions makes it difficult to unambiguously determine whether the observed effects can be attributed to the remote viewers' (paranormal) ability, to the ability of the judge to interpret ambiguous information, or to the combination or interaction of the viewers and the judge. Furthermore, given the Principal Investigator's familiarity with the viewers, the target set, and the experimental procedures, it is possible that subtle, unintentional factors may have influenced the results obtained in these studies. Thus, until it can be shown that independent judges agree, and similar effects are obtained in studies using independent judges, it cannot be said that adequate evidence has been provided for existence of the remote viewing phenomenon.

Both reviewers agree that no compelling explanation has been provided for the observed effects. One reviewer considers the investigation of the determinants of remote viewing as the next necessary step, but does not see it as essential to continue to conduct experiments designed solely to demonstrate the existence of the phenomenon. The other reviewer argues that without identifying causal mechanisms and explicitly providing evidence that alternative explanations cannot account for the observed effects, we cannot say we have convincing evidence for the existence of a phenomenon. Essentially, this position holds that an observed effect may arise for many reasons; to say a phenomenon has been demonstrated we must know the reasons for its existence.

Our conclusion is that at this juncture it would be premature to assume that we have a convincing demonstration of a paranormal phenomenon. In fact, until a plausible causal mechanism has been identified, and competing explanations carefully investigated, we cannot interpret the set of anomalous observations localized to one laboratory with one set of methods. Given these observations, and the methodological problems noted above, we must conclude that

- ***Adequate experimental and theoretical evidence for the existence of remote viewing as a parapsychological phenomenon has not been provided by the research component of current program. A significant change in focus and methods would be necessary to justify additional laboratory research within the current program.***

This is not to say definitively that paranormal phenomena do not exist. At some point in time, adequate evidence might be provided for the existence of remote viewing. With this point in mind, we considered the potential applications of remote viewing in intelligence gathering.

The first consideration involves the conditions under which remote viewing occurs and if those conditions constrain its application for intelligence purposes. Prior research suggests that distance is not a constraint and indeed that a sender or "beacon" may not be necessary. However, other characteristics of intelligence gathering indicate that remote viewing is of little value. Intelligence operations do not provide targets of a fixed bandwidth; rather, targets and target types are highly variable. Moreover, the apparent necessity for feedback to the remote viewers would preclude its use in intelligence gathering operations. Finally, intelligence information is most valuable if it is concrete and specific, and reliably interpretable. Unfortunately, the research conducted to date indicates that the remote viewing phenomenon fails to meet those preconditions. Therefore, we conclude that

- ***Remote viewing, as exemplified by the efforts in the current program, has not been shown to have value in intelligence operations.***

This point was also graphically illustrated in the user interviews, where it was found that remote viewings have **never** provided an adequate basis for "actionable" intelligence

operations — that is, information sufficiently valuable or compelling enough so that action was taken as a result. If a phenomenon does not contribute to intelligence operations, it is difficult to see what justification exists for its continued application. This is particularly true in the case of remote viewing, where a large amount of irrelevant, erroneous information is provided and little agreement is observed among viewers' reports.

Particularly troublesome from the perspective of the application of paranormal phenomena is the fact that the remote viewers and project managers reported that remote viewing reports were changed to make them consistent with know background cues. While this was appropriate in that situation, it makes it impossible to interpret the role of the paranormal phenomenon independently. Also, it raises some doubts about some well-publicized cases of dramatic hits, which, if taken at face value, could not easily be attributed to background cues. In at least some of these cases, there is reason to suspect, based on both subsequent investigations and the viewers' statement that reports had been "changed" by previous program managers, that substantially more background information was available than one might at first assume. Give these observations, it is difficult to argue that available evidence justifies application of remote viewing in intelligence operations.

In summary, two clear-out conclusion emerge from our examination of the operational component of the current program. First, as stated above, evidence for the operational value of remote viewing is not available, even after a decade of attempts. Second, it is unlikely that remote viewing — as currently understood — even if existence can be unequivocally demonstrated, will prove of any use in intelligence gathering due to the conditions and constraints applying in intelligence operations and the suspected characteristics of the phenomenon. We conclude that:

- ***Continued support for the operational component of the current program is not justified.***

REFERENCES

Bem, D.J. & Honorton, C. (1994) Does PSI Exist? Replicable Evidence for an Anomalous Process of Information Transfer, Psychological Bulletin, 115, 4-18.

Cook, T. D., & Campbell, P.T. (1979) Quasi-experimentation: Design and Analysis Issues for Field Settings. Chicago, Il; Rand Mc Nally

Druckman, D., & Swets, G.A. (Eds.) (1988) Enhancing Human Performance: Issues, Theories, and Techniques. Washington, DC: National Academy Press.

Honorton, C. (1985) Meta-Analysis of PSI Ganzfeld Research: A Response to Hyman. Journal of ParaPsychology, 49, 51-91.

Hyman, R., (1994) Anomaly or Artifact? Comments on Bem and Honorton. Psychological Bulletin, 115, 19-24.

James, L.R., Muliak, S.A., & Brett, J.M. (1982). Causal Analysis: Assumptions, Models, and Data. Beverly Hills, CA: SAGE

Lantz, N. D., Luke, W.L., & May, E.C. (1994) Target Sender Dependencies in Anomalous Cognition Experiments. Menlo Park, CA: Science Applications International Corporation.

May, E.C., Lantz, N.D., & Piantineda, J. (1994) Feedback Considerations in Anomalous Cognition Experiments. Menlo Park, CA: Science Applications International Corporation.

May, E.C., Luke, W.L., & James, C.C. (1994) Phenomological Research and Analysis. Menlo Park, CA: Science Applications International Corporation.

May, E.C., Luke, W.L., & Lantz, N.D. (1993) Phenomological Research and Analysis, Menlo Park, CA: Science Applications International Corporation.

May, E.C., Spottiswoode, S.J. & James C. C. (1994) Managing the Target Pool Bandwidth: Noise Reduction for Anomalous Cognition Experiments. Menlo Park, CA: Science Applications International Corporation.

Messick, S. (1989) Validity. In R.L. Linn (Ed) Educational Measurement (PP 13-104). New York, Macmillian

Swets, J.A., & Bjork, R.A. (1990) Enhancing Human Performance: An Evaluation of New Age Techniques Considered by the U.S. Army. Psychological Science, 1, 85-96.

Approved For Release 2002/05/22 : CIA-RDP96-00791R000200180005-5

APPENDIX A: REVIEWER VITAS
CURRICULUM VITAE[1]

September 1994

Name: Ray Hyman

Present position: Professor of Psychology
Department of Psychology
University of Oregon
Eugene, Oregon 97403

Date and place of birth: June 23, 1928
Chelsea, Massachusetts

Academic degrees: 1950 A.B., Boston University
Phi Beta Kappa
Honors in Psychology
1952 M.A., The Johns Hopkins University
1953(FEB) PhD., The Johns Hopkins Univ.

Positions held:

Jul 1953-Jun 1961 Assistant Professor of Social Psychology, Harvard University

Jul 1958-Jun 1961 Consultant in Behavioral Research, General Electric Company

Sep 1961-Aug 1964 Associate Professor of Psychology, University of Oregon

Sep 1964- Professor of Psychology, University of Oregon

Sep 1967-Aug 1968 Fulbright-Hays Research Scholar, University of Bologna, Italy

Sep 1976-Dec 1976/Jun 1977-Sep 1977/Jun 1978-Sep 1978 NSF Faculty Fellowship in Science Applied to Societal Problems

Sep 1982-Jun 1983 Visiting Professor of Psychology, Stanford University (Thomas Welton Stanford Chair for Psychical Research)

[1]This is a very abbreviated and somewhat modified version of my full resume.

Additional Positions and Experience[2]

Mar 1949-Jul 1949 Statistician for antihistamine study, Allergy Fund of Boston

Jan 1953-Jul 1953 Consultant to Systems Coordination Division, Naval Research Laboratory, Washington, D.C.

Jul 1974-Aug 1974 Subcommittee on Preparing Problems and Examples for Committee on the Mathematical Training of Social Scientists, Social Science Research Council

Oct 1955-Jan 1956/Oct 1956-Jan 1957/Oct 1957-Jan 1958 Director of Statistical Workshop for Psychology Department, Clark University

Oct 1957-Jun 1958 Statistical Consultant, Age Center of New England

1970-1975 Statistical Consultant, Roseburg VA Hospital

Some Current or Recent Positions:

Executive Council, Committee for the Scientific Investigation of Claims of the Paranormal

Editorial Board, The Skeptical Inquirer

Member of National Research Council's Committee on Techniques for the Enhancement of Human Performance (1985-1991)

Publications:

BOOKS 7 MONOGRAPHS:

Bush, R.R., Abelson, R.P., & Hyman, R. (1956). Mathematics for psychologists: examples and problems. New York: Social Science Research Council.

Vogt E.Z., & Hyman, R. (1979, Revised Edition). Water Witching U.S.A. Chicago: Univ. of Chicago Press.

Hyman, R. (1960). Some experiments in creativity. New York: General Electric Company (101 pages).

[2]A Partial listing.

Hyman, R. (1960). Methods for the study of creativity: an evaluation of current research. New York: General Electric Company (204pp).

Hyman, R. (1964). The nature of psychological inquiry. Englewood Cliffs, New Jersey: Prentice-Hall.

Hyman, R. (1965). Creativity and the prepared mind: preconceptions in creative achievement and creativity research. National Art Education Association.

Hyman, R. (1989). The elusive quarry: a scientific appraisal of psychical research. Buffalo, NY: Prometheus Books.

Hyman, R. (in preparation). How smart people go wrong: cognition and human error.

ARTICLES:

I have published over 200 articles in professional journals on perception, pattern recognition, creativity, problem-solving and critiques of the paranormal. I consider all my publications as dealing with aspects of my major interest in human error and deception. I have also published articles in magic journals and have won awards for inventing new conjuring effects.

Public Talks and Media Presentations

During this period I have given talks to public schools, civic groups, and other organizations. I also have appeared on several television and radio programs. I serve as a resource to the media on various topics related to the paranormal, deception, mysticism, the occult, and human error. In this connection I have appeared on all the major networks, on Cable Network News, the Larry King Live show, Italian Television, Canadian Television, BBC television, and Nova. I see these appearances as part of my program to help educate the public about how human cognition both enables us to cope with problems and makes us susceptible to illusion and deception.

My experience and Credentials Relevant to Alleged Psychic Phenomena

Since 1953, I have been called upon by various governmental agencies to investigate or evaluate paranormal claims. Some examples would be the evaluation of a lady who allegedly could read with her finger tips, the evaluation of the claim by a group of engineers that they could teach psychics to gather information from photographs; the assessment of parapsychological research by American and foreign investigators; and the observation of metal bending and other alleged miracles by Uri Geller. I have also served as consultant and expert witness in court cases involving psychics or related paranormal claims. I have

appeared on several radio and television programs in the United States, Great Britain, Canada and Italy to comment upon paranormal claims and to demonstrate how alleged paranormal phenomena can be duplicated by simple trickery and psychological manipulation.

I earned my way through college performing mentalism and reading palms. I have demonstrated psychic readings on several television and radio shows. I have done research and written articles on why people can falsely believe that their psychic readings were accurate and depended upon occult information. I have been invited to the Euroskeptics' conference in Ostende, Belgium to conduct a workshop on psychic readings at the end of September, 1994.

JESSICA M. UTTS

Current Employment:

Professor, Division of Statistics
University of California, Davis
Davis, CA 95616

916-752-6496 (office phone)
916-752-7099 (FAX)
jmutts@ucdavis.edu

Education

BA State University of New York at Binghamton, Math and Psychology, 1973
MA Pennsylvania State University, Statistics, 1975
PhD. Pennsylvania State University, Statistics, 1978

Previous Titles and Visiting Positions:

Visiting Senior Research Fellow, Department of Psychology, University of Edinburgh, Scotland, Winter 1994.

Director, Statistical Laboratory, University of California, Davis, 7/88-6/93

SRI International, Cognitive Sciences Program, Menlo Park, CA, Visiting Scientist, 6/87-8/88

Stanford University, Dept. of Statistics, Visiting Professor, 6/83-9/83 and 9/84-6/85

University of California, Davis, Division of Statistics, Assistant and Associate Professor, 7/79-6/79

University of California, Dais, Dept. of Mathematics, Assistant Professor, 7/78-6/79

Penn State University, Dept. of Statistics, Instructor, 3/78-8/78

Academic Honors:

Fellow, American Association for the Advancement of Science, 1992
Fellow, Institute of Mathematical Statistics, 1991
Fellow, American Statistical Association, 1990
Academic Senate Distinguished Teaching Award, University of California, Davis, 1984
Magnar Ronning Award for Teaching Excellence, University of California, Davis, 1981
Phi Beta Kappa, State University of New York at Binghamton, 1973

Professional Affiliations and Offices (alphabetical order):

American Association for the Advancement of Science: various roles in Section U (Statistics)
American Statistical Association: President, State College PA Chapter, 1977-78
Biometric Society, western North American Region: President, 1986; Reg. Comm., 1982-84; Program Chair, 1983
Caucus for Women in Statistics: President, 1988
Institute of Mathematical Statistics: Treasurer, 1988-1994; Assistant Program Secretary, 1980, 1989
Parapsychological Association: Representative to AAAS Section X, 1989-Phi Beta Kappa: President of UC Davis Chapter, 1984-85, Vice President, 1983-84
Society for Scientific Exploration: Council Member, 1987-93

Major Consultations and Panels:

National Academy of Sciences, Panel on the Evaluation of AIDS Interventions
Congressional Office of Technology Assessment, Panel to Assess Defense Technologies
National Park Service, Statistics Short Course for Resource Management Trainees
California Department of Health Services, Course on Statistics for Groundwater
SRI International Cognitive Sciences Program, Consultant
California Public Utilities Commission, Consultant
Hershey Medical Center, Sudden Infant Death Syndrome Study, Consultant
ABC News 20/20 Program, Interview (appeared July 4, 1985); various other national television shows

Editorial Positions:

Associate Editor, *Journal of the American Statistical Association* (Theory & Method)
Statistical Editor, *Journal of the American Society for Psychical Research*

JESSICA M. UTTS

PUBLICATIONS

1. 1976 Utts, J.M. and T.A. Ryan, Jr. Lack of fit in regression. *American Statistical Association Proceedings of the Statistical Computing Section,*285-287

2, 1976 Varner, L. and J. Utts. Parallel prediction lines: a test for interaction. *The Journal of Educational Research,* 70(2):63-66

3. 1977 Hettmansperger, T.P. and J.M. Utts. Robustness properties for a simple class of rank estimates. *Communication in Statistics - Theory and Methods,* A6(9): 855-868

4. 1977 Naeye, R.L., W.L. Harkness and J. Utts. Abruptio placentae and perinatal death: a prospective study. *American Journal of Obstetrics and Gynecology,* 128(7): 740-746

5. 1980 Utts, J.M. and T.P. Hettmansperger. A robust class of tests and estimates for multivariate location. *Journal of the American Statistical Association,* 75(372): 939-946

6. 1981 Ainley, D.G., C.R. Grau, T.E. Roudybush, S.H. morrell and J.M. Utts. Petroleum ingestion reduces reproduction i Cassin's anklets. *Marine Pollution Bulletin,* 12(9): 314-317.

7. 1982 Utts, J.M. The rainbow test for lack of fit in regression. *Communications in Statistics-Theory and Methods,* 11(24): 2801-2815

8. 1983 Samaniego, F.J. and J.M. utts. Evaluating performance in continuous experiments with feedback to subjects. *Psychometrika,* 48(2): 195-209

9. 1985 Rucker, M.K. McGee, M. Hopkins, A. Harrison and J. Utts. Effects of similarity and consistency of style of dress on impression formation. In *The Psychology of Fashion,* ED. M.R. Solomon, Lexington Books, Lexington, MASS, 309-318

10. 1985 May, E.C.D.I. Radin, G.S. Hubbard, B.S. Humphrey and J.M. Utts. Psi experiments with random number generators: an informational model.

Proceedings of the Parapsychological Association Annual Convention, 235-266

11. 1985 Johnson, W.O., J.M. Utts, and L.M. Pearson. Bayesian robust estimation of the mean. *The Journal of the Royal Statistical Society Series C (Applied Statistics),* 35(1):63-72

12. 1986 Utts, J.M. The Ganzfel Debate: a statistician's perspective. *Journal of Parapsychology,* 50: 393-402

13. 1986 Utts, J.M. Comment on "Computers in statistical research." *Statistical Sciuence,* 1(4): 437-439

14. 1986 Johnson, W.O., L.M. Pearson, and J.M. Utts. A Monte Carlo comparison of Bayesian estimators and trimmed means. *Journal of Statistical Computation and Simulation,* 25:167-192

15. 1986 Utts, J.M. and E.C. May. An exact method for combining P-values. *Research in Parapsycholgy,* 99-103, Scarecrow Press, Metuchen, N.J.

16. 1987 G.P. Hansen and J. Utts. Use of both sum of ranks and direct hits in free-response psi experiments. *Journal of Parapsycholgy,* 51:321-335

17. 1987 Utts, J. Psi, statistics, and society. *Behavioral and Braiun Sciences,* 10:615-616

18. 1988 Utts, J. Successful replication versus statistical significance. *Proceedings of the Parapsychological Association,* 31:148-162

19. 1988 Humphrey, B.S., E.C. May and J. Utts. Fuzzy set technology in the analysis of remote viewing, *Proceedings of the Parapsychology Association,* 31:378-394

20. 1988 Palmer, J.A., C. Honorton and J. Utts. Reply to the National Research Council study on parapsychology, *Journal of the American Society for Psychical Research,* 83(1):31-49. (Also published in Proceedings of the Parapsychology Association, 31:424-451

21. 1988 Utts, J. Successful replication versus statistical significance.

Journal of Parapsychology, 52(4):305-320

22. 1989 Radin, D. and J. Utts. Experiments investigating the influence of intention on random and pseudorandom events. *Journal of Scientific Exploraiton,* 3(1):65-79

23. 1989 Flay, B.R., R.C. Kessler, and J.M. Utts. Evaluating media campaigns. In *Evaluating AIDS Prevention Programs,* Coyle, Boruch, and Turner, Eds., National Academy Press, Washington, DC

24, 1989 Utts, J. Randomness and randomization tests: A reply to Gilmore. *Journal of Parapsychology,* 53(4):345-35

25. 1990 Utts, J. Use hammers for nails and corkscrews for wine: In defence of defendable statistical methods. *The Skeptic,* 4(5): SEpt./Oct. 1990, 16-18

26. 1990 May, E.C., J.M. Utts, B.S. Humphrey, W. Luke, T.J. Frivold, and V.V. Trask. Advances in Remote Viewing, *Journal of Parapsychology,* 54(3), 193-228

27. 1991 Utts, J.M. Replication and meta-analysis in parapsychology (with discussion). *Statistical Science,* 6(4), 363-403

28. 1992 Christensen, R. and J.M. Utts Testing for nonadditivity in log-linear and logit models. *Journal of Statistical Planning and Inference,* 33, 333-343

29. 1992 Christensen, R. and J.M. Utts. Bayesian resolution of the exchange paradox. *The American Statistician,* November 1992, 274-276

30. 1992 Hansen G.P., J.M. Utts and B. Markwick. Critique of the PEAR remote viewing experiments. *Journal of Parapsychology,* 56(2), 97-113

31. 1993 Utts, J. Analyzing free-response data-a pregress report. In *PSI Research Methodology-A Reexamination,* ed. L. Coly and J.D.S. McMahon, Parapsychology Foundation, New York, 71-83

32. 1993 Utts, J. Honorton the Meta-Analyst. *Journal of Parapsychology,*

57(1), 89-100

33. 1993 Utts, J. Obituary: Florence Nightengale David, 1909-1993. *Biometrics,* 49, 1289-1291

34. 1993 Krippner, S., W. Braud, I.L. Child, J. Palmer, K.R. Rao, M. Schlitz, R.A. White and J.M. Utts. Demonstration research and meta-analysis in parapsychology. *Journal of Parapsychology,* 57(3), 275-286

35. 1994 Utts, J. Social, Institutional, and Cultural Influenmces of Gender on Science. In *Women and Parapsychology,* ed. L. Coly and R.A. White, Parapsychology Foundation, New York, 28-44

36. 1994 Murphy, T.M. and J.M. Utts. A retrospective analysis of peer review at Physiologia Plantarum. *Physiologia Plantarum,* 92, 535-542

BOOKS

1. 1996 Utts, Jessica *Seeing Through Statistics,* Belmont, CA: Wadsworth(In Press)

LIMITED DISTRIBUTION

1. 1975 Utts, J.M. A test for lack of fit in regression models. Technical Report No. 29, Department of Statistics, Teh Pennsylvania State University.

2. 1979 Utts, J.M. and T.P. Hettmansperger. A robust class of tests and estimates for multivariate location. Technical Report No. 1, Division of Statistics, University of California, Davis. (see Publ. #5A.)

3. 1980 Utts, J.M. and L.L. Varner. A preocedure for analyzing a large data set using log-linear models. Technical Report No. 14, University of California, Davis.

4. 1982 Samaniego, F.J. and Utts, J.M. Evaluating performance in continuous experiments with feedback to subjects. Technical Report No. 30, University of California, Davis. (See Publ. #8)

5. 1982 Utts, J.M. A Regional test for fit in regression. Technical Report No. 38, University of California, Davis.

6. 1982 Johnson, W.O., L.M. Pearson and J.M. Utts. Assessing the performance of some robust point and interval estimators in the presence of maean-shift contamination. Technical Report NO. 44, Division of Statistics, University of California, Davis.

7. 1983 Utts, J.M. A sensible approach to hypothesis testing. Technical Report No. 51, University of California, Davis.

8. 1985 Utts, J.M. Inference for an experiment based on repeated majority vote. Technical Report, Dept. of Statistics, Stanford University

9. 1987 Utts, J.M., E.C. May and T.J. Frivold. Intuitive data sorting. Technical Report, SRI International Cognitive Sciences Program, Menlo Park, CA, December 1987.

10. 1987 Hubbard, G.S., J.M. Utts and W.W. Braud. Experimental protocol for hemolysis: Confirmation experiment. Technical Report, SRI International Cognitive Sciences Program, Menlo Park, CA, December 1987.

11. 1987 Humphrey, B.S., E.C. May, J.M. Utts, T.J. Frivold, W.W. Luke, and V.V. Trask. Fuzzy set applications in remote viewing analysis. Technical Report, SRI International Cognitive Sciences Program, Menlo Park, CA, December 1987.

12. 1991 Christensen, R. and J. Utts. Bayesian resolution of classical paradoxes: two examples. Technical Report #220, Division of Statistics, University of California, Davis.

BOOK REVIEWS

1. 1987 Utts, J.M. Nonparametric Methods for Quantitative Analysis(2nd ed.) by Jean Dickinson Gibbons. *Journal of the American Statistical Association* 82(397):364-367

2. 1988 Utts, J.M. Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work, Vol. 1: Discrete Models, by G.P. Patil, M.T. Boswell and

M.V. Ratnaparkhi; Vol. 2: Continuous Univariate Models, by G.P. Ratnaparkhi; Vol. 3: Multivariate Models, by G.P. Patil, M.T. Boswell, M.V. Ratnaparkhi, and J.J.J. Roux. *Journal of the American Statistical Association,* 83:562

3. 1993 Utts, J. Perspectives on Contemporary Statistics, ed. by D.C. Hoaglin and D.S. Moore. *Journal of the American Statistical Association,* 88, 695.

APPENDIX B: BIBLIOGRAPHY

DeGraff, E., May, E.C. (1993). Phenomenological Research and Analysis. Periodic Status Report -- Covering 12/30/92 - 04/23/93, Contract MDA908-93-C-0004, SAIC, Menlo Park, CA.

DeGraff, E., May E.C. (1993). Phenomenological Research and Analysis. Periodic Status Report -- Covering 4/23/93 - 6/30/93, Contract MDA908-93-C-0004, SAIC, Menlo Park, CA.

DeGraff, E., May, E.C. (1993). Phenomenological Research and Analysis. Periodic Status Report -- Covering 6/22/93 - 9/10/93, Contract MDA908-93-C-0004, SAIC, Menlo Park, CA.

Hubbard, G.S. (1987). Initial Protocol for Remote Action Interactions with α-Particles. Interim Report -- Objective H, Task 1, Deliverable a, SRI International, Menlo Park, CA.

Hubbard, G.S., Bentley, P.P., Pasturel, P.K., Isaacs, J. (1987). A Remote Action Experiment With a Piezoelectric Transducer. Final Report -- Objective H, Tasks 3 and 3a, SRI International, Menlo Park, CA.

Hubbard, G.S., Isaacs, J.D. (1986). An Experiment to Examine the Possible Existence of Remote Action Effects in Piezoelectric Strain Gauges. Final Report -- Objective E, Task 8, SRI International, Menlo Park, CA.

Hubbard, G.S., Langford, G.O. (1986). A Suggested Remote Viewing Training Procedure (U). Final Report -- Objective D, Task 1, SRI International, Menlo Park, CA.

Hubbard, G.S., Utts, J.M., Braud, W.W. (1987). Experimental Protocol for Hemolysis: Confirmation Experiment. Final Report -- Objective E, Task 2, SRI International, Menlo Park, CA.

Lantz, N.D. (1989). An Effort to Improve Remote Viewing Quality Using Hypnosis. Final Report -- Task 6.0.6, Covering Period 1 October 1988 to September 1989, SRI Project 1291, SRI International, Menlo Park, CA.

Lantz, N.D. (1987). Review of the Personality Assessment System. Final Report -- Objective C, Tasks 2 and 3, SRI International, Menlo Park, CA.

Lantz, N.D., Kierman, R.J. (1986). Neuropsychological Assessment of Participants in Psychoenergetic Tasks. Final Report -- Objective C, Task 5, SRI International, Menlo Park, CA.

Lantz, N.D., Luke, W.L.W., May, E.C. (1994). Target and Sender Dependencies in Anomalous Cognition Experiments. SAIC, Menlo Park, CA.

* Lantz, N.D., Luke, W.L.W., May, E.C. (1994). Target and Sender Dependencies in Anomalous Cognition Experiments. Journal of Parapsychology, 58, 285-302.

Lantz, N.D., May, E.C. (1988). Mass Screening for Psychoenergetic Talent Using a Remote Viewing Task. Final Report -- Objective B, Task 1, Covering the Period 1 October 1987 to 30 September 1988, SRI Project 1291, SRI International, Menlo Park, CA.

Luke, W.L.W., Frivold, T.J., May, E.C., Trask, V.V. (1989). A Prototype Analysis System For Special Remote Viewing Tasks. Final Report -- Task 6.0.3, Covering the Period 1 October 1988 to 30 September 1989, SRI Project 1291, SRI International, Menlo Park, CA.

* Luke, W.L.W., May, E.C. (1993). Phenomenological Research and Analysis. Interim Technical Report -- Covering 30 December 1992 thru 30 April 1993, Contract MDA908-93-C0004, SAIC, Menlo Park, CA.

Luke, W.L.W., May, E.C. (1993). Travel Report: The Third Meeting of The Chinese Society of Somatic Science (CSSS), Bejing China. Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

* May, E.C. (1989). An Application Oriented Remote Viewing Experiment. Final Report -- Covering the Period 1 May 1988 to April 1989, SRI Project 2740, SRI International, Menlo Park, CA.

May, E.C. (1986). Enhanced Human Performance Investigation. Final Technical Report -- Covering the Period 1 July 1986 to 15 November 1986, SRI Project 1291, SRI International, Menlo Park, CA.

May, E.C. (1988). Enhanced Human Performance Investigation (U). Final Technical Report -- Covering the Period 1 October 1987 to 30 September 1988, SRI Project 1291, SRI International, Menlo Park, CA.

May, E.C. (1993). Phenomenological Research and Analysis. Periodic Status Report

-- Covering 12/16/92 - 04/05/93, Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

May, E.C. (1992). Phenomenological Research and Analysis. Periodic Status Report -- Covering 6/19/92 - 09/11/92, Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

May, E.C. (1992). Phenomenological Research and Analysis. Periodic Status Report -- Covering 9/11/92 - 11/16/92, Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

May, E.C. (1992). Phenomenological Research and Analysis. Periodic Status Report -- Covering 3/31/92 - 06/09/92, Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

May, E.C. (1992). Phenomenological Research and Analysis. Periodic Status Report -- Covering 06/11/92 - 12/16/92, Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

May, E.C. (1992). Phenomenological Research and Analysis. Periodic Status Report -- Covering 10/01 - 11/30/91, Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

May, E.C. (1991). Research of Anomalous Mental Phenomena. Periodic Status Report -- Covering 2/4 - 9/30/91, Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

May, E.C., Lantz, N.D., Piantineda, J. (1994). Feedback Considerations in Anomalous Cognition Experiments. SAIC, Menlo Park, CA.

* May, E.C., Luke W.L.W. (1992). Phenomenological Research and Analysis. Final Report: Support, Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

* May, E.C., Luke, W.L.W., James, C.L. (1994). Phenomenological Research and Analysis. Final Report -- Contract MDA908-93-C-0004, SAIC, Menlo Park, CA.

* May, E.C., Luke, W.L.W., Lantz, N.D. (1993). Phenomenological Research and Analysis. Final Report: 6.2 and 6.3, Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

May, E.C., Luke W.L.W., Lantz, N.D. (1992). Phenomenological Research and Analysis. Interim Technical Report -- Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

May, E.C., Luke, W.L.W. (1991). Technical Protocol for the MEG Investigation. SAIC, Menlo Park, CA.

* May, E.C., Luke W.L.W. (1992). Phenomenological Research and Analysis. Interim

Report: 6.2, 6.3, 6.4, Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

* May, E.C., Luke W.L.W. (1992). Phenomenological Research and Analysis. Final Report 6.5: Support, Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

* May, E.C., Luke W.L.W. (1992). Phenomenological Research and Analysis. Interim Report Covering 2/4/91 - 12/31/91, Contract MDA908-91-C-0037, SAIC, Menlo Park, CA.

May, E.C., Luke, W.L.W., Trask, V.V., Frivold, T.J. (1989). Observation of Neuromagnetic Fields in Response to Remote Stimuli. Final Report -- Task 6.0.2, Covering the Period 1 October 1988 to 30 September 1989, SRI Project 1291, SRI, Menlo Park, CA.

May, E.C., Luke, W.L.W., Trask, V.V., Frivold, T.J. (1990). Observation of Neuromagnetic Fields in Response to Remote Stimuli. Final Report (Rev.) -- Task 6.0.2, SRI Project 1291, SRI International, Menlo Park, CA.

May, E.C., MacGowan, D., DeGraff, E. (1993). Phenomenological Research and Analysis. Periodic Status Report -- Covering 9/11/93 to 12/31/93, Contract MDA908-93-C-0004, SAIC, Menlo Park, CA.

May, E.C., Pleass, C.M. (1987). A Remote Action Investigation with Marine Animals. Final Report -- Objective E, Task 2, SRI International, Menlo Park, CA.

May, E.C., Puthoff, H.E. (1981). Feasibility Study on the Use of RV Detection Techniques to Determine Location of Military Targets. SRI International, Menlo Park, CA.

May, E.C., Spttiswoode, S.J., James, C.C. (1994). Managing the Target Pool Bandwidth: Noise Reduction for Anomalous Cognition Experiments. SAIC, Menlo Park, CA.

* May, E.C., Trask, V.V. (1988). Forced Choice Remote Viewing. Final Report -- Objective E, Task 4, Covering the Period 1 October 1987 to 30 September 1988, SRI Project 1291, SRI International, Menlo Park, CA.

* May, E.C., Utts, J.M., Trask, V.V., Luke, W.W., Frivold, T.J., Humphrey, B.S. (1989). Review of the Psychoenergetic Research Conducted at SRI International (1973-1988). Final Report -- Task 6.0.1, Covering the Period 1 October to 15 February 1989, SRI Project 1291, SRI International, Menlo Park, CA.

Puthoff, H.E., Targ, R. (1978). Standard Remote-Viewing Protocol (Local Targets). SRI International, Menlo Park, CA.

Trask, V.V., Lantz, N.D., Luke, W.W., May, E.C. (1989). Screening for Remote Viewing Talent. Final Report -- Task 6.0.5, Covering the Period 1 October 1988 to 30 September 1989, SRI Project 1291, SRI International, Menlo Park, CA.

Trask, V.V., Lantz, N.D., Luke, W.L.W., May, E.C. (1989). Screening for Remote Viewing Talent. Final Report -- Task 6.0.5, Covering the Period 1 October 1988 to 30 September 1989, SRI Project 1291, SRI, Menlo Park, CA.

Utts, J.M., May, E.C., Frivold, T.J. (1987). Intuitive Data Sorting. Final Report -- Objective I, Task 2, SRI International, Menlo Park, CA.

Listings with "*" are from Ed May's top 10 list.

SAIC Papers

Phenomenological Research and Analysis Technical Proposal (U). (1993). MDA908-93-C-0004, Contract Modification Number P00002, SAIC, Menlo Park, CA.

Phenomenological Research and Analysis Technical Proposal (U). (1993). MDA908-93-C-0004, Contract Modification Number P00003, SAIC, Menlo Park, CA.

SRI Papers

A Remote Viewing Evaluation Protocol (U). (1983). Final Report, SRI International, Menlo Park, CA.

An Application Oriented Remote Viewing Experiment (U). (1988). Final Report, Covering the Period 1 May 1987 to April 1988, SRI Project 8339, SRI International, Menlo Park, CA.

An Automated RV Evaluation Procedure (U). (1985). Final Report -- Covering the Period October 1983 to October 1984, SRI Project 7408-12, SRI International, Menlo Park, CA.

Applications of Fuzzy Sets to Remote Viewing Analysis (U). (1988). Final Report -- Objective F, Task 1, Covering the Period 1 October 1987 to 30 September 1988, SRI Project 1291, SRI International, Menlo Park, CA.

Audiolinguistic Correlations With The Quality Of Remote Viewing Sessions (U). (1982). Final Report -- Covering the Period October 1980 to October 1981, SRI International, Menlo Park, CA.

Computer-Assisted Search (U). (1987). Final Report -- Objective D, Task 1 and Objective G, Task 1, Covering the Period 1 October 1985 to 30 September 1987, SRI Project 1291, SRI International, Menlo Park, CA.

Co-Ordinate Remote Viewing (CRV) Technology 1981 - 1983 Three-Year Project. (1983). Draft Report.

Countermeasures: A Survey and Evaluation (U). (1982). Final Report -- Covering the Period October 1980 to September 1981, SRI International, Menlo Park, CA.

Enhanced Human Performance Investigation (U). (1988). Final Report, Covering the Period November 1983 to October 1985, SRI International, Menlo Park, CA.

Enhanced Human Performance Investigation (U). (1987). Final Technical Report -- Covering the Period 1 October 1986 to 30 September 1987, SRI Project 1291, SRI International, Menlo Park, CA.

Feedback and Precognition Dependent Remote Viewing Experiments (U). (1987). Final Report -- Objective F, Tasks 1a and 1b, Covering the Period 1 October 1986 to 30 September 1987, SRI Project 1291, SRI International, Menlo Park, CA.

Feedback and Target Dependencies in Remote Viewing Experiments (U). (1988). Final Report -- Objective E, Tasks 1 and 2, Covering the Period 1 October 1987 to 30 September 1988, SRI International, Menlo Park, CA.

Free World Psychoenergetics Research Survey (U). (1983). Final Report -- Covering the Period October 1982 to September 1983, SRI Project 4028-2, SRI International, Menlo Park, CA.

Geophysical Effects Study (U). (1984). Interim Report -- Covering the Period 15 November 1983 to 15 July 1984, SRI Project 6600, SRI International, Menlo Park, CA.

Geophysical Effects Study (U). (1984). Final Report -- Covering the Period 15 November 1983 to 15 December 1984, SRI Project 6600, SRI International, Menlo Park, CA.

Location of Target in Space and Time. (1986). Interim Report -- Covering the Period 1 October 1985 to 30 September 1986, SRI International, Menlo Park, CA.

Neurophysiological Correlates to Remote Viewing (U). (1988). Final Report -- Objective D, Task 1, Covering the Period 1 October 1987 to 30 September 1988, SRI Project 1291, SRI International, Menlo Park, CA.

NIC Techniques (U). (1980). Quarterly Progress Report -- Covering the Period 1 October to 31 December 1979, SRI Project 7560, SRI International, Menlo Park, CA.

Personnel Identification and Selection (U). (1984). Final Report -- Covering the Period 15 November 1983 to 15 December 1984, SRI Project 6600, SRI International, Menlo Park, CA.

Photon Production (Chinese Replication) (U). (1985). Final Report -- Covering the Period October 1983 to October 1984, SRI International, Menlo Park, CA.

Remote Viewing Evaluation Techniques (U). (1986). Final Report -- Covering the Period 1 October 1985 to 30 September 1986, SRI International, Menlo Park, CA.

RV Reliability, Enhancement , and Evaluation (U). (1982). Final Report -- Covering

the Period October 1981 to September 1982, SRI International, Menlo Park, CA.

RV Reliability, Enhancement, and Evaluation (U). (1982). Final Report -- Covering the Period October 1980 to September 1981, SRI International, Menlo Park, CA.

Special Orientation Techniques (U). (1980). Final Report -- Covering the Period 1 May 1979 to 31 March 1980, SRI International, Menlo Park, CA.

Special Orientation Techniques: S-I, S-II, S-III (U). (1984). Final Report -- Covering the Period 15 November 1983 to 15 December 1984, SRI Project 6600, SRI International, Menlo Park, CA.

Special Orientation Techniques: S-V,S-VI (U). (1984). Final Report -- Covering the Period 15 November 1983 to 15 December 1984, SRI Project 6600, SRI International, Menlo Park, CA.

Special Orientation Techniques: S-IV (U). (1984). Final Report -- Covering the Period 1 February 1983 to 30 April 1984, SRI Project 5590, SRI International, Menlo Park, CA.

Targeting Requirements Tasks (U). (1981). Final Report -- Covering the Period October 1980 to October 1981, SRI International, Menlo Park, CA.

Target Search Techniques (U). (1984). Final Report -- Covering the Period 15 November 1983 to 15 December 1984, SRI Project 6600, SRI International, Menlo Park, CA.

APPENDIX C
ATTACHMENT 1
INTERVIEWER INSTRUCTIONS

In this interview, you will be attempting to gather information bearing on the use of remote viewing in the intelligence community.

Attached you will find the questions to be asked in the interview. You should begin the interview by noting that all information, particularly the respondents' identity, will be treated as confidential.

You should begin the interview by stating your name, clearance, and affiliation. After ensuring confidentiality, you should provide a brief description of the reasons for conducting these interviews, describing both the origins of the work, including program transfer, and the ensuing congressional mandate. You should also note the program evaluation is being conducted by the CIA in conjunction with a not-for-profit research organization.

Once you have provided this background information, you may proceed to the interview questions. Twelve general questions are presented. Please try to work through each of these questions in turn. The early questions focus on background, the middle questions on process, and the final questions on evaluation. To help you follow this structure one intended to feed the discussion objective, it might be useful to ask the person being interviewed to hold their evaluation until the end of the interview.

Once you ask a question, allow the persons being interviewed to answer the question in their own terms. After they have finished you might want to follow-up on their answers. You should only ask follow-up questions when the answer provided initially was not clear.

Please sketch out the answers provided in the space below the question and prompts. Note, you should record only what the interviewee says not you impressions. If you are unsure about a point, ask the interviewee only after you have finished working through the question and prompts.

It should take you about ninety minutes to complete this interview. All material should be returned to Dr. Michael Mumford at the American Institutes for Research.

ATTACHMENT 2
INTERVIEW QUESTIONS

Users:

1. Have you used the support of remote viewers?
2. Had you used remote viewers before?
3. What information did you request from the remote viewer?
4. What information did you receive from the remote viewer?
5. How did you use the remote viewing products?
6. Did the remote viewing product seem to confirm your initial approach?
7. Did you receive any subsequent information that confirmed/disconfirmed the remote viewing information (e.g., other intelligence sources)?
8. Could the remote viewing products be used without information from other sources?
9. Were the remote viewing products accurate?
10. How much relevant information was included in the remote viewing product?
11. Would you use this again and if so, under what circumstances?
12. Would you pay for the services of a remote viewer?

Viewers:

1. What lead you to become interested in the remote viewing program?
2. How were you selected to become a remote viewer?
3. Were you provided with any training and if so, was it helpful?
4. How much familiarity did you have with the remote viewing research?
5. Was this research helpful in doing your work?

6. What were the major types of taskings you were assigned?

7. How much background information did you have about the nature of the target?

8. Was this background information useful in producing a viewing?

9. How did you typically go about generating viewings?

10. When multiple viewers worked on a tasking how were you viewings similar to or different from each other?

11. What could have been done to improve the accuracy and usefulness of viewings?

12. What were the conditions that led to the best viewings?

13. What were the conditions that led to poor viewings?

14. How could the program be managed differently to improve the accuracy and usefulness of viewings?

15. What actions could be taken to make you a better viewer?

Manager:

1. Why were you selected to manage the remote viewing program?

2. How did you identify potential users of the remote viewing services?

3. Why did users request viewings?

4. What were the major types of taskings requested by the users?

5. How many user groups request multiple viewings?

6. How did you assign viewers to taskings?

7. How much information did the viewers have about the nature or background of the tasking?

8. How useful was the research background in structuring the viewers' activities?

9. What kind of information was typically provided to the viewers about the tasking?

10. How accurate was the information provided by the viewers relative to other sources of information?

11. When were viewers particularly likely to produce useful and accurate products?

12. What influenced users' acceptance of or use of the remote viewings?

13. How often did viewings have a major influence on operational decisions?

14. What organizational influences contributed to the successes of the program?

15. What organizational influences limited the success of the program?

16. How should the program be managed differently in the future?

ATTACHMENT 3
INTERVIEW REPORT
STAR GATE OPERATIONAL USE INTERVIEW #1
DATE: JULY 7, 1995

CONTEXT: The Unit Commander in the organization had not previously used the services of remote viewers. A single tasking was his only experience in using remote viewing information. The decision to use remote viewers was in part based on contact with the Star Gate Program Director and awareness of publicized incidents where remote viewings were used in police cases. The primary motivation for use of the remote viewers appears to simply have been to try out a low cost approach that might pay off.

TARGET: The target in this case was a person rather than a site. The target person was suspected of potential involvement in espionage. The primary evidence bearing on this assessment was access, finances, and reported comments of which the most important evidence was financial data.

REQUEST: The remote reviews were asked to provide a variety of information about the target person. The requested information included descriptions of the person, likely travel locations, and events occurring during travel. Four, sequential, apparently "independent" remote viewings were obtained.

NATURE OF INFORMATION: The four sequential viewings were provided and accompanied by reports. The information provided in these reports included both verbal descriptions and drawings. A high degree of agreement was not observed among the four remote viewing reports. Furthermore, the narrative descriptive information was provided in broad, highly ambiguous terms.

USE OF INFORMATION: The information provided by these viewings was not held to be useful in any operational sense. The reasons stated for reaching this conclusion were: 1) the information was too broad and too vague to direct relevant observations: 2) crucial elements of the case, particularly financial concerns, did not appear in any of the reports; 3) the information provided could be interpreted in too many different ways; 4) hits were often stereotypic given the available cues in the tasking; 5) there were a large number of demonstrably wrong conclusions.

Given the preceding reactions of the user, no attempt was made to use the information provided by the remote viewers. It did not in any way contribute to the actions taken in the case of interpretation of other available data. It is of note that the data apparently were used objectively without forcing them into preconceptions about the case.

MERITS RELATED TO OTHER SOURCES: It was noted that the users relied more on other sources of information (e.g., financial, human, etc...) than the reports of the remote viewers. The viewings were, apparently, discredited due to the number of inaccuracies and failure to identify known key aspects of the case.

UTILITY: The remote viewings were not held to be of substantial value due to the inaccurate described above. The user noted however, that would be given further consideration use of remote viewers if the "situation were desperate, no costs were entailed to the user, and the viewers were more intimately involved in the case for some period of time." It was further noted that remote viewers could be viewed as another source of manpower. In the case at hand, however, the viewings proved of no practical value.

ATTACHMENT 3
INTERVIEW REPORT
STAR GATE OPERATIONAL USER INTERVIEW #2
DATE: JULY 26, 1995

CONTEXT: The manager of this group has used the services of remote viewers a number of times. In the course of this interview, the unit manager stressed the importance of keeping an open mind. However, the primary motivation for using remote viewing was that it provided a way to obtain a simple, clear terms information that could not be easily obtained from other sources, particularly human intelligence, archival records, or financial records. This manager required analysis to provide targets on a regular basis.

TARGET: Three major types of targets were involved. One set of targets consisted of people and their roles in criminal organizations. The second set of targets were ships or the location of material on ships. The third and final type of target was the site of manufacturing operations or shipping. Targets were selected by analysis based on current concerns and generally were described in terms of names and locations.

REQUESTS: The type of information remote viewers were asked to provide was consistent with the nature of the targets. When the target was a person the viewers were given a name and asked to indicate the person's role and position in a criminal organization. When the target was a ship, they might be asked to indicate its location or type of cargo. When the target was a plant they might be asked to indicate what was being produced.

NATURE OF INFORMATION: Multiple viewings were made of each target. The results of these viewings were provided in reports presenting both verbal descriptions and drawings. The report material was typically stated in rather broad terms and was synthesized by the analysts. The viewers had some knowledge of the target organizations and their operations but not the background of the particular tasking at hand.

USE OF INFORMATION: The manager indicated that the information provided by the viewers was useful for filling in background information about a person and his or her likely role in a criminal organization (e.g., who is Mr. Chin?). The information provided by the viewers was not deemed useful when they were required to address specific operational issues such as the location of objects on a ship. As a result, the information was not used to guide operations. Instead, it was used to identify the roles of individuals within targeted criminal organizations.

Generally, this background information was felt to be accurate, although it did not permit immediate action. However, the manager being interviewed offered that some degree of accuracy could be expected if the viewers had a knowledge of the sponsoring organization

and its areas of interest. The manager also noted that some analysts felt far more comfortable than others about using this type of information implying that some analysts had to be persuaded to use the information.

MERITS RELATED TO OTHER SOURCES: The nature of the sponsoring organization was such that relatively few alternative sources of information were available. Much of the intelligence information used was indirect, such as shipping patters or records of business ownership. In comparison to this type of indirect information, the remote viewings were viewed as useful, particularly in identifying backgrounds and roles of people in criminal organizations when direct information was not available.

UTILITY: Although the viewings were deemed useful for filling in background information, they were not deemed useful for operations. It was also pointed out that analysts doubts about the value of the source might limit use of the resulting information. Further, it was noted that a viable program could only be maintained if it were an in house Government activity. If this provision was met, the manager was willing to devote operational resources to pay for the services of remote viewers.

ATTACHMENT 3
INTERVIEW REPORT
STAR GATE OPERATIONAL USER INTERVIEW #3
DATE: AUGUST 3, 1995

CONTEXT: This interview consisted of a group manager and three analysts responsible for identifying lost or missing Department of Defense personnel. The group was interested in remote viewing as another possible tool for identifying the location of lost or missing personnel when more conventional methods (e.g., radios) could not be used. The remote viewings had not been used operationally. Instead, the group was using the services of the remote viewers in a series of experiments to determine whether it had any potential operational value. They had found out about the potential applications of remote viewing through research reports acquired by a fellow manager. These experiments were conducted with the cooperation of the remote viewing program manager. Both viewers and the program manager had met with members of this operational group of a number of times.

TARGET: In these experiments, two major types of targets were involved: 1) the location of a person at a predetermined time; 2) the nature of the location or its physical characteristics. Targets were selected on the basis of convenience, typically in proximity to office, although more distant locations were sometimes selected. In the last experiment, the viewers were asked to indicate the map coordinates of a person who had moved to an area around the office site.

REQUEST: The viewers were asked to indicate the location of the person serving as beacon and describe the location. The person playing the role of beacon evaluated accuracy.

NATURE OF INFORMATION: Three viewers provided independent viewings at the designated time. The material was presented in a three to four-page report which included drawings and a verbal description of the location. Information contained in the reports often included a number of vague general statements (e.g., "There is water nearby"). The nature of the information was sufficiently broad and vague that the analysts typically had to force an interpretation. These interpretations often involved actions and locations occurring at another time in the day. It was only with this extended interpretation that hits were identified by the analysts.

In the experiment where a specific map location was identified by the viewers, none of the viewers correctly identified the target site. In fact, most of the locations were not in the vicinity of the target site which was a park near the group's office. However, one viewer located the beacon at another nearby park and did describe general characteristics of the park, including trash cans, wooden steps, etc.

USE OF INFORMATION: As noted above, these experiments were not intended to provide information for operational use. However, all members of the group stated that the information was too vague and ambiguous for operations, noting that unless specific map locations could be identified, the information could not be used in operational decision making.

MERITS RELATED TO OTHER SOURCES: Because viewings were being used on an experimental basis, it was not possible to obtain direct comparisons of the viewings with other potential sources of information. However, all members of the group took the position that the information provided was so vague that it could not be used without other sources of information. If used at all, it would be treated as supplemental information that could not by itself provide a basis for operational decisions.

UTILITY: Because the viewers could not identify specific locations in unambiguous terms, the group did not feel the procedures had any current operational utility. They did note, however, that if further research were to yield more specific consistent information from viewers, it might eventually have some utility. Even under these conditions, however, the information would have to be obtained under the auspices of a formal Government program to ensure that it was viewed as credible.

ATTACHMENT 4
INTERVIEW REPORT
STAR GATE VIEWERS INTERVIEW
DATE: AUGUST 4, 1995

CONTEXT: This interview was conducted with the three viewers employed in the Star Gate program at the time of its suspension. All three of the viewers attended the group interview. The viewers had worked in the program for at least five years. Generally, all viewers had in one way or another been affiliated with the sponsoring organization prior to becoming viewers.. They were recruited by more senior officials based on background factors indicating an interest in parapsychology or psychology more generally.

SELECTION AND TRAINING: Systematic selection procedures or pretesting were not used in selecting two of the three viewers. The third viewer, however, was selected based on a series of tests given by a contractor organization. This viewer reported scoring relatively high in a relatively large pool of candidates.

When viewers entered the program they were provided with formal training. The training included three component stages moving from simple to complex targets. They began with beacon-based viewing on site and progressed to operational targets. This initial training was six to 18 months long. Follow-up or refresher training did not occur. However, all of these viewers had initiated and maintained personal self-development program which included attempts to keep up with the literature in parapsychology in general and remote viewing in particular.

With regard to the remote viewing research, the viewers noted that the research paradigm did not directly correspond to operational assignments in terms of conditions and requirements (e.g., the availability of a beacon). Thus a variety of different sources and techniques was used in self development and operational assignments aside form those found in the remote viewing research literature.

TECHNIQUES: Each of the three viewers used different techniques to generate viewings. One viewer used coordinate viewing coupled with meditation. The second viewer relied on relaxation techniques and meditation. The third viewer relied on channelling and automatic writing.

Although the viewers tended to rely on different techniques, they noted that these techniques were not necessarily incompatible. Further, it was pointed out that all of these techniques were physically and emotionally draining and resulted in products which were somewhat ambiguous or vague in nature. As a result, the ability of users to correctly interpret the results of a viewing was considered essential to successful application. Generally, the viewers felt that successful use was influenced by acceptance of the phenomenon.

TASKINGS: The viewers were presented with a number of different types of taskings. The targets included people, locations, intentions, and sites. The concerns in initiating taskings included locating people, counter intelligence, locating objects, and identifying the activities at various sites.

With regard to these taskings, viewers differed in the amount of background information they preferred to have before producing viewing. Some viewers preferred substantial background information while others wanted minimal information about the target. The viewers noted, however, that when they had background information it sometimes distorted the process. More specifically, the viewers indicated that they sometimes changed the content of their reports to bring the information presented into line with the known characteristics of the target. They noted that reports tended to be more specific when substantial background information was available. Typically, each viewer generated independent reports for a target. These reports presented into line with the known characteristics of the target. They noted that reports tended to be more specific when substantial background information was available. Typically, each viewer generated independent reports for a target. These reports presented both verbal descriptions and drawings. Frequently, different viewers produced rather different reports. This result was attributed in part to the use of different techniques, and, in part, to the tendency of viewers to focus on different aspects of the target.

CONDITIONS PROMOTING ACCURACY: The viewers noted a number of conditions that influenced the accuracy and quality of the information contained in their reports. Generally, they felt that the quality of viewing suffered when they were presented with a large number of repetitive taskings over a short period of time. They also noted the quality suffered when the targets lacked intrinsic interest and when external evaluations placed too much pressure on the viewers. The viewers noted that it was particularly difficult to produce accurate viewings for some types of targets. It also was unclear exactly what characteristics of targets influenced accuracy.

The viewers also identified a number of organizational factors that influenced accuracy. First, they field that knowledge supportive users contributed to accuracy and effective use of the reported information. Second, they felt more judicious use of background information might contribute to the preparation of more accurate reports, less biased by logical cues. Third, they pointed out that earlier program managers had edited or changed viewers' reports. The viewers felt that this could lead to the loss of important of potential significant to the users.

MANAGEMENT: The viewers noted a number of actions that could be taken to improve the success of the program. To begin with, they indicated that it was important to have a manager who understood the process, the demands it made on people, and the need for a supportive work environment. Along related lines, they indicated that more balanced work-load management, rather than periods of boom and bust, would contribute to program success.

In addition to these issues bearing on management style and management strategy, the viewers made a number of other suggestions. Onc involved the research program. Here they suggested that research be more focused on the work rather than on narrow technical demonstrations of the existence of the phenomenon. They also suggested that consistent organizational support would play a key role in the long-term success of the program.

ATTACHMENT 5
INTERVIEW REPORT
STAR GATE OPERATIONAL MANAGER INTERVIEW
DATE: AUGUST 4, 1995

CONTEXT: This interview was conducted with the last manager of the remote viewing program. The manager become responsible for this program in 1991 when the previous program manager retired. He was tasked, in accordance with accordance with congressional guidelines, with initiating and managing operational use of the remote viewing service while addressing the two other major program elements. Foreign assessment and research and development. The manager's background before accepting this position was in human intelligence. He did not specifically have a background in parapsychology or remote viewing.

CLIENTS: In accordance with his charter, the manager explicitly sought to expand operational use of remote viewing services. He primarily accomplished this through the use of personal contacts and his own background in intelligence. The manager established contact with ten potential user groups. Of these ten groups, two used the service one time while five used the service multiple times. Typically, groups who used the service had at one time or another employed viewers early on in the program's history.

The users or clients typically requested viewings on an experimental basis. More specifically, they were interested in seeing whether there was any potential operational payoff from the program. It was noted that one especially attractive feature of the service was that information could very rapidly be obtained in a concise form.

TASKINGS: In general, users presented four major types of taskings. The viewers were asked to provide information about: 1) people, 2) objects, 3) locations, and 4) intentions. The amount of background information viewers had about the nature of the targets varied. In some cases, particularly when there were multiple sequential taskings, the viewers might have a great deal of information. In most cases, however, the viewers had the name of the sponsoring organization and a one or two-word description of the target. It is of note that the viewers also differed in the amount of background information requested.

All viewers worked on all taskings although they were allowed to opt out of certain taskings.

For each tasking, the viewers produced a three- or four-page report which included pictures and a verbal description. The accuracy and usefulness of viewings was assessed by customers and this evaluative information was provided as feedback to the viewers by the program manager who instituted this formal evaluation procedure.

In managing the group and responding to the taskings, the manager, although familiar with the research literature, did not explicitly consider the findings obtained in this research. He noted that the research was more useful for general background, establishing the existence of the phenomenon, and foreign assessment, than in managing the viewers' activities.

INFORMATION ACCURACY AND UTILITY: The manager necessarily relied on users to evaluate the accuracy and utility of the viewings provided in response to a specific tasking. He generally recommended that clients use viewings provided in response to a specific tasking. He generally recommended that clients use viewings as supplemental information. In the manager's experience, no viewings served as the primary input into an operational decision. Principally, the viewings were used to validate or extend the accuracy of already available information or, alternatively, as a vehicle to stimulate further intelligence gathering. Application of this information depended on the function of the sponsoring organization.

ORGANIZATIONAL INFLUENCES: In the manager's view, the original Congressional directive along with support of the designated organization were crucial to the successes of the program. However, prior controversy surrounding the program and the existence of the phenomenon set constraints on the success of the program.

With regard to potential improvements in program management, a number of points were mentioned. First, the manager noted that the program should not necessarily reside in the intelligence community. Instead, he recommended that the program be declassified and moved to a more appropriate "open source" sponsor such as the Department of Justice or the National Science Foundation. It was noted that this kind of move would not hurt and indeed might help the foreign assessment component of the program. Further, it was suggested that the program involve research and foreign assessment, with viewer services being obtained only, as needed on a contractual basis.

APPENDIX D
SUMMARY REPORT
STAR GATE OPERATIONAL TASKING AND EVALUATION

1.0 EXECUTIVE SUMMARY

From 1986 to the first quarter of FY 1995, the DoD paranormal psychology program received more than 200 tasks from operational military organizations requesting that the program staff apply a paranormal psychological technique know as "remote viewing" (RV) to attain information unavailable from other sources. The operational tasking comprised "targets" identified with as little specificity as possible to avoid "telegraphing" the desired response.

In 1994, the DIA Star Gate program office created a methodology for obtaining numerical evaluations from the operational tasking organizations of the accuracy and value of the products provided by the Star Gate program. By May 1, 1995, the three remote viewers assigned to the program office had responded, i.e., provided RV product, to 40 tasks from five operational organizations. Normally, RV product was provided by at least two viewers for each task.

Ninety-nine accuracy scores and 100 value scores resulted from these product evaluations by the operational users. On a 6-point basis where "1" is the most accurate, accuracy scores cluster around "2's" and "3's" (55 of the entries) with 13 scores of "1". Value scores, on a 5-point basis with "1" the highest, cluster around "3's" and "4's" (80 of the entries); there are no "l's" and 1 1 scores of "2".

After careful study of the RV products and detailed analysis of the resulting product evaluations for the 40 operational tasks, we conclude that the utility of RV for operational intelligence collection cannot be substantiated. The conclusion results from the fact that the operational utility to the Intelligence Community of the information provided by this paranormal RV process simply cannot be discerned. Furthermore, this conclusion is supported by the results of interviews conducted with representatives of the operational organizations that provided tasking to the program. The ambiguous and subjective nature of the process actually creates a need for additional efforts of questionable operational return on the part of the intelligence analyst. Assuming that the subjective nature of the psychic process cannot be eliminated, one must determine whether the information provided justifies the required resource investment.

2.0 GENERIC DESCRIPTION OF OPERATIONAL TASKING

Over the period from 1986 to first quarter of FY 1995, the Star Gate program received more than 200 tasks from operational military organizations. These tasks requested that the program staff apply their paranormal psychological technique know as "remote viewing" (RV) in the hope of attaining information unavailable from other sources. The operational tasking comprised "targets" which were "identified" in some manner, normally with as little specificity as possible (see discussion below) to avoid excessively "telegraphing" the desired

response. However, until 1994, the results from this tasking were not evaluated by the tasking organizations by any numerical **method** that would identify the accuracy and value of the provided information (for a few cases in prior years narrative comments were provided by some organizations).

In 1994, this situation changed when the Program Office developed a methodology for obtaining numerical evaluations from the tasking organizations of the Star Gate inputs; this methodology is described briefly in Section 3.0. By May 1, 1995, 40 tasks assigned by five operational organizations had been evaluated under this process.[1] Section 4.0 describes the numerical evaluations performed by evaluators from the tasking organizations. The descriptions presented below regarding the tasking and the related targets refer principally to the operational tasks that were numerically evaluated.

The process for a typical tasking, RV response and subsequent evaluation is as follows:

- The tasking organization provides information to the Star Gate Program Manager (PM) describing the problem to be addressed.

- The PM provides a Tasking Form delineating only the most rudimentary information to one or more of the three Star-Gate RV's[2] for their use during the RV session (a typical Tasking Form is presented in Figure 2-1). In addition, the RV's are appraised of the identity of the tasking organization.

- Subsequently the RV's hold individual "viewing" sessions recording their comments, observations, feelings, etc. and including line drawings or sketches of things, places, or other items "observed" during the session.

- The individual RV inputs are collected and provided to the tasking organization for their review with a request for completing a numerical evaluation of the individual RV inputs for accuracy and for value.

- Finally, for those organization who comply with the request, the evaluation scores are returned to the Star Gate Program Office.

[1] Evaluation of additional 1994-95 tasks continued after 5/1/95; three tasks since evaluated were reviewed. They caused only insignificant changes to the statistical information provided in Table 4-1 and did not alter any of the Conclusions and Recommendations in Section 7.0.

[2] (U) All three RV's were full time government employees.

FIGURE 2-1

TASKING SHEET

SOURCE 140: 079

DATE: IS Jul 94

SUSPENSE: 18 Jul 94

1600 Hrs

1. PROJECT NUMBER: 94-252-0

2. METHOD/TECHNIQUE: Method of Choice

3. BACKGROUND:

4. ESSENTIAL ELEMENTS OF INFORMATION: Access and describe target.

5. COMMENTS:

Twenty-six (26) of the 40 operational tasks originated from DIA in support of two joint Task Forces, Org. B and Org. C, (see Section 4.0). Typical tasking targets for these organizations comprised the name of a person or thing (e.g., vessel) with a generic request to describe the target, his/herAts activities, location, associations, etc. as appropriate. No specific information (e.g., what is the height/weight/age of the target?) was requested in the tasking. As noted above, the identity of the supported organizations also was provided. For these tasks that identification provides the RV's with knowledge regarding the specific operational interests of these organizations. Thus, any information provided by the RV's which describes or relates to those interests "could be" relevant; and, therefore, could be interpreted by the evaluators as having some level of "accuracy" and "value" depending upon the information described and the evaluator's interests and beliefs.

The tasking provided by the organization denoted as Org. A comprised targets that were "places" visited by "beacons", i.e., an individual from Org. A who visited and ,'viewed" the site of interest to assist the RV in "visualizing" and describing the site. Targets could be a general vista in or around a particular location, a particular facility at a selected location or, perhaps, a particular item at a location (in the one case where this type of target was used, the item was a particular kind of boat). Usually, no specifics regarding the type of target or its location were provided.

Tasking by Org. D comprised two generic types of targets that related to military interests/concerns current at the time of the tasking, e.g., North Korean (NK) capabilities and leadership. The first type of target focused upon then-current military concerns while the second type required "precognitive" (predictive) capabilities since it required a prognoses of future intentions and actions.[3]

The tasking from Org. E was similar in scope, albeit quite different in context, from the tasks noted earlier for Org. B and Org. C , i.e., describe a person, his activities, location, etc..

As mentioned at the beginning of this section, the descriptions noted above relate to operational tasks that were numerically scored. During the summer/fall period of 1993, eight operational tasks were levied on the program pertaining to North Korean (NK) tunnels. The target information provided to the RV's typically comprised a map of a large section of NK with a request to identify tunnels within the map area. Evaluation of the results from these tasks was in narrative form only; discussion regarding this narration is presented at the end of Section 3.0.

[3]Some operational tasks from the period Oct. 1990 to Jan 1991 regarding Middle East issues were of a similar types, albeit these were not numerically evaluated. They would provide some data for an after-the-fact check of the accuracy of the RV predictions - see Section 7.0 for a discussion of this possibility.

3.0 EVALUATION MEASURES

The numerical evaluation measures that were given to the evaluators of the tasking organizations to score the accuracy and value of the Star Gate inputs were extracted from the Defense Intelligence Agency Manual (DIAM) 58-13. These measures are shown in Table 3-1. Most of the stipulated measures include modifiers such as "may", "possibly', "high", "low", etc. which are subjective and open to individual interpretation by each evaluator. The DIAM 58-13 definitions for the ratings under "Value" are presented in Table 3-2; whether the individual evaluators reviewed these definitions prior to their scoring is unknown. There was no clarification of what was intended by the generic headings of "Accuracy" and "Value", e.g., in the evaluator's estimation how much of the RV's response to the tasking had to qualify for a particular measure, 1 10%, 90%, to be granted the related score?

Table 3-1 Numerical Evaluation Measures

Category	Score
Accuracy - Is the information accurate?	
Yes (true)	1
May be true	2
Possible true	3
No	4
Possibly not true[4]	5
Unsure	6
Value - what is the value of the sources' information?	
Major significance	1
High Value	2
Of Value	3
Low Value	4
No Value	5

As noted in Section 2.0, one series of tasks were evaluated by a narrative discussion only. While much of the final narrative evaluation for this series was complimentary, it lacked any real specifics regarding the usefulness or relevance of the Star Gate inputs and much of the narrative was replete with modifiers and other hedges. A sanitized extract from the final evaluation report for these tasks is presented in Appendix A illustrating the subjective, "uncertain" nature of the comments.

[4]Note that Accuracy scores 5 and 6 actually rank "higher" than 4 since both imply that there may be something accurate in the information. Changing the scoring order to accommodate this observation causes insignificant changes to both the averages and the standard deviations shown on Table 4-1.

TABLE 3-2 - VALUE RATING DEFINITIONS FROM DIAM 58-13

MAJOR SIGNIFICANCE - Intelligence Information Report (IIR) provided information which will alter or significantly influence national policy, perceptions, or analysis; or provided unique or timely indications and warning of impending significant foreign military or political actions having a national impact.

HIGH VALUE - IIR(s) was best report to date or first report on this important topic, but did not significantly influence policy or change analyses.

OF VALUE - IIR(s) provided information which supplements, updates, confirms, or aids in the interpretation of information in data bases, intelligence production, policy research and analysis, or military operations and plans; most DoD HUMINT System reporting falls into this category.

LOW VALUE - IIR was not a good report because the information was not reported in a timely manner, or was of poor quality/of little substance. Nevertheless, it satisfied some of the consumer's informational needs.

NO VALUE - IIR provided no worthwhile information to support data base maintenance, intelligence production, policy research and analysis, or military operations and planning; or its information had no utility, was erroneous, or misleading.

4.0 EVALUATION SUMMARY AND COMMENTS

Thirty-nine (39) of the 40 numerically evaluated, operational tasks were performed in 1994 and one in 1995. The information provided by the Star Gate RV's for each task was evaluated by staff of the tasking organization. The complete compilation of evaluated scores is presented in Table 4-1 which includes a designation of the tasking organization and, where known, a numerical designator for the individual from that organization who signed the response to the evaluation request (in some instances, this was also an evaluator). Also presented are the individual and collective scores for Accuracy (A) and Value (V) for each of the three RV's and the related average and standard deviations for the compiled scores. (Note that the total number of scoring entries for either Accuracy or Value is not equal to the maximum of 120, i.e., 3x4O, since all three RV's did not participated in all tasks). Table 4-2 presents the same scoring data by tasking organization.

Histograms of the scores from Table 4-1 are shown below. Note that "Accuracy" scores tend to cluster around 2's and 3's (55 of the 99 entries) while "Value" scores cluster around 3's and 4's (80 of the 1 00 entries). This is not too surprising as the nonspecific, nebulous nature of the individual task)target requests permits the RV to "free associate" and permits the evaluator to pick and choose from the RV commentary anything that he thinks "may" or "possibly" is related to his problem (and score accordingly) regardless of how much of the RVcommentary

OPERATIONAL TASKS THAT HAVE NUMERICAL EVALUATIONS

1	Doc.	Date	Tasking	Evaluato	Remote Viewer & Scores						Totals	
2			Org.	No.	1A	1V	2A	2V	3A	3V	A	V
3	250	7/13/94	Org. A	1	3.0	3.0	2.0	3.0	4.0	5.0		
4	264	9/6/94	Org. A	2			2.0	3.0	5.0	4.0		
5	270	11/3/94	Org. B	3	5.0	4.0			5.0	4.0		
6	271	11/3/94	Org. B	3	3.0	4.0			5.0	4.0		
7	273	11/3/94	Org. B	3	4.0	5.0	5.0	4.0	4.0	5.0		
8	267	11/3/94	Org. B	3	3.0	4.0			3.0	4.0		
9	268	11/3/94	Org. B	3	3.0	4.0	4.0	3.0	5.0	4.0		
10	269	11/3/94	Org. B	3			3.0	3.0	5.0	5.0		
11	272	11/3/94	Org. B	3				3.0				
12	258	8/3/94	Org. C	4	1.0	3.0	2.0	3.0				
13	257	8/1/94	Org. C	4	3.0	5.0			3.0	5.0		
14	256	7/28/94	Org. C	4	2.0	3.0			5.0	4.0		
15	249	7/11/94	Org. D	5	1.0	4.0	2.0	2.0	2.0	4.0		
16	248	7/6/94	Org. D	5	3.0	3.0	2.0	2.0	1.0	4.0		
17	245	6/24/94	Org. D	5	3.0	3.0			1.0	4.0		
18	252	7/18/94	Org. C	4	4.0	4.0			2.0	3.0		
19	251	7/15/94	Org. C	4	2.0	3.0	1.0	3.0	2.0	3.0		
20	243	5/13/94	Org. C	4	3.0	3.0	5.0	4.0	1.0	4.0		
21	242	5/25/94	Org. C	4	1.5	3.0			1.5	3.0		
22	244	6/6/94	Org. A	1	4.0	5.0	2.0	3.0	1.0	2.0		
23	239	6/12/94	Org. A	6	2.0	2.0	1.0	2.0	2.0	2.0		
24	230	4/1/94	Org. A	7	2.0	2.0	2.0	2.0	1.0	2.0		
25	240	5/17/94	Org. C		2.0	3.0			3.0	4.0		
26	235	4/18/94	Org. C	8	3.0	4.0	3.0	3.0	3.0	4.0		
27	234	4/14/94	Org. C	8	2.0	3.0	5.0	3.0	6.0	5.0		
28	233	4/11/94	Org. C	8	3.0	3.0	3.0	3.0	3.0	3.0		
29	229	3/29/94	Org. C	4	2.0	4.0	2.0	4.0	5.0	4.0		
30	228	3/28/94	Org. C	4	1.0	2.0	3.0	4.0	3.0	3.0		
31	227	3/24/94	Org. C	4	3.0	3.0	4.0	5.0	3.0	3.0		
32	226	3/22/94	Org. C	4	5.0	4.0	5.0	4.0	2.0	3.0		
33	225	3/21/94	Org. C	4	2.0	3.0	3.0	3.0	2.0	3.0		
34	232	4/11/94	Org. E	9	2.0	4.0	5.0	4.0	5.0	4.0		
35	236	4/26/94	Org. E	9	6.0	4.0			6.0	2.0		
36	237	4/26/94	Org. E	9	5.0	4.0	5.0	4.0				
37	241	4/27/94	Org. E	9	3.0	4.0			2.0	4.0		
38	247	6/29/94	Org. D	10/11	1.0	3.0	3.0	3.0	3.0	3.0		
39	265	7/6/94	Org. D	10/11	1.0	3.0	2.0	3.0	2.0	4.0		
40	259	'7/15/94	Org. C	4	5.0	4.0			2.0	2.0		
41	262	8/23/94	Org. C	4(?)	6.0	4.0			4.0	5.0		
42	287	4/23/95	Org. C	12	2.0	4.0			1.0	4.0		
43			Score sums=		106.	130	76.0	83.0	113.5	135.0	296	348
44			Number of entries=		37	37	25					
45			Avg score=		2.9	2.5	2.9					
46			Std.Deviation=		1.4	0.8	1.3					

TABLE 4-1

OPERATIONAL TASKS THAT HAVE NUMERICAL EVALUATIONS

1	Doc.	Date	Tasker	Evaluator	Remote Viewer & Scores						Totals	
2			Org.	No.	1A	1V	2A	2V	3A	3V	A	V
3												
4		**By Tasking Organization**										
5	**Org. C**											
6	258	8/3/94	Org. C	4	1.0	3.0	2.0	3.0				
7	257	8/1/94	Org. C	4	3.0	5.0			3.0	5.0		
8	256	7/28/94	Org. C	4	2.0	3.0			5.0	4.0		
9	252	7/18/94	Org. C	4	4.0	4.0			2.0	3.0		
10	251	'7/15/94	Org. C	4	2.0	3.0	1.0	3.0	2.0	3.0		
11	243	5/31/94	Org. C	4	3.0	3.0	5.0	4.0	1.0	4.0		
12	242	5/25/94	Org. C	4	1.5	3.0			1.5	3.0		
13	240	5/17/94	Org. C		2.0	3.0			3.0	4.0		
14	235	4/18/94	Org. C	8	3.0	4.0	3.0	3.0	3.0	4.0		
15	234	4/14/94	Org. C	8	2.0	3.0	5.0	3.0	6.0	5.0		
16	233	4/11/94	Org. C	8	3.0	3.0	3.0	3.0	3.0	3.0		
17	229	3/29/94	Org. C	4	2.0	4.0	2.0	4.0	5.0	4.0		
18	228	3/28/94	Org. C	4	1.0	2.0	3.0	4.0	3.0	3.0		
19	227	3/24/94	Org. C	4	3.0	3.0	4.0	5.0	3.0	3.0		
20	226	3/22/94	Org. C	4	5.0	4.0	5.0	4.0	2.0	3.0		
21	225	3/21/94	Org. C	4	2.0	3.0	3.0	3.0	2.0	3.0		
22	259	7/15/94	Org. C	4	5.0	4.0			2.0	2.0		
23	262	8/23/94	Org. C	4(?)	6.0	4.0			4.0	5.0		
24	287	4/3/95	Org. C	12	2.0	4.0			1.0	4.0		
25				Score sums=	52.5	65.0	36.0	39.0	51.5	65.0	140	169
26				No. of entries=	19	19	11	11	18	18	48	48
27				Avg score=	2.8	3.4	3.5	3.5	2.9	3.6	2.9	3.5
28	**Org. B**											
29	270	11/3/94	Org. B	3	5.0	4.0			5.0	4.0		
30	271	11/3/94	Org. B	3	3.0	4.0			5.0	4.0		
31	273	11/3/94	Org. B	3	4.0	5.0	5.0	4.0	4.0	5.0		
32	267	11/3/94	Org. B	3	3.0	4.0			3.0	4.0		
33	268	11/3/94	Org. B	3	3.0	4.0	4.0	3.0	5.0	4.0		
34	269	11/3/94	Org. B	3			3.0	3.0	5.0	5.0		
35	272	11/3/94	Org. B	3				3.0				
36				Score sums=	18.0	21.0	12.0	13.0	27.0	26.0	57	60
37				No. of entries=	5	5	3	4	6	6	14	15
38				Avg score=	3.6	4.2	4.0	3.2	4.5	4.3	4.1	4.0

Table 4-2

OPERATIONAL TASKS THAT HAVE NUMERICAL EVALUATIONS

Doc.	Date	Tasker Org.	Evaluator No.	Remote Viewer & Scores 1A	1V	2A	2V	3A	3V	Total A	V
By Tasking Organization											
Org. D											
249	7/11/94	Org. D	5	1.0	4.0	2.0	2.0	2.0	4.0		
248	7/6/94	Org. D	5	3.0	3.0	2.0	2.0	1.0	4.0		
245	6/24/94	Org. D	5	3.0	3.0			1.0	4.0		
247	6/29/94	Org. D	10/11	1.0	3.0	3.0	3.0	3.0	3.0		
265	7/6/94	Org. D	10/11	1.0	3.0	2.0	3.0	2.0	4.0		
			Score sums=	9.0	16.0	9.0	10	9.0	19.0	27	45
			No. of entries=	5	5	4	4	5	5	14	14
			Avg score=	1.8	3.2	2.2	2.5	1.8	3.8	1.9	3.2
Org. A											
102	7/13/94	Org. A	1	3.0	3.0	2.0	3.0	4.0	5.0		
101	9/6/94	Org. A	2			2.0	3.0	5.0	4.0		
82	6/6/94	Org. A	1	4.0	5.0	2.0	3.0	1.0	2.0		
81	6/12/94	Org. A	6	2.0	2.0	1.0	2.0	2.0	2.0		
79	4/1/94	Org. A	7	2.0	2.0	2.0	2.0	1.0	2.0		
			Score sums=	11.0	12.0	9.0	13.0	13.0	15.0	33	44
			No. of entries=	4	4	5	5	5	5	14	14
			Avg score=	2.8	3.0	1.8	2.6	2.6	3.0	2.4	2.9
Org. E											
232	4/11/94	Org. E	9	2.0	4.0	5.0	4.0	5.0	4.0		
236	4/26/94	Org. E	9	6.0	4.0			6.0	2.0		
237	4/26/94	Org. E	9	5.0	4.0	5.0	4.0				
241	4/27/94	Org. E	9	3.0	4.0			2.0	4.0		
			Score sums=	16.0	16.0	10.0	8.0	13.0	10.0	39	34
			No. of entries=	4	4	2	2	3	3	9	9
			Avg score=	4.0	4.0	5.0	4.0	4.3	3.3	4.3	3.8

Comparison - Average Scores by Organization

Organization	Average Scores						
Org. C	2.8	3.4	3.3	3.5	2.9	3.6	
Org. B	3.6	4.2	4.0	3.2	4.5	4.3	
Org. D	1.8	3.2	2.2	2.5	1.8	3.8	
Org. A	2.8	3.0	1.8	2.6	2.6	3.0	
Org. E	4.0	4.0	5.0	4.0	4.3	3.3	

may satisfy the particular measure. If the Accuracy of the information is somewhat uncertain, its Value must be vaguer still, i.e., scored lower. This presumption is supported by review of the scored "pairs" for all cases, e.g., 1 A and 1 V; only rarely does the "V" score equal or exceed the "A" score for a specific RV and target. Note further that of the 1 00 "V" scores shown on Table 4-1, there are no "1" scores[5], while the 99 "A" scores include 13 "l's". Regarding the latter, a detailed review of the evaluator comments and/or the tasking suggests that the importance of these l's is less than the score would imply in all but four cases since:

- the evaluator of Document 243 stated that the RV 3A score "...though vague, is probably correct."

- the tasking and targets for Documents 245,247, 248, 249 and 265[6] concern topics widely publicized in the open media during the same. period, hence the "source" of the RV 1 A and 3A comments, intended or not, is suspect, and

- for Documents 230, 239 and 244, the evaluator's supporting narrative is between the RV(s) and the evaluator:

- has a very narrow information bandwidth, i.e., the RV-derived information cannot be embellished by a dialogue with the evaluator without substantially telegraphing the evaluator's needs and interests, thereby biasing any RV information subsequently derived , and
- is extremely "noisy" as a result of the unidentifiable beliefs, intentions, knowledge, biases, etc. that reside in the subconsciousness of the RV(s) and/or the evaluator .

As a result, the potential for self-deception on the part of the evaluator exists, i.e, he/she "reads" into the RV information a degree of validity that in truth is based upon fragmentary, generalized information and which may have little real applicability to his/her problem. The relevant question in the overall evaluation process is who and what is being evaluated, i.e., is the score a measure of the RV's paranormal capabilities or of the evaluators views, beliefs and concepts?

---

[5]The significance of this omission is further enhanced if one assumes that the evaluators were familiar with the definitions in Table 3-2 since even those 11 instances scored as #2 ("High value") merely require that the input be the "best report to date or first report on this important topic, but [it] did not significantly influence policy or change analyses."

[6](U) The evaluation of Document 265 is actually a second evaluation of the same RV inputs provided many months after the first evaluation of Document 248 and probably done by a different evaluator.

One of the RV's expressed a concern to the author that the protocols that were followed in conducting the RV process in response to the operational tasking were not consistent with those that are generally specified for the study of paranormal phenomena. Whether the claimed discrepancy was detrimental to the information derived by the RV's,or to its subsequent evaluation or use cannot be determined from the available data.

The operational tasking noted earlier concerning activities in North Korea which required precognitive abilities on the part of the RV's provides an opportunity for a post-analysis by comparing the RV predictions against subsequent realities. Additional comparative data of this type is available from operational tasking during the period 11/90 through 1/91 regarding the Middle East situation (this tasking was not numerically evaluated).

6.0 SUMMARY FROM USER INTERVIEWS (U)

Subsequent to the review and analysis of the numerically scored tasking described in the previous sections of this report, the author participated in interviews with representatives of all of the tasking organizations presented in Table 4-1 except Org. D. Only a brief summary of the results from those interviews is presented here; more detailed synopses are presented in Appendix B. In all cases except for Org. C, the interviewees were the actual personnel who had participated directly in the tasking and evaluation of the Star Gate program. For Org. C, the sole interviewee was the Chief of the Analysis Branch; the staff who defined the tasking and performed the evaluations was comprised of his lead analysts.

A brief summary of the salient points which appeared consistently throughout these interviews follows:

- the principal motivation for using Star Gate services was the hope that something useful might result; the problems being addressed were very difficult and the users were justifiably (and admittedly) "grasping at straws" for anything that might be beneficial.
- the information provided by the Star Gate program was never specific enough to cause any operational user to task other intelligence assets to specifically corroborate the Star Gate information.
- while information that was provided did occasionally contain portions that were accurate, albeit general, it was - without exception - never specific enough to offer substantial intelligence value for the problem at hand.
- two of the operational user organizations would be willing to pay for this service if that was required and if it was not too expensive (although one user noted that his organization head would not agree). However, the fact that Star Gate service was free acted as an incentive to obtain "it might be useful - who knows" support for the program from the user organizations.

The reader is referred to Appendix B for additional information resulting from these interviews. However, two inconsistencies noted during the discussion of the numerical evaluations in Section 4.0 were supported by information obtained from the interviews.

On the average, the Org. C evaluators scored higher that those of Org. B. One cause for this discrepancy may be due to the fact that the Org. B evaluators were, in general, skeptical of the process while the lead person at Org. C claimed to be a believer in parapsychology and, in addition, had the last say in any evaluations that were promulgated back to the Star Gate PM. This comment is in no way intended to impugn the honesty or motivation of any of these personnel, merely to point out that this difference in the belief-structure of the staff at these two organizations may have resulted in the perceived scoring bias. As noted above, the subjectivity inherent in the entire process is impossible to eliminate or to account for in the results.

The higher average scoring, especially Accuracy scores, from the Org. A evaluators appears to be explained by the procedure they used to task and evaluate the experiments they were performing with the Star Gate program. Namely, they used a staff member as a "beacon" to "assist" the RV's in "viewing" the beacon's location. Subsequently, the same Org. A staff member evaluated the RV inputs. However, since he/she had been at the site, he/she could interpret anything the appeared to be related to the actual site as accurate. When asked if the information from the multiple RV's was sufficiently accurate and consistent such that a "blind" evaluator, i.e., one who did not know the characteristics of the site, would have been able to identify information from the RV inputs that they could interpret to be accurate, they

all answered in the negative and agreed that the score would have been lower. Again the subjectivity of the process appears - the evaluator could interpret the admittedly general comments from any RV that seemed to relate to the actual site as "accurate", e.g., consider an RV input "there is water nearby", the evaluator knows this it true of almost anyplace especially if one does not or cannot define what kind of water, i.e., is it a lake, a water line, a commode, a puddle?

7.0 CONCLUSIONS AND RECOMMENDATIONS

7.1 CONCLUSIONS

The single conclusion that can be drawn from an evaluation of the 40 operational tasks is that the value and utility to the Intelligence Community of the information provided by the process cannot be readily discerned. This conclusion was initially based solely upon the analysis of the numerical evaluations presented in Section 4.0, but strong confirmation was provided by the results of the subsequent interviews with the tasking organizations (Ref. Section 6.0 and Appendix B). While, if one believes the validity of parapsychological phenomena, the potential for value exists in principal, there is, nonetheless, an alternative view of the phenomenology that would disavow any such value and, in fact, could claim that the ambiguous and subjective nature of the process actually creates a need for additional efforts with questionable operational return on the part of the intelligence analyst.

Normally, much of the data provided by the RV(s) is either wrong or irrelevant although one cannot always tell which is which without further investigation. Whether this reality reduces or eliminates the overall value of the totality of the information can only be assessed by the intelligence analyst. It clearly complicates his/her problem in two ways: 1) it adds to the overburden of unrelated data which every analyst already receives on a daily basis, i.e., the receipt of information of dubious authenticity and accuracy is not an uncommon occurrence for intelligence analysts, and 2) since the analyst does not n(Jrmally know which information is wrong or irrelevant, some of it is actually "disinformation" and can result in wasted effort as the analyst attempts to verify or discount those data from other sources.

The review of the operational tasking and its subsequent evaluation does not provide any succinct conclusions regarding the validity of the process (or the information provided by it). First and foremost, as discussed in Section 5.0, the entire process, from beginning to end, is highly subjective. Further, as noted in Section 3.0, the degree of consistency in applying the scoring measures, any guidance or training provided to the evaluators by any of the tasking organizations and/or the motivation of the evaluators are either unknown or, in the case of the latter, may be highly polarized (see Appendix B) The lack of information regarding these items could account for some of the variability in the scores across organizations noted in Table 4-2, but this cannot be certified and is, at most, a suspicion.

Whether the information provided by the Star Gate source is of sufficient value to overcome the obvious detriment of accommodating the irrelevant information included therein is an open question? More precisely, whether the Star Gate information is of sufficient value to continue this program - vis-a-vis other sources of information and other uses of resources - is an important question for the Intelligence Community to address, irrespective of one's personal views and/or beliefs regarding this field of endeavor, i.e., does the information provided justify the required resource investment?

One method that might assist this evaluation is to develop a means for scoring the complete input from the RV process, i.e., evaluate all information and determine how much is truly relevant, how much is of undeterminable value and how much is completely irrelevant. One could then analyze how much information is being handled to achieve the relevant information (along with some measure of the relevancy) and make judgments on its value vis-a-vis the investment in time and money. Other, less technical methods, for adjudicating this issue also exist.

7.2 RECOMMENDATIONS

Considering the statements above, the only sensible recommendation in this author's mind is to bring some "scientific method" into this process (if it is continued). As evidenced by more than 20 years of research into paranormal psychology, much of it done by institutions of higher education or others with excellent credentials in related fields, validation of parapsychological phenomena may never be accredited in the sense that is understood in other scientific and technical fields of endeavor . Control in any rigorous scientific sense of the multitude of human and physical variables which could, and probably do, influence this process is difficult - perhaps impossible - for any except the most mundane types of experiments, e.g., blind "reading" of playing cards. Even these restricted experiments have led to controversy among those schooled in the related arts.

One of the foundation precepts of scientific endeavor is the ability to obtain repeatable data from independent researchers. Given the subjective nature of RV activities, it is difficult to believe that this aspect of parapsychology will ever be achieved. As an admitted neophyte in this area of endeavor, I categorize the field as a kind of religion, i.e., you either have "faith" that it indeed is something real, albeit fleeting and unique, or you "disbelieve" and attribute all positive results to either chicanery or pure chance.[1]

[1]Practicioners in the field, including those funded under government contracts, would argue with these observations, perhaps vehemently; some would argue further that the phenomenology has been verified beyond question already. This reviewer disagrees; albeit, these observations are not intended to discard the possibility of such phenomena.

Thus, one must recognize at the start that any attempt to bring scientific method into the operational tasking aspects of this project may not succeed. Others with serious motives and intentions have attempted to do this with the results noted above. However, as a minimum, one could try to assure that the scoring measures are succinctly defined and promulgated such that different organizations and evaluators would have a better understanding of what is intended and, perhaps could be more consistent in their scoring. The use of independent, multiple evaluators on each task could aid in reducing some of the effects of the subjective nature of the evaluation process and the possible personal biases (intentional or otherwise) of the evaluators.

Since, according to some parapsychologists, the time of the remote viewing is not relevant to the attainment of the desired information, controlled "blind tests" could be run by requesting tasking for which the accurate and valuable information is already known to determine statistics on RV performance (clearly one key issue in such tests is what information is given to the RV in the task description to avoid any semblance of compromise, not a casual problem). Controlled laboratory experiments of parapsychology have done this type of testing and the results, usually expressed in terms of probability numbers that claim to validate the parapsychological results, have done little to quell the controversy that surrounds this field. Thus it may be naive and optimistic to believe that such additional testing would help resolve the question of the"value of the process" (or its utility for operational intelligence applications), but it might assist in either developing "faith" in those who use it, or conversely "disbelief".

Before additional operational tasks are conceived, some thought could be given to how and what one defines as a "target'. Broad generic target descriptions permit unstructured discourse by the RV which - especially if there is a knowledge (or even a hint) of the general area of interest - leads to data-open to very subjective, perhaps illusionary, interpretation regarding both accuracy and value. If some specificity regarding the target could be defined such that the relevance and accuracy of the RVderived data could be evaluated more readily, some of the uncertainties might be eliminated. In this context, note that the cases where targets were more specific, e.g., the North Korean targets , the resulting scores were generally higher.

Finally, it was noted in Section 5.0 that some of the RV information obtained from operational tasks regarding North Korea (and others concerning the Middle East) depended upon the precognitive ability of the RV's in predicting events yet to occur. These data provide an opportunity for a post-analysis of the accuracy of these predictions by making a comparison with subsequent information regarding actual events (some data for this comparison might require access to classified information from other sources). Such a post-anaiysis would provide data for evaluating the ability of the RV's to perform precognitive tasks and of the related operational value of the predictions. Performance of this post-analysis lies beyond the scope of this paper, but is a topic for a subsequent study if any sponsor is interested.

Next 1 Page(s) In Document Exempt

APPENDIX D

STAR GATE OPERATIONAL USER INTERVIEWS

APPENDIX D:
STAR GATE OPERATIONAL USER INTERVIEW
ORGANIZATION: A
USER POC: #7
DATE: 3 August 1995

OPERATIONAL TASK: SG was asked to participate in a series of experiments to determine if their paranormal service could assist in locating someone who was at an unknown location and had no radio or other conventional method for communicating. Members of the user organization acted as "beacons" for the RV's by visiting sites unknown to the RV's at specified times. The RV's were requested to identify any information that would assist in determining the site location by "envisioning" what the beacons were seeing.

MOTIVATION FOR EMPLOYING STAR GATE: The previous head of the user's group was aware of the program from other sources and requested that SG participate in these experiments in the hopes that some information might be obtained to assist in locating the sites and/or people given the scenario above This situation is similar to that noted from other user interviews, namely, the difficulty of obtaining relevant information from any other source renders the use of the paranormal approach as a worthwhile endeavor from the user's perspective "just in case" it provides something of value

USER ATTITUDE: All of the interviewees were positive regarding the application of this phenomenology to their problem, albeit they all agreed that the RV information provided from the experiments performed to date were inadequate to define the utility of the phenomena and that additional experiments were needed.

RESULTS - VALUE/UTILITY: For each user task, the evaluator was the same individual who had acted as the beacon, i.e., the person who had actually been at the candidate location. Each evaluator noted that some of the information provided by the RV's could be considered to be accurate. When asked if the accuracy of the information would be ranked as high if the evaluator did not know the specifics of the site, i.,e., had not be the "beacon, which is the real "operational situation", all answered in the negative. Several interviewees indicated that their interpretation of the RV data led them to believe that the RV's had witnessed other items or actions the beacon was engaged in but not related to the site of interest. As a result of the experiments done to date, the user decided that the approach being pursued was not providing information of operational utility since it was too general. However, the user was convinced of the possible value of the paranormal phenomena and was planning a new set of experiments using a substantially modified approach in the hope of obtaining useful results.

FUTURE USE OF SG SERVICES: As inferred above, the user would continue to use SG-type services, albeit in a new set of experiments. The user would be willing to pay for this service if it was not too expensive and requested that they be contacted if the program was reinitiated. When advised that they could obtain services of this type from commercial

sources, they noted that this would be difficult due to the highly classified nature of some of their activities.

STAR GATE OPERATIONAL USER INTERVIEW
ORGANIZATION: B
USER POC: #3, et al
DATE: 14 July 1995

OPERATIONAL TASK: Most tasking requested information about future events, usually the time and/or place (or location) of a meeting. Some tasking requested additional information describing a person or a thing, e.g., a vessel. In one instance, after previous "blind" requests had yielded no useful information, the user met with the RV's and provided a picture and other relevant information about an individual in hope of obtaining useful information about his activities.

MOTIVATION FOR EMPLOYING STAR GATE: Gate- SG PM briefed RV activities and his desire to expand customer base. User was willing to "try" using SG capabilities since there was no cost to the user and, given the very difficult nature of user business, "grasping at straws" in the hope of receiving some help is not unreasonable. Note that this organization had tasked the program in the '91 time frame but had not continued tasking in '92-'93 until briefed by the new Star Gate PM.

USER ATTITUDE: DIA POC was openly skeptical, but was willing to try objectively. Members of the organization he supports (Org. B) had varied levels of belief, one individual appear very supportive noting the successful use of psychics by law enforcement groups (based upon media reporting). Evaluation of the tasking was accomplished collectively by the DIA POC and three other Org. B members.

RESULTS - VALUE/UTILITY: None of the information provided in response to any of the tasks was specific enough to be of value or to warrant tasking other assets. SG data was too vague and generic, information from individual RV's regarding the same task were conflicting, contained many known inaccuracies and required too much personal interpretation to warrant subsequent action. User would be more supportive of process if data provided was more specific and/or closely identified with known information. In one instance, a drawing was provided which appeared to have similarity with a known vessel, but information was not adequate to act on.

FUTURE USE OF SG SERVICES: User would be willing to use SG-type services in future. However, in current budget environment, demonstrated value and utility are not adequatetojustifyfundingfromuserresources. Wouldnotfundinanycaseunless program could demonstrate a history of successful and useful product. User believes that RV's working directly with his analysts on specific problems would be beneficial in spite of the obvious

drawbacks. Individual quoted above suggested recruiting RV's from other sources, noting his belief that the government RV's may not be best qualified, i.e., have best psychic capabilities.

STAR GATE OPERATIONAL USER INTERVIEW
ORGANIZATION: C
USER POC: #4
DATE: 26 July 1995

OPERATIONAL TASK: Most tasking requested information describing a person, a location or a thing, e.g., a vessel. Occasionally, the tasking would provide some relevant information about the target or "his/herfits" associates in hope of obtaining useful information about its activities.

MOTIVATION FOR EMPLOYING STAR GATE: In circa 1993, the SG PM briefed RV activities and his desire to expand the customer base. This desire conjoined with the user's[1] belief that it provided an alternate source of information led to the subsequent tasking. User was willing to "try" using SG capabilities since there was no cost to the user and, as noted in other interviews, given the very difficult nature of the user's business, "grasping at straws" in the hope of receiving some help is not unreasonable. Thisorganization had tasked the program in th e (circa) '86-'90 time frame but had terminated tasking since there was no feedback mechanism.

USER ATTITUDE: User was a believer in the phenomena based upon his "knowledge of what the Soviets were doing" and his perceptions from the media regarding its use by law enforcement agencies. He noted that his lead analysts, who generated the tasking, were very skeptical, as was his management. User insisted that analysts be objective in spite of their skepticism. In general,numerical evaluation of the task was performed by the individual who had defined it.

RESULTS - VALUE/UTILITY: This interviewee claimed value and utility for the information provided by the RV's, noting that information regarding historical events was always more accurate that information requiring predictions. RV's were "fairly consistent' in identifying the "nature" of the target, e.g., is it a person or a thing, but not always. On occasions where RV inputs were corroborated, additional data were requested, but these data usually could not be corroborated. User commented that all reports had some accurate

[1]Only one person provided all of the information at this review. Where the "user" or "interviewee" is cited, it reflects the remarks of that single individual.

information[2], 2 however, the SG data provided was either not specific enough and/or not timely enough to task other assets for additional information. Some SG data was included in "target packages" given to field operatves- however, there was no audit trail so there is no evidence regarding the accuracy or use of these data. User also noted that classification prohibited data dissemination as did concerns about skepticism of others regarding the source and the potential for a subsequent negative impact on his organization.

FUTURE USE OF SG SERVICES: User desires to continue using SG-type service if the program continues. In addition, the user stated that he would be willing to pay for the service if necessary. However, subsequent discussion indicated that his management would not fund the activity unless the credibility could be demonstrated better and the phenomenology legitimized. User went on to claim that only the sponsorship of a government agency could "legibmize " this activity and its application to operational problems. User believes that RV's working directly with his analysts on specific problems would not be beneficial due to the skepticism of his analysts and the deleterious impact that would have on the RV's. The views provided by the user - note none of the actual evaluators were present - appeared to be unique to him and his belief in the phenomenology, i.e., his remarks indicated that the use of this process was not actively supported by anyone else in his organization. The numerical evaluations of the 19 tasks performed in 1994/95 certainly do not indicate, on the average, either a high degree of accuracy or value of the data provided.

[2]User was unaware that the tasking organization and its primary mission were known to the RV's. Portions of the data provided by the RV's could have been predicted from this knowledge.

STAR GATE OPERATIONAL USER INTERVIEW
ORGANIZATION-. E
USER POC: #9
DATE: 7 July 1995

OPERATIONAL TASK: Request to assist in determining if a suspect was engaged in espionage activities, e.g., who is he meeting? where? about what? are these activities related to espionage or criminal actions? Tasking comprised a series of four sequential tasks, each time a bit more information was provided to the RV's, including at one point the name of the suspect. (Note-. this "sequential tasking" is unique. Each of the tasks assigned from other operational organizations was a "singular" or "stand alone" event.)

MOTIVATION FOR EMPLOYING STAR GATE: SG PMO briefed RV activities and his desire to expand customer base. User was willing to "try" using SG capabilities since there was no cost to the user and, given the very difficult nature of user business, "grasping at straws" in the hope of receiving some help is not unreasonable.

USER ATTITUDE:

Pre-SG experience - User (#9) had a perception of beneficial assistance allegedly provided to domestic police by parapsychologists; thereby he was encouraged to try using the SG capabilities and hopeful of success.

Post-SG experience - Still very positive in spite of the lack of value or utility from SG efforts (see below). User is "willing to try anything" to obtain assistance in working his very difficult problems.

RESULTS - VALUE/UTILITY: None of the information provided in any of the four sequential tasks was specific enough to be of value or to warrant tasking his surveillance assets to collect on-site information as a result of SG information. SG data was too generic and while it may have contained accurate information, it required too much personal interpretation to warrant subsequent actions by his assets. Much of the SG information was clearly wrong so there was no way to ascertain the validity of the rest. One major deficiency noted in the SG responses was the lack of any RV data regarding large fund transfers that the suspect was known to be engaged in and which the user believes would have been uppermost in the suspect's mind. User would be more supportive of process if data provided was more specific and/or closely identified with known information.

FUTURE USE OF SG SERVICES: User would be willing to use SG-type services in future. However, in current budget environment, demonstrated value and utility are not adequate to justify funding from user resources. User would be willing to have a joint activity whereby RV's work directly with his analysts on specific problems if: a) user did not pay for RV services and b) commitment for joint RV's services was long term , i.e., several years.