

Evaluation of a Program on Anomalous Mental Phenomena

RAY HYMAN

1227 University of Oregon, Department of Psychology, Eugene, OR 97403

Abstract — Jessica Utts and I were commissioned to evaluate the research on remote viewing and related phenomena which was carried out at Stanford Research Institute (SRI) and Scientific Applications International Corporation (SAIC) during the years from 1973 through 1994. We focussed on the ten most recent experiments which were conducted at SAIC from 1992 through 1994. These were not only the most recent but also the most methodologically sound. We evaluated these experiments in the context of contemporary parapsychological research. Professor Utts concluded that the SAIC results, taken in conjunction with other parapsychological research, proved the existence of ESP, especially precognition. My report argues that Professor Utts' conclusion is premature, to say the least. The reports of the SAIC experiments have become accessible for public scrutiny too recently for adequate evaluation. Moreover, their findings have yet to be independently replicated. My report also argues that the apparent consistencies between the SAIC results and those of other parapsychological experiments may be illusory. Many important inconsistencies are emphasized. Even if the observed effects can be independently replicated, much more theoretical and empirical investigation would be needed before one could legitimately claim the existence of paranormal functioning.

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 report, this issue, p. 3)

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. (Utts report, this issue, p. 3)

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) — are 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 scien-

tific 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 modern 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 para-

psychologists 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 meta-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-Goat 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 meta-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 ex-

periments 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. Parapsychology 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 admit 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 SAIC did benefit from the input of a distinguished oversight committee. But this still falls short of what could have taken place in an open forum.

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 in 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. 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 nonparapsychologists.

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* (1977, 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 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 may 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 ap-

parent 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'être* 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 ex-

periments. 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 *aparadigm*. 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" 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. Be-

cause 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' in 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 that 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 they 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 *The Psychology of the Psychic* (Marks & Kamman, 1981) in which the authors 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 be-

tween 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 misled 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 statistical 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 accelerate 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 discovery 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. The 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.

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

JESSICA UTTS

Division of Statistics, University of California, Davis 95616

Ray Hyman's report (this issue, p. 31), written partially in response to my report (this issue, p. 3) elucidates the issues on which he and I agree and disagree. I basically concur with his assessment of where we agree and disagree, 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." (this issue, p. 35)

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 (this issue, p. 36)" 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 done 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 parapsychologists did not simply spring out of thin air. The original ganzfeld experiments followed from Honorton's observation at Maimonides 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." (this issue, p.38)

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 1995, 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 now 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 variability 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 is 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" (this issue, p. 49).

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 studied 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., 295.