Preliminary Evaluation of SRI/SAIC Anomalous Mental Phenomena Program

Ray Hyman
August 29, 1995

For twenty-five years I have been in touch with the literature of psychical research, and have had acquaintance with numerous "researchers." I have also spent a good many hours (though far fewer than I ought to have spent) in witnessing (or trying to witness) phenomena. Yet I am theoretically no "further" than I was at the beginning; and I confess that at times I have been tempted to believe that the Creator has eternally intended this department of nature to remain baffling, to prompt our curiosities and hopes and suspicions all in equal measure, so that, although ghosts and clairvoyances, and raps and messages from spirits, are always seeming to exist and can never be fully explained away, they also can never be susceptible of full corroboration.

The peculiarity of the case is just that there are so many sources of possible deception in most of the observations that the whole lot of them may be worthless, and yet that in comparatively few case can aught more fatal than this vague general possibility of error be pleaded against the record. Science meanwhile needs something more than bare possibilities to build upon; so your genuinely scientific inquirer--I don't mean your ignoramus "scientist"--has to remain unsatisfied.

--William James, 1909.

Jessica Utts and I have each been given the task to evaluate the results of the research program on anomalous mental phenomena carried out at SRI and SAIC from 1973 through 1992. Because of the limited time allotted for this task, we have focussed on the reports of the work selected to best convey the outcome of this program. However, even this selective focus places severe constraints on the adequacy of our evaluations. A fully comprehensive evaluation of the program would require a minimum of several months and would include visits to the sites of the experiments as well as some reanalysis of the raw data. Consequently, my present assessment should be considered the tentative outcome of a quick first pass.

We were asked to assess how well the results meet scientific standards as well as how well the alleged anomalous mental phenomena can be harnessed for intelligence gathering. On the basis of our conclusions we were further requested to recommend whether investigations into this subject should continue and, if so, in what manner.

The Scientific Status of the Program

In their final report (1994) the investigators conclude that they have clearly demonstrated anomalous cognition, but not anomalous perturbation. According to common scientific standards, I would judge such a conclusion as premature. If, for example, a scientist announces the discovery of a new element, the claim is not recognized until the element's existence has been carefully documented in two or more independent laboratories. The reason for such caution is obvious. The history of science contains many examples of discoveries that subsequently could not be replicated and eventually had to be attributed to some artefact--known or unknown. Certainly, the claim that anomalous cognition exists is much more revolutionary in its consequences than the claim that a new element has been observed. So, at the very least, we would want to see the claim of anomalous cognition supported by independent replications in other laboratories. We do not have anything like this at this time. Possibly, this could be an unfortunate consequence of the results having been classified as secret until very recently.

My first scan through the reports impressed me with the apparent consistency with which the best percipients or subjects produced significant evidence for anomalous cognition. I was also impressed, in many instances, with the apparent sophistication in methodology and data analysis. However, as was my experience in dealing with the ganzfeld database, further examination began to raise questions and doubts. I also began noticing inconsistencies, incompleteness of documentation, and other problematic signs. Again, I suspect that some of these drawbacks can be attributed to the secret auspices under which the research was conducted.

I was unimpressed by the results of meta-analysis on the psychoenergetic research conducted at SRI International from 1973-1988. Indeed, this particular report illustrates the drawbacks of relying on meta-analysis to draw conclusions. The meta-analysis is based on a total of 25, 449 trials. The probability of the observed hit rate for this total to have occurred by chance is vanishingly small. Obviously, the departure from chance expectation is real. The authors of this report conclude that, "Using accepted criteria set forth in the standard behavioral sciences, we conclude that this constitutes convincing, if not conclusive, evidence for the existence of psychoenergetic functioning." The "accepted criteria" that they mention refer to rejecting the null hypothesis; these same criteria are silent on the reasons for the departure from this hypothesis.

The problem here is that plausible, mundane alternative explanations exist for this departure from the null hypothesis. The vast majority--of these trials were collected under the original protocol developed by Targ and Puthoff. In this protocol, a subject would be closeted with an experimenter at SRI. A target team would visit a randomly selected site within a half-hour drive of SRI. While the target team was at the site, the subject would describe his/her impressions for 15 to 30 minutes. When the target team returned to SRI, all the participants, including the subject, would visit the site and discuss the correspondences between the target and the subject's impressions. On a second day, the same subject would go through a similar procedure. An experimental series typically consisted of nine such trials with a given subject. At the conclusion of the series, the transcripts of the subjects' impressions were given to a judge. The judge visited the sites and, at each site, ranked the nine impressions from 1 to 9 in order of how well he/she thought they described the site. If the average ranking was significantly better than that expected by chance, the outcome was declared to be evidence for anomalous cognition.

This protocol has several problems. They all derive from the fact that each successive trial is not independent of the preceding ones. The first problem is that the rank statistic that Targ and Puthoff originally used assumed that the trials were independent. The consequence was that the statistical outcomes exaggerated the degree of significance actually present. More serious problems arose when David Marks discovered that the transcripts given to the judge sometimes contained clues that were sufficient for the judge to correctly match impressions against target site without assuming anomalous cognition. Even if such clues could be edited out of the transcripts, I pointed out that a fatal flaw still existed.

Rather than to go into technical detail, I will give one illustration how leakage can occur with this protocol. Assume that the target on the first day was the Hoover Tower on the Stanford campus. Because the subject has been given feedback immediately after the first session, he/she may reasonably avoid describing anything that resembles the Hoover Tower during the second session. Supposed that the target for the second day was the Palo Alto train station. On the third day, we can assume that the subject will not describe anything that closely resembles either the Hoover Tower or the train station. The impression for the third day, then, might be judged as being closer to the target for the third day than to either of the first two targets simply because it clearly does not correspond to either of the targets for the first day. This problem is compounded as the trials progress through the entire series.

Thus, a skeptic can easily imagine non-paranormal reasons why the judge might consistently match impressions against target sites significantly better than chance. In this case, the possible artefact is obvious. In much scientific research, biases and artefacts can be much more subtle and elusive. New protocols, instrumentation, methodologies, and analytic techniques require long periods of debugging. Often a new field of inquiry might proceed for years before it is discovered that hidden flaws have biased the outcomes. This is why independent replications and consistent and lawful outcomes across a variety of conditions are so crucial in the sciences.

The problem I have with the outcomes from the present program is that we are dealing with novel protocols and methodologies which have not had time to be sufficiently debugged and have not been independently replicated. In addition, the results that have been obtained so far suggest that anomalous cognition still comes and goes in mysterious ways. Consistency both within the program and with other findings in parapsychology is not impressive.

I could go into similar detail for each of the other reports. Instead, I will deal with them as a unit. The reason is that these remaining reports deal with experiments that were clearly conducted with a better protocol. If alternative explanations exist to account for the results of these latter experiments, they are not as obvious as the explanations for the earlier remote viewing experiments. Although no obvious alternative explanations come to mind, warning signs abound. Of the 10 independent experiments conducted at SAIC, several are described as pilot attempts. Some fail to replicate previous experiments in the program. For example, the first attempt to replicate the Chinese photon production experiment succeeded. However, a second, more careful attempt at replication failed. The investigators conclude that the original Chinese effect was an artifact. What will happen if the experimenters try to replicate each of the other "successful" experiments with more elaborately controlled designs?

We know from the descriptions in the reports that some of the experiments provided possibilities for sensory leakage and other biases. The investigators either discount these possibilities or believe they adequately compensated for them. Not enough documentation exists in the reports to be sure that all necessary controls were consistently in place. In some of the experiments, for example, the experienced subjects operated from their own homes, hundreds of miles from the laboratory and the principal investigator. These experienced subjects presumably know the procedures, the target pool, the laboratory personnel and the judge.

The judge raises another problem. The investigators do not try to explain adequately why the judging procedure in the ganzfeld procedure succeeds best when the subject does the judging while it apparently succeeds best in the remote viewing situation when someone other than the subject does the judging. Apparently only one or two judges consistently give good results. From what Ed May told us, I gather they typically use one judge and this judge is the same one across several experiments. Even if we assume the judge is honest, conscientious and otherwise free from suspicion, the scientific community will not readily accept conclusions that depend upon the use of particular individual. This is like the experimenter effect. Many parapsychologists argue that only certain experimenters are capable of obtaining evidence of anomalous cognition. If this is true, parapsychology faces serious obstacles in its attempts to gain scientific recognition. Scientific credibility depends critically on the ability of any conscientious observer to obtain a given outcome.

From its inception in the late 1800s, parapsychology has been plagued by such problems as non-replicability, non-cumlativeness, lack of robustness, and inability to specify boundary conditions. The decline effect, which was the subject of one of the experiments in the current program, is a case in point. When Rhine announced the discovery of this effect it was presented as a strong argument for the reality of ESP. Rhine argued that he had discovered the decline effect in experiments whose investigators had not been looking for it. Rhine believed that the decline effect also explained why so many ESP experiments yielded overall results consistent with chance. Because of the decline effect the first half of many experiments typically showed an excess of hitting above chance. The second half, on the other hand, would show hitting below chance. The two halves, when pooled over the entire experiment would cancel each other out and yield an overall result that seemed to be due to chance.

As the present investigators point out, the decline effect can show itself in multitudinous ways. Investigators have reported decline effects within a run, within a series, within a collection of studies, and even across subjects. When decline effects are found in a body of data, the parapsychologists do not hesitate to declare this evidence for anomalous cognition. However, when decline effects are *not* found, investigators, including the present ones, are still willing to assert the existence of anomalous cognition if other departures from a chance baseline can be found. The problem this raises is that we have no way of specifying conditions under which psi will not be found. Just about any departure from a statistical model can be evidence for psi. We have no way of telling when psi is not present.

I will here briefly mention other signs of potential inconsistencies. The central claim for the autoganzfeld experiments is that, as Honorton allegedly predicted, evidence for anomalous cognition was obtained for dynamic targets and not for static targets. In their first replication attempt, the present experiments obtained evidence for anomalous cognition only with the static targets and not with the dynamic targets. As always, they can generate a quasi-plausible explanation. They do this in terms of bandwidth. Although, the second experiment to test this idea does apparently support their conjecture, the results are not altogether compelling and more needs to be done. Honorton and his colleagues claim that the most consistent personality correlate of anomalous cognition is extroversion. Yet, the major replication of Honorton's work, which was subcontracted to the present project, shows the introverts, if anything, doing somewhat better.

I can go on and list other inconsistencies and possible problems. However, I will stop at this point so that I can get this draft into the mail. The quotation from William James at the beginning of this report captures my feelings about the scientific status of the present project.

My advice is that, if the project is continued, that serious effort be made to contract with a number of independent parapsychological laboratories as well as some non-parapsychological, neutral investigators to replicate the key findings from the present project. The data from the present project should be sufficient, given the claims being made for it, to allow us to specify the appropriate conditions, the effect size, and the number of cases necessary to get a significant effect across different laboratories given that anomalous cognition exists. Presumably, the labs could either use the best subjects from the SAIC experiments or use a similar screening device to find those individuals who belong the one percent of the population who supposedly have AC abilities.

I have not discussed potential utility of remote viewing. If we accept the conclusions made by the investigators on the current project, the potential for utility is bleak. Although they accept the reality of anomalous cognition, they state or imply, in several places, that operational applications of anomalous cognition do not look very promising.

Dr. Verona called 7 Sept 1995

SG1J

To complain about the Schnabel article.

[redacted]

Veronas says don't hire these people in-house or don't do it under any circumstances an gov. sponsored.

slanderous slime