Report of a
Workshop on Experimental Parapsychology

International Security and Commerce Program
Office of Technology Assessment
United States Congress

February 22, 1989

Approved For Release 2003/09/09 : CIA-RDP96-00792R000100130001-2

OFFICE OF TECHNOLOGY ASSESSMENT
International Security and Commerce Program

Workshop

September 30, 1988

Professor James Alcock
Glendon College
Toronto, Canada

Daryl Bem
Professor of Psychology
Cornell University
Ithaca, NY

Charles Honorton
Director
Psychophysical Research Laboratories
Princeton, NJ

Ray Hyman
Professor of Psychology
University of Oregon
Eugene, OR

Robert G. Jahn
Professor of Aerospace Studies
Princeton University
Princeton, NJ

John Palmer
Institute of Parapsychology
College Station
Durham, NC

Theodore Rockwell
Chevy Chase, MD

Marcello Truzzi
Department of Sociology
Eastern Michigan University
Ann Arbor, MI

Jessica Utts
Professor
University of California, Davis
Davis, CA

Rapporteur
Courtland Lewis
Arlington, VA

Chairman
Alan Shaw
Office of Technology Assessment

Approved For Release 2003/09/09 : CIA-RDP96-00792R000100130001-2

CONTENTS

I. INTRODUCTION

HISTORICAL BACKGROUND

The modern field of study known as parapsychology grew directly out of 19th-century "psychical research" into reported psychic phenomena such as telepathy and clairvoyance. The field had its earliest roots in various investigations into a wide range of supernatural, occult, and mystical topics reaching far back into human cultural history. However, the field did not become formalized as an arena of organized scholarly research until 1882, with the establishment in London of the Society for Psychical Research (SPR). A U.S. counterpart, the American Society for Psychical Research (ASPR), was founded in 1885. Both societies still exist today.[1]

From the beginning, the field represented by these organizations has been highly controversial. Although it has occasionally attracted and engaged well-known scientists, it has generally encountered substantial resistance and criticism from the scientific "establishment" on the basis of its aims and methods. To its detriment, parapsychology has had difficulty in freeing itself of association, in the minds of those outside the field, with extreme and unsubstantiated claims, commercial ventures of questionable validity, and a certain amount of quasi-scientific "research" carried out in its general area of inquiry.

NATURE OF RESEARCH

Throughout this century a considerable body of parapsychology research has been conducted in a manner that attempts deliberately to follow scientific methodology. The research approaches most often derive from and resemble research in psychology, although a few programs based on engineering and applied science have recently appeared. A seminal example of academic research in the field was the work of J.B. Rhine and Louisa Rhine at Duke University, beginning in the 1920s. The Rhines established many of the basic concepts and protocols of modern parapsychology. In 1937, they founded The Journal of Parapsychology, which is perhaps the leading journal in the field today.

The general focus of parapsychological research is on "psychic phenomena," or "psi," which is defined as "processes of information and/or energy exchange which involve animate consciousness in a manner not currently explicable in terms of known science."[2] The field can be divided into several major categories and subcategories, as shown in Table 1. Most (but not all) of the parapsychological research in laboratory settings involves categories I and II.

---

1 For a detailed examination of the development of parapsychology, see: Jahn, R.G. & B.J. Dunne. Margins of Reality: The Role of Consciousness in the Physical World. New York: Harcourt, Brace, Jovanovich, 1987.

2 Jahn, R.G. The persistent paradox of psychic phenomena: An engineering perspective. Proceedings of the IEEE, 1982, 70(2): 136-170.

RECENT CONTROVERSY

In 1957 a professional organization in the field, the Parapsychological Association (PA), was formed. (Unlike the SPR, membership in the PA requires formal recognition of professional status by the Association's Council.) It was accepted as an affiliate by the American Association for the Advancement

TABLE 1

CATEGORIES OF PSYCHIC PHENOMENA

I. Extrasensory Perception (ESP)

- A. Telepathy
- B. Clairvoyance
- C. Precognition/Retrocognition
- D. Animal ESP

II. Psychokinesis (PK)

- A. Physical Systems (equipment, etc.)
- B. Biological Systems

III. Survival

- A. Reincarnation
- B. Apparitions
- C. Mediumship

IV. Out-of-Body Experiences (OBE)

SOURCE: Jahn, R.G. The persistent paradox of psychic phenomena: An engineering perspective. Proceedings of the IEEE, 1982, 70(2): 136-170.

of Science (AAAS) in 1969, signaling a tentative new status and recognition of the field. In 1976 a group of philosophers, magicians, science writers, scientists, and others concerned about a widespread increase in popular interest in the occult formed a Committee for the Scientific Investigation of Claims of the Paranormal (CSICOP).[3] CSICOP and a number of affiliated local groups have since become a primary source of criticism and skepticism directed primarily at highly publicized claims of the paranormal, but often extending more generally to the field of parapsychology. A highly charged and long-running debate has emerged, featuring charges and countercharges between members of CSICOP and the scientific community, on the one hand, and members of the parapsychological community, on the other.

One recent chapter of that debate appeared in the March 1985 issue of The Journal of Parapsychology, in which a leading parapsychology researcher (Charles Honorton) and a leading psychologist and critic (Ray Hyman) presented their views on the nature of one category of parapsychology experiments.[4] In the December 1986 issue of the same journal, the two published a "joint communique" in which they explored in a constructive manner areas of agreement and disagreement regarding experiment design, documentation, and data analysis.[5] (Both individuals participated in the OTA workshop.) Another landmark publication appeared in 1987: a substantial portion of one issue of Behavioral and Brain Sciences was devoted to articles and letters on the subject.[6]

The most recent episode in the continuing controversy over the field appeared in the form of a December 1987 report by the National Research Council (NRC), which evaluated parapsychology along with other possible mechanisms for enhancing the performance of U.S. Army personnel.[7] The NRC's evaluation of the potential of parapsychology was generally negative, prompting strenuous and detailed claims of bias and unfairness from the parapsychological community.[8]

PURPOSE OF THE OTA WORKSHOP

It was in this context that the OTA workshop took place. OTA examined parapsychology at the request of its oversight body, the Technology Assessment Board, which expressed interest in the human potential aspect of

---

3 It should be noted that few scientists were involved in the creation of CSICOP. Although today many of its Fellows are scientists, the organization has always been led and operated by non-scientists.

4 Hyman, R. The ganzfeld psi experiment: A critical appraisal. and Honorton, C. Meta-analysis of psi ganzfeld research: A response to Hyman. The Journal of Parapsychology, 1985, 49(1): 3-91.

5 Hyman, R. & C. Honorton. A joint communique: The psi ganzfeld controversy. The Journal of Parapsychology, 1986, 50(4): 351-364.

6 Behavioral and Brain Sciences, 1987, 10(4).

7 National Research Council. Enhancing Human Performance: Issues, Theories, and Techniques. Report of the Committee on Techniques for the Enhancement of Human Performance, Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press, 1988.

8 See, for example: Palmer, J.A., C. Honorton, & J. Utts. "Reply to the National Research Council Study on Parapsychology." Special report prepared for the Board of Directors of the Parapsychological Association, Inc., Research Triangle Park, N.C., 1988.

parapsychological research. This workshop and report are OTA's response to that interest. OTA's intention in holding the workshop was not to continue the debate described above, but to illuminate it by identifying the main points of contention and the reasons for them. For that reason, background materials were researched to provide a fuller elucidation of some issues not adequately aired in the course of the one-day meeting.

Discussion was limited to the area of laboratory parapsychology. Participants represented three groups: researchers in parapsychology, constructive critics of the field, and knowledgeable individuals who follow the field from a position of more neutral interest or involvement. Although the meeting was to some extent an occasion for spirited debate, no attempt has been made to "score" that debate. This report will describe the controversies only to the extent that they bear on OTA's inquiry into the status of the field.

II. DESIGN AND CONDUCT OF EXPERIMENTS

Controversy in the area of experiment design and conduct appears to center around the two issues of flaws in methodology and replicability of experiments.

METHODOLOGY

One of the most persistent and perhaps most damaging charges leveled by the critics is that various flaws in the design and conduct of nearly all parapsychology experiments render their results scientifically less meaningful than they are reported to be -- and, in fact, often account for the results. Parapsychologists dispute this charge, saying that most of the alleged flaws have little or no effect on the experimental findings that indicate "psi," and that much research in psychology (the home discipline of many critics) -- and indeed in many other fields of science -- exhibits similar flaws without encountering comparable criticism and questioning of the results. They say the question of flaws is highly subjective, with individual views differing widely over what distinguishes an experimental "flower" from an experimental "weed." Further, the parapsychologists claim, this criticism is a red herring intended to tarnish the image of parapsychology within the scientific community and the public and to impede broader awareness of findings and progress in the field. Those at the workshop asserted strenuously that they and their colleagues make every effort to conduct their experiments in the most rigorous manner possible. The real issue, they say, is what constitutes a flaw. The types of flaws that have been brought into question include purported <u>procedural</u> flaws such as:

- inadequate precautions against "sensory leakage" (for example, allowing subjects in a telepathy experiment to examine the same "target", such as a photograph, that had been handled earlier by a person serving as "sender")
- inadequate security provisions (e.g., to prevent tampering with equipment)
- improper randomization techniques (such as hand shuffling of cards)
- feedback (such as failure to randomize targets before presenting them to the subject for judging)
- incomplete documentation of experimental procedures
- inconsistency of conditions and procedures used in an experiment.

Also at issue are various purported <u>analytical and statistical</u> flaws, including:

- multiple testing or analysis (i.e., to find the test or analysis that gives the most positive "effect size" for psi)

- o underestimation of the effective error rate and overestimation of the actual significance level
- o erroneous use of particular statistical procedures.

In their joint communique, Honorton argued that there is "no significant correlation between indices of study quality [flaws] and study outcome." Hyman agreed in most cases, but insisted that there is a positive correlation in the case of poor randomization, feedback, and inadequate documentation.[9] (This assertion was rebutted in a later analysis.)[10] Workshop participants generally agreed that in any event the "flawless," perfect experiment is very rare in any field, if it exists at all. The point of the issue, critics say, is that the presence of a flaw usually implies inadequate controls for error. The most stringent among them believe that an apparently successful experiment cannot be considered to have demonstrated a true anomaly (i.e., a possible psi effect) unless it can be shown that the experiment is completely flawless. The parapsychologists counter that any such flaws should be an issue only if they could have systematically influenced the result.

Another point made by parapsychologists is that, by and large, the critics have done very little experimentation in parapsychology. Therefore, they have little awareness of the difficulties encountered or of what the truly significant flaws might be. Indeed, the parapsychologists note, there could be additional flaws that have not yet been considered and which critics could identify through actual experimentation.

The NRC report distinguished three types of criticisms relating to flaws in experiment design and conduct.[11] First is what is called the "smoking gun," in which observed findings are said to be definitely attributable to a specific flaw or other factor (such as deliberate fraud) that is shown to be present. Second, and more common, is the "plausible alternative" allegation that a particular flaw is present and could have produced the reported results. Third is the "dirty test tube" concept, which alleges not that a particular discovered flaw produced the positive result but simply that the presence of flaws in the experiment demonstrates a general sloppiness which brings the results into question. In other words, if results have been obtained under conditions that fail to meet "generally accepted standards" for scientific research (such as clean test tubes), that fact alone casts doubt on the accuracy and validity of the results. The authors of the NRC report claimed that, while they could find no instances of the "smoking gun" or "plausible alternative," even the best parapsychology experiments exemplify the "dirty test tube" problem (an allegation that has been heatedly disputed by parapsychologists).

Participants agreed that the burden of proof in the first two categories should be on the critic making the allegation, as it is in other sciences where an experiment is challenged on these grounds. In the case of the "dirty test tube," the critics feel that the burden of proof should be on the experimenter. This position brings from the parapsychologists the objection that it is non-

9 Hyman & Honorton, op. cit., p. 353.
10 Harris, M.J. & Rosenthal, R. "Interpersonal Expectancy Effects and the Human Performance Research." Washington, DC: National Academy Press, 1988. (p. 3)
11 National Research Council, op. cit., pp. 199-200.

falsifiable -- that is, that there is no definitive defense against the charge of "general sloppiness". Without detailed criteria for what constitutes acceptable research, they say, criticism of an experiment cannot be anticipated or refuted. In their view, it is an example of the unfairly high standards that are set for parapsychology research, in which no amount of care is ever "enough."

The critics present at the workshop tended to agree in principle with the unfairness of "non-falsifiability," but they believe that the "dirty test tube" charge is nevertheless valid because the success of the experiments rests on whether a small but significant departure from some statistical norm or observed norm has in fact been seen, making it essential to eliminate extraneous factors. Parapsychologists object that the charge itself is irrelevant -- that if the alleged flaws or "dirt" have no demonstrable connection with the experimental outcome, it is capricious to argue that the experiment is invalid.

The litany of charge and countercharge could be detailed much further. Suffice it to say that the controversy surrounding parapsychology is nowhere more heated and complex than in the area of experimental methodologies.

REPLICABILITY

For critics of parapsychology, a key argument against the existence of psi phenomena is the difficulty of replicating (repeating) successful experiments -- a difficulty that also plagues many accepted areas of the social and behavioral sciences. If the same effect could be produced many times by different experimenters using different equipment but the same procedures, then the likelihood that an extraneous variable is producing the effect would be much lower and the validity level of the anomalous results would be higher. However, positive results are in most cases maddeningly (to parapsychologists) difficult to replicate, even with the same operator. (Because the effects seem to be highly operator-specific, results are even more difficult to replicate with different operators or groups of operators.)

Parapsychologists attribute this replicability problem to the elusiveness and weakness of the psi "signal," and to its vulnerability to subtle subjective factors, most of them unknown. Critics contend that the near non-replicability merely demonstrates the effects of randomness combined with multiple flaws in experiment design and conduct.

The parapsychologists are well aware that experimental flaws, both actual and alleged, become less important as a phenomenon becomes more replicable. Therefore, in recent years they have placed great emphasis on replicability. As one noted in his presentation, "The final criterion of the success or failure of my research program is the extent to which I'm able to develop procedures that I can articulate with sufficient precision that other people are able to obtain similar results."

In the case of ESP-oriented experiments (such as remote perception or "ganzfeld" psi experiments), the number of these more standardized experiments is still not large. There, parapsychologists have relied upon what is termed "meta-analysis" of many studies, taken as a group, to give statistical confirmation for a pattern of positive results. The rationale is that "an

experiment or an effect may be considered replicated if a series of replication attempts provides statistically significant evidence for the original effect when analyzed as a series."[12] It is also contended that this approach allows the analyst to assess methodological flaws to determine their effect empirically across the data base.

Using the meta-analytic approach, parapsychologists have calculated an overall "success" rate, in achieving statistically significant effects ($p \leq .05$), of 45% for ganzfeld ESP studies and 21% for random event generator (REG) studies. However, critics have objected strongly to these findings, saying that many of the studies included are not directly comparable due to differences in methodology. They also claim that many unsuccessful experiments have not been published and are therefore not in the data base. This "file drawer problem", they assert, artificially increases the success ratio and the overall effect size.[13] They also cite a corresponding "retrospective study" problem, in which small exploratory studies are included in the data base after the fact if their results turn out positive.[14] Furthermore, they believe, the proliferation of flaws in methodology simply perpetuates the same erroneous results.[15]

The debate over replicability, particularly in the context of meta-analysis, has been heated and continues to be so. However, the nature of the controversy is currently being outlined in detail and areas of agreement are being identified. For example, the recent joint communique by two of the workshop participants refers to "the growing role we believe meta-analysis will play, both in the evaluation of research quality and in the assessment of moderating variables. We urge parapsychological investigators to plan and report their experiments with the idea that their single experiment will contribute to a future meta-analysis."[16]

This idea relies in part on guidelines for greater uniformity in experiment design and conduct which the two authors specified in their article. Such uniformity or consistency is viewed as crucial by many moderate critics, who point to the fact that there has never been a successful parapsychology experiment which any and all competent researchers, following the same procedure, can replicate exactly. As one of the workshop participants noted in a recent article, "Although one cannot set precise standards that evidence of psi must meet, judgment should be suspended until there is at least some consistency among research findings from a body of methodologically irreproachable experiments, at least some of which are repeatable [by any other researcher]."[17]

---

12 Rao, K.R. & J. Palmer. The anomaly called psi: Recent research and criticism. Behavioral and Brain Sciences, 1987, 10(4): 539-551 (p. 544).

13 It should be noted that at least one critic (Hyman) has acknowledged the file drawer problem to be less evident in parapsychology than in other areas of science. (Hyman, R. & C. Honorton, op. cit., p.359)

14 The selective reporting of "positive" results by either nonreporting or retrospective reporting is prohibited by Parapsychological Association policy.

15 Hyman, R., op. cit., pp. 8-15.

16 Hyman, R. & C. Honorton, op. cit., p. 361.

17 Alcock, J.E. Parapsychology: Science of the anomalous or search for the soul? Behavioral and Brain Sciences, 1987, 10(4): 553-565 (p. 559).

CRITERIA FOR ACCEPTABLE EXPERIMENTS

One challenge that parapsychologists face as they strive for greater uniformity and more cleanly defined experimental methodologies is the variety of (putative) phenomena and corresponding experiment types that are being pursued. In the psi ganzfeld experimental regime,[18] for example -- which focuses on ESP effects -- the experimental setup, equipment, and methodology are quite different from those employed in REG experiments, which attempt to measure psychokinesis. Other types of remote viewing and PK experiments have still other setups and protocols. Within each category, moreover, different researchers use different setups, equipment, and methodologies from each other and over time. This is true, of course, in any field of research; but the critical scrutiny applied to parapsychology militates against the customary flexibility.

The current movement in the direction of greater uniformity and cleaner methodologies is thus, in large part, an attempt to strengthen the case for parapsychology in defense against its critics. In their joint communique, Hyman and Honorton made several recommendations for the design and conduct of future psi ganzfeld experiments. In general, they said, such experiments should strive for:

- elimination of possible means of sensory leakage;
- better procedures and documentation for the randomization of targets;
- more thorough documentation of judging and feedback procedures;
- avoidance of arbitrary multiple analyses;
- avoidance of selective reporting of statistically significant results;
- correct use of valid statistical tests and methods; and
- painstakingly thorough documentation of experimental procedures and other information on experimenters and subjects as well as certain raw data.

Detailed suggestions were made in each of these areas.[19] The authors noted that many of the recommendations would apply equally well to other types of parapsychology research besides ganzfeld.

During the workshop, Robert Jahn described several criteria or "tactics" that he and his colleagues had found to be "useful in helping us to understand the phenomena that we are examining." While some were general suggestions for improving the standing of the field as a whole, others were specific recommendations for strengthening parapsychology experiments, particularly in the PK area. The latter are:

- employ impeccable experimental equipment and technique, using the highest possible precision;
- work with the largest possible data base (to separate the marginal effects from background noise);
- concentrate on operator-specific, statistical replicability;

18 Ganzfeld, or "whole field," ESP experiments place the receiver in an environment of reduced external stimuli for attempted receipt of a visual image "transmitted" telepathically by a sender located remotely.

19 Hyman, R. & C. Honorton, op. cit., pp. 355-361.

- o use a "tripolar protocol," i.e., collect data under three interspersed conditions (e.g., intention to raise the number of events, intention to lower the number of events, and baseline); and
- o within the experimental program, maintain equal respect for the aesthetic/impressionistic and rigorous/analytical aspects of the research on this unusual topic.

One additional point that was made repeatedly in the workshop regarded the necessity of distinguishing between exploratory and confirmatory (proof-oriented) experiments. The former are generally shorter, more broadly defined, and not reported. There was general agreement that the explorational objectives should be clear and the temptation to report positive results as "proof" of a hypothesis should be resisted. Instead, exploratory studies should be used to suggest and develop hypotheses for later proof-oriented experiments which should be highly defined and constrained. Exploratory studies can also be used to develop and refine experimental procedures.

III. ANALYSIS OF DATA

As was noted earlier, the scientific evidence for psi consists chiefly of statistically significant deviations from a theoretical chance expectation or experiential baseline. Consequently, the validity of statistical analyses employed is crucial to the strength and acceptability of parapsychological claims.

According to Jessica Utts, a statistician participating in the OTA workshop, parapsychologists in general tend to be more careful with statistical methodology than researchers in other fields. They are generally more responsive to criticism on this score than other researchers are, she said. There is general understanding among them, Utts said, that human responses cannot be assumed to be statistically random. It is also well understood that selection of targets with replacement is preferable because it allows independent trials. She also referred to the formal policy of parapsychology for reporting negative results. However, it was noted, not all statistical analysis by parapsychologists (or by their critics) is of high caliber.

Much of the discussion centered around a single broad topic: how to inject the maximum robustness and power into the empirical data base. It was recommended that parapsychologists should confine themselves to methodologies that yield such robust data that the outcome does not depend largely on subtle statistical points. There was general agreement on the various suggestions that were made to this end.

For example, there was considerable discussion of the traditional notion that a "successful" experiment has a significance level (p value) of .05 or less (meaning that there is no more than a 5 percent chance that the results are due to chance). Partly as a result of this concept, Utts noted, nonsignificant results (that is, where $p > .05$) are too often interpreted erroneously by critics as insignificant results. An example was given to demonstrate that sample size (n) has a substantial impact on the statistical outcome; if n is small there is very little chance for a "successful" outcome, even if relatively high actual probabilities of success based on ESP are assumed. Meta-analysis of all the forced-choice precognition experiments conducted from the 1930s on showed that

those with significant results had, on average, two and one-half times larger sample sizes than those with nonsignificant results. Thus, large values of n impart considerably greater "power" to the analysis. Power is, essentially, the probability that the experiment is going to succeed. (This concept of power is now being taught at the graduate level in statistics.) Table 2 illustrates how power increases with increasing n.

At least one of the critics present argued that there is no correlation between sample size and the probability of obtaining a significant outcome, but Utts refuted that claim. It was agreed that the spread of data -- for example, one or two "outliers" (data points far outside the general cluster) -- can greatly weaken the correlation.

The statistician recommended that parapsychologists stop focusing on the arbitrary, "sacrosanct" $p < .05$ and begin focusing, instead, on more meaningful measures such as estimating the magnitude of the observed effect, looking for consistency, and studying patterns across many studies. Meta-analysis is a way to obtain much larger sample sizes and "hits." In an analysis of ganzfeld psi studies with hit probability of .25 but which were nonsignificant (and thus "failures"), combining the studies gave a p of .02 ("success"). This recommendation met with general agreement.

TABLE 2

"POWER" AS A FUNCTION OF SAMPLE SIZE
FOR DIRECT HIT GANZFELD ANALYSIS
WITH CHANCE p=.25

	Power	
True p	n=10	n=100
.30	.047	.29
.33	.073	.54
.40	.166	.94
.50	.377	.9998
	(= .02)	(.04)

It is well established that some individuals achieve greater effects than others in psi experiments. Figure 1 shows the distribution of results across 33 operators (subjects) in a series of REG runs.[20] Given this correlation of individual operator identities and characteristics with some of the alleged psi phenomena, it is also important to understand how individual operator effects are related to collective effects. The point was made that, in seeking larger data bases and greater (collective) significance levels, researchers should not lose sight of individual effects by combining many operators. Scientifically, the individual and group effects pose two distinct sets of questions.

Effect size is another crucial issue in parapsychology. From the standpoint of practical significance, the effect appears to be quite small. In the largest REG data base of over 750,000 trials (150 million bits), compiled at the Princeton Engineering Anomalies Research Laboratory over the past decade, the overall means were shifted in the intended direction by only 0.035%. Figure 2 plots graphically the total cumulative deviations from baseline of all 33 operators in the program (accounting for 250,000 trials in each direction).[21] It can be seen that, while the deviation is well beyond the .05 significance level, it is not an enormous effect -- although it is said to be beyond the tolerance level of many contemporary microelectronics processors, communications discriminators, and dedicated computational arrays.

Nevertheless, parapsychologists point out that psi effects, if indeed that is what they are, could be amplified by information theory techniques such as "majority vote counting." It is said that the effect size in the best ganzfeld psi research is approximately equal to that found in the recent studies of aspirin in heart attack prevention. And small effect sizes in physical phenomena (e.g., the bending of light by gravity) have been very important in the evolution of human knowledge in the past. Indeed, the evolution of life itself depends on genetic mutations which occur at very low rates. In other words, significance (in the "real world" sense) and effect size are two separate considerations. It is not the magnitude of an event, but its centrality and relevance to established beliefs or processes, that determines its importance.

Effect size means different things in the context of different types of experiments. The average effect sizes found in psi ganzfeld experiments (about .17 in one series of experiments described at the workshop) are several times larger than those found in the REG work. The anomalies being displayed (if indeed they are true anomalies) may well be of fundamentally different types. The primary value of using effect sizes appears to be in compensating for differences in sample size.

The key to satisfying proponents and critics of parapsychology may, in the end, be replicability. If a "recipe" for replication of successful parapsychology experiments could be developed, it would render moot many of the issues of statistical analysis. However, failure to replicate does not disprove the existence of psi; and the criteria for replicability can be difficult to

20 Dunne, B.J., et al. Individual operator contributions in large data base anomalies experiments (Technical Note PEAR 88002). Research Report. Princeton Engineering Anomalies Research, July 1988.

21 Jahn, R.G. & B.J. Dunne, op. cit., p. 110.

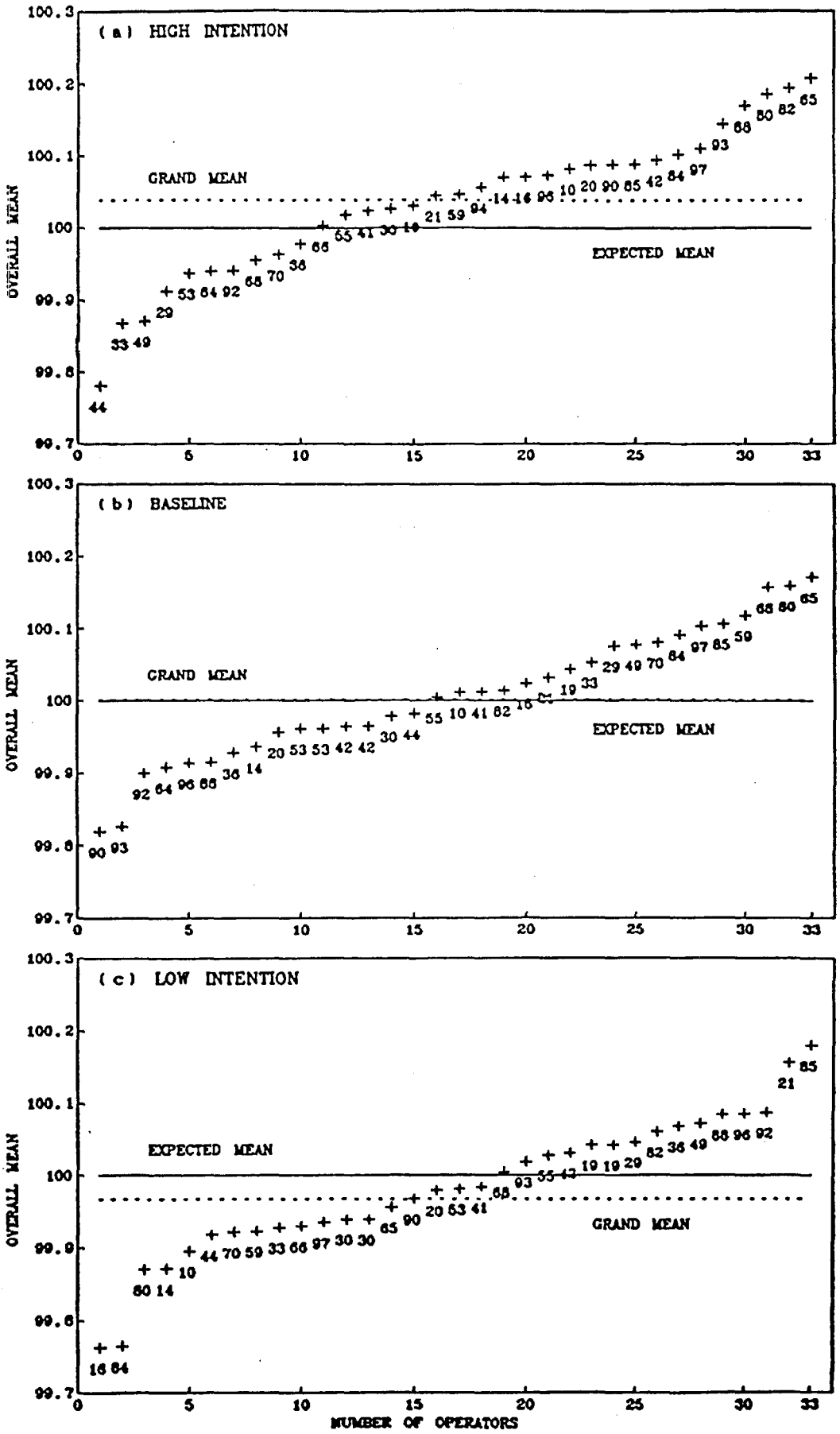

FIGURE 1: 200-SAMPLE REG. ASCENDING OPERATOR MEANS

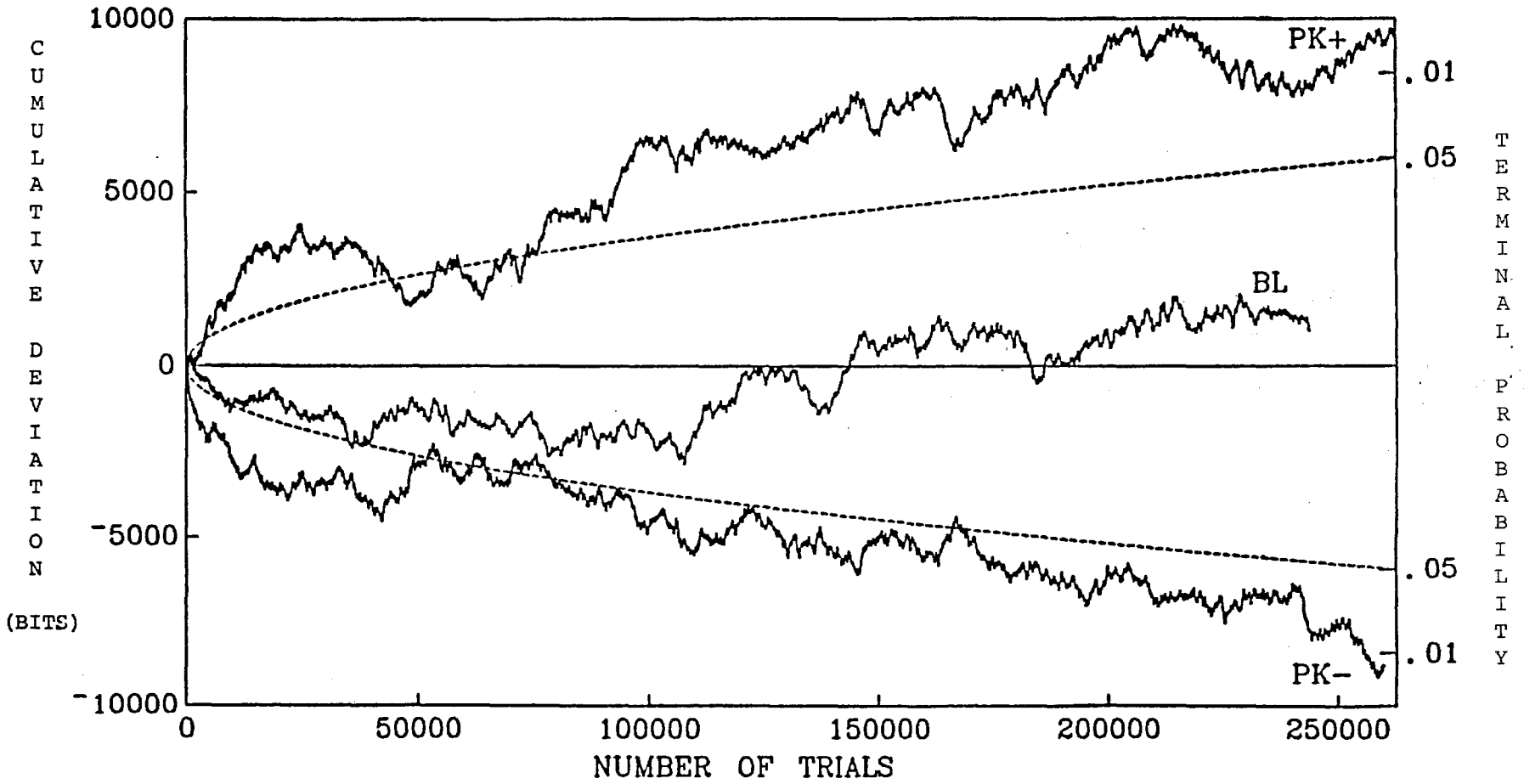

FIGURE 2: CUMULATIVE REG RESULTS, ALL OPERATORS (33), ALL SERIES (87)

establish. What is needed is consensus among the various camps on the ground rules for replicability. For example, individuals must rigorously apply the same criteria to themselves that they do to others, and apply to others the same criteria that they themselves use elsewhere.

One approach that might move parapsychology out of the quagmire of statistical significance levels could be to control for desired conditions -- i.e., for large effect size -- stripping away in advance the "noise" injected by those subjects with low capability. It may be that replicability will hinge on getting large effect sizes. But even replicability in operators with large effect sizes is often problematical across time and changing experimental conditions.

In the end, some parapsychologists believe, it may be that replicability will follow from the development of theory in parapsychology. In this view, the priority should be to pursue process-oriented research that can elucidate the mechanisms controlling these anomalies. It will then be possible to design experiments and conditions to produce the desired effects, rather than relying on experiments designed to produce statistical significance and satisfy a critical audience. (Critics, if they are given the last word, retort that parapsychologists should not bother presenting their case to the scientific establishment until they have succeeded in developing an adequate theory.)

IV. THE MEANING OF EXPERIMENTAL RESULTS

ASSESSMENTS OF MEANING AND MEANINGFULNESS

Leading off the "Findings and Conclusions" on parapsychology in the recent NRC report was a provocative assertion:
"The Committee finds no scientific justification from research conducted over a period of 130 years for the existence of parapsychological phenomena."[22]
Although the treatment of parapsychology in the body of the report was generally more moderate in tone, this finding was taken by the media and many readers to be the "bottom line" of the report with respect to parapsychology. It produced considerable distress and resentment among parapsychologists and others, who felt that the field had been dealt an unwarranted blow.[23] [24]

However, the statement did clearly delineate the view of many of those who monitor the field. Some critics are willing to concede that parapsychology research has demonstrated definite anomalies, but they believe that these reflect statistical or experimental artifacts, not psi. In their view, the

22 National Research Council, op. cit., p. 22.
23 See, for example: Palmer, J.A., C. Honorton, & J. Utts, op. cit.
24 This statement was based on a previously published assertion by Ray Hyman, chairman of the NRC subcommittee which authored the report (Proc. IEEE, June 1986, 74(6): 823-849). At the workshop, Hyman said that the precise wording of the NRC's statement was not his. Discussion ensued over the semantic differences between Hyman's statement and the NRC report statement. Hyman said that he would have phrased the statement differently. He later referred to any differences between the wording he would have used and the wording in the report as a matter of style, not substance (personal communication with Dr. Alan Shaw, 28 December 1988).

anomalies are just anomalies -- mere "puzzles" -- and nothing more. The more conservative parapsychologists tend to agree that nothing more than anomalies have been demonstrated; but they point out that the anomalies are structured statistically in very interesting ways, suggesting that something more significant is at work. Critics, in turn, counter that if a "successful" experiment has been cleanly run and the results are well above chance, then most scientists would agree that something truly puzzling has been manifested. That is why methodology and replicability are so important for parapsychology experiments. But, they conclude, it is not a puzzle that has high priority when viewed against the other questions facing scientists. They assume that the explanation for the puzzle, when found, will not be very interesting after all.

One reason (of several) for the critics' skepticism is the "experimenter effect," which is seen in the fact that certain investigators obtain consistently positive experimental results while others get only chance results. So pervasive is this phenomenon that it has been called "the only confirmed result of parapsychology."[25] The experimenter effect lends force to criticisms of parapsychology, seeming as it does almost to imply fraud -- or at least a lack of objectivity. Parapsychologists point out that the identity of the experimenter is also positively correlated with the placebo effect in clinical medicine. It may be, they suggest, a function of belief in the phenomenon under investigation. According to this view, some of those who believe in psi are somehow able to elicit the phenomenon in their subjects. But, if true, this presents a problem for the experiments, since the investigator can influence the experiment subjectively and perhaps involuntarily. In other words, it might demonstrate the existence of psi while making it even more difficult to prove scientifically.

Stringent requirements for proof are placed on parapsychology. The editor of Science has stated that "extraordinary claims require extraordinary proof."[26] Some parapsychologists object that the burdens of proof in this field are unique, and unfairly so. In their view, the critics of the field have an arbitrarily high threshold for what constitutes scientifically acceptable evidence, and everything falling below that threshold is dismissed as insignificant. Indeed, the NRC committee recognized only two successful experiments over 15 years in all of parapsychology -- one in remote viewing[27] and one using the random number generator[28] -- as meeting "most of the minimal

---

25 Parker, A. A pilot study of the influence of experimenter expectancy on ESP scores. In J.D. Morris et al. (Eds.) Research in Parapsychology 1974. Metuchen, NJ: Scarecrow Press, 1975.

26 Abelson, P.H. "A stepchild of science starts to win friends." U.S. News & World Report, 31 July 1978.

27 Schlitz, M. & J.M. Haight. Remote viewing revisited: An intrasubject replication. Journal of Parapsychology, 1984, 48: 39-49.

28 May, E.C., B.S. Humphrey, & G.S. Hubbard. Electronic System Perturbation Techniques (Final Report). Menlo Park, CA: SRI International, 1980.

criteria of scientific acceptability."[29] [30] In this connection, parapsychologists point to the unreasonableness of requiring extraordinary proofs and, when such proofs are provided, dismissing the results as a "mere puzzle."

Critics, for their part, reply that the stringent requirements are warranted because the field itself is unique. One workshop participant listed several ways in which, in his view, parapsychology is different from other fields:

- o It lacks boundary conditions specifying when psi phenomena should and should not be expected to be present.
- o Only the experimenter effect has been definitively established.
- o There is no easily replicable paradigm for experiments.
- o Debates are about primary phenomena (i.e., does psi exist or not) rather than about its "lawfulness."
- o The concept of demonstrating anomalies is different -- in parapsychology the "successful" results are not strikingly different from the baseline model.

In general, then, from the perspective of critics the criticisms of the field and its results can be summarized as follows: (1) there has been no "conclusive" experiment; (2) there is no truly repeatable parapsychology experiment; (3) the "significant" results are isolated and disconnected observations of no obvious scientific interest; (4) the results are nonsensical in that they do not suggest any lawful relationships or progressive research programs; and (5) even if real, psi is too weak to be of any practical importance.[31]

Each of these points is countered by parapsychologists. Nevertheless, given that harsh assessment by the critics, it is relevant to ask them what would comprise acceptable proof of psi.

WHAT WOULD CONSTITUTE PROOF OF PSI?

Parapsychologists often maintain that the critics have been unwilling or unable to specify detailed criteria that would lead them to accept the psi hypothesis. In their view, this enables the critics to continually shift their ground in generating criticism -- in effect, presenting a "moving target" that moves faster than parapsychologists can adjust their research protocols and aims.

---

29 National Research Council, op. cit., pp. 184, 190.

30 Some critics point to a third study as another example of a well-run, tightly controlled experiment giving significant results that are difficult to explain. (Schmidt, H., R.L. Morris, & L. Rudolph. Channeling evidence for a psychokinetic effect to independent observers. Journal of Parapsychology, 1986, 50: 1-6.) This study was favorably reviewed in a background paper prepared for the NRC committee, but was omitted from the report's conclusions. Parapsychologists claim that there are many other more recent studies, not reviewed by the committee, which would also satisfy any reasonable standard of acceptability.

31 Rao, R.K. & J. Palmer, op. cit., p. 541

Based on the list of differences above, one workshop participant presented a list of criteria that, in his view, would need to be satisfied before "proof" of psi could be achieved. These included both local criteria (i.e., those focused on the particular experiment) and global criteria (those that cut across experiments). Local criteria have to do with statistical significance and power, and with the experimental methodology and controls. In this critic's view, if a proof relies entirely on statistics, it is probably not science. (It was noted, however, that a full application of this position would eliminate several well-established fields of science, such as particle physics.)

Global criteria are those which, taken together, present an overall pattern that the scientific community is forced to recognize and accept. They include replicability, lawfulness, consistency of effects, and cumulativeness across time. The latter idea -- that parapsychology represents a series of fits and starts over the long term, with each generation of researchers beginning anew -- generated considerable discussion. A number of the parapsychologists described how their work is built on the foundation of work that came before. For example, Charles Honorton's ganzfeld work evolved out of earlier research on ESP and altered states of consciousness such as dreaming and hypnosis. Robert Jahn's research is built on the earlier work of Schmidt (with random number generators) and Puthoff (in remote viewing).

CONFLICTING ORIENTATIONS

A major challenge that parapsychologists face is the insistence by many of their critics (particularly the psychologists among them) that all "normal" explanations -- those based on known science, such as neurological or psychological mechanisms, or those based on the charge of poor methodology -- for experimental results must first be considered and discarded before psi, the paranormal effect, can be entertained. In this view, psi is the residuum when all other possible explanations have been eliminated.

This issue goes to the heart of the problem that skeptical scientists have with parapsychology, which is that, in their view, parapsychologists begin with a hypothesis in which they believe (viz., that there is interaction between consciousness and the physical world) and then run experiments until they get results that tend to confirm the hypothesis and the belief. Critics contend that, since the existence of psi has not been demonstrated, and since, indeed, there is not even an adequate theory that could support its existence, the presumption that psi explains any anomalous result cannot be supported.

Yet, the parapsychologists insist, this tendency to believe in a certain explanation for the effect they are demonstrating through experiment does not disqualify the experiment. Indeed, such belief is quite common among scientists in every field. It is the interpretation of results that is open, they say, and it is not their job (nor are they qualified) to eliminate all possible physiological and other explanations outside their field of expertise. They have simply demonstrated an anomaly in which science ought to be interested.

NEED FOR A THEORY OF PSI

Although many theoretical models of psi have been proposed,[32] both groups agree that a fundamental problem in psi is the lack of an adequate theory of paranormality to test by experiment. It leaves parapsychologists in the unscientific position, their critics contend, of trying to demonstrate a phenomenon that they have no real basis for seeking. Conversely (or perversely, in the view of parapsychologists), they are also criticized for premature theorizing without having first demonstrated that a phenomenon exists.

It is here, the participants agreed, that more process-oriented research is needed to better define the nature of the phenomenon -- when it is manifested, under what conditions, in individuals of what characteristics, etc.. With a positive theory of psi in hand, parapsychologists would escape the present demand that they eliminate every other plausible explanation for their results until only the residuum, psi, remains.

V. RELATION OF PARAPSYCHOLOGY TO THE BROADER WORLD OF SCIENCE

IS PARAPSYCHOLOGY A SCIENCE?

Disagreements about the meaning of experimental results in parapsychology and its aims, and the general agreement that a theory of psi is sorely needed, all converge on the basic question of whether or not parapsychology can be considered a science. The discussion in the previous section is neatly encapsulated in a single sentence from a recent critical article:[33] "Although many parapsychological research projects have been carried out under what have been described as well-controlled conditions, this does not by itself make a science, for unless and until it can be demonstrated that paranormal phenomena really exist, there is no subject matter around which a science can develop."

As was noted earlier, parapsychologists consider this view a form of Catch-22, since conditions, they believe, are preset so that no such demonstration will be acknowledged. But the underlying implication of the assertion -- that science cannot study nothing -- also draws objections from parapsychologists. In their view, one of the purposes of science is to investigate whether there is nothing (or something) present in an empirical situation. Whether the motivation for the investigation comes from prior observation or subjective expectation is irrelevant, they maintain. They cite sociology, economics, and even astronomy as analogous fields in which "experimentation" consists essentially of observation and measurement guided by an imperfect or non-existent theory. While it may be argued that each of these fields, unlike parapsychology, has real and obvious subject matter, parapsychologists cite human experience and the voluminous anecdotal record of paranormal occurrences. They point out that gravity, like psi phenomena, has long been observed without benefit of an accepted scientific explanation.

---

32 See, for example: Jahn, R.G. & B.J. Dunne. On the quantum mechanics of consciousness with application to anomalous phenomena. Foundations of Physics, August 1986, 16: 721-772 and the bibliographies of works listed as Notes #1 and #2 above.

33 Alcock, J.E., op. cit., p. 553.

At bottom, the question of whether parapsychology is a science runs up against philosophical questions that challenge the prevailing world-view of contemporary science. To quote again from the article cited above:[34] "The dispute about psi reflects the clash of two fundamentally different views of reality. The first of these is the materialistic, monistic view that the human mind is some sort of emergent manifestation of brain processes, whereas the second is the dualistic position that maintains that the human mind/personality is something beyond the stuff of atoms and molecules." The issue of psi thus seems, to many, to transcend ordinary scientific controversy and to take on overtones of philosophy and even religion. This fact seems to account, in part, for the unusually strong resistance the field has met with from the scientific establishment.

RESISTANCE BY "ESTABLISHMENT" SCIENCE

This report has made repeated reference to the difficulty that parapsychology has encountered, throughout its long history, in gaining acceptance within the scientific community. This resistance by the scientific establishment appears to have taken three forms: (1) the loosely defined but stringent criteria for proof imposed on parapsychology experiments; (2) a nearly total rejection by mainstream scientific journals of articles or research reports supporting parapsychology; and (3) the more derisive and more public form of criticism practiced by some members of the CSICOP.

One workshop participant documented a long series of exchanges with the management of the journal Science and its parent organization, the American Association for the Advancement of Science, over the issue of publication. No report of research conducted by a parapsychologist has ever been published in Science, although criticisms of parapsychological research have been published from time to time. According to this participant, parapsychology research reports submitted to Science have been reviewed by reviewers from outside the field and rejected on often arbitrary and irrelevant grounds. Positive reviews by parapsychologists are evidently discounted on the basis that the reviewers are biased by their belief in psi. Repeated efforts to bring this situation to the attention of Science and AAAS management apparently failed to gain a sympathetic hearing. This individual recommends (1) that editors and grant-givers be required to define publicly the qualifications of their reviewers for any given paper or proposal and (2) that reviewers be required to sign their evaluations.

Also heard during the workshop was the suggestion that there is a hierarchy of sciences, in which psychology and the other social sciences rank below the physical sciences and, indeed, from the most conservative perspective are still only barely established as being legitimate. According to this view, psychologists (generally the most active in criticism of parapsychology) do not wish to be tarnished by association with a field that resembles psychology but which they consider unscientific and "flaky." They are therefore anxious to act as the "gatekeepers" of science with respect to parapsychology, partly in order to reinforce their own sense of legitimacy and acceptability. Those who hold this rather harsh view suggest that there may also be an underlying motivation to suppress potential competitors for the same limited pool of research funds.

34 Alcock, J.E., op. cit., p. 565.

Thus, parapsychologists claim that the issue is not the validity of psi research, but the integrity of the scientific process as it deals with controversial findings. They believe that the scientific gate-keepers should be willing to trust the normal scientific process to arrive at scientific truth, and not rely on reviewers who are not scientifically qualified in the discipline being reviewed. They note that the idea of disqualifying an entire discipline from reviewing work in its field is unprecedented in modern science.

CHOICES FOR THE FUTURE

Regardless of the reasons, it is clear that parapsychology continues to face strong resistance from the scientific establishment. The question is, How can the field improve its chances of obtaining a fairer hearing across a broader spectrum of the scientific community, so that emotionality does not impede objective assessment of experimental results? Whether the final result of such an assessment is positive, negative, or something in between, the field appears to merit such consideration.

From the standpoint of the experiments themselves, some of the answers to that question have already been outlined in Section IV (i.e., replicability, better controls, more process-oriented research, etc.). On a higher level, the need for an acceptable theory of psi (one accepted by a large number of parapsychologists) has been emphasized repeatedly. It would probably improve the outlook for the field if such a theory were more hospitable to current models of mechanisms in physics and psychology than the idea that the mind can operate outside the realm of known physical laws -- although, indeed, some aspects of psi suggest recent advances in quantum theory such as nonlocality.

On an individual level, it was suggested that researchers in parapsychology must begin to make certain basic choices. Workshop participants referred several times to the fact that parapsychologists began in the 19th century by investigating something experiential and "real" -- instances of paranormal phenomena. Gradually, they adapted their experimental approaches to the methodologies of conventional science. But in doing so, and in attempting to demonstrate the phenomena scientifically and statistically, they have lost touch with the powerful experiential base that gave the field its impetus. By "looking under the lamppost" using a rigorous scientific methodology, it was suggested, parapsychologists may be ignoring more promising avenues of investigation. Thus, each experimenter must decide whether to pursue research designed primarily to satisfy the critics of the field, or to follow his or her instincts as to where the most fruitful and powerful findings might be made. It was suggested that one way parapsychology might be released from the constrained situation in which it currently operates would be to demonstrate practical applications of psi. (Some studies in this direction have already been attempted.[35]) This would required working toward a better understanding of the conditions and characteristics under which different kinds of psi operate, finding or cultivating individuals with unusually strong capabilities, and mounting controlled demonstrations of the uses of ESP and PK. If such a strategy were to be successful, public opinion and the interest of potential users -- rather than the strictures of the scientific establishment -- would begin driving progress in the field.

---

35 See, for example, the discussion of "Potential Applications" in Palmer, J.A., C. Honorton, & J. Utts, op. cit., p. 17.

Another, less risky strategy or improving the status and acceptance of parapsychology is to encourage a constructive dialogue with the critics of the field. Indeed, this process has already begun with the published exchanges between leading proponents and critics, the joint communique of Hyman and Honorton, and the OTA workshop itself. A more concrete step was suggested in the joint communique and elaborated upon in the workshop, where it received enthusiastic support.[36] This was to appoint a committee of parapsychologists and critics who would jointly evaluate the ongoing work in parapsychology, establish guidelines for acceptable experimental protocols and, possibly, even design and conduct an independent experiment. Although there would be obvious difficulties to be overcome, they would pose small barriers compared to those the field now faces. This would certainly represent a large step toward evaluating parapsychology and possibly bringing the field further inside the edifice of established science.

---

36 Hyman, R. & C. Honorton, op. cit., p. 363.