Journal of Parapsychology, Vol. 57, June 1993

RHETORIC OVER SUBSTANCE: THE IMPOVERISHED STATE OF SKEPTICISM

BY CHARLES HONORTON

Few of us can afford to take time to familiarize ourselves with the detailed and often technical arguments underlying new knowledge claims that would enable us to evaluate properly the merits of such claims for ourselves. Most of the time we have to rely on "experts" to do this for us. We are poorly served when only one side of a controversy is presented and benefit most when all perspectives are vigorously debated by knowledgeable protagonists. The Comitato Italiano per il Controllo delle Aftermazioni sul Paranormale has provided an invaluable service by presenting a balanced forum for discussing the status of parapsychology by six leading researchers and critics. CICAP's initiative in this regard is probably unique and one that its American counterpart, the Committee for the Scientific Investigation of Claims of the Paranormal, would do well to emulate. To fill out this innovative format, CICAP asked a skeptical parapsychologist, Susan Blackmore, to critique the contributions of the parapsychologists, and I have been asked to comment on the critics' contributions.

1. *What the Critics No Longer Claim*

Before examining the current arguments made by Hyman, Alcock, and Randi, it is important to understand what they are *not now* claiming but *have* claimed in the past. First, they no longer claim that the results of the major lines of experimental psi research are consistent with the null hypothesis (mere chance fluctuation). They now concede that at least some parapsychological effects are, to use Hyman's words, "astronomically significant." This concession is important because it shifts the focus of the debate from the *existence* of effects to their *interpretation*. Second, they no longer claim to have

This article originally appeared in *Scienza & Paranormale,* Vol. 1, No. 3, June 1993, and is reprinted here with permission of the publishers.

demonstrated a relationship between methodological flaws and study outcomes. These concessions, which are documented in Section 3, did not come quickly or easily and the critics are obviously not eager to advertise them. Over the past decade Hyman and other critics tried very hard to show that psi effects are either not really significant or that their significance is systematically related to the presence of flaws in the experiments. Having failed on both counts, the critics now face a serious dilemma: they have been forced to admit parapsychology has demonstrated anomalous effects that need to be explained and they have run out of plausible conventional explanations.

2. *What the Critics Now Claim*

A Century of Failure?

Instead, they offer a caricature of the history of parapsychology and present polemical arguments designed to convince us that there is really nothing in parapsychology that warrants scientific interest, except, perhaps, for the motivations of those who persist in studying it. Hyman's use of absolutist language to characterize parapsychologists' data claims seems designed to turn off scientists. Unlike the formal sciences such as mathematics, empirical science does not deal with "irrefutable proof" or "foolproof evidence." Empirical evidence is always a matter of degree and remains subject to later reinterpretation. It is in this sense that science represents a unique self-correcting approach to knowledge. Scientific truth always carries the caveat, "until further notice."

At the core of the critics' current arguments is the rhetorical claim that 100 years of research has failed to provide convincing evidence for parapsychological phenomena. When parapsychologists have not been given an opportunity to respond, they have claimed that *130* years of research has produced *no* evidence for psi (e.g., Druckman & Swets, 1988). An English critic, who was recently appointed to a four-year £100,000 psychical research fellowship at Darwin College, Cambridge, to write a book about why people believe impossible things, has been quoted in *The New Scientists* as saying that after *150* years of psychical research "there is no evidence at all of there being any phenomena" (Bown, 1992). Such statements are themselves extraordinary claims inasmuch as psychical research did not exist until 1882 and systematic laboratory research using quantitative methods did not begin until the early 1930s. Throughout its history, research in parapsychology has been sustained through extremely meager resources. Utrecht University psychologist Sybo Schouten (in press) compared funding patterns in parapsychology with those of American psychology; he found that the total human and financial resources devoted to parapsychology since 1882 might, at best, equal the expenditures for *two months* of conventional psychological research in the United States in the year 1983!

Is psychology a "failed" science? If we were to apply the "century of failure" arguments of Hyman and Alcock to academic psychology, we might well conclude that psychology has failed in its mission: after a hundred years of relatively well-funded research, vigorous controversies continue over such basic phenomena as memory, learning, and perception. The simple act of human facial recognition, for example, remains a mystery and is currently a hot research topic in cognitive psychology. And while it is widely assumed that consciousness is a by-product of brain activity, neither psychology nor physiology has produced, over the past 100 years, even an intelligible model of how biochemical processes could be transformed into conscious experience. Are psychology and physiology failed sciences? Of course not. The most successful sciences such as physics deal with relatively simple and invariant processes: electrons, for example, are interchangeable; they do not have individual personalities, intentions, emotional states, or motivations. The behavioral sciences must contend with extremely complex and variable biological systems that possess these and many other individual attributes. Nevertheless, these sciences have produced many achievements, and so has parapsychology, even though it has been forced to exist on the outskirts of established science with marginal resources. The papers by Broughton, Krippner, and Morris summarize some of parapsychology's accomplishments.

The lack of research by critics and its consequences. There is, however, one important difference between the psi controversy and more conventional scientific disputes. Controversies in science normally occur between groups of *researchers* who formulate hypotheses, develop research methods, and collect empirical data to test their hypotheses. When disputes arise over the interpretation of experimental findings, or when critics suspect that the findings were caused by artifacts, they design new experiments to test alternative explanations or the impact of suspected artifacts. It is through this process that scientific controversies are resolved. In contrast, the psi controversy is largely characterized by disputes between a group of researchers, the parapsychologists, and a group of critics who do not do experimental research to test psi claims or the viability of their counterhypotheses. Psi critics argue the plausibility of various alternative hypotheses (or the implausibility of the psi hypothesis) but they rarely feel obliged to test them. This has been especially true of the current generation of psi critics, most of whom have made no original research contributions. Exceptions like Susan Blackmore and David Marks prove the rule. The lack of research by critics may surprise you, especially if your primary source of information about parapsychology has come from the skeptical literature where you may have encountered statements such as the following by the well-known American skeptic Martin Gardner (1983).

> How can the public know that for fifty years skeptical psychologists have been trying their best to replicate classic psi experiments, and with notable unsuccess? It is this fact more than any other that has led to parapsychology's perpetual stagnation. Positive evidence keeps coming from a tiny group of enthusiasts, while *negative evidence keeps coming from a much larger group of skeptics.* (p. 60, my emphasis)

Gardner does not attempt to document this assertion, nor could he. It is pure fiction. Look for the skeptics' experiments and see what you find. (To his credit, Gardner did get one thing right: half a century is a more accurate time-frame than 100, 130, or 150 years.) The lack of research by critics serves to perpetuate the psi controversy by enabling them to shift continually from one line of criticism to another as each is successively answered through new research conducted by parapsychologists. It is clear from their statements that Hyman, Alcock, and Randi expect the controversy to extend into the indefinite future. Whatever time-frame one chooses to adopt, I think we can all agree on two points: the psi controversy has gone on for a long time, and its lack of resolution represents a very unsatisfactory state of affairs.

Lack of Cumulativeness?

How can we reconcile the "century of failure" argument with the critics' admission that there are "astronomically significant" effects and their failure to demonstrate even plausible alternative explanations for those effects? The answer, they say, is that parapsychology lacks "cumulativeness." "Every science, except parapsychology," Hyman says, "builds upon its previous data. The data base continually expands with each new generation but the original investigations are still included. In parapsychology, the data base expands very little because previous experiments are continually discarded and new ones take their place." The "astronomically significant" effects for which they have no plausible alternative explanations are, Hyman says, based upon "retrospective" meta-analyses of many similar experiments. Truly skeptical readers should be alarmed by the logical contradiction in this argument: if parapsychology is "noncumulative," and if each new generation of parapsychologists discards the findings of earlier generations, how could there be "astronomically significant" effects in meta-analyses that are, by definition, the cumulation of findings from many earlier studies? Hyman refers only to meta-analyses of two relatively recent research areas, the ganzfeld and random number generator experiments (Honorton, 1985; Hyman, 1985; Radin & Nelson, 1989). He overlooks other meta-analyses, such as those discussed by Broughton and Morris involving precognition experiments (Honorton & Ferrari, 1989) and psychokinesis research with dice (Radin & Ferrari, 1991), both of which involve the cumulation of research findings going back to the 1930s. In Section 3, I present a detailed example of a line of parapsychological research that has systematically built upon earlier research. There are other inconsistencies in Hyman's historical analysis that are also self-documenting:

In 1940 J. B. Rhine and his colleagues published a book entitled *Extra-Sensory Perception After Sixty Years* which summarized all quantitative ESP studies since the founding of the Society for Psychical Research in 1882 (Pratt, Rhine, Smith, Stuart, & Greenwood, 1940/1966). Known within the field as *ESP-60,* this book is the central classic of experimental parapsychology. How can we reconcile *ESP-60* with Hyman's claim that each successive generation of parapsychologists claims evidence for psi "*without any reference to the data used by the preceding generation*" (Hyman's emphasis)?

"By the 1940s," Hyman claims, "even parapsychologists admitted that Rhine's experiments possessed too many flaws to qualify as foolproof evidence for psi." How can we reconcile this statement with the fact that as late as 1980 the English critic C. E. M. Hansel was still trying to account for the results of these experiments on the basis of speculative and elaborate fraud scenarios (Hansel, 1966/1980)?

3. Historical Overview of the Psi Controversy

I will now summarize an alternative view of the history of the psi controversy that suggests a very different conclusion, namely, that over the past 60 years of active experimental research by parapsychologists, critics have consistently failed to demonstrate plausible alternative explanations of psi effects. In examining the psi controversy, it is useful to note the order in which various types of criticism have occurred. During each major phase of the controversy, the criticisms have followed this pattern:

- *Statistical criticisms seeking to demonstrate that the claimed effects are not really significant.* This type of criticism has usually been championed by psychologists and refuted by statisticians. If critics could sustain their case at this point, the controversy would end here.
- *Methodological criticisms asserting that the effects are caused by procedural flaws.* As I have already stated, advocates of flaw hypotheses have seldom subjected their flaw hypotheses to empirical test, but have tended instead to argue for their plausibility. In response, parapsychologists have conducted new experiments that eliminate the suspected flaws.
- *Speculative criticisms based on* a priori *and* ad hominem *arguments.* This form of criticism has usually been founded on the assumption that the existence of psi phenomena is incompatible with fundamental scientific principles, but the proponents of *a priori* arguments have never successfully demonstrated the nature of such incompatibilities.

The ESP Controversy of the 1930s

The first major phase of the psi controversy occurred between 1934 and 1939, and was stimulated by publication of the ESP card-guessing experiments initiated by J. B. Rhine and his colleagues at Duke University (Rhine, 1934/1964). During this period, approximately 60 critical articles appeared, primarily in the American psychological literature. Elsewhere (Honorton, 1975) I have presented a more detailed review of this controversy with references to most of the critical papers. Figure 1 summarizes the major issues raised during this phase of the psi controversy.

In most of the early card-guessing experiments, subjects were asked to guess the order of concealed decks of 25 randomized cards containing five each of five geometrical symbols. Since subjects usually did not receive feedback of the actual target order until after one or more runs of 25 trials, statistical analysis of the card exper-

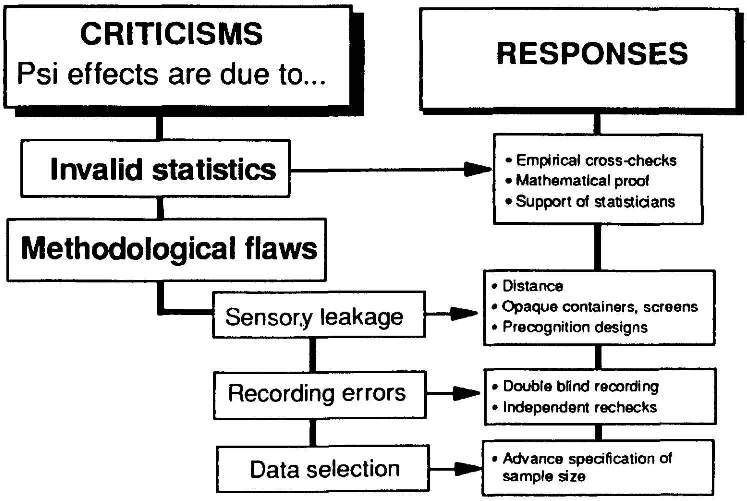

Figure 1. The ESP card-guessing controversy of the 1930s.

iments assumed that the probability of success on each trial was 1/5. The first major criticism of Rhine's work questioned the validity of this assumption. This issue was resolved by mathematical proof and through empirical "cross-checks," a type of control series in which subjects' guesses were deliberately compared with target orders for which they were not intended. For example, the guesses intended for target cards in run 1 were compared with the targets for run 2, and so on. Empirical cross-checks were reported for 24 separate experimental series and while the actual experimental run scores (i.e., guesses for Run 1 compared to targets for Run 1) were highly significant (average: 7.23/25), the control cross-check results were in all cases nonsignificant (average: 5.04/25). (See Pratt et al., 1940/1966.) Other technical statistical issues were raised and eventually abandoned. In a 1938 article, E. V. Huntington asked, "If mathematics has successfully disposed of the hypothesis of chance, what has psychology to say about the hypothesis of ESP?"

By far the most serious methodological criticism of the early card-guessing experiments concerned the possibility of sensory cues. It is clear that some of the early studies reported in Rhine's 1934 monograph did not adequately control against possible sensory leakage. Rhine did not base any major conclusions on these early studies, but their inclusion in his monograph provided a basis for legitimate criticism and sidetracked discussion away from the better controlled studies which were not susceptible to explanation by sensory cues. These later studies used one of four methods to eliminate potential sensory contact between the subjects and target cards: (a) use of targets enclosed in sealed opaque envelopes, (b) use of opaque screens to conceal targets from subjects, (c) separation of subjects and use of targets in different buildings, and (d) use of precognition designs in which the targets were randomly selected only *after* subjects registered their guesses. Between 1934 and 1939, 33 experiments involving nearly one million experimental trials were reported using these methods, and highly significant results were obtained with each method. (For study references, see Honorton, 1975; Pratt et al., 1940/1966).

Another line of criticism suggested that significant ESP results might result from motivated recording errors. This represents one of the few instances in which critics attempted to provide empirical evidence for an alternative explanation. Kennedy and Uphoff (1939) had 28 observers record 11,125 mock ESP trials. While only 1.13 percent were misrecorded overall, both "believers" and "skeptics" systematically erred in the direction of their biases: 71.5 percent of the errors by observers favorable to ESP spuriously increased the ESP scores, and 100 percent of the errors by those unfavorable to ESP decreased the ESP scores. Many years later, Robert Rosenthal (1978) summarized 27 different recording error studies in the behavioral sciences and again found the average error rate to be about 1 percent. An error-rate of this magnitude could not explain the results of the ESP card-guessing experiments, but investigators quickly adopted controls against recording errors. By the end of the 1930s, double-blind data recording and checking had become routine. The results were still "astronomically significant." (See Pratt et al., 1940/1966, Table 9, p. 102.)

The final major issue to arise during this period concerned the possibility of improper data selection. By convention, the criterion of significance for statistical tests is usually set at $p = .05$. When the outcome of a study reaches this criterion it means that the odds are 20:1 against the likelihood that the observed result arose purely by chance. This, of course, does happen. If an investigator conducts 100 experiments, we would on the average expect five to yield spuriously significant results. When chance alone is operating, these pseudosignificant results will be cancelled out by the other experiments. Now consider an extreme case of data selection where the investigator discards the 95 "unsuccessful" experiments and attempts to draw conclusions only from the "successful" ones. This would be highly improper and the investigator's conclusions would be meaningless. As we shall see later, there are various other ways in which data selection problems could compromise research findings. In the 1930s, the issue was addressed by parapsychology researchers through studies which specified the number of trials in advance or explicitly stated that all of the data collected was used in the analysis. (See Pratt et al., 1940/1966, pp. 118–124, for an extensive discussion of data selection issues in the 1930s.)

By the end of the decade, there was general agreement that the various methodological counterhypotheses raised by critics during this period could not explain the outcomes of the more rigorously controlled experiments. (See comments by leading critics of the day in Pratt et al., 1940/1966, chap. 8.) One final point is in order concerning this phase of the psi controversy. It is still widely believed that most of the successful ESP card-guessing experiments came from Rhine and his Duke University group while most of the independent replications were unsuccessful. This is not true. Independent investigators contributed 33 of the 50 studies published during this period, and 61 percent of these studies reported significant ESP effects. Moreover, the difference in success rate between Duke and other investigators was not significant (Honorton, 1975, Table 2).

An Era of Speculative Criticism (1950–1980)

Virtually no new substantive criticisms appeared between 1950 and 1980. This phase of the psi controversy centered instead on two speculative claims. Figure 2 summarizes the issues raised during this period. The first line of speculative attack, championed by Spencer Brown (1953, 1957), was that the card-guessing experiments provided evidence not for ESP, but rather that there were fundamental defects in probability theory. Spencer Brown's arguments, based upon irregularities in early random number tables, were refuted by Scott (1958). This approach never attracted serious support, and it requires little imagination to see why. Much of modern science relies upon probability theory, and acceptance of Spencer Brown's claims would have far greater consequences for science than would ESP. In any case, his arguments do not explain the ESP results. They do not explain the empirical cross-check controls I summarized in the preceding section, and they are incapable of explaining systematic variations in performance such as "sheep/goat" experiments where psi believers consistently score higher than psi skeptics, studies showing correlations between psi performance and personality var-

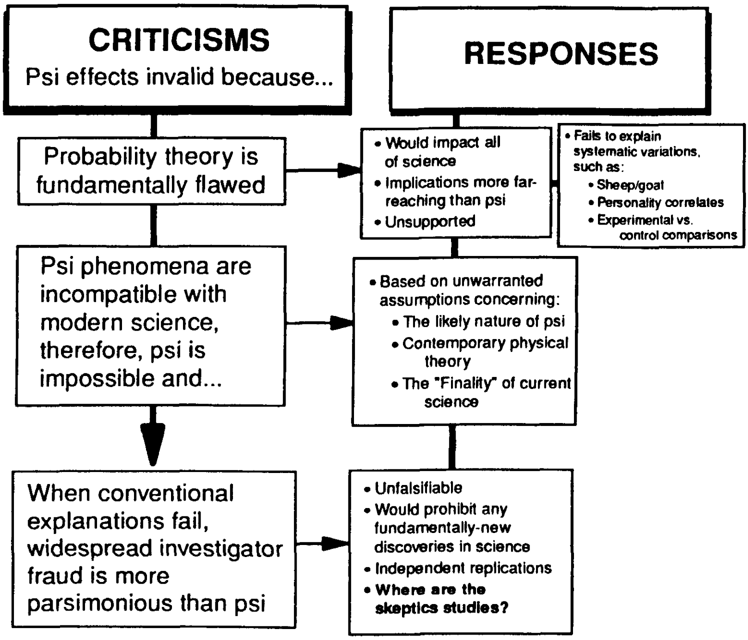

Figure 2. An era of speculative criticism.

iables such as extraversion, or those in which psi performance systematically varies in relation to different experimental conditions as when subjects are instructed to alternate producing high versus low scores.

The second line of attack during this period centers on the hypothesis of widespread investigator fraud. It was most forcefully presented in a lead article in *Science*, entitled "Science and the Supernatural," by Price (1955), who began with the following observations:

> Believers in psychic phenomena...appear to have won a decisive victory and virtually silenced opposition....This victory is the result of an impressive amount of careful experimentation and intelligent argumentation....Against all this evidence, almost the only defense remaining to the skeptical scientist is ignorance, ignorance concerning the work itself and concerning its implications. The typical scientist contents himself with retaining...some criticism that at most applies to a small fraction of the published studies. But these findings (which challenge our very concepts of space and time) are—if valid—of enormous importance...so they ought not to be ignored. (p. 359)

Price then went on to assert that ESP is "incompatible with current scientific theory," and that it is therefore more parsimonious to believe that parapsychologists cheat than that ESP is a real phenomenon. He based this argument on philosopher David Hume's essay on miracles. Hume argued that since we know people lie but have no independent evidence of miracles it is more reasonable to believe that claims of miracles are based on lies than that miracles actually occur. Price concluded, "My opinion concerning the findings of the parapsychologists is that many of them are dependent on clerical and statistical errors and unintentional use of sensory clues, *and that all extrachance results not so explicable are dependent on deliberate fraud or mildly abnormal mental condition*" (p. 360). Since it was given such prominence in one of the scientific world's leading interdisciplinary journals, this remarkable critique was widely reviewed. Responses came not only from parapsychologists but also from other scientists as well. One of the most effective responses was a joint paper by psychologist Paul Meehl and philosopher of science Michael Scriven (1956), who pointed out that Price's argument rests on two highly questionable assumptions: that contemporary scientific knowledge is complete and that ESP necessarily conflicts with it.

The most prominent critic of this period was the English psychologist C. E. M. Hansel (1966/1980). Hansel pursued the line of attack initiated by Price.[1] "It is wise," Hansel wrote, "to adopt initially the assumption that ESP is impossible, since there is a great weight of knowledge supporting this point of view" (Hansel, 1980, p. 22). He provided no documentation whatsoever for this assumption. Neither Hansel, nor any other critic has ever, to the best of my knowledge, shown that the existence of psi phenomena necessarily conflicts with established knowledge. Consider, for example, the following comment by physicist Gerald Feinberg (1975), concerning what is probably the most intuitively distressing parapsychological phenomenon—precognition:

> Instead of forbidding precognition from happening, [accepted physical] theories typically have sufficient symmetry (between past and future) to suggest that phenomena akin to precognition should occur.... Indeed, phenomena involving a reversed time order of cause and effect are generally excluded from consideration on the ground that they have not been observed, rather than because the theory forbids them. This exclusion itself introduces an element of asymmetry into the physical theories, which some physicists have felt was improper or required further

[1] Citation to Hansel in this section are to Hansel's (1980) revision.

explanation.... Thus, if such phenomena indeed occur, *no change in the fundamental equations of physics would be needed to describe them.* (pp. 54–55, emphasis added)

Asserting that psi is *a priori* extremely unlikely has permitted Hansel and other psi critics extraordinary latitude in the types of alternative explanations they allow themselves to entertain: "A possible explanation other than [ESP], provided it involves only well-established processes," he said, "should not be rejected on the grounds of its complexity" (p. 21). "If the result *could have arisen through a trick,* the experiment must be considered unsatisfactory proof of ESP, *whether or not it is finally decided that such a trick was, in fact, used*" (p. 21, my emphasis). Hansel admitted that "no single experiment can be conclusive," and that replications of "an ESP experiment by independent investigators could render the possibility of deception or error extremely unlikely... if the original result is repeatedly confirmed,... ESP becomes increasingly likely" (p. 21). Hansel then proceeded to examine the evidence in a manner that was logically inconsistent with these statements. His critique, which focused on a small number of the classic card-guessing experiments, consisted of showing how each individual experiment could be dismissed if one were willing to adopt complex and elaborate fraud scenarios. He succeeded only in reaffirming his initial proposition that no single experiment should be regarded as conclusive.

There have been two documented cases of investigator fraud in parapsychology (Markwick, 1978; Rhine, 1974), and the scientific community has, in recent years, been forced to confront the unpleasant fact that scientific fraud is more common than we earlier believed (Broad & Wade, 1982; Kohn, 1988). Surely the most effective solution to this problem is, as Hansel says, to require independent replication of studies believed to have important practical or theoretical consequences before their findings are accepted. Unsubstantiated fraud accusations are not merely unethical, they are incompatible with scientific progress. New discoveries in science would be impossible if scientists rejected unexpected findings on the ground that "if the result could have arisen through a trick, the experiment must be considered unsatisfactory evidence of X, *whether or not it is finally decided that such a trick was, in fact, used.*"

Unfortunately, replication research is neither strongly encouraged nor highly valued in mainstream science. A recent study of social and behavioral science journal editors' attitudes toward publication of replication studies found a strong bias against publishing replications (Neuliep & Crandall, 1991). Other studies of behavioral science publication practices show similar biases against publication of studies that do not produce statistically significant results (Bozarth & Roberts, 1972; Sterling, 1959). In their survey of 1,334 articles from psychological journals, Bozarth and Roberts found that while 94 percent of the articles using statistical tests reported significant results, less than one percent involved replications. In contract, parapsychologists have long recognized the importance of replications and of reporting nonsignificant results. The Parapsychological Association (PA) has had an official policy against selective reporting of "positive" results since 1975. The PA is, to the best of my knowledge, the *only* professional scientific organization that has adopted such a policy. If you examine the PA-affiliated journals and conference proceedings, you will find many replication attempts, both successful and unsuccessful

The "Ganzfeld Debate" of the 1980s

The ESP ganzfeld paradigm provides an excellent counter to Hyman's ~~central theme,~~ that parapsychology lacks cumulativeness. I will precede discussion of the psi ganzfeld controversy with a brief account of the background and rationale underlying psi ganzfeld research to show how it has systematically built upon earlier research. (See Figure 3.)

Historically, apparent psi effects have been frequently associated with dreaming, hypnosis, meditation, and other naturally occurring or deliberately induced internal attention states. This generalization is based on converging evidence from spontaneous case studies, claims associated with various cultural practices, clinical observations, and experimental studies. I have presented this background material in detail elsewhere (Honorton, 1977). To recapitulate:

Dreaming. Cross-cultural surveys of spontaneous cases indicate that approximately 2 out of 3 reported "real-life" psi interactions are mediated through dreams rather than waking experiences (Green, 1960; Prasad & Stevenson, 1968; L. E. Rhine, 1962; Sannwald, 1959). Of course spontaneous cases are anecdotal and no conclusions should be based upon them; but they can (and should) serve as the basis of hypotheses to be tested experimentally. Experimental evidence supporting these spontaneous case trends was first provided by the ESP dream studies at Maimonides Medical Center in New York (Child, 1985; Ullman & Krippner with Vaughan, 1973). Using electrophysiological sleep-monitoring techniques to de-

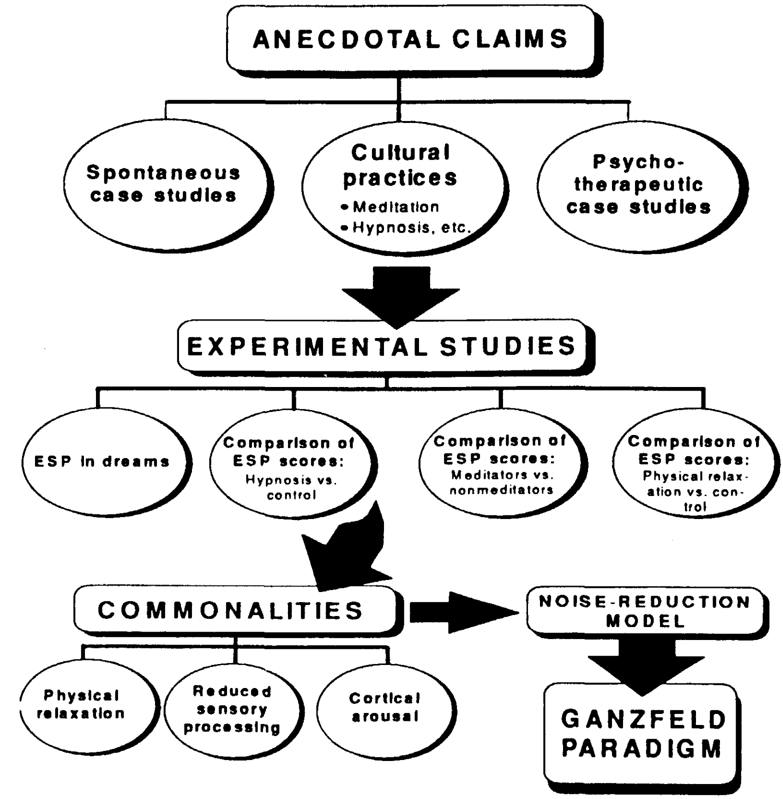

Figure 3. Origins of the ESP ganzfeld paradigm.

tect dream (REM) periods, investigators awoke physically remote senders who concentrated on randomly selected target pictures while the subjects dreamed. The subject was awakened and a dream report was recorded after each dream period. Following an experimental series, outside judges read transcripts of each night's dream reports and attempted, on a blind basis, to match them to the actual target picture used that night. The judges successfully matched the dream reports to their correct targets to a highly significant degree.

Hypnosis. The association between hypnosis and ostensible psi effects dates back to the claims of "travelling clairvoyance" and "community of sensation" in early mesmerism (Dingwall, 1968). Experimental support for a relationship between hypnosis and ESP comes from a variety of experimental studies, perhaps most persuasively from modern experimental studies comparing the effects of hypnotic induction versus nonhypnotic control conditions on ESP card-guessing performance (Schechter, 1984; Van de Castle, 1969). Schechter, for example, reported a meta-analysis of 25 experiments carried out between 1945-1981 by investigators in 10 different laboratories. ESP scores in the hypnotic induction condition were consistently (and significantly) higher than in the control condition of these experiments.

Meditation/relaxation. Claims of ostensible psychic phenomena occurring during the practice of meditation occur in most of the classical texts on meditation. A variety of modern experimental studies have indicated that meditation and relaxation exercises facilitate ESP test performance relative to control conditions (e.g., Braud & Braud, 1973, 1974; Dukhan & Rao, 1973; Stanford & Mayer, 1974).

Commonalities and a provisional model. The psi ganzfeld paradigm emerged as an attempt to explain the apparent psi-conduciveness of these and similar conditions. The question was asked: "What do dreaming, hypnosis, and meditation have in common that would lead each of them to facilitate ESP test performance?" While differing in many ways, each of these states involves physical relaxation, a reduction in ordinary perceptual processing (sensory deprivation), and a sufficient level of cortical arousal to sustain conscious awareness. This led to the development of a low-level descriptive model of ESP functioning, according to which internal attention states facilitate psi detection by reducing sensory and somatic stimuli that normally mask weaker psi input. This "noise-reduction" model thus identified sensory deprivation as a key to the frequent association between ostensible psi communication and internal states, and the ESP ganzfeld procedure was specifically developed to test the impact of perceptual isolation on psi performance. Thus we can see that the ganzfeld paradigm systematically built upon a diverse range of evidence including four different lines of experimental findings.

The ganzfeld debate consisted of a set of exchanges between Hyman (1983, 1985) and myself (Honorton, 1983, 1985), involving meta-analyses of 42 ganzfeld studies reported between 1974 and 1981. This phase of the psi controversy is unique because it resulted in a joint collaboration between critics and researchers who agreed upon specific methodological guidelines for future research that would be mutually acceptable (Hyman & Honorton, 1986). The various issues are summarized in Figure 4.

As in earlier phases of the psi controversy, the first issue concerned whether there was any effect requiring explanation. Initial estimates of ganzfeld replication rates suggested that around 50 per-

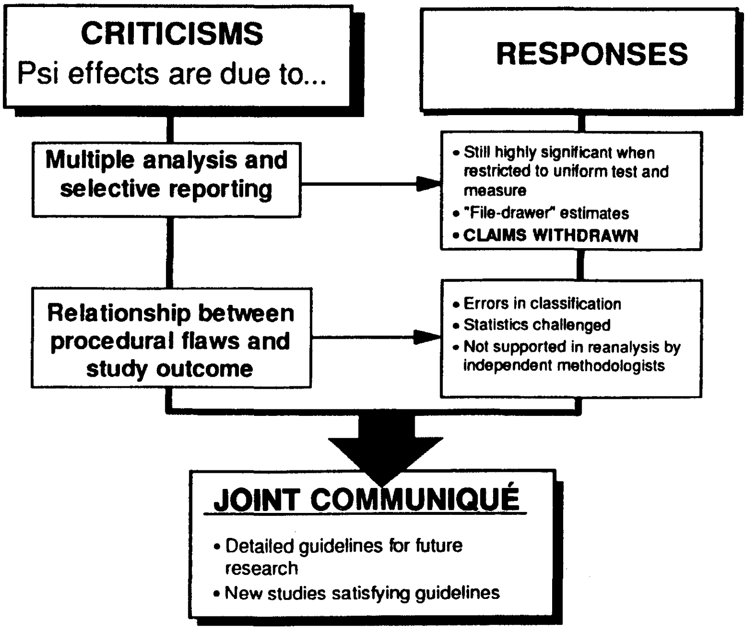

Figure 4. The ganzfeld debate of the 1980s.

cent of the reported ganzfeld studies yielded significant results compared to the expected chance rate of 5 percent. These estimates were challenged by Hyman. He pointed out that a number of the ganzfeld investigators had either applied multiple statistical tests or multiple measures of success to the results of their studies, creating a multiple analysis problem that *could* have inflated the estimates of significance; in fact, Hyman argued that the effects of multiple analysis were such as to increase the chance rate from 5 percent to 25 percent. He also argued that biased reporting of positive results (the "file-drawer" problem) might have exaggerated the significance of the known studies. In response, I restricted my analysis to 28 of the 42 studies examined by Hyman for which a uniform measure (direct hits) and test could be applied. The results were still "astronomically significant," with odds against chance of a billion to one, and would require 15 unknown studies averaging chance results for every known study in order to reduce the overall results to nonsignificance. Hyman subsequently agreed that the significance of the ganzfeld studies could not be explained through multiple analysis or selective reporting:

Although we probably still differ on the magnitude of the biases contributed by multiple testing, retrospective experiments, and the file-drawer problem, we agree that the overall significance observed in these studies cannot reasonably be explained by these selective factors. Something beyond selective reporting or inflated significance levels seems to be producing nonchance outcomes. Moreover, we agree that the significant outcomes have been produced by a number of different investigators. (Hyman & Honorton, 1986, p. 352)

The next line of criticism concerned the effects of procedural flaws on the study outcomes. In our meta-analyses of the ganzfeld studies, Hyman and I independently coded each study's procedures with respect to potential flaws involving sensory cues, randomization method, security, and so on. Here Hyman and I did not agree: my analysis showed no significant relationship between these variables and study success, while Hyman claimed that some of the flaw variables, such as the type of randomization, did correlate with results. In his initial assessment, Hyman claimed there was a nearly perfect linear correlation between the number of flaws in a study and its success (Hyman, 1982); this analysis contained a large number of errors that Hyman later attributed to typing errors (communication to Honorton, November 29, 1982). Later, Hyman (1985) claimed a significant relationship between study flaws and outcomes based on a complex multivariate analysis. However, an independent psychological statistician described this analysis as "meaningless" (Saunders, 1985). Finally, Hyman agreed that "the present data base does not support any firm conclusion about the relationship between flaws and study outcome" (Hyman & Honorton, 1986, p. 353). Were our differences in flaw assessment simply reflections of our respective biases? Perhaps, but independent examination of the issue by nonparapsychologists has unanimously failed to support Hyman's conclusions (Atkinson, Atkinson, Smith, & Bem, 1990; Harris & Rosenthal, 1988a, 1988b; Saunders, 1985; Utts, 1991). In an independent analysis using Hyman's own flaw codings, two behavioral science methodologists concluded, "Our analysis of the effects of flaws on study outcome lends no support to the hypothesis that Ganzfeld research results are a significant function of the set of flaw variables" (Harris & Rosenthal, 1988b, p. 3).

Rather than continue the debate, Hyman and I collaborated on a "joint communiqué" in which we agreed that:

the best way to resolve the controversy between us is to await the outcome of future ganzfeld psi experiments. These experiments, ideally,

> will be carried out in such a way as to circumvent the file-drawer problem, problems of multiple analysis, and the various defects in randomization, statistical application, and documentation pointed out by Hyman. If a variety of parapsychologists and other investigators continue to obtain significant results under these conditions, then the existence of a genuine communications anomaly will have been demonstrated. (Hyman & Honorton, 1986, pp. 353–354)

The joint communiqué presented detailed methodological guidelines for the conduct and reporting of future ganzfeld experiments. Four years later, my colleagues and I reported an extensive series of ganzfeld experiments using an automated methodology that satisfied these guidelines (Honorton et al., 1990). These experiments are discussed from several different perspectives in the contributions by Broughton, Krippner, and Morris, but are curiously omitted from the contributions of Alcock, Hyman, and Randi. Elsewhere, however, Hyman has commented on these studies:

> Honorton's experiments have produced intriguing results. If, as Utts suggests, independent laboratories can produce similar results with the same relationships and with the same attention to rigorous methodology, then parapsychology may indeed have finally captured its elusive quarry. (Hyman, 1991, p. 392)

In this paper I have focused at length on Hyman's contribution because he more than the other skeptical contributors to this series has taken the trouble to familiarize himself with the material he is criticizing and has been willing, in the face of contrary evidence, to modify his position. His critical evaluations of various areas of psi research have been hard-hitting and I believe they have often been mistaken. But his involvement has contributed to a more accurate appraisal of the status of the areas in question and, most importantly, to the development of better experiments. It is therefore disappointing that he does not, in the spirit of our joint communiqué, actively encourage replication attempts by a broader range of scientists outside of parapsychology. Instead, here as in his other recent writings, Hyman appears to discourage replication efforts by scientists outside of parapsychology.

The automated ganzfeld experiments provide further evidence of cumulativeness in addition to confirmation of an overall effect consistent with the earlier studies. As Morris points out in his contribution, several additional hypotheses derived from trends in the previous meta-analysis were tested and supported, including superior performance in trials using dynamic rather than static targets and the use of senders who were friends of the subjects rather than relative strangers. It is important to draw attention to the fact that for each of the above hypotheses—overall success rate, impact of target type, and sender type—the actual *magnitudes* of the effects were consistent with the meta-analytic estimates. This was also true of another hypothesis based on a meta-analysis of the relationship between ESP performance and the psychological trait extraversion (Honorton, Ferrari, & Bem, in press); the magnitude of the correlation between psi performance and extraversion in the automated ganzfeld studies was significant and very close to the estimate from that meta-analysis. Findings such as these are important because they indicate the operation of a systematic process, not just an anomalous departure from a chance baseline, and they demonstrate that it is possible to build systematically upon earlier findings. They validate the meta-analytic estimates and provide the kind of "prospective evidence" Hyman calls for.

4. *Other Issues*

Parapsychology's Hidden Agenda?

Alcock's contribution does not address the scientific issues and therefore provides little basis for substantive comment. As in his earlier writings (e.g., Alcock, 1981; 1987), Alcock continues to focus on what he perceives to be a hidden agenda of religious or philosophical belief among parapsychologists—the desire to justify some form of spiritual belief. Most parapsychologists are motivated by a desire to increase fundamental understanding of human nature, but so too are most other scientists. Parapsychology is a scientific problem area, not a belief system. There are parapsychologists who believe that the findings of psi research will ultimately require accepting some form of mind-body dualism. Others believe the findings can be accommodated within a monistic framework. And there are still others—I suspect the majority of the contemporary researchers—who believe that a satisfactory scientific understanding of the psi data must await theoretical developments in other areas, especially physics and neurophysiology. Contrast this with the sort of appeals to religious belief that one sees in the popular writings of certain modern cosmologists and even prominent skeptics. Consider, for example, the final paragraph of Stephen Hawking's widely-acclaimed book, *A Brief History of Time*:

> However, if we do discover a complete theory, it should in time be understandable in broad principle by everyone.... Then we shall all, philosophers, scientists, and just ordinary people, be able to take part in the discussion of the question of why it is that we and the universe exist. If we find the answer to that, it would be the ultimate triumph of human reason—for then we would know the mind of God. (Hawking, 1988, p. 175)

And what would Alcock say of the "hidden agenda" of psi-skeptic Martin Gardner?

> As for empirical tests of the power of God to answer prayer, I am among those theists who, in the spirit of Jesus' remark that only the faithless look for signs, consider such tests both futile and blasphemous. (Gardner, 1983, p. 239)

Should the sentiments expressed in these and similar statements cast doubt on Hawking's physics or Gardner's skeptical acumen? I think not. They are entitled to their personal beliefs. Such beliefs should be considered irrelevant to the assessment of their scientific accomplishments unless there is ample reason to suspect that their science has been compromised by those beliefs.

Randi as Methodologist and Statistician

Randi's contribution is pure polemic and fails to deal in any substantive way with the scientific issues underlying the psi controversy. His disparaging comments about meta-analysis suggest that he does not understand meta-analysis and is unaware of its widespread use in medicine and the behavioral sciences. Randi's skill as a magician is well-known; but despite well-publicized claims to methodological expertise, his ability to design scientifically adequate psi experiments is not at all apparent from an examination of his public efforts. Serious methodological weaknesses and statistical errors occur, for example, in his book on testing ESP and in his televised tests of psychics (e.g., Morris, 1992; Rao, 1984).

5. *Skepticism, Science, and the "Paranormal"*

I believe the concept of the "paranormal" is an anachronism and should be abandoned. The term is usually used to imply that psi interactions must necessarily, if real, represent an order of reality outside the natural realm. The term emerged within the context of Newtonian physics and has, in my view, clearly outlived whatever usefulness it ever had. It has not served to guide the development of constructive research programs; indeed, its primary effect has been to create an artificial schism between psi researchers and the broader scientific community. A more empirically fruitful conceptualization is that parapsychology involves the study of currently anomalous communication and energetic processes. This approach guides the efforts of most of the parapsychological researchers I know, who work on the assumption that they are dealing with unexplained—anomalous, but not unexplainable—natural processes.

I believe in science, and I am confident that a science that can boldly contemplate the origin of the universe, the nature of physical reality 10^{-33} seconds after the Big Bang, anthropic principles, quantum nonlocality, and parallel universes, can come to terms with the implications of parapsychological findings—whatever they may turn out to be. There is no danger for science in honestly confronting these issues; it can only be enriched by doing so. But there is a danger for science in encouraging self-appointed protectors who engage in polemical campaigns that distort and misrepresent serious research efforts. Such campaigns are not only counterproductive, they threaten to corrupt the spirit and function of science and raise doubts about its credibility. The distorted history, logical contradictions, and factual omissions exhibited in the arguments of the three critics represent neither scholarly criticism nor skepticism, but rather counteradvocacy masquerading as skepticism. True skepticism involves the suspension of belief, not disbelief. In this context, we would do well to recall the words of the great nineteenth century naturalist and skeptic, Thomas Huxley: "Sit down before fact like a little child, be prepared to give up every preconceived notion, follow humbly to wherever and to whatever abysses nature leads or you shall learn nothing."

References

ALCOCK, J. E. (1981). *Parapsychology: Science or magic?* Oxford: Pergamon Press.

ALCOCK, J. E. (1987). Parapsychology: Science of the anomalous or search for the soul? *Behavioral and Brain Sciences,* **10,** 553–565.

ATKINSON, R. L., ATKINSON, R. C., SMITH, E. E., & BEM, D. J. (1990). *Introduction to psychology* (10th ed.). New York: Harcourt Brace Jovanovich.

BOWN, W. (1992). Unholy row rages over Trinity's ghostbuster. *New Scientist,* **135** (No. 1831), 9.

BOZARTH, J. D., & ROBERTS, R. R. (1972). Signifying significant significance. *American Psychologist,* **27,** 774–775.

BRAUD, L. W., & BRAUD, W. G. (1974). Further studies of relaxation as a psi-conducive state. *Journal of the American Society for Psychical Research,* **68,** 229–245.

BRAUD, W. G., & BRAUD, L. W. (1973). Preliminary explorations of psi-conducive states. Progressive muscular relaxation. *Journal of the American Society for Psychical Research,* **67,** 27–46.

BROAD, W., & WADE, N. (1982). *Betrayers of the truth: Fraud and deceit in the halls of science.* New York: Simon & Shuster.

CHILD, I. L. (1985). Psychology and anomalous observations: The question of ESP in dreams. *American Psychologist,* **40,** 1219–1230.

DINGWALL, E. J. (Ed.). (1968). *Abnormal hypnotic phenomena* (4 vols.). London: Churchill.

DRUCKMAN, D., & SWETS, J. A. (Eds.). (1988). *Enhancing human performance: Issues, theories, and techniques.* Washington, DC: National Academy Press.

DUKHAN, H., & RAO, K. R. (1973). Meditation and ESP scoring. In W. G. Roll, R. L. Morris, & J. D. Morris (Eds.) *Research in parapsychology 1972* (pp. 148–151). Metuchen, NJ: Scarecrow Pres.

FEINBERG, G. (1975). Precognition—a memory of things future. In L. Oteri (Ed.), *Quantum physics and parapsychology* (pp. 54–73). New York: Parapsychology Foundation.

GARDNER, M. (1983). *The whys of a philosophical scrivener.* New York: Quill.

GREEN, C. E. (1960). Analysis of spontaneous cases. *Proceedings of the Society for Psychical Research,* **53,** 97–161.

HAWKING, S. W. (1988). *A brief history of time.* New York: Bantam.

HANSEL, C. E. M. (1966/1980). *ESP and parapsychology: A critical re-evaluation.* Buffalo, NY: Prometheus.

HARRIS, M. J., & ROSENTHAL, R. (1988a). *Interpersonal expectancy effects and human performance research.* Washington, DC: National Academy Press.

HARRIS, M. J., & ROSENTHAL, R. (1988b). *Postscript to interpersonal expectancy effects and human performance research.* Washington, DC: National Academy Press.

HONORTON, C. (1975). "Error some place!" *Journal of Communication,* **25,** 103–116.

HONORTON, C. (1977). Psi and internal attention states. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 435–472). New York: Van Nostrand Reinhold.

HONORTON, C. (1983). Response to Hyman's critique of psi ganzfeld studies. In W. G. Roll, J. Beloff, & R. A. White (Eds.), *Research in parapsychology 1982* (pp. 23–26). Metuchen, NJ: Scarecrow Press.

HONORTON, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology,* **49,** 51–91.

HONORTON, C., BERGER, R. E., VARVOGLIS, M. P., QUANT, M., DERR, P., SCHECHTER, E. I., & FERRARI, D. C. (1990). Psi communication in the ganzfeld: Experiments with an automated testing system and a comparison with a meta-analysis of earlier studies. *Journal of Parapsychology,* **54,** 99–139.

Honorton, C., & Ferrari, D. C. (1989). "Future telling": A meta-analysis of forced-choice precognition experiments, 1935–1987. *Journal of Parapsychology,* **53,** 281–308.

Honorton, C., Ferrari, D. C., & Bem, D. J. (in press). Extraversion and ESP performance: Meta-analysis and a new confirmation. In *Research in parapsychology 1990.* Metuchen, NJ: Scarecrow Press.

Huntington, E. V. (1938). Is it chance or ESP? *American Scholar,* **7,** 201–210.

Hyman, R. (1982). Hyman's tally of flaws in ganzfeld/psi experiments. Written communication to C. Honorton, July 29, 1982.

Hyman, R. (1983). Does the ganzfeld experiment answer the critics' objections? In W. G. Roll, J. Beloff, & R. A. White (Eds.), *Research in parapsychology 1982* (pp. 21–23). Metuchen, NJ: Scarecrow Press.

Hyman, R. (1985). The ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology,* **49,** 3–49.

Hyman, R. (1991). Comment. *Statistical Science,* **6,** 389–392.

Hyman, R., & Honorton, C. (1986). A joint communiqué: The psi ganzfeld controversy. *Journal of Parapsychology,* **50,** 351–364.

Kennedy, J. L., & Uphoff, H. F. (1939). Experiments on the nature of extrasensory perception: III. The recording error criticism of extra-chance results. *Journal of Parapsychology,* **3,** 226–245.

Kohn, A. (1988). *False prophets: Fraud and error in science and medicine.* Oxford: Blackwell.

Markwick, B. (1978). The Soal-Goldney experiments with Basil Shackleton; New evidence of data manipulation. *Proceedings of the Society for Psychical Research,* **56,** 250–281.

Meehl, P. E., & Scriven, M. (1956). Compatibility of science and ESP. *Science,* **123,** 14–15.

Morris, R. L. (1992). Reply to Randi. *The Psi Researcher,* No. 5, 16–18.

Neuliep, J. W., & Crandall, R. (1991). Editorial bias against replication research. In J. W. Neuliep (Ed.), *Replication research in the social sciences* (pp. 85–90). Newbury Park, CA: Sage.

Prasad, J., & Stevenson, I. (1968). A survey of spontaneous psychical experiences in school children of Uttar Pradesh, India. *International Journal of Parapsychology,* **10,** 241–261.

Pratt, J. G., Rhine, J. B., Smith, B. M., Stuart, C. E., & Greenwood, J. A. (1940/1966). *Extra-sensory perception after sixty years.* Boston: Bruce Humphries.

Price, G. R. (1955). Science and the supernatural. *Science,* **122,** 359–367.

Radin, D. I., & Ferrari, D. C. (1991). Effects of consciousness on the fall of dice: A meta-analysis. *Journal of Scientific Exploration,* 61–83.

Radin, D. I., & Nelson, R. D. (1989). Evidence for consciousness-related anomalies in random physical systems. *Foundation of Physics,* **19,** 1499–1514.

Rao, K. R. (1984). Review of *Test your ESP potential: A complete kit with instructions, scorecards, and apparatus* by James Randi. *Journal of Parapsychology,* **48,** 356–358.

RHINE, J. B. (1934/1964). *Extra-sensory perception.* Boston: Bruce Humphries.

RHINE, J. B. (1974). Comments: A new case of experimenter unreliability. *Journal of Parapsychology,* **38,** 215–225.

RHINE, L. E. (1962). Psychological processes in ESP experiences: I. Waking experiences. *Journal of Parapsychology,* **26,** 88–111.

ROSENTHAL, R. (1978). How often are our numbers wrong? *American Psychologist,* **33,** 1005–1008.

SANNWALD, G. (1959). Statistische untersuchungen an Spontanphänomene. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie,* **3,** 59–71.

SAUNDERS, D. R. (1985). On Hyman's factor analyses. *Journal of Parapsychology,* **49,** 86–88.

SCHECHTER, E. I. (1984). Hypnotic induction vs. control conditions: Illustrating an approach to the evaluation of replicability in parapsychology. *Journal of the American Society for Psychical Research,* **78,** 1–27.

SCHOUTEN, S. A. (in press). Are we making progress? In *Psi research methodology: A re-examination* (37th Annual International Conference of the Parapsychology Foundation). New York: Parapsychology Foundation, Inc.

SCOTT, C. (1958). G. Spencer Brown and probability: A critique. *Journal of the Society for Psychical Research,* **39,** 217–234.

SPENCER BROWN, G. (1953). Statistical significance in psychical research. *Nature,* **172,** 154–156.

SPENCER BROWN, G. (1957). *Probability and scientific inference,* NY: Longmans.

STANFORD, R. G., & MAYER, B. (1974). Relaxation as a psi-conducive state: A replication and exploration of parameters. *Journal of the American Society for Psychical Research,* **68,** 182–191.

STERLING, T. C. (1959). Publication decisions and their possible effects on inference drawn from tests of significance—or vice versa. *Journal of the American Statistical Association,* **54,** 30–34.

ULLMAN, M., KRIPPNER, S., & VAUGHAN, A. (1973). *Dream telepathy: Experiments in noctural ESP.* New York: Macmillan.

UTTS, J. M. (1991). Replication and meta-analysis in parapsychology. *Statistical Science,* **6,** 363–403.

VAN DE CASTLE, R. L. (1969). The facilitation of ESP through hypnosis. *American Journal of Clinical Hypnosis,* **12,** 37–56.

Psychology Department
University of Edinburgh
Edinburgh, EH8 902
Scotland