To be published in *Psychological Bulletin*
Still subject to final editing.

Does Psi Exist? Replicable Evidence for an Anomalous Process of Information Transfer

Daryl J. Bem
Cornell University

Charles Honorton
University of Edinburgh

Most academic psychologists do not yet accept the existence of psi, anomalous processes of information or energy transfer (like telepathy or other forms of extrasensory perception) that are currently unexplained in terms of known physical or biological mechanisms. We believe that the replication rates and effect sizes achieved by one particular experimental method, the *ganzfeld* procedure, are now sufficient to warrant bringing this body of data to the attention of the wider psychological community. We review competing meta-analyses of the ganzfeld database, one by Hyman (1985), a skeptical critic of psi research, the other by Honorton (1985), a parapsychologist and major contributor to the ganzfeld database. Next we summarize the results of 11 new ganzfeld studies that comply with guidelines jointly authored by Hyman and Honorton (1986). Finally, we discuss issues of replication and theoretical explanation.

The term *psi* denotes anomalous processes of information or energy transfer, processes like telepathy or other forms of extrasensory perception that are currently unexplained in terms of known physical or biological mechanisms. The term is purely descriptive: It neither implies that such anomalous phenomena are paranormal nor connotes anything about their underlying mechanisms.

Does psi exist? Most academic psychologists don't think so. A survey of over 1,100 college professors in the United States found that 55% of natural scientists, 66% of social scientists (excluding psychologists), and 77% of academics in the arts, humanities, and education believed that extrasensory perception is either an established fact or a likely possibility. The comparable figure for psychologists was only 34%. Moreover, an equal number of psychologists declared extrasensory perception to be an impossibility, a view expressed by only 2% of all other respondents (Wagner & Monnet, 1979).

Psychologists are probably more skeptical about psi for several reasons. First, we believe that extraordinary claims require extraordinary proof. And although our colleagues from other disciplines would probably agree with this dictum, we are more likely to be familiar with the methodological and statistical requirements for sustaining such claims—as well as with previous claims that failed either to meet those requirements or to survive the test of successful replication. Even for ordinary claims, our conventional statistical criteria are conservative. The sacred $p = .05$ threshold is a constant reminder that it is far more sinful to assert that an effect exists when it does not (the Type I error) than to assert that an effect does not exist when it does (the Type II error).

Second, most of us distinguish sharply between phenomena whose explanations are merely obscure or controversial (e.g., hypnosis) and phenomena like psi, which would appear to fall outside our current explanatory framework altogether. (Some would characterize this as the difference between the unexplained and the inexplicable.) In contrast, many laypersons treat all exotic psychological phenomena as epistemologically equivalent—many even consider déjà vu to be a psychic phenomenon. The blurring of this critical distinction is aided and abetted by the mass media, "new age" books and mind-power courses, and by "psychic" entertainers who present both genuine hypnosis and fake "mindreading" in the course of a single performance. Accordingly, most laypersons would not have to revise their conceptual model of reality as radically as we would in order to assimilate the existence of psi. For us, psi is simply more extraordinary.

And finally, research in cognitive and social psychology has sensitized us to the errors and biases that plague intuitive attempts to draw valid inferences from the data of everyday experience (Gilovich, 1991; Nisbett & Ross, 1980; Tversky & Kahneman, 1971). This leads us to give virtually no probative weight to anecdotal or journalistic reports of psi—the main source cited by our academic colleagues as evidence for their beliefs about psi (Wagner & Monnet, 1979).

Ironically, however, psychologists are probably *not* more familiar than others with recent experimental research on psi. Like most psychological research, parapsychological research is reported primarily in specialized journals; unlike most psychological research, however, contemporary parapsychological research is not usually reviewed or summarized in psychology's textbooks, handbooks, or mainstream journals. For example, only 1 of 64 introduc-

---

Sadly, Charles Honorton died of a heart attack on November 4, 1992, nine days before this article was accepted for publication. He was 46. Parapsychology has lost one of its most valued contributors. I have lost a valued friend.

This collaboration had its origins in a 1983 visit I made to Honorton's Psychophysical Research Laboratories (PRL) in Princeton, New Jersey, as one of several outside consultants brought in to examine the design and implementation of the experimental protocols.

Preparation of this article was supported, in part, by grants to Charles Honorton from the American Society for Psychical Research and the Parapsychology Foundation, both of New York City. The work at PRL summarized in the second half of this article was supported by the James S. McDonnell Foundation of St. Louis, Missouri, and by the John E. Fetzer Foundation of Kalamazoo, Michigan.

Helpful comments on earlier drafts were received from Deborah Delanoy, Edwin May, Donald McCarthy, Robert Morris, John Palmer, Robert Rosenthal, Lee Ross, Jessica Utts, Philip Zimbardo, and two anonymous reviewers.

Correspondence should be addressed to Daryl J. Bem, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853.

tory psychology textbooks recently surveyed even mentions the experimental procedure reviewed in this article, a procedure that has been in widespread use since the early 1970s (Roig, Icochea, & Cuzzucoli, 1991). Other secondary sources for nonspecialists are frequently inaccurate in their descriptions of parapsychological research (Child, 1985; Palmer, Honorton, & Utts, 1989).

This situation may be changing. Discussions of modern psi research have recently appeared in a widely-used introductory textbook (Atkinson, Atkinson, Smith, & Bem, 1990, 1993), two mainstream psychology journals (Child, 1985; Rao & Palmer, 1987), and a scholarly but accessible book for nonspecialists (Broughton, 1991). The purpose of the present article is to supplement these broader treatments with a more detailed, meta-analytic presentation of evidence issuing from a single experimental method: the *ganzfeld* procedure. We believe that the replication rates and effect sizes achieved with this procedure are now sufficient to warrant bringing this body of data to the attention of the wider psychological community.

The Ganzfeld Procedure

By the 1960s, a number of parapsychologists had become dissatisfied with the familiar ESP testing methods pioneered by J. B. Rhine at Duke University in the 1930s. In particular, they felt that the repetitive forced-choice procedure in which a subject repeatedly attempts to select the correct "target" symbol from a set of fixed-alternatives failed to capture the circumstances that characterize reported instances of psi in everyday life.

Historically, psi has often been associated with meditation, hypnosis, dreaming, and other naturally occurring or deliberately induced altered states of consciousness. For example, the view that psi phenomena can occur during meditation is expressed in most classical texts on meditative techniques; the belief that hypnosis is a psi-conducive state dates all the way back to the days of early mesmerism (Dingwall, 1968); and cross-cultural surveys indicate that most reported "real-life" psi experiences are mediated through dreams (Green, 1960; Prasad & Stevenson, 1968; Rhine, 1962; Sannwald, 1959).

There are now reports of experimental evidence consistent with these anecdotal observations. For example, several laboratory investigators report that meditation facilitates psi performance (Honorton, 1977). A meta-analysis of 25 experiments on hypnosis and psi conducted between 1945–1981 in 10 different laboratories suggests that hypnotic induction also facilitates psi performance (Schechter, 1984). And dream-mediated psi was reported in a series of experiments conducted at Maimonides Medical Center in New York, and published between 1966 and 1972 (Child, 1985; Ullman, Krippner, & Vaughan, 1973).

In the Maimonides dream studies, two subjects—a "receiver" and a "sender"—spent the night in a sleep laboratory. The receiver's brainwaves and eye movements were monitored as he or she slept in an isolated room. When the receiver entered a period of REM sleep, the experimenter pressed a buzzer that signaled the sender—under the supervision of a second experimenter—to begin a sending period. The sender would then concentrate on a randomly chosen picture (the "target") with the goal of influencing the content of the receiver's dream.

Toward the end of the REM period, the receiver was awakened and asked to describe any dream just experienced. This procedure was repeated throughout the night with the same target. A transcription of the receiver's dream reports was given to outside judges who blindly rated the similarity of the night's dreams to several pictures, including the target. In some studies similarity ratings were also obtained from the receivers themselves. Across several variations of the procedure, dreams were judged to be significantly more similar to the target pictures than to the control pictures in the judging sets. (Failures to replicate the Maimonides results are also reviewed by Child (1985).)

These several lines of evidence suggested a working model of psi in which psi-mediated information is conceptualized as a weak signal that is normally masked by internal somatic and external sensory "noise." By reducing ordinary sensory input, these diverse psi-conducive states are presumed to raise the signal-to-noise ratio, thereby enhancing a person's ability to detect the psi-mediated information (Honorton, 1969, 1977). In order to test the hypothesis that a reduction of sensory input itself facilitates psi performance, investigators turned to the ganzfeld procedure (Braud, Wood, & Braud, 1975; Honorton & Harper, 1974; Parker, 1975), a procedure originally introduced into experimental psychology during the 1930s to test propositions derived from Gestalt theory (Avant, 1965; Metzger, 1930).

Like the dream studies, the psi ganzfeld procedure has most often been used to test for telepathic communication between a sender and a receiver. The receiver is placed in a reclining chair in an acoustically-isolated room. Translucent ping-pong ball halves are taped over the eyes and headphones are placed over the ears; a red floodlight directed toward the eyes produces an undifferentiated visual field and white noise played through the headphones produces an analogous auditory field. It is this homogeneous perceptual environment that is called the *Ganzfeld* ("total field"). In order to reduce internal somatic "noise," the receiver typically also undergoes a series of progressive relaxation exercises at the beginning of the ganzfeld period.

The sender is sequestered in a separate acoustically-isolated room, and a visual stimulus (art print, photograph, or brief videotaped sequence) is randomly selected from a large pool of such stimuli to serve as the target for the session. While the sender concentrates on the target, the receiver provides a continuous verbal report of his or her ongoing imagery and mentation, usually for about 30 minutes. At the completion of the ganzfeld period, the receiver is presented with several stimuli (usually four) and, without knowing which stimulus was the target, is asked to rate the degree to which each matches the imagery and mentation experienced during the ganzfeld period. If the receiver assigns the highest rating to the target stimulus, it is scored as a "hit." Thus, if the experiment employs judging pools containing four stimuli (the target and three "decoys" or control stimuli), then the hit rate expected by chance is .25. The ratings can also be analyzed in other ways; for example, they can be converted to ranks or standardized scores within each set and analyzed parametrically across sessions. And, as with the dream studies, the similarity ratings can also be made by outside judges using transcripts of the receiver's mentation report.

Meta-analyses of the Ganzfeld Database

In 1985 and 1986, the *Journal of Parapsychology* devoted two entire issues to a critical examination of the

ganzfeld database. The 1985 issue comprised two contributions: (a) a meta-analysis and critique by Ray Hyman (1985), a cognitive psychologist and skeptical critic of parapsychological research; and (b) a competing meta-analysis and rejoinder by Charles Honorton (1985), a parapsychologist and major contributor to the ganzfeld database. The 1986 issue contained four commentaries on the Hyman-Honorton exchange, a joint communiqué by Hyman and Honorton, and six additional commentaries on the joint communiqué itself. We summarize the major issues and conclusions here.

Replication Rates

By study. Hyman's meta-analysis covered 42 psi ganzfeld studies reported in 34 separate reports written or published from 1974 through 1981. One of the first problems he discovered in the database was multiple analysis. As noted above, it is possible to calculate several indices of psi performance in a ganzfeld experiment and, further, to subject those indices to several kinds of statistical treatment. Many investigators reported multiple indices or applied multiple statistical tests without adjusting the criterion significance level for the number of tests conducted. Worse, some may have "shopped" among the alternatives until finding one that yielded a significantly successful outcome. Honorton agreed that this was a problem.

Accordingly, Honorton applied a uniform test on a common index across all studies from which the pertinent datum could be extracted, regardless of how the investigators had analyzed the data in the original reports. He selected the proportion of hits as the common index because it could be calculated for the largest subset of studies: 28 of the 42 studies. The hit rate is also a conservative index because it discards most of the rating information; a second place ranking—a "near miss"—receives no more credit than a last place ranking. Honorton then calculated the exact binomial probability and its associated z score for each study.

Of the 28 studies, 23 (82%) had positive z scores ($p = 4.6 \times 10^{-4}$ exact binomial test with $p = q = .5$). Twelve of the studies (43%) had z scores that were independently significant at the 5% level ($p = 3.5 \times 10^{-9}$, binomial test with $n = 28$ studies, $p = .05$, and $q = .95$) and 7 of the studies (25%) were independently significant at the 1% level ($p = 9.8 \times 10^{-9}$). The composite Stouffer z score across the 28 studies was 6.60 ($p = 2.1 \times 10^{-11}$).[1] A more conservative estimate of significance can be obtained by including 10 additional studies that also used the relevant judging procedure but did not report hit rates. If we assign these studies a mean z score of zero, then the Stouffer z across all 38 studies becomes 5.67 ($p = 7.3 \times 10^{-9}$).

Thus, whether one considers only the studies for which the relevant information is available or includes a null estimate for the additional studies where the information is not available, the aggregate results cannot reasonably be attributed to chance. And by design, the cumulative outcome reported here cannot be attributed to the inflation of significance levels through multiple analysis.

By laboratory. One objection to estimates like those above is that studies from a common laboratory are not independent of one another (Parker, 1978). Thus it is possible for one or two investigators to be disproportionately responsible for a high replication rate while other, independent investigators are unable to obtain the effect.

The ganzfeld database is vulnerable to this possibility. The 28 studies providing hit rate information were conducted by investigators in 10 different laboratories. One laboratory contributed nine of the studies; Honorton's own laboratory contributed five; two other laboratories contributed three each; two contributed two each; and the remaining four laboratories each contributed one. Thus half of the studies were conducted by only two laboratories, one of them Honorton's own.

Accordingly, Honorton calculated a separate Stouffer z score for each laboratory. Significantly positive outcomes were reported by 6 of the 10 laboratories and the combined result across laboratories yielded a z of 6.16 ($p = 3.6 \times 10^{-10}$). Even if all the studies conducted by the two most prolific laboratories are discarded from the analysis, the Stouffer z across the eight other laboratories remains significant ($z = 3.67, p = 1.2 \times 10^{-4}$). Four of these studies are significant at the 1% level ($p = 9.2 \times 10^{-6}$, binomial test with $n = 14$ studies, $p = .01$, and $q = .99$), and each was contributed by a different laboratory.

Thus, even though the total number of laboratories in this database is small, a majority of them have reported significant studies, and the significance of the overall effect does not depend on just one or two of them.

Selective Reporting

In recent years, behavioral scientists have become increasingly aware of the "file-drawer" problem, the likelihood that successful studies are more likely to be published than unsuccessful studies—which are more likely to be consigned to the file drawers of their disappointed investigators (Bozarth & Roberts, 1972; Sterling, 1959). Parapsychologists were among the first to become sensitive to the problem; and, in 1975, the Parapsychological Association Council adopted a policy opposing the selective reporting of positive outcomes. As a consequence, negative findings have been routinely reported at the Association's meetings and in its affiliated publications for almost two decades now. As we have already seen, more than half of the ganzfeld studies included in the meta-analysis yielded outcomes whose significance falls short of the conventional .05 level.

There is a variant of the selective reporting problem which arises from what Hyman (1985) has termed the "retrospective study." An investigator conducts a small set of exploratory trials. If they yield null results, they remain "exploratory" and never become part of the official record; if they happen to yield positive results, they get defined as a study after the fact and are submitted for publication. In support of this possibility, Hyman notes that there are more significant studies in the database with fewer than 20 trials than one would expect under the assumption that, all other things being equal, statistical power should increase with the square root of the sample size. Although Honorton questions the assumption that "all other things" are in fact equal across the studies and disagrees with Hyman's particular statistical analysis, he does agree that there is an apparent clustering of significant studies with fewer than 20 trials. (Out of the complete ganzfeld

[1] Stouffer's z is computed by dividing the sum of the z scores for the individual studies by the square root of the number of studies (Rosenthal, 1978).

database of 42 studies, 8 have fewer than 20 trials, and 6 of these report statistically significant results.)

Because it is impossible, by definition, to know how many unknown studies—exploratory or otherwise—are languishing in file drawers somewhere, the major tool for estimating the seriousness of selective reporting problems has become some variant of Rosenthal's "file drawer" statistic, an estimate of how many unreported studies with z scores of zero would be required to exactly cancel out the significance of the known database (Rosenthal, 1979). For the 28 direct-hit ganzfeld studies alone, this estimate is 423 fugitive studies, a ratio of unreported-to-reported studies of approximately 15 to 1. When it is recalled that a single ganzfeld session takes over an hour to conduct, it is not surprising that—despite his concern with the retrospective study problem—Hyman concurs with Honorton and other participants in the published debate that selective reporting problems cannot plausibly account for the overall statistical significance of the psi ganzfeld database (Hyman & Honorton, 1986).[2]

Methodological Flaws

If the most frequent criticism of parapsychology is that it has not produced a replicable psi effect, the second most frequent criticism is that many, if not most, psi experiments have inadequate controls and procedural safeguards. A frequent charge is that positive results emerge primarily from initial, poorly controlled studies and then vanish as better controls and safeguards are introduced.

Fortunately, meta-analysis provides a vehicle for empirically evaluating the extent to which methodological flaws may have contributed to artifactual positive outcomes across a set of studies. First, ratings are assigned to each study that index the degree to which particular methodological flaws are or are not present; these ratings are then correlated with the studies' outcomes. Large positive correlations constitute evidence that the observed effect may be artifactual.

In psi research, the most fatal flaws are those that might permit a subject to obtain the target information in normal sensory fashion, either inadvertently or through deliberate cheating. This is called the problem of *sensory leakage*. Another potentially serious flaw is inadequate randomization of target selection.

Sensory leakage. Because the ganzfeld is itself a perceptual isolation procedure, it goes a long way toward eliminating potential sensory leakage during the ganzfeld portion of the session. There are, however, potential channels of sensory leakage following the ganzfeld period. For example, if the experimenter who interacts with the receiver knows the identity of the target, he or she could bias the receiver's similarity ratings in favor of correct identification. Only one study in the database contained this flaw, a study in which subjects actually performed slightly below chance expectation. Second, if the stimulus set given to the receiver for judging contains the actual physical target handled by the sender during the sending period, there might be cues (e.g., fingerprints, smudges, or temperature differences) that could differentiate the target from the decoys. Moreover, the process of transferring the stimulus materials to the receiver's room itself opens up other potential channels of sensory leakage. Although contemporary ganzfeld studies eliminate both of these possibilities by using duplicate stimulus sets, some of the earlier studies did not.

Independent analyses by Hyman and Honorton agreed that there was no correlation between inadequacies of security against sensory leakage and study outcome. Honorton further reported that if studies that failed to use duplicate stimulus sets were discarded from the analysis, the remaining studies are still highly significant (Stouffer $z = 4.35, p = 6.8 \times 10^{-6}$)

Randomization. In many psi experiments, the issue of target randomization is critical because systematic patterns in inadequately randomized target sequences might be detected by subjects during a session or might match subjects' pre-existing response biases. In a ganzfeld study, however, randomization is a much less critical issue because only one target is selected during the session and most subjects serve in only one session. The primary concern is simply that all targets be sampled about equally over the course of the study. Similar considerations govern the second randomization, which takes place after the ganzfeld period and determines the sequence in which the target and decoys are presented to the receiver (or external judge) for judging.

Nevertheless, Hyman and Honorton disagreed over the findings here. Hyman claimed there was a correlation between flaws of randomization and study outcome; Honorton claimed there was not. The sources of this disagreement were in conflicting definitions of flaw categories, in the coding and assignment of flaw ratings to individual studies, and in the subsequent statistical treatment of those ratings.

Unfortunately, there have been no ratings of flaws by independent raters who were blind to the studies' outcomes (Morris, 1991). Nevertheless, none of the contributors to the subsequent debate concurred with Hyman's conclusion whereas four nonparapsychologists—two statisticians and two psychologists—explicitly concurred with Honorton's conclusion (Harris & Rosenthal, 1988b; Saunders, 1985; Utts, 1991a). For example, Harris and Rosenthal (one of the pioneers in the use of meta-analysis in psychology) used Hyman's own flaw ratings and failed to find any significant relationships between flaws and study outcomes in each of two separate analyses: "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" (1988b, p. 3). (For a more recent exchange over Hyman's analysis, see Hyman (1991), Utts (1991a), and Utts (1991b).)

Effect Size

Some critics of parapsychology have argued that even if current laboratory-produced psi effects turn out to be replicable and non-artifactual, they are too small to be of theoretical interest or practical importance. We do not believe this to be the case for the psi ganzfeld effect.

In psi ganzfeld studies, the hit rate itself provides a straightforward descriptive measure of effect size, but this cannot be compared directly across studies because they do not all use a four-stimulus judging set and, hence, do

[2] A 1980 survey of parapsychologists uncovered only 19 completed but unreported ganzfeld studies. Seven of these had achieved significantly positive results, a proportion (.37) very similar to the proportion of independently significant studies in the meta-analysis (.43) (Blackmore, 1980).

not all have a chance baseline of .25. The next most obvious candidate, the difference in each study between the hit rate observed and the hit rate expected under the null hypothesis, is also intuitively descriptive but is not appropriate for statistical analysis because not all differences between proportions that are equal are equally detectable (e.g., the power to detect the difference between .55 and .25 is different from the power to detect the difference between .50 and .20).

In order to provide a scale of equal detectability, Cohen (1988) devised the effect size index h, which performs an arcsine transformation on the proportions before calculating their difference. Cohen's h is quite general and can assess the difference between any two proportions drawn from independent samples or between a single proportion and any specified hypothetical value. For the 28 studies examined in the meta-analyses, h is .28, with a 95% confidence interval from .11 to .45.

But because values of h do not provide an intuitively descriptive scale, Rosenthal and Rubin (1989; Rosenthal, 1991) have recently suggested a new index, π, which applies specifically to one-sample, multiple-choice data of the kind obtained in ganzfeld experiments. In particular, π expresses all hit rates as the proportion of hits that would have been obtained if there had been only two equally likely alternatives—essentially a coin flip. Thus, π ranges from 0 to 1, with .5 expected under the null hypothesis. The formula is:

$$\pi = \frac{P(k-1)}{P(k-2) + 1}$$

where P is the raw proportion of hits and k is the number of alternative choices available. Because π has such a straightforward intuitive interpretation, we will use it (or its conversion back to an equivalent four-alternative hit rate) throughout this article whenever it is applicable.

For the 28 studies examined in the meta-analyses, the mean value of π is .62, with a 95% confidence interval from .55 to .69. This corresponds to a four-alternative hit rate of 35%, with a 95% confidence interval from 28% to 43%.

Cohen (1988, 1992) has also categorized effect sizes into *small*, *medium*, and *large*, where *medium* denotes an effect size that should be apparent to the naked eye of a careful observer. For a statistic like π, which indexes the deviation of a proportion from .5, Cohen considers .65 to be a medium effect size: A statistically unaided observer should be able to detect the bias of a coin that comes up heads on 65% of the trials. Thus, at .62, the psi ganzfeld effect size falls just short of Cohen's naked-eyeball criterion. From the phenomenology of the ganzfeld experimenter, the corresponding hit rate of 35% implies that he or she will see a subject obtain a hit approximately every third session rather than every fourth.

It is also instructive to compare the psi ganzfeld effect with the results of a recent medical study that sought to determine whether aspirin can prevent heart attacks (Steering Committee of the Physicians' Health Study Research Group, 1988). The study was discontinued after six years because it was already clear that the aspirin treatment was effective ($\chi^2 = 25.01, p < .00001$) and it was considered unethical to keep the control group on placebo medication. The study was widely publicized as a major medical breakthrough. But despite its undisputed reality and practical importance, the size of the aspirin effect is quite small: Taking aspirin reduces the probability of suffering a heart attack by only 0.008. The corresponding effect size (h) is .068—about 1/3 to 1/4 the size of the psi ganzfeld effect (Atkinson et al., 1993, p. 236; Utts, 1991b).

In sum, we believe that the psi ganzfeld effect is large enough to be of both theoretical interest and potential practical importance.

Experimental Correlates of the Psi Ganzfeld Effect

We saw above that the technique of correlating variables with effect sizes across studies can help to assess whether methodological flaws might have produced artifactual positive outcomes. The same technique can be used more affirmatively to explore whether an effect varies systematically with conceptually relevant variations in experimental procedure. The discovery of such correlates can help to establish an effect as genuine, suggest ways of increasing replication rates and effect sizes, and enhance the chances of moving beyond the simple demonstration of an effect to its explanation. This strategy is only heuristic, however. Any correlates discovered must be considered quite tentative, both because they emerge from post hoc exploration and because they necessarily involve comparisons across heterogeneous studies that differ simultaneously on many interrelated variables, known and unknown. Two such correlates emerged from the meta-analyses of the psi ganzfeld effect.

Single versus multiple-image targets. Although most of the 28 studies in the meta-analysis used single pictures as targets, 9 of the studies (conducted by three different investigators) used *View Master* stereoscopic slide reels which presented multiple images focused on a central theme. Studies using the *View Master* reels produced significantly higher hit rates than did studies using the single-image targets (50% vs. 34%), t (26) = 2.22, p = .035, two-tailed.

Sender/receiver pairing. In 17 of the 28 studies, participants were free to bring in friends to serve as senders. In 8 studies, only laboratory-assigned senders were employed. (Three studies used no sender.) Unfortunately, there is no record of how many participants in the former studies actually brought in friends. Nevertheless, those 17 studies (by six different investigators) had significantly higher hit rates than did the studies that used only laboratory-assigned senders (44% vs. 26%), t (23) = 2.39, p = .025, two-tailed.

The Joint Communiqué

Following their published exchange in 1985, Hyman and Honorton agreed to contribute a joint communiqué to the subsequent discussion which was published in 1986. First they set forth their areas of agreement and disagreement:

> We agree that there is an overall significant effect in this data base that cannot reasonably be explained by selective reporting or multiple analysis. We continue to differ over the degree to which the effect constitutes evidence for psi, but we agree that the final verdict awaits the outcome of future experiments conducted by a broader range of investigators and according to more stringent standards. (Hyman & Honorton, 1986, abstract, p. 351)

They then spelled out in detail the "more stringent standards" they believed should govern future experiments. These include strict security precautions against sensory leakage, testing and documentation of randomization methods for selecting targets and sequencing the judging pool, statistical correction for multiple analyses, advance specification of the status of the experiment (e.g., pilot study, confirmatory experiment), and full documentation in the published report of the experimental procedures and the status of statistical tests (e.g., pre-planned or post hoc).

The NRC Report

In 1988, the National Research Council (NRC) of the National Academy of Sciences released a widely publicized report commissioned by the U. S. Army which assessed several controversial technologies for enhancing human performance, including accelerated learning, neurolinguistic programming, mental practice, biofeedback, and parapsychology (Druckman & Swets, 1988; summarized in Swets & Bjork, 1990). The report's conclusion concerning parapsychology was quite negative: "The Committee finds no scientific justification from research conducted over a period of 130 years for the existence of parapsychological phenomena" (p. 22).

An extended refutation strongly protesting the Committee's treatment of parapsychology has been published elsewhere (Palmer et al., 1989). The pertinent point here is simply that the NRC's evaluation of the ganzfeld studies does not reflect an additional, independent examination of the ganzfeld database but is based on the same meta-analysis by Hyman that we have discussed in this article.

Hyman chaired the NRC's Subcommittee on Parapsychology; and, although he had concurred with Honorton two years earlier in their joint communiqué that "there is an overall significant effect in this data base that cannot reasonably be explained by selective reporting or multiple analysis" (p. 351) and that "significant outcomes have been produced by a number of different investigators" (p. 352), neither of these points is acknowledged in the Committee's report.

The NRC also solicited a background paper from Harris and Rosenthal (1988a), who provided the Committee with a comparative methodological analysis of the five controversial areas listed above. Harris and Rosenthal noted that of these areas, "only the Ganzfeld ESP studies [the only psi studies they evaluated] regularly meet the basic requirements of sound experimental design" (p. 53), and they concluded that "it would be implausible to entertain the null given the combined p from these 28 studies. Given the various problems or flaws pointed out by Hyman and Honorton...we might estimate the obtained accuracy rate to be about 1/3...when the accuracy rate expected under the null is 1/4" (p. 51).[3]

The Autoganzfeld Studies

In 1983, Honorton and his colleagues initiated a new series of ganzfeld studies designed to avoid the methodological problems he and others had identified in earlier studies (Honorton, 1979; Kennedy, 1979). These studies complied with all the detailed guidelines that he and Hyman were to publish later in their joint communiqué. The program continued until September of 1989, when a loss of funding forced the laboratory to close.

The major innovations of the new studies were the computer control of the experimental protocol—hence the name "autoganzfeld"—and the introduction of videotaped film clips as target stimuli.

Method[4]

The basic design of the autoganzfeld studies was the same as that described earlier: A receiver and sender were sequestered in separate, acoustically-isolated chambers. Following a 14-minute period of progressive relaxation, the receiver underwent ganzfeld stimulation while describing his or her thoughts and images aloud for 30 minutes. Meanwhile, the sender concentrated on a randomly selected target. At the end of the ganzfeld period, the receiver was shown four stimuli and, without knowing which of the four had been the target, rated each stimulus for its similarity to his or her mentation during the ganzfeld.

The targets consisted of 80 still pictures (*Static Targets*) and 80 short video segments complete with soundtracks (*Dynamic Targets*), all recorded on videocassette. The static targets included art prints, photographs, and magazine advertisements; the dynamic targets included excerpts of approximately one minute duration from motion pictures, TV shows, and cartoons. The 160 targets were arranged in judging sets of four static or four dynamic targets each, constructed to minimize similarities among targets within a set.

Target selection and presentation. The VCR containing the taped targets was interfaced to the controlling computer, which selected the target and controlled its repeated presentation to the sender during the ganzfeld period, thus eliminating the need for a second experimenter to accompany the sender. Following the ganzfeld period, the computer randomly sequenced the four-clip judging pool and presented it to the receiver on a TV monitor for judging. The receiver used a computer game paddle to make his or her ratings on a 40-point scale which appeared on the TV monitor after each clip was shown. The receiver was permitted to see each clip and to change the ratings repeatedly until he or she was satisfied. The computer then wrote these and other data from the session into a file on a floppy disk. At that point, the sender moved to the receiver's chamber and revealed the identity of the target to both the receiver and the experimenter. Note that the experimenter did not even know the identity of the four-clip judging pool until it was displayed to the receiver for judging.

[3]In a troubling development, the chair of the NRC Committee phoned Rosenthal and asked him to delete the parapsychology section of the paper (R. Rosenthal, private communication, September 15, 1992). Although Rosenthal refused to do so, that section of the Harris-Rosenthal paper is nowhere cited in the NRC report.

[4]Because Honorton and his colleagues have complied with the Hyman-Honorton specification that experimental reports be sufficiently complete to permit others to reconstruct the investigator's procedures, readers who wish to know more detail than we provide here are likely to find whatever they need in the archival publication of these studies in the *Journal of Parapsychology* (Honorton et al., 1990).

Randomization. The random selection of the target and sequencing of the judging pool were controlled by a noise-based random number generator interfaced to the computer. Extensive testing confirmed that the generator was providing a uniform distribution of values throughout the full target range (1–160), a uniform distribution of targets from among the four alternatives within each judging set, and a uniform distribution of judging sequences from among the 24 permutations of 4 stimuli.

Additional control features. Both the receiver's and sender's rooms were sound-isolated, electrically-shielded chambers with single-door access that could be continuously monitored by the experimenter. There was two-way intercom communication between the experimenter and the receiver but only one-way communication into the sender's room; thus neither the experimenter nor the receiver could monitor events inside the sender's room. The archival record for each session includes an audio tape containing the receiver's mentation during the ganzfeld period and all verbal exchanges between the experimenter and the receiver throughout the experiment.

The automated ganzfeld protocol has been examined by several dozen parapsychologists and behavioral researchers from other fields, including well-known critics of parapsychology. Many have participated as subjects or observers. All have expressed satisfaction with the handling of security issues and controls.

Parapsychologists have often been urged to employ magicians as consultants to ensure that the experimental protocols are not vulnerable either to inadvertent sensory leakage or to deliberate cheating. Two "mentalists," magicians who specialize in the simulation of psi, have examined the autoganzfeld system and protocol. Ford Kross, a professional mentalist and officer of the mentalist's professional organization, the Psychic Entertainers Association, provided the following written statement "In my professional capacity as a mentalist, I have reviewed Psychophysical Research Laboratories' automated ganzfeld system and found it to provide excellent security against deception by subjects" (private communication, May, 1989).

The first author of this article has also performed as a mentalist for many years and is a member of the Psychic Entertainers Association. As noted in the author footnote, this article has its origins in a 1983 visit he made to Honorton's laboratory, where he was asked to critically examine the research protocol from the perspective of a mentalist, a research psychologist, and a subject. Needless to say, this article would not exist if he did not concur with Ford Kross's assessment of the security procedures.

Experimental Studies

Altogether 100 men and 141 women participated as receivers in 355 sessions during the research program. The participants ranged in age from 17 to 74 years (mean = 37.3, *SD* = 11.8), with a mean formal education of 15.6 years (*SD* = 2.0). Eight separate experimenters, including Honorton, conducted the studies.

The experimental program included three pilot and eight formal studies. Five of the formal studies employed novice (first-time) participants who served as the receiver in one session each. The remaining three formal studies employed experienced participants.

Pilot Studies. Sample sizes were not preset in the three pilot studies. Study 1 comprised 22 sessions and was conducted during the initial development and testing of the autoganzfeld system. Study 2 comprised 9 sessions testing a procedure in which the experimenter, rather than the receiver, served as the judge at the end of the session. Study 3 comprised 36 sessions and served as practice for participants who had completed the allotted number of sessions in the ongoing formal studies but who wanted additional ganzfeld experience. This study also included several demonstration sessions when TV film crews were present.

Novice Studies. Studies 101–104 were each designed to test 50 participants who had had no prior ganzfeld experience; each participant served as the receiver in a single ganzfeld session. Study 104 included 16 of 20 students recruited from the Juilliard School of Music in New York City in order to test an artistically gifted sample. Study 105 was initiated to accommodate the overflow of participants who had been recruited for Study 104, including the four remaining Juilliard students. Sample size for this study was set to 25, but only 6 sessions had been completed when the laboratory closed. For purposes of exposition, we have divided the 56 sessions from Studies 104 and 105 into two parts: Study 104/105 (a) comprises the 36 non-Juilliard participants; Study 104/105 (b) comprises the 20 Juilliard students

Study 201. This study was designed to retest the most promising participants from the previous studies. The number of trials was set to 20, but only 7 sessions with 3 participants had been completed when the laboratory closed.

Study 301. This study was designed to compare static and dynamic targets. Sample size was set to 50 sessions. Twenty-five experienced participants each served as the receiver in two sessions. Unknown to the participants, the computer control program was modified to ensure that they would each have one session with a static target and one session with a dynamic target.

Study 302. This study was designed to examine a dynamic target set which contained one target that had often been correctly identified in the previous studies and another target that had never been correctly identified. The study involved experienced participants who had had no prior experience with this particular target set and who were unaware that only one target set was being sampled. Each served as the receiver in a single session. The design called for the study to continue until 15 sessions were completed with each of the targets, but only twenty-five sessions had been completed when the laboratory closed.

The 11 studies just described comprise all sessions conducted during the 6–1/2 years of the program. There is no "file drawer" of unreported sessions.

Results

Overall hit rate. As in the earlier meta-analysis, receivers' ratings were analyzed by tallying the proportion of hits achieved and calculating the exact binomial probability for the observed number of hits compared with chance expectation of .25. As noted above, 241 participants contributed 355 sessions. For reasons discussed below, Study 302 is analyzed separately, reducing the number of sessions in the primary analysis to 330.

As Table 1 shows, there were 106 hits in the 330 sessions, a hit rate of 32% ($z = 2.85$, $p = .002$, one-tailed), with a 95% confidence interval from 30% to 35%. This cor-

Table 1
Outcome by Study

Study	Study type	*N* subjects	*N* trials	*N* hits	% hits	Effect size π	Effect size *z*
1	Pilot	19	22	8	36	.62	.99
2	Pilot	4	9	3	33	.60	.25
3	Pilot	25	36	10	28	.54	.22
101	Novice	50	50	12	24	.47	−.30
102	Novice	50	50	18	36	.63	1.60
103	Novice	50	50	15	30	.55	.67
104/105 (a)	Novice	36	36	12	33	.60	.97
104/105 (b)	Juilliard Sample	20	20	10	50	.75	2.20
201	Experienced	3	7	3	43	.69	.69
301	Experienced	25	50	15	30	.56	.67
302	Experienced	25	25	16	44[a]	.70[a]	2.02[a]
	Overall (Studies 1–301)	241	330	106	32	.59	2.85

Note. *Z* scores are based on the exact binomial probability with $p = .25$ and $q = .75$.
[a]Corrected for maximum possible response bias. Hit rate actually observed was 64%.

responds to an effect size (π) of .59, with a 95% confidence interval from .53 to .64.

Table 1 also shows that when Studies 104 and 105 are combined and re-divided into the non-Juilliard and Juilliard samples, 9 of the 10 studies yield positive effect sizes, with a mean effect size (π) of .61, t (9) = 4.35, p = .0009, one-tailed. This effect size is equivalent to a four-alternative hit rate of 34%. Alternatively, if we retain Studies 104 and 105 as separate studies, 9 of the 10 studies again yield positive effect sizes, with a mean effect size (π) of .62, t (9) = 3.67, p = .003, one-tailed. This effect size is equivalent to a four-alternative hit rate of 35% and is identical to that found across the 28 studies of the earlier meta-analysis.[5]

Considered together, sessions with novice participants (Studies 101–105) yielded a statistically significant hit rate of 32.5% (p = .009), which is not significantly different from the 31.6% hit rate achieved by experienced participants in Studies 201 and 301. And finally, each of the 8 experimenters also achieved a positive effect size, with a mean π of .60, t (7) = 3.44, p = .005, one-tailed.

[5]As noted above, the laboratory was forced to close before three of the formal studies could be completed. If we assume that the remaining trials in Studies 105 and 201 would have yielded only chance results, this would reduce the overall z for the first 10 autoganzfeld studies from 2.85 to 2.73 (p = .003). Thus, inclusion of the two incomplete studies does not pose an optional stopping problem. The third incomplete study, Study 302, is discussed below.

The Juilliard sample. There are several reports in the literature of a relationship between creativity or artistic ability and psi performance (Schmeidler, 1988). In order to explore this possibility in the ganzfeld setting, 10 male and 10 female undergraduates were recruited from the Juilliard School of Music in New York City. Of these, 8 were music students, 10 were drama students, and 2 were dance students. Each served as the receiver in a single session in Studies 104 or 105. As shown in Table 1, these students achieved a hit rate of 50% (p = .014), one of the five highest hit rates ever reported for a single sample in a ganzfeld study. The musicians were particularly successful: Six of the eight (75%) successfully identified their targets (p = .004). Further details about this sample and their ganzfeld performance are reported in Schlitz and Honorton (1992).

Study size and effect size. There is a significant negative correlation across the 10 studies listed in Table 1 between the number of sessions in a study and its effect size (π): r = −.64, t (8) = 2.36, $p < .05$, two-tailed. This is reminiscent of Hyman's discovery that the smaller studies in the original ganzfeld database were disproportionately likely to report statistically significant results. He interpreted this finding as evidence for a bias against the reporting of small studies that fail to achieve significant results. A similar interpretation cannot be applied to the autoganzfeld studies, however, because there are no unreported sessions.

One reviewer of this article suggested that the negative correlation might reflect a decline effect in which earlier sessions of a study are more successful than later ses-

Table 2
Study 302: Proportion of Sessions in which Each Video Clip was Ranked First when it was a Target and when it was a Decoy

Video Clip	Relative Frequency as Target	Ranked First when Target	Ranked First when Decoy	Difference	Fisher's Exact *p*
Tidal Wave	.28 (7/25)	.57 (4/7)	.11 (2/18)	+.46	.032
Snakes	.12 (3/25)	.67 (2/3)	.05 (1/22)	+.62	.029
High-Speed Sex	.16 (4/25)	.25 (1/4)	.05 (1/21)	+.20	.300
Bugs Bunny	.44 (11/25)	.82 (9/11)	.36 (5/14)	+.46	.027
	Means	.58	.14	+.44	–

sions. If there were such an effect, then studies with fewer sessions would show larger effect sizes because they would end before a decline could set in. To check this possibility, we computed point-biserial correlations between hits (= 1) or misses (= 0) and the session number within each of the 10 studies. All of the correlations hovered around zero; six were positive, four were negative, and the overall mean was +.03, indicating that hits were actually slightly more likely to occur in later sessions of a study than in earlier ones.

An inspection of Table 1 reveals that the negative correlation derives primarily from the two studies with the largest effect sizes: the 20 sessions with the Juilliard students and the 7 sessions of Study 201, the study specifically designed to retest the most promising participants from the previous studies. Accordingly, we believe that the larger effect sizes of these two studies—and hence the significant negative correlation between number of sessions and effect size—reflects a genuinely higher psi ability of participants in these two small but highly selected samples.

Study 302. All of the studies except Study 302 randomly sampled from a pool of 160 static and dynamic targets. Study 302 sampled from a single, dynamic target set which contained one video clip that had been correctly identified several times in the previous ten studies—a scene of a tidal wave from the movie *Clash of the Titans*—and one clip that had never been correctly identified—a high-speed sex scene from *Clockwork Orange.* The set also contained a scene of crawling snakes from a TV documentary and a scene from a Bugs Bunny cartoon.

The experimental design called for this study to continue until each of the clips had served as the target 15 times. Unfortunately, the premature termination of this study at 25 sessions left an imbalance in the frequency with which each clip had served as the target. This means that the chance expectation of .25 used as the baseline for evaluating the results in the other studies cannot be applied to this study. More importantly, it means that the high hit rate observed (64%) could well be inflated by response biases.

To illustrate, water imagery is frequently reported by receivers in ganzfeld sessions whereas sexual imagery is rarely reported. (It is reasonable to suppose that some participants might be reluctant both to report sexual imagery and to give the highest rating to the sex-related clip.) If a video clip containing popular imagery (like water) happens to appear as a target more frequently than a clip containing unpopular imagery (like sex), a high hit rate might simply reflect the coincidence of those frequencies of occurrence with participants' response biases. And, as the first column of Table 2 reveals, the tidal wave clip did in fact appear more frequently as the target than did the sex clip. The Bugs Bunny clip was even more frequent, appearing as the target in 11 of the 25 (44%) sessions.

We can assess the damage in two ways. First, we can ask for each of the four clips whether it was selected as the target (i.e., ranked in first place) significantly more frequently when it was the target than when it was one of the three control clips (decoys). This comparison, which controls for the baseline popularity of the themes and images within each clip, is shown in the remaining columns of Table 2.

As can be seen, each of the four clips was selected as the target relatively more frequently when it was the target than when it was a decoy, a difference that is significant for three of the four clips. On average, a clip was identified as the target 58% of the time when it was the target compared with only 14% of the time when it was a decoy. In other words, there is an effect over and above that produced by response bias.

Next, we can calculate a conservative estimate of that remaining effect by imagining a hypothetical set of informed receivers who were able to maximally exploit the unequal target frequencies to achieve the highest possible hit rate. Their optimal (non-psi) strategy would be to iden-

tify the most frequent clip (Bugs Bunny) as the target in every session, thereby achieving a hit rate of 44%. This figure thus represents the hit rate expected if response biases were making their maximum possible contribution to the hit rate—a worst case baseline for evaluating any remaining psi effect. When the observed hit rate of 64% is compared with a hypothetical hit rate of 44%, the effect size (h) is .40. As shown in Table 1, this is equivalent to a four-alternative hit rate of (coincidentally) 44% or a π value of .70 and is statistically significant (z = 2.02, p = .022).

Dynamic versus static targets. The success of Study 302 raises the question of whether dynamic targets are, in general, more effective than static targets. This possibility was also suggested by the earlier meta-analysis, which found that studies using multiple-image targets (*View Master* stereoscopic slide reels) obtained significantly higher hit rates than did studies using single-image targets. By adding motion and sound, the video clips might be thought of as high-tech versions of the *View Master* reels.

The ten autoganzfeld studies that randomly sampled from both dynamic and static target pools yielded an even split of 165 sessions with each target type. As predicted, sessions using dynamic targets yielded significantly more hits than did sessions using static targets (37% vs. 27%), Fisher's exact $p < .04$.

Sender/receiver pairing. The earlier meta-analysis found that studies in which participants were free to bring in friends to serve as senders produced significantly higher hit rates than studies that used only laboratory-assigned senders. As noted, however, there is no record of how many of the participants in the former studies actually did bring in friends.

Whatever the case, sender/receiver pairing was not a significant correlate of psi performance in the autoganzfeld studies: The 198 sessions in which the sender and receiver were friends did not yield a significantly higher proportion of hits than did the 132 sessions in which they were not (34% vs. 29%), Fisher's exact p = .34.

Correlations between receiver characteristics and psi performance. Most of the autoganzfeld participants were strong believers in psi: On a 7-point scale, where "1" indicates strong disbelief and "7" indicates strong belief in psi, the mean was 6.2 (SD = 1.03); only two participants rated their belief in psi below the midpoint of the scale. In addition, 88% of the participants reported personal experiences suggestive of psi and 80% had some training in meditation or other techniques involving internal focus of attention.

All of these appear to be important variables. The correlation between belief in psi and psi performance is one of the most consistent findings in the parapsychological literature (Palmer, 1978). And within the autoganzfeld studies, successful performance of novice (first-time) participants is significantly predicted by reported personal psi experiences, involvement with meditation or other mental disciplines, and high scores on the Feeling and Perception factors of the Myers-Briggs Type Inventory (Honorton, 1992; Honorton & Schechter, 1987). This recipe for success has now been independently replicated in another laboratory (Broughton, Kanthamani, & Khilji, 1990).

The personality trait of extraversion is also associated with better psi performance. A meta-analysis of 60 independent studies with nearly 3,000 subjects revealed a small but reliable positive correlation between extraversion and psi performance—especially in studies that used free-response methods of the kind employed in the ganzfeld experiments (Honorton, Ferrari, & Bem, 1992). Across 14 free-response studies by 4 independent investigators, the correlation for 612 subjects was .20 (z = 4.82, $p = 1.5 \times 10^{-6}$). This correlation was replicated in the autoganzfeld studies, where extraversion scores were available for 221 of the 241 subjects: r = .18, t (219) = 2.67, p = .004, one-tailed.

And finally, there is the strong psi performance of the Juilliard students, discussed above, which is consistent with other studies in the parapsychological literature suggesting a relationship between successful psi performance and creativity or artistic ability.

Discussion

Earlier in this article we quoted from the abstract of the Hyman-Honorton communiqué: "We agree that the final verdict awaits the outcome of future experiments conducted by a broader range of investigators and according to more stringent standards" (1986, p. 351).

We believe that the "stringent standards" requirement has been met by the autoganzfeld studies. The results are statistically significant and consistent with those in the earlier database. The mean effect size is quite respectable when compared with other controversial research areas of human performance (Harris & Rosenthal, 1988a). And, there are reliable relationships between successful psi performance and conceptually relevant experimental and subject variables—relationships that also replicate previous findings. Hyman has also commented on the autoganzfeld studies:

> Honorton's experiments have produced intriguing results. If...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. (1991, p. 392)

Issues of Replication

As Hyman's comment implies, the autoganzfeld studies by themselves cannot satisfy the requirement that replications be conducted by a "broader range of investigators." Accordingly, we hope the findings reported here will be sufficiently provocative to prompt others to try replicating the psi ganzfeld effect.

We believe it is essential, however, that future studies comply with the methodological, statistical, and reporting standards set forth in the joint communiqué and achieved by the autoganzfeld studies. It is not necessary for studies to be as automated or as heavily instrumented as the autoganzfeld studies in order to satisfy the methodological guidelines, but they are still likely to be labor intensive and potentially expensive.[6]

[6]As the closing of the autoganzfeld laboratory exemplifies, it is also difficult to obtain funding for psi research. The traditional, peer-refereed sources of funding familiar to psychologists have almost never funded proposals for psi research. The widespread skepticism of psychologists toward psi is almost certainly a contributing factor.

Statistical Power and Replication

Would-be replicators also need to be reminded of the power requirements for replicating small effects. Although many academic psychologists do not believe in psi, many apparently do believe in miracles when it comes to replication. Tversky and Kahneman (1971) posed the following problem to their colleagues at meetings of the Mathematical Psychology Group and the American Psychological Association:

> Suppose you have run an experiment on 20 subjects and have obtained a significant result which confirms your theory ($z = 2.23$, $p < .05$, two-tailed). You now have cause to run an additional group of 10 subjects. What do you think the probability is that the results will be significant, by a one-tailed test, separately for this group? (p. 105)

The median estimate was .85, with nine out of ten respondents giving an estimate greater than .60. The correct answer is approximately .48.

As Rosenthal (1990) has warned: "Given the levels of statistical power at which we normally operate, we have no right to expect the proportion of significant results that we typically do expect, even if in nature there is a very real and very important effect" (p. 16).

In this regard, it is again instructive to consider the medical study that found a highly significant effect of aspirin on the incidence of heart attacks. The study monitored over 22,000 subjects. Had the investigators monitored 3,000 subjects, they would have had less than an even chance of finding a conventionally significant effect. Such is life with small effect sizes.

Given its larger effect size, the prospects for successfully replicating the psi ganzfeld effect are not quite so daunting, but they are probably still grimmer than intuition would suggest. If the true hit rate is in fact about 34% when 25% is expected by chance, then an experiment with 30 trials (the mean for the 28 studies in the original meta-analysis) has only about 1 chance in 6 of finding an effect significant at the .05 level with a one-tailed test. A 50 trial experiment boosts that to about 1 in 3. One must escalate to 100 trials in order to come close to the break even point—where one has a 50–50 chance of finding a statistically significant effect (Utts, 1986). (Recall that only 2 of the 11 autoganzfeld studies yielded results that were individually significant at the conventional .05 level.) Those who require that a psi effect be statistically significant every time before they will seriously entertain the possibility that an effect really exists know not what they ask.

Significance vs. Effect Size

But the preceding discussion is unduly pessimistic because it perpetuates the tradition of worshipping the significance level. Regular readers of this journal are likely to be familiar with recent arguments imploring behavioral scientists to overcome their slavish dependence on the significance level as the ultimate measure of virtue and to focus more of their attention on effect sizes instead: "Surely, God loves the .06 nearly as much as the .05" (Rosnow & Rosenthal, 1989, p. 1277). Accordingly, we suggest that achieving a respectable effect size with a methodologically tight ganzfeld study would be a perfectly welcome contribution to the replication effort—no matter how untenurable the p level renders the investigator.

Career consequences aside, this suggestion may seem quite counterintuitive. Again, Tversky and Kahneman (1971) provide an elegant demonstration. They asked several of their colleagues to consider an investigator who runs 15 subjects and obtains a significant t value of 2.46. Another investigator attempts to duplicate the procedure with the same number of subjects and obtains a result in the same direction but with a nonsignificant value of t. Tversky and Kahneman then asked their colleagues to indicate the highest level of t in the replication study they would describe as a failure to replicate. The majority of their colleagues regarded $t = 1.70$ as a failure to replicate. But if the data from two such studies ($t = 2.46$ and $t = 1.70$) were pooled, the t for the combined data would be about 3.00 (assuming equal variances):

> Thus, we are faced with a paradoxical state of affairs, in which the same data that would increase our confidence in the finding when viewed as part of the original study, shake our confidence when viewed as an independent study. (p. 108)

Such is the iron grip of the arbitrary .05. Pooling the data, of course, is what meta-analysis is all about. Accordingly, we suggest that two or more laboratories could collaborate in a ganzfeld replication effort by conducting independent studies and then pooling them in meta-analytic fashion—what we might call real-time meta-analysis. (Each investigator could then claim the pooled p level for his or her own curriculum vitae.)

Maximizing Effect Size

Rather than buying or borrowing larger sample sizes, those who seek to replicate the psi ganzfeld effect might find it more intellectually satisfying to attempt to maximize the effect size by attending to the variables associated with successful outcomes. Thus researchers who wish to enhance the chances of successful replication should use dynamic rather than static targets. Similarly we advise using participants with the characteristics we have reported to be correlated with successful psi performance. Random college sophomores enrolled in introductory psychology do not constitute the optimal subject pool.

And finally, we urge, ganzfeld researchers to read carefully the detailed description of the warm social ambiance that Honorton et al. (1990) sought to create in the autoganzfeld laboratory. We believe that the social climate created in psi experiments is a critical determinant of their success or failure.

The Problem of "Other" Variables

This caveat about the social climate of the ganzfeld experiment prompted one reviewer of this article to worry that this provided "an escape clause" which weakens the falsifiability of the psi hypothesis: "Until Bem and Honorton can provide operational criteria for creating a warm social ambiance, the failure of an experiment with otherwise adequate power can always be dismissed as due to a lack of warmth."

Alas, it's true; we devoutly wish it were otherwise. But the operation of unknown variables in moderating the success of replications is a fact of life in all the sciences. Consider, for example, an earlier article in this journal by Spence (1964). He reviewed studies testing the straightforward derivation from Hullian learning theory that high anxiety subjects should condition more

strongly than low anxiety subjects. This hypothesis was confirmed 94% of the time in Spence's own laboratory at the University of Iowa but only 63% of the time in laboratories at other universities. In fact, Kimble and his associates at Duke and North Carolina obtained results in the opposite direction in two out of three experiments.

In searching for a post hoc explanation, Spence noted that "a deliberate attempt was made in the Iowa studies to provide conditions in the laboratory that might elicit some degree of emotionality. Thus, the experimenter was instructed to be impersonal and quite formal ... and did not try to put [subjects] at ease or allay any expressed fears." Moreover, he pointed out, his subjects sat in a dental chair whereas Kimble's subjects sat in a secretarial chair. Spence even considered "the possibility that cultural backgrounds of southern and northern students may lead to a difference in the manner in which they respond to the different items in the [Manifest Anxiety] scale."

If this was the state of affairs in an area of research as well established as classical conditioning, then the suggestion that the social climate of the psi laboratory might affect the outcome of ganzfeld experiments in ways not yet completely understood should not be dismissed as a devious attempt to provide an "escape clause" in case of replication failure.

The best the original researcher can do is provide as complete a description of the experimental conditions as possible in an attempt to anticipate what some of the relevant moderating variables might be. The detailed description of the autoganzfeld procedures provided by Honorton et al. (1990) comes as close as current knowledge permits in providing the "operational criteria for creating a warm social ambiance."

Theoretical Considerations

Up to this point, we have confined our discussion to strictly empirical matters. We are sympathetic to the view that one should establish the existence of a phenomenon, anomalous or not, before attempting to explain it.

So let us suppose for the moment that it's true. Let us suppose that we have a genuine anomaly of information transfer before us. How can it be understood or explained?

The Psychology of Psi

In attempting to understand psi, parapsychologists have typically begun with the working assumption that, whatever its underlying mechanisms, it should behave like other, more familiar psychological phenomena. In particular, they typically assume that target information behaves like an external sensory stimulus that is encoded, processed, and experienced in familiar information-processing ways. Similarly, individual psi performances should covary with experimental and subject variables in psychologically sensible ways. These assumptions are embodied in the model of psi that motivated the ganzfeld studies in the first place.

The ganzfeld procedure. As noted in the introduction, the ganzfeld procedure was designed to test a model in which psi-mediated information is conceptualized as a weak signal that is normally masked by internal somatic and external sensory "noise." Accordingly, any technique that raises the signal-to-noise ratio should enhance a person's ability to detect the psi-mediated information. This noise-reduction model of psi organizes a large and diverse body of experimental results, particularly those demonstrating the psi-conducive properties of altered states of consciousness like meditation, hypnosis, dreaming, and, of course, the ganzfeld itself (Rao & Palmer, 1987).

Alternative theories propose that the ganzfeld (and altered states) may be psi-conducive because it lowers resistance to accepting alien imagery, diminishes rational or contextual constraints on the encoding or reporting of information, stimulates more divergent thinking, or even just serves as a placebo-like ritual that participants perceive as being psi conducive (Stanford, 1987). At this point, there are no data that would permit us to choose among these alternatives, and the noise-reduction model remains the most widely accepted.

The target. There are also a number of plausible hypotheses that attempt to account for the superiority of dynamic targets over static targets: Dynamic targets contain more information, involve more sensory modalities, evoke more of the receiver's internal schemata, are more lifelike, have a narrative structure, are more emotionally evocative, and are "richer" in other, unspecified ways. Several psi researchers have attempted to go beyond the simple dynamic/static dichotomy to more refined or theory-based definitions of a good target. Although these efforts have involved examining both psychological and physical properties of targets, there is not much progress to report yet (Delanoy, 1990).

The receiver. Some of the subject characteristics associated with good psi performance also appear to have psychologically straightforward explanations. For example, garden-variety motivational explanations seem sufficient to account for the relatively consistent finding that those who believe in psi perform significantly better than those who do not. (Less straightforward, however, would be an explanation for the frequent finding that nonbelievers actually perform significantly worse than chance (Broughton, 1991, p. 109).)

The superior psi performance of creative or artistically gifted individuals—like the Juilliard students—may reflect individual differences that parallel some of the hypothesized effects of the ganzfeld, mentioned above: Artistically gifted individuals may be more receptive to alien imagery, be better able to transcend rational or contextual constraints on the encoding or reporting of information, or be more divergent in their thinking. It has also been suggested that both artistic and psi abilities might be rooted in superior right-brain functioning.

The observed relationship between extraversion and psi performance has been of theoretical interest for many years. Eysenck (1966) reasoned that extraverts should perform well in psi tasks because they are easily bored and respond favorably to novel stimuli. In a setting like the ganzfeld, extraverts may become "stimulus starved" and thus be highly sensitive to any stimulation, including weak incoming psi information. In contrast, introverts would be more inclined to entertain themselves with their own thoughts and thus continue to mask psi information despite the diminished sensory input. Eysenck also speculated that psi might be a primitive form of perception antedating cortical developments in the course of evolution, and, hence, cortical arousal might suppress psi functioning. Because extraverts have a lower level of cortical arousal than introverts, they should perform better in psi tasks. (The evolutionary biology of psi is also discussed by Broughton (1991, pp. 347–352).)

But there are more mundane possibilities. Extraverts might perform better than introverts simply because they are more relaxed and comfortable in the social setting of the typical psi experiment (e.g., the "warm social ambiance" of the autoganzfeld studies). This interpretation is strengthened by the observation that introverts outperformed extraverts in a study in which subjects had no contact with an experimenter but worked alone at home with materials they received in the mail (Schmidt & Schlitz, 1989). In order to help decide among these interpretations, ganzfeld experimenters have begun to use the extraversion scale of the NEO personality inventory (Costa & McCrae, 1985), which assesses six different facets of the extraversion-introversion factor.

The sender. In contrast to all this information about the receiver in psi experiments, virtually nothing is known about the characteristics of a good sender or about the effects of the sender's relationship to the receiver. As we have seen, the initial suggestion from the meta-analysis of the original ganzfeld database that psi performance might be enhanced when the sender and receiver are friends was not replicated at a statistically significant level in the autoganzfeld studies.

A number of parapsychologists have entertained the more radical hypothesis that the sender may not even be a necessary element in the psi process. In the terminology of parapsychology, the sender-receiver procedure tests for the existence of *telepathy*, anomalous communication between two individuals; but if the receiver is somehow picking up the information from the target itself, it would be termed *clairvoyance*, and the presence of the sender would be irrelevant (except for possible psychological reasons like expectation effects).

At the time of his death, Honorton was planning a series of autoganzfeld studies that would systematically compare sender with no-sender conditions while keeping both the receiver and the experimenter blind to the condition of the ongoing session. In preparation, he conducted a meta-analytic review of ganzfeld studies that used no sender. He found 12 studies with a median of 33.5 sessions conducted by 7 investigators. The overall effect size (π) was .56, which corresponds to a four-alternative hit rate of 29%. But this effect size does not reach statistical significance (Stouffer $z = 1.31$, $p = .095$). So far, then, there is no firm evidence for psi in the ganzfeld in the absence of a sender. (There are, however, non-ganzfeld studies in the literature that do report significant evidence for clairvoyance, including a classic card-guessing experiment by Rhine (Rhine & Pratt, 1954).)

The Physics of Psi

The psychological level of theorizing discussed above does not, of course, address the conundrum that makes psi phenomena anomalous in the first place: their presumed incompatibility with our current conceptual model of physical reality. Parapsychologists differ widely from one another in their taste for theorizing at this level, but several whose training lies in physics or engineering have proposed physical (or biophysical) theories of psi phenomena. (An extensive review of theoretical parapsychology is provided by Stokes (1987).) Only some of these would force a radical revision in our conception of physical reality.

Those who follow contemporary debates in modern physics, however, will be aware that several phenomena predicted by quantum theory and confirmed by experiment are themselves incompatible with our current conceptual model of physical reality. Of these, it is the 1982 empirical confirmation of Bell's theorem that has created the most excitement and controversy among philosophers and the few physicists who are willing to speculate on such matters (Cushing & McMullin, 1989; Herbert, 1987). In brief, Bell's theorem states that any model of reality that is compatible with quantum mechanics must be *nonlocal*. This implies, among other things, that any model of reality compatible with quantum mechanics must allow for the possibility that information about an event at one location can be instantaneously available at some other arbitrarily distant location, unattenuated and without the mediation of any kind of transmitting signal (Herbert, 1987).

Several possible models of reality that incorporate nonlocality have been proposed by both philosophers and physicists. Some of these clearly rule out psi-like information transfer; others permit it; and some actually require it. Thus, at a grander level of theorizing, some parapsychologists believe that one of the more radical models of reality compatible with both quantum mechanics and psi will eventually come to be accepted. If and when that occurs, psi phenomena would cease to be anomalous.

But we have learned that all such talk provokes most of our colleagues in psychology and in physics to roll their eyes and gnash their teeth. So let's just leave it at that.

More generally, we have learned that our colleagues' tolerance for *any* kind of theorizing about psi is strongly determined by the degree to which they have been convinced by the data that psi has been demonstrated. We have further learned that their diverse reactions to the data themselves are strongly determined by their a priori beliefs about and attitudes toward a number of quite general issues, some scientific, some not. In fact, several statisticians believe that the traditional hypothesis testing methods used in the behavioral sciences should be discarded in favor of Bayesian analyses, which take into account a person's a priori beliefs about the phenomenon under investigation (e.g., Bayarri & Berger, 1991; Dawson, 1991).

But in the final analysis, we suspect that both one's Bayesian a prioris *and* one's reactions to the data are ultimately determined by whether one was more severely punished in childhood for Type I or Type II errors.

References

Atkinson, R., Atkinson, R. C., Smith, E. E., & Bem, D. J. (1990). *Introduction to psychology* (10th ed.). San Diego: Harcourt Brace Jovanovich.

Atkinson, R., Atkinson, R. C., Smith, E. E., & Bem, D. J. (1993). *Introduction to psychology* (11th ed.). San Diego: Harcourt Brace Jovanovich.

Avant, L. L. (1965). Vision in the ganzfeld. *Psychological Bulletin, 64*, 246-258.

Bayarri, M. J., & Berger, J. (1991). Comment. *Statistical Science, 6*, 379-382.

Blackmore, S. (1980). The extent of selective reporting of ESP Ganzfeld studies. *European Journal of Parapsychology, 3*, 213-219.

Bozarth, J. D., & Roberts, R. R. (1972). Signifying significant significance. *American Psychologist, 27*, 774-775.

Braud, W. G., Wood, R., & Braud, L. W. (1975). Free-response GESP performance during an experimental

hypnagogic state induced by visual and acoustic ganzfeld techniques. A Replication and extension. *Journal of the American Society for Psychical Research, 69*, 105-113.

Broughton, R. S. (1991). *Parapsychology: The controversial science*. New York: Ballantine Books.

Broughton, R. S., Kanthamani, H., & Khilji, A. (1990). Assessing the PRL success model on an independent ganzfeld data base. In L. Henkel, & J. Palmer (Eds.), *Research in parapsychology 1989* (pp. 32-35). Metuchen, NJ: Scarecrow Press.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*, 98-101.

Costa, P. T. J., & McCrae, R. R. (1985). *The NEO Personality Inventory Manual*. Odessa, FL: Psychological Assessment Resources.

Cushing, J. T., & McMullin, E. (Eds.). (1989). *Philosophical consequences of quantum theory: Reflections on Bell's theorem* . Notre Dame, IN: University of Notre Dame Press.

Dawson, R. (1991). Comment. *Statistical Science, 6*, 382-385.

Delanoy, D. L. (1990). Approaches to the target: A time for reevaluation. In L. A. Henkel, & J. Palmer (Eds.), *Research in Parapsychology 1989* (pp. 89-92). Metuchen, NJ: Scarecrow Press.

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

Druckman, D., & Swets, J. A. (Eds.). (1988). *Enhancing human performance. Issues, theories, and techniques*. Washington, D. C.: National Academy Press.

Eysenck, H. J. (1966). Personality and extra-sensory perception. *Society for Psychical Research, 44*, 55-71.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York: Free Press.

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

Harris, M. J., & Rosenthal, R. (1988a). Human performance research: An overview. Background paper commissioned by the National Research Council. Washington, DC: National Academy Press.

Harris, M. J., & Rosenthal, R. (1988b). Postscript to "Human performance research: An overview." Background paper commissioned by the National Research Council. Washington, DC: National Academy Press.

Herbert, N. (1987). *Quantum reality: Beyond the new physics*. Garden City, NY: Anchor.

Honorton, C. (1969). Relationship between EEG alpha activity and ESP card-guessing performance. *Journal of the American Society for Psychical Research, 63*, 365-374.

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. (1979). Methodological issues in free-response experiments. *Journal of the American Society for Psychical Research, 73*, 381-394.

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

Honorton, C. (1992, August). The ganzfeld novice: Four predictors of initial ESP performance. *Proceedings of the Parapsychological Association 35th Annual Convention, Las Vegas, NV*, 51-58.

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., & Bem, D. J. (1992). Extraversion and ESP performance: Meta-analysis and a new confirmation. In L. A. Henkel, & G. R. Schmeidler (Eds.), *Research in Parapsychology 1990* (pp. 35-38). Metuchen, NJ: Scarecrow Press.

Honorton, C., & Harper, S. (1974). Psi-mediated imagery and ideation in an experimental procedure for regulating perceptual input. *Journal of the American Society for Psychical Research, 68*, 156-168.

Honorton, C., & Schechter, E. I. (1987). Ganzfeld target retrieval with an automated testing system: A model for initial ganzfeld success. In D. B. Weiner, & R. D. Nelson (Eds.), *Research in parapsychology 1986* (pp. 36-39). 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. E. (1979). Methodological problems in free-response ESP experiments. *Journal of the American Society for Psychical Research, 73*, 1-15.

Metzger, W. (1930). Optische Untersuchungen am Ganzfeld: II. Zur phanomenologie des homogenen Ganzfelds. *Psychologische Forschung, 13*, 6-29.

Morris, R. L. (1991). Comment. *Statistical Science, 6*, 393-395.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Palmer, J. (1978). Extrasensory perception: Research findings. In S. Krippner (Ed.), *Advances in parapsychological research* (Vol. 2, pp. 59-243). New York: Plenum.

Palmer, J. A., Honorton, C., & Utts, J. (1989). Reply to the National Research Council Study on Parapsychology. *Journal of the American Society for Psychical Research, 83*, 31-49.

Parker, A. (1975). Some findings relevant to the change in state hypothesis. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology, 1974* (pp. 40-42). Metuchen, NJ: Scarecrow Press.

Parker, A. (1978). A holistic methodology in psi research. *Parapsychology Review, 9*, 1-6.

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.

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

Rhine, J. B., & Pratt, J. G. (1954). A review of the Pearce-Pratt distance series of ESP tests. *Journal of Parapsychology, 18*, 165-177.

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

Roig, M., Icochea, H., & Cuzzucoli, A. (1991). Coverage of parapsychology in introductory psychology textbooks. *Teaching of Psychology, 18*, 157-160.

Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin, 85*, 185-193.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638-641.

Rosenthal, R. (1990). Replication in behavioral research. *Journal of Social Behavior and Personality, 5*, 1-30.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Revised ed.). Newbury Park, CA: Sage.

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin, 106*, 332-337.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44*, 1276-1284.

Sannwald, G. (1959). Statistische untersuchungen an Spontanphänomene. *Zeitschrift für Parapsychologie and 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.

Schlitz, M. J., & Honorton, C. (1992). Ganzfeld psi performance within an artistically gifted population. *Journal of the American Society for Psychical Research, 86*, 83-98.

Schmeidler, G. R. (1988). *Parapsychology and psychology: Matches and Mismatches*. Jefferson, NC: McFarland.

Schmidt, H., & Schlitz, M. J. (1989). A large scale pilot PK experiment with prerecorded random events. In L. A. Henkel, & R. E. Berger (Eds.), *Research in Parapsychology 1988* (pp. 6-10). Metuchen, NJ: Scarecrow Press.

Spence, K. W. (1964). Anxiety (Drive) level and performance in eyelid conditioning. *Psychological Bulletin, 61*, 129-139.

Stanford, R. G. (1987). Ganzfeld and hypnotic-induction procedures in ESP research: Toward understanding their success. In S. Krippner (Ed.), *Advances in parapsychological research* (Vol. 5, pp. 39-76). Jefferson, NC: McFarland.

Steering Committee of the Physicians' Health Study Research Group (1988). Preliminary report: Findings from the aspirin component of the ongoing Physicians' Health Study. *New England Journal of Medicine, 318*, 262-264.

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

Stokes, D. M. (1987). Theoretical parapsychology. In S. Krippner (Ed.), *Advances in parapsychological research* (Vol. 5, pp. 77-189). Jefferson, NC: McFarland.

Swets, J. A., & Bjork, R. A. (1990). Enhancing human performance: An evaluation of "new age" techniques considered by the U. S. Army. *Psychological Science, 1*, 85-96.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 2*, 105-110.

Ullman, M., Krippner, S., & Vaughan, A. (1973). *Dream telepathy*. New York: Macmillan.

Utts, J. (1986). The ganzfeld debate: A statistician's perspective. *Journal of Parapsychology, 50*, 393-402.

Utts, J. (1991a). Rejoinder. *Statistical Science, 6*, 396-403.

Utts, J. (1991b). Replication and meta-analysis in parapsychology. *Statistical Science, 6*, 363-403.

Wagner, M. W., & Monnet, M. (1979). Attitudes of college professors toward extra-sensory perception. *Zetetic Scholar, 5*, 7-17.