Psychological Bulletin
2010, Vol. ●●, No. ●, 000–000

0033-2909/10/$12.00 DOI: 10.1037/a0019457

Meta-Analysis of Free-Response Studies, 1997–2008: Assessing the Noise Reduction Model in Parapsychology

Lance Storm
University of Adelaide

Patrizio E. Tressoldi and Lorenzo Di Risio
Università di Padova

We report the results of meta-analyses on 3 types of free-response study: (a) ganzfeld (a technique that enhances a communication anomaly referred to as "psi"); (b) nonganzfeld noise reduction using alleged psi-enhancing techniques such as dream psi, meditation, relaxation, or hypnosis; and (c) standard free response (nonganzfeld, no noise reduction). For the period 1997–2008, a homogeneous data set of 29 ganzfeld studies yielded a mean effect size of 0.142 (Stouffer $Z = 5.48$, $p = 2.13 \times 10^{-8}$). A homogeneous nonganzfeld noise reduction data set of 16 studies yielded a mean effect size of 0.110 (Stouffer $Z = 3.35$, $p = 2.08 \times 10^{-4}$), and a homogeneous data set of 14 standard free-response studies produced a weak negative mean effect size of −0.029 (Stouffer $Z = -2.29$, $p = .989$). The mean z size value of the ganzfeld database were significantly higher than the mean ES of the nonganzfeld noise reduction and the standard free-response databases. We also found that selected participants (believers in the paranormal, meditators, etc.) had a performance advantage over unselected participants, but only if they were in the ganzfeld condition.

Keywords: ESP, free response, ganzfeld, meta-analysis, psi

The 20th century was an era in which a number of scientific revolutions took place, indicated not only by technological breakthroughs but also by changes in, and challenges to, our understanding of the nature of the universe. Major conventional perspectives in physics (classical and mechanical) were challenged by models such as relativity theory and quantum mechanics that gave more accurate pictures of the physical world. Likewise, in the sciences of the mind (i.e., psychology, neuropsychology, and philosophy), there were revolutions that upturned conventional viewpoints in our understanding of mental–cognitive processes and brain events. A paradigmatic shift in consciousness usually requires a huge leap of faith, and it is noted that such changes in mindset are never made immediately or easily in any given scientific community or epoch (Collins & Pinch, 1982).

In that same century, relegated to the sidelines of these major events, was the equally controversial subject of an anomaly known as "psi"—a shorthand term for psychic functioning, specifically categorized as either extrasensory perception (i.e., ESP, which covers telepathy, clairvoyance, and precognition) or psychokinesis (i.e., PK, which refers to paranormal mental influence on matter). Telepathy refers to the "paranormal acquisition of information concerning the thoughts, feelings or activity of another conscious being" (Thalbourne, 2003, p. 125). Clairvoyance is defined as "paranormal acquisition of information concerning an object or contemporary physical event; in contrast to telepathy, the information is assumed to derive directly from an external physical source" (Thalbourne, 2003, p. 18). Precognition is defined as "a form of extrasensory perception in which the target is some future event that cannot be deduced from normally known data in the present" (Thalbourne, 2003, p. 90). Related to telepathy is "remote viewing," which Thalbourne (2003) defines as "a neutral term for general extrasensory perception . . . especially in the context of an experimental design wherein a percipient [perceiver/receiver] attempts to describe the surroundings of a geographically distant agent [sender]" (p. 107).

Major laboratory efforts to determine the likelihood and extent of both hypothesized effects (ESP and PK) were first undertaken by J. B. Rhine in the 1930s, with his card-guessing (Rhine et al., 1940/1966) and dice-throwing studies (Rhine, 1948/1954). Card guessing is described as a form of "forced-choice" experiment because there are a limited number of choices, and the participant is forced to guess the target from a limited set of symbols (usually one of five in the card-guessing design: square, cross, circle, star, and wavy lines). Years later, meta-analytic studies on card-guessing and dice-throwing databases yielded statistically significant effects (Honorton & Ferrari, 1989; Radin & Ferrari, 1991).

Departures from the Rhinean paradigms in the 1960s and 1970s saw a shift in interest toward "free-response" designs (e.g., Honorton, 1977). *Free response* is a term that "describes any test of ESP in which the range of possible targets is relatively unlimited and is unknown to the percipient [perceiver/receiver]" (Thalbourne, 2003, p. 44). Picture guessing is typically free response, although the perceiver is usually presented with, say, four randomly selected pictures (target plus three decoys), none of which have been sighted by the perceiver until he or she is required to

Lance Storm, School of Psychology, University of Adelaide, Adelaide, Australia; Patrizio E. Tressoldi and Lorenzo Di Risio, Dipartimento di Psicologia Generale, Università di Padova, Padua, Italy.

We thank Richard Broughton, Dean Radin, Chris Roe, Serena Roney-Dougal, Simon Sherwood, Jerry Solfvin, and Fiona Steinkamp for their collaboration in data retrieval. We also thank Dick Bierman, Nicola Holt Hutchinson, Eva Lobach, Vitor Moura, Alejandro Parra, Dean Radin, and Jessica Utts for their helpful comments.

Correspondence concerning this article should be addressed to Lance Storm, School of Psychology, University of Adelaide, Adelaide SA 5005, Australia. E-mail: lance.storm@adelaide.edu.au

identify the target among the four pictures. The perceiver is restricted to four pictures, but the free-response component manifests in the "mentation" (i.e., the stream of images, thoughts, and ideas in the mind of the perceiver that are recorded to assist the judging process). It is the variability in content of the mentation report itself that suggests that the range of possible targets is virtually unlimited and unknown to the perceiver. Along with card guessing and dice throwing, the free-response domain has also proved successful for parapsychologists. For example, Milton (1997a) conducted a meta-analysis of studies on participants who were in a normal waking state of consciousness during a free-response task. Her period of analysis was 1964–1992. For a database of 75 studies, Milton found a weak but significant effect.

Skeptics have disputed claims for paranormal phenomena (see Alcock, 1981; Blackmore, 1985; Hansel, 1966, 1980; Hyman, 1989). Alcock (1981) insisted that parapsychologists must "demonstrate the existence of psi independently of . . . non-chance results" (p. 147). Blackmore (1985), once a parapsychologist herself, was disturbed that she found so little evidence in her PhD thesis for the paranormal, but as Berger (1989) pointed out, "'Flaws' were invoked [by Blackmore] to dismiss significant results while other flaws were ignored when studies produced nonsignificant results" (p. 123). Berger also claimed that Blackmore "misreport[ed]" and "incorrectly reported" experimental details and statistical results that supported the psi hypothesis in Blackmore's studies.

Hansel (1966, 1980) argued that because psi cannot be convincingly and consistently demonstrated (or better, induced at call) in the laboratory, it does not exist. He even argued that if fraud is possible in a psi experiment, it is reasonable to assume that fraud did take place.

Hyman (1989) has admitted that he attempts to "justify withholding any attention to the claims for the paranormal on the part of orthodox science" (p. 206). However, in regard to the success of free-response studies, he has acknowledged that the critics have not "demonstrated a plausible alternative" that explains the significant effects (p. 157).

Akers (1984) drew attention to "explanations" that skeptics use to justify their doubt regarding the existence of psi, such as randomization failure, sensory leakage, cheating, and procedural errors. His point is that even the seemingly best parapsychological studies have flaws that preclude their findings as evidence of paranormal effects. However, that was over two decades ago, before standards were markedly improved. What might have been true then has little bearing today.

Added to the skeptic's list of explanations for the psi communication anomaly are "delusion, . . . coincidence, [and] unconscious inference" (Stokes, 1987, p. 84). Because these claims abound, it remains for the parapsychologist not only to eliminate all possibility of these explanations but also to design experiments that furnish repeatable evidence of psi. One such design may be the ganzfeld.

The Origins of the Ganzfeld

The most renowned and certainly the most controversial domain in parapsychology is a procedure for testing telepathy in a form of free-response condition known as *Ganzfeld* (German for "total field"). The ganzfeld is a "special type of environment (or the technique for producing it) consisting of homogenous [*sic*], unpatterned sensory stimulation" to the eyes and ears of the participant, who is usually in "a state of bodily comfort" (Thalbourne, 2003, p. 45). Traditionally, the ganzfeld is a procedure whereby an agent in one room is required to "psychically communicate" one of four randomly selected picture targets or movie film targets to a perceiver in another room, who is in the ganzfeld condition of homogeneous sensory stimulation. The ganzfeld environment involves setting up an undifferentiated visual field by viewing red light through halved translucent ping-pong balls taped over the perceiver's eyes. Additionally, an analogous auditory field is produced by listening to stereophonic white or pink hissing noise. As in the free-response design, the perceiver's mentation is recorded and accessed later in order to facilitate target identification. At this stage of the session, the perceiver ranks from 1 to 4 the four pictures (one target plus three decoys; Rank 1 = "hit"). This condition follows the noise reduction model, which is considered "psi conducive" because it allegedly reduces irrelevant background noise, leaving mainly the psi signal.

A number of investigators pioneered the ganzfeld technique in the 1970s (W. G. Braud, Wood, & Braud, 1975; Honorton & Harper, 1974; Parker, 1975). The technique arose from a consideration of the putative psi-conducive (noise reducing) effects of dreaming, hypnosis, relaxation, and meditation. Honorton (1977), one of the first discussants of "internal attention states" in ESP research, justified the importance of the ganzfeld setting as a crucial component favorable to psi by conceptualizing psi as a weak cognitive signal normally masked by internal cognitive and external "noise." W. G. Braud (2002) saw the ganzfeld state as one that brought about stillness of mind and cognitive quietude in accordance with the model of (mental) noise reduction. Theoretically, by increasing the signal-to-noise ratio (i.e., reducing the noise), the presumed psi information could be better detected.

The Ganzfeld Meta-Analyses

Honorton (1985) undertook one of the first meta-analyses of the many ganzfeld studies that had accrued by the mid-1980s. Twenty-eight studies yielded a collective hit rate of 38%, where 25% would be expected by chance. Honorton noted that of the 28 studies, 23 (82%) had positive z scores. Various flaws in his approach were pointed out by Hyman (1985), but ultimately they came to an agreement (Hyman & Honorton, 1986). Their words from the so-called Joint Communiqué were as follows: "We agree that there is an overall significant effect in this database that cannot reasonably be explained by selective reporting or multiple analysis" (p. 351). They differed only on the "degree to which the effect constitutes evidence for psi" (p. 351).

A second major meta-analysis on 11 "autoganzfeld" studies followed (Honorton et al., 1990). These studies adhered to the guidelines laid down in the Joint Communiqué. The autoganzfeld procedure avoids methodological flaws by using a computer-controlled target randomization, selection, and judging technique. Subsequently, Bem and Honorton (1994) reduced the Honorton et al. (1990) database to 10 studies by removing one very successful study that was not methodologically comparable to the others. They reported a hit rate of 32.2% for these 10 studies.

Milton and Wiseman (1999) followed up with the intention of replicating the result of Bem and Honorton's (1994) 10-study meta-analysis. For the period 1987–1997, 30 studies (including seven autoganzfeld studies) produced an overall nonsignificant effect size (ES) of 0.013 (Stouffer $Z = 0.70$, $p = .242$). Milton and Wiseman concluded that a significant communication anomaly for the ganzfeld had not been replicated by a "broader range of researchers" (p. 391).[1] However, Storm and Ertel (2001) compared Milton and Wiseman's database with Bem and Honorton's database of 10 studies and found the two did not differ significantly. A homogeneous post-Communiqué database of 40 studies was formed (ES = 0.05, Stouffer $Z = 1.88$, $p = .03$).

Going one step further, Storm and Ertel (2001) found 11 pre-Communiqué studies not previously meta-analyzed and combined them with Honorton's (1985) database of 28 studies as well as the 40-study database just described. After a minor adjustment (see Storm & Ertel, 2002), a 79-study database was compiled, which had a significant mean ES of 0.138 (Stouffer $Z = 5.59$, $p = 1.14 \times 10^{-8}$).

In a study with a different agenda, Bem, Palmer, and Broughton (2001) showed that if one considered all the available published articles up to and including 1997 (i.e., Milton and Wiseman's 30 studies, plus 10 new studies), 29 studies that used standard ganzfeld protocols yielded a cumulative hit rate that was significantly above chance (31%), comparable to the results presented by Bem and Honorton (1994). In contrast, the nine studies that followed nonstandard protocols obtained an overall nonsignificant hit rate of only 24%. The two types of studies were significantly different from each other.

Rationale for the Present Study

Over the course of the last two decades, many parapsychologists abandoned proof-oriented research in favor of process-oriented research, which is the major convention in all scientific fields. This change of focus was necessary to help raise psi beyond its status as a "statistical anomaly" that needed explaining. Thus, whereas Bem et al. (2001) indicated the merits of "continuing to conduct exact replications of the ganzfeld procedure" (p. 215), they also advised process-oriented researchers that they "must be willing to risk replication failure" (p. 215) in their modifications to the standard ganzfeld procedure in order to learn more about paranormal processes.

Because of the great deal of attention that has been lavished on the ganzfeld technique, it may be the case that the ganzfeld design has been partly responsible for the lack of attention paid to other noise reduction studies that do not follow the ganzfeld protocol. These studies, which involve dreaming, hypnosis, relaxation, and meditation, should be meta-analyzed and contrasted with the ganzfeld meta-analyses to assess the real benefit to parapsychology of the ganzfeld design. The ganzfeld may not be the final word on how best to elicit an anomalous communication effect or to reach an understanding as to its nature.

It may also be the case that clues as to the extent and/or optimal strength of the ESP effect could be unearthed through the investigation of individual differences in participants. Statistician Jessica Utts (Utts, 1991) made this claim some years ago: "A promising direction for future process-oriented research is to examine the causes of individual differences in psychic functioning" (p. 377). The evidence that certain psychological factors are psi conducive was clearly illustrated by Honorton (1997), who claimed that psi could be enhanced with training, or might be better elicited from special personality types or those with specific psi-experiential characteristics. W. G. Braud (2002) also stressed the advantages in using participants who (a) underwent prior psi testing, (b) reported psi experiences, or (c) had personality traits measured in conventional personality inventories. Also, on the basis of past findings, Honorton and Ferrari (1989) and Morris (1991) have argued that unselected participants somehow do not perform as well as selected participants on psi tasks.

Various personality traits (e.g., extraversion; see Palmer, 1977) and personal characteristics (i.e., psi belief, psi experience; see Lawrence, 1993) have been investigated over the decades as psi-conducive variables, suggesting that most researchers share the view that these variables play an important role in ESP performance. For example, Bem et al. (2001) advocated for psi testing the selection of participants who have had "previous psi experiences, or practiced a mental discipline such as meditation" (p. 215).

We note that there are a number of reasons why researchers avoid special participants and sample only normal populations (see Discussion). However, if paranormal ability is generally distributed throughout the population, or, conversely, it exists only in subpopulations, skepticism about psi is not justified either way. In the present article we follow a line of inquiry that assesses whether selected participants (i.e., participants who had previous experience in ESP experiments, or were psi believers, or had special psi training, or were longtime practitioners of meditation or relaxation exercises, etc.) have a psi advantage.

Design of the Present Study

The meta-analysis by Bem et al. (2001) included 10 studies (1997–1999), which were not assessed by Storm and Ertel (2001) because the latter's retrospective study was not an update but mainly a critique of the meta-analysis by Milton and Wiseman (1999). Similarly, the meta-analysis by Bem et al. was not comprehensive but merely comparative. A comprehensive, up-to-date meta-analysis of the ganzfeld domain is therefore long overdue. Also, we would argue that the noise reduction model per se tends to be associated exclusively with the ganzfeld, but we find that a considerable number of noise reduction studies (1997–2008) using dream ESP, hypnosis, relaxation, and meditation protocols have not been meta-analyzed. Given the alleged importance of noise reduction, we see no reason not to include these marginalized studies in separate meta-analyses. Also, any psi performance differences between selected and unselected participants have not been fully tested, although past research does suggest there are differences. We note here that there is no justification in comparing forced-choice studies in our meta-analysis with free-response studies, as the two designs are incompatible, whereas the three types of study we do compare are all free response. Therefore, the following hypotheses are proposed:

[1] In fact, Milton and Wiseman were premature in their conclusion. Jessica Utts (personal communication, December 11, 2009) used the exact binomial test only on trial counts in Milton and Wiseman's database ($N = 1{,}198$, hits = 327) and found a significant hit rate of 27% ($z = 1.80$, $p = .036$).

Hypothesis 1: Ganzfeld studies yield a stronger mean ES than the nonganzfeld (non-Gz) noise reduction studies, which, in turn, yield a stronger mean ES than the standard free-response studies (i.e., non-Gz, no noise reduction).

Hypothesis 2: Selected participants perform better than unselected participants in all three databases.

Method

Study Retrieval

The following major English-language peer-reviewed journals and other publications were accessed for studies: *Journal of Parapsychology*, *European Journal of Parapsychology*, *Journal of the Society for Psychical Research*, *Journal of the American Society for Psychical Research*, *Proceedings of the Annual Convention of the Parapsychological Association*, *Journal of Scientific Exploration*, *Australian Journal of Parapsychology*, *Journal of Cognitive Neuroscience*, and *Dreaming*.

The period of analysis for free response was from 1992 to 2008, which continues from Milton's (1997a) cutoff date of January 1992. The period of analysis for noise reduction studies (including ganzfeld) dates from March 1997 to 2008, which continues from Storm and Ertel's (2001) cutoff date of February 1997. We note that this period includes six ganzfeld articles from the meta-analysis by Bem et al. (2001),[2] but we merged their data set with ours because of the nature of the hypotheses we proposed to test.

To find appropriate research articles online, we conducted exhaustive Internet searches through EBSCOhost of the relevant databases, including PsycINFO, PsycARTICLES, and CINAHL, as well as other relevant databases (i.e., Medline, Web of Science, Lexscien, and InformIT). The following keywords and subject headings were entered in the search: *extrasensory perception*, *ESP*, *ganzfeld*, *PK*, *telepathy*, *clairvoyance*, *precognition*, *anomalous cognition*, *parapsychology*, *paranormal*, and *psi*. Most of these Internet searches yielded studies already found in the above-listed journals. We adopted the following criteria:

- Only ESP (i.e., telepathy, clairvoyance, and precognitive) studies are to be assessed.
- The number of participants must be in excess of two to avoid the inherent problems such as is typical of case studies.
- Target selection must be randomized by using a random number generator in a computer or otherwise, or a table of random numbers.
- Studies must provide sufficient information (e.g., z scores or number of trials and outcomes) for the authors to calculate the direct hit rates and to apply appropriate statistical tests and calculate ES as $z/\sqrt{n}$.

Procedure

For each study, we checked the following factors: (a) the criteria adopted for selecting participants, (b) number of participants, (c) number of trials, (d) type of ESP task (clairvoyance, telepathy, or precognition), (e) number of alternatives in the tasks, and (f) total number of hits. Despite Milton's (1997b) finding of a nonsignificant difference between direct hits and sum-of-ranks statistics, we preferred the former measure, as it provides a more "conservative" result (see Honorton, 1985, p. 54), and it is easier to grasp intuitively.

With these data we derived the proportion of hits. For studies in which z scores were not given, we calculated z scores using the binomial exact probability from the number of hits and trials (http://faculty.vassar.edu/lowry/binomialX.html).

In parapsychological studies each participant is usually given one trial, and each will either get a hit or not get a hit. The "true" hit rate, then, is the probability of success for that trial. We therefore have a true null hypothesis, so the exact binomial test is used because there is an obvious null value associated with that null hypothesis.

Studies were grouped into one of three categories according to the following criteria:

- Category 1: ganzfeld (unselected and selected participants),
- Category 2: non-Gz noise reduction (unselected and selected participants), and
- Category 3: standard free response (no ganzfeld or similar noise reduction techniques to alter the normal waking cognitive state through hypnosis, meditation, dreaming, or relaxation).

We classified remote-viewing studies as Category 3 standard free-response studies in alignment with Milton's (1997a) claim that remote-viewing studies "are just a subset of a larger group of ESP studies in which free-response methods are used without the participant being in an altered state of consciousness" (p. 280).

Studies were rated for quality by two judges who were graduate students of the second author and were kept blind during the judging phase. That is, following Bem et al. (2001, p. 209), the judges saw only the method sections from which all identifiers had been deleted, such as article titles, authors' hypotheses, and references to results of other experiments in the article. Other criteria used were adapted from Milton (1997a). These criteria are

- appropriate randomization (using electronic apparatuses or random tables),
- random target positioning during judgment (i.e., target was randomly placed in the presentation with decoys),
- blind response transcription or impossibility to know the target in advance number of trials preplanned,
- sensory shielding from sender (agent) and receiver (perceiver),
- target independently checked by a second judge, and
- experimenters blind to target identity.

Two judges answered "Yes" or "No" to each of the criteria. The study score is the ratio of points awarded with respect to the items applicable (minimum score is 1/7 = .143; maximum score is 7/7 = 1.00). ~~There were 65 studies in the total pool, but 10 studies (Bem et al., 2001) were already judged. Of the 55 studies we had judged, 17 studies (31%) received a perfect score from at least one judge. However, most criteria (i.e., five or more out of seven) were met in 58 of 65 studies (i.e., 89%);~~ We stress that failure to make explicit declaration does not mean that any given criterion was not

[2] These six articles account for 10 studies (Cases 1, ~~2,~~ 9, 10, 11, 25, 26, 27, 28, and 29 in Appendix A, Category 1).

incorporated into any given experiment. For example, a minority of studies (approximately 20%) did not declare explicitly "Random target positioning during judgment," but we would assume that meeting this criterion is a prerequisite in any given parapsychological study dating from the time of implementation of the Joint Communiqué (Hyman & Honorton, 1986). In the last 10–15 years, we note that all studies have used automated procedures to randomize trials, and they used strict controls to avoid any conventional (i.e., normal modality-based) communication between researchers, agents, and perceivers.

Results

Descriptive Statistics and Quality Ratings

Appendix A shows all noise reduction and free-response studies in their respective categories. Sixty-seven studies were reported in 48 articles conducted by 61 experimenters (see References for articles; articles marked with asterisks indicate articles included in the meta-analyses). Sixty-three studies (94%) used a four-choice design (the remaining four studies used a three-, five-, or eight-choice design). For the four-choice designs only, there were 4,442 trials and 1,326 hits, corresponding to a 29.9% hit rate where mean chance expectation (MCE) is equal to 25%. Across the three categories, 31 studies (46.3%) tested telepathy, 26 studies (38.8%) tested clairvoyance, and 10 studies (14.9%) tested precognition. Using a Type III univariate analysis of variance (ANOVA) test to control for the imbalance in group sizes, we found no significant difference between the three psi modalities, $F(2, 66) = 0.36$, $p = .697$ (two-tailed).

Cronbach's alpha for the two judges' ratings was .79, suggesting a high degree of interrater reliability. Because the correlation between mean quality scores and ES values (the latter are given in Appendix A) was extremely weak and not significant, $r_s(65) = .08$, $p = .114$ (two-tailed), we claim that ES is not likely to be an artifact of poor experimental design.

Z Statistics and Effect Sizes

T1 Table 1 lists by category all the various statistics: mean z scores, mean ES values, Stouffer Z, and corresponding p values.

Category 1 (ganzfeld). This database consisted of 30 ganzfeld studies (47.7% of all studies in our search). For all these 30 studies, paranormal tasks were performed exclusively during the ganzfeld condition. Sixteen studies tested unselected participants only, and 14 studies tested selected participants only (e.g., they believed in the possibility of psi, or had prior spontaneous psi experiences or prior psychic training, or regularly practiced some type of mental discipline).

The full data set of ganzfeld studies ($N = 30$) yielded mean $z = 1.16$, mean ES = 0.152, and Stouffer $Z = 6.34$ ($p = 1.15 \times 10^{-10}$; see Table 1 for details). However, the skew of the z-score distribution was not normal, although the skew of the distribution of ES values was normal.[3] One outlier (Case 2 in Appendix A, Category 1) was excluded from further analyses in the present article for having an extremely high z score (≥ 5.20). Fn3

A homogeneous data set of 29 ganzfeld studies yielded mean $z = 1.02$ ($SD = 1.36$; range: -1.45 to 4.32), mean ES = 0.142 ($SD = 0.20$; range: -0.26 to 0.48), and Stouffer $Z = 5.48$ ($p = 2.13 \times 10^{-8}$). Ninety-five percent confidence intervals (CIs) are as follows: z scores, [0.50, 1.54]; ES values, [0.07, 0.22]. Note that neither of these 95% CIs includes values of MCE (i.e., zero). Of the 29 studies, 19 (65.5%) had positive z scores. Nine (31%) of the 29 studies are independently significant ($\alpha \leq .05$).

Following the example set by Utts (see Footnote 1), we conducted a binomial exact test on trial counts only. One study (Roe & Flint, 2007) used an eight-choice design (i.e., one target plus seven decoys), so it could not be included, but all other studies used the four-choice design. For 29 ganzfeld studies ($N = 1{,}498$, hits = 483), we found a 32.2% hit rate (binomial $z = 6.44$, $p < .001$).

Although we report a significant Stouffer Z of 5.48 for this database, a more stringent approach to testing a database of studies is provided by Darlington and Hayes (2000), who regard "mean(z) as the real test statistic" (p. 505). Darlington and Hayes criticized the alternatives: "Stouffer's method and all other well-known probability poolers suffer from three major but avoidable limitations: vulnerability to criticisms of individual studies . . . difficulty in handling the file drawer problem . . . [and] vague conclusions" (pp. 497–498). Darlington and Hayes's "Stouffer-max" test provides a "MeanZ(s,k)" value, which is the "mean of the s highest of k mutually independent values of z" (p. 506), which is then compared with a critical MeanZ. If we take $s = 9$ (i.e., the nine studies with the significant z scores) and $k = 29$ (i.e., where $k = N = 29$), our MeanZ is 2.52. Darlington and Hayes (p. 506, Table 3) give critical MeanZ = 2.52. In other words, the mean z for the ganzfeld database is sufficiently higher than is required by the Stouffer-max test.

With the file-drawer formula given by Rosenthal (1995, p. 189; see Appendix C for formula), there would have to be no fewer than 293 unpublished articles in existence with overall nonsignificant results to reduce our significant finding to a chance result. Darlington and Hayes (2000) offered a more conservative and reliable test that determines the number of unpublished nonsignificant studies needed to reduce a database to nonsignificance, but these may also be "highly negative" (or "psi-missing" in the case of parapsychology). Darlington and Hayes claimed that "the [fail-safe N] derived with the binomial method is a lower limit on the number of missing studies that would have to exist to threaten the significance of the pooled p value" (p. 500).

Using Darlington and Hayes's (2000, p. 503, Table 2) tabled data, for nine studies with significant positive outcomes, we find the pooled p less than or equal to .05 if there is a total of up to 95 studies. In other words, we find a "fail-safe N" of up to 95 unpublished studies must exist in total for this category alone, and 86 of these (i.e., 95 minus 9) could all be negative (psi-missing) studies, yet our nine significant studies would still constitute a proof against the null hypothesis. The existence of such a large number of hypothesized unpublished studies (i.e., up to 86) is unlikely.

Category 2 (non-Gz; noise reduction). This database was composed of 16 studies (23.9% of all studies in our search). These 16 studies were classified as noise reduction studies (i.e., they were non-Gz but involved treatment conditions such as dream psi, hypnosis, relaxation, or meditation). Samples were composed en-

[3] Outlier studies that cause significant deviations in the distribution may deflate or inflate the mean z score and/or mean ES. Outliers are identified from SPSS stem-and-leaf and box-and-whiskers plots as significantly deviated ("extreme") cases. Such indicators of heterogeneity are standard practice in meta-analyses.

tapraid5/z2r-psybul/z2r-psybul/z2r00410/z2r2194d10z | xppws | S=1 | 4/7/10 | 14:22 | Art: 2009-0253

Table 1
Three Free-Response Databases by Category

Category	*Z*				Effect Size				Sums of *Z* (Σ*Z*)	Stouffer *Z*	*p*[a]
	M	*SD*	Skew	*SE*	*M*	*SD*	Skew	*SE*			
1 (*N* = 30)[b]	1.157	1.54	0.71	0.43	0.152	0.20	−0.04	0.43	34.70	6.34	1.15×10^{-10}
2 (*N* = 16)[c]	0.831	1.22	−0.33	0.56	0.110	0.19	−0.24	0.56	13.30	3.35	2.08×10^{-4}
3 (*N* = 21)[d]	0.474	1.48	1.47	0.50	0.102	0.27	1.16	0.50	9.96	2.17	1.50×10^{-2}

[a] One-tailed. [b] Ganzfeld. [c] Nonganzfeld noise reduction. [d] Standard free response.

tirely of either selected participants ($n = 10$) or unselected participants ($n = 6$). The full data set yielded mean $z = 0.83$ ($SD = 1.23$; range: −1.55 to 3.06), mean ES = 0.110 ($SD = 0.19$; range: −0.22 to 0.47), and Stouffer $Z = 3.35$ ($p = 2.08 \times 10^{-4}$). The database was significant. The distributions for z scores and ES values were both normal. CIs (95%) are as follows: z scores, [0.18, 1.48]; ES values, [−0.010, 0.21]. We note that neither of these 95% CIs includes values of MCE. Of the 16 studies, 11 (68.8%) had positive z scores, with three (18.8%) of the 16 studies independently significant ($\alpha = .05$).

Again following Utts's example, we conducted a binomial exact test on trial counts only. For 16 studies ($N = 870$, hits = 269), we found a 30.9% hit rate ($z = 3.99$, $p = 3.30 \times 10^{-5}$).

Once again we applied Darlington and Hayes's (2000) Stouffer-max test, given $s = 3$ and $k = 16$. Darlington and Hayes (p. 506, Table 3) give critical MeanZ = 1.99. Our MeanZ is 2.57, so again the mean z for the non-Gz noise reduction database is sufficiently higher than is required in the Stouffer-max test.

With Rosenthal's (1995) file-drawer formula, there would have to be approximately 49 unpublished and nonsignificant articles in existence to reduce to chance our significant finding to a chance result. It is arguable whether 49 unpublished and nonsignificant studies exist in this category. However, using Darlington and Hayes's (2000, p. 503, Table 2) tabled data, for three significant positive studies, pooled $p \leq .05$ only if the total number of studies is 16 or less, based on a 16-study database with four significant studies. In other words, 13 studies (i.e., 16 minus 3) might be unpublished, and these are all permitted to be psi-missing studies.

Category 3 (free response). This database consisted of 21 standard free-response studies (28.4% of all studies in our search). These studies did not involve altered states of consciousness or noise reduction techniques. Twelve studies (57.1%) were composed completely of unselected participants, and nine (42.9%) were composed of selected participants. The full data set yielded mean $z = 0.47$, mean ES = 0.102, and Stouffer $Z = 2.17$ ($p = 1.50 \times 10^{-2}$), which is significant. However, the skews of the z-score and ES distributions were not normal. Six outliers (Cases 2, 3, 6, 11, 17, 20, and 21 in Appendix A, Category 3) were excluded from further analyses in the present article because ES values and z scores for these studies were extremely high ($z \geq 3.50$, ES ≥ 0.27) or extremely low (ES ≤ −0.29).

A homogeneous data set of 14 free-response studies yielded mean $z = -0.21$ ($SD = 0.47$; range: −0.94 to 0.65), mean ES = −0.029 ($SD = 0.07$; range: −0.15 to 0.09), and a nonsignificant Stouffer Z of −2.29 ($p = .989$). A binomial exact test on trial counts only ($N = 1{,}872$, hits = 455), yielded a 24.3% hit rate (binomial $z = -0.67$, $p = .749$).

Of the 14 studies, only four (28.6%) had positive z scores, and none of these were significant ($\alpha = .05$). CIs (95%) are as follows: z scores, [−0.50, 0.09; ES values, [−0.07, 0.01]. Both intervals include MCE.

Differences Between Databases and Participants

Hypothesis 1. We hypothesized that ganzfeld studies yield a stronger mean ES than the non-Gz noise reduction studies, which, in turn, yield a stronger mean ES than the standard free-response studies. From our analysis above, we can see that the ganzfeld studies produced the strongest mean ES values (mean ES = 0.142), followed by the non-Gz noise reduction studies (mean ES = 0.110), followed by the standard free-response studies (mean ES = −0.029).

An ANOVA test showed a significant difference in mean ES values, $F(2, 53) = 6.08$, $p = .004$ (two-tailed). A post hoc Tukey's test revealed that only Categories 1 and 3 differed significantly from each other (ES mean difference = 0.17, $p = .005$). Noise reduction studies (ganzfeld and nonganzfeld) were significantly different from standard free-response studies, which do not feature noise reduction. Therefore, Hypothesis 1 was partially supported.

Hypothesis 2. We hypothesized that selected participants perform better than unselected participants in all three databases. For all studies ($N = 59$), samples were composed entirely of either unselected participants ($n = 30$) or selected participants ($n = 29$).

The same univariate ANOVA test above showed that there was no difference on mean ES values between unselected and selected participants, $F(1, 53) = 2.25$, $p = .140$ (two-tailed). However, the ANOVA also shows a significant Category × Participant interaction effect, $F(2, 53) = 3.50$, $p = .037$ (two-tailed). Figure 1 shows F1
that the effect is attributable to participants in the ganzfeld only. In a separate t test on ganzfeld participants, we found a significant difference between selected (ES = 0.26) and unselected participants (ES = 0.05), $t(27) = -3.44$, $p = .002$ (two-tailed).

Given that we have a significant interaction effect, we tested for two simple effects (unselected participants only and selected participants only) between the three conditions. For unselected participants ($n = 30$), there was no significant difference between conditions, $F(2, 27) = 1.44$, $p = .255$ (two-tailed). However, for selected participants ($n = 29$), there was a significant difference between conditions, $F(2, 26) = 6.83$, $p = .004$ (two-tailed). A post hoc Tukey's test revealed that the difference was between Category 1 (mean ES = 0.26) and Category 3 (mean ES = −0.34, ES mean difference = 0.29, $p = .004$).

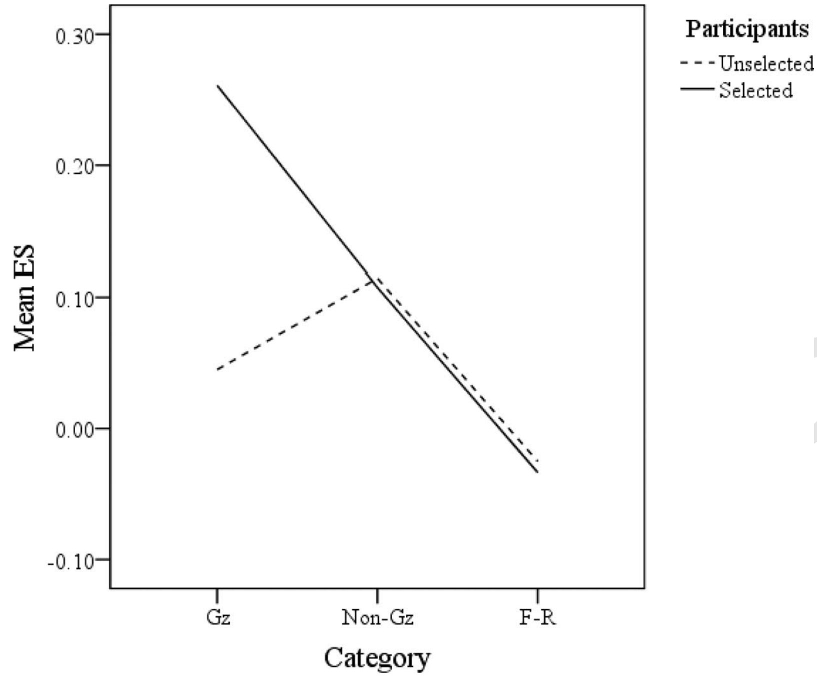

Figure 1. There is a category difference on effect size (ES) between groups—ganzfeld (Gz), nonganzfeld (non-Gz) noise reduction, and standard free response (F-R)—and an interaction effect between group and participant type.

ES Comparisons

To ascertain whether the two significant databases (i.e., ganzfeld and non-Gz/noise reduction) were the result of extremely positive ES values for a limited pool of experimenters and/or laboratories, we conducted a one-way ANOVA test on the pooled data ($N = 45$) for those two databases after dividing them into experimenter and laboratory groups. We could not test Experimenter × Laboratory interaction, as we found that a number of experimenters had worked in more than one specific laboratory.

We formed seven mutually exclusive experimenter and laboratory groups with at least two studies in each: "Morris," "Parker," "Parra," "Roe," "Roney-Dougal," "Tressoldi," and "Wezelman." ES values were not significantly different between laboratory and experimenter groups, $F(6, 32) = 1.97$, $p = .315$ (two-tailed).

Comparisons With Other Databases

To gauge the importance and relevance of the above results, we contrasted our ganzfeld and free-response databases with two databases compiled from two previous meta-analyses: Storm and Ertel (2001) and Milton (1997a), respectively.

Ganzfeld studies. When we compared our newly formed 29-study ganzfeld database with Storm and Ertel's (2001) 79-study database (see The Ganzfeld Meta-Analyses section), we found no significant differences on z scores, $t(106) = -1.01$, $p = .314$ (two-tailed), or ES values, $t(106) = 0.19$, $p = .848$ (two-tailed). We combined the two databases to form a database of 108 studies: mean $z = 0.80$ ($SD = 1.36$), mean ES $= 0.142$ ($SD = 0.27$), and Stouffer $Z = 8.31$ ($p < 10^{-16}$).

However, this database is heterogeneous. We found three outliers with extremely high ES values (≥0.70; Cases 7, 16, and 31 as listed in Honorton, 1985, p. 84, Appendix A) and three that were extremely low (≤−0.46: Case 4 as listed in Honorton, 1985, p. 84, Table A1; and Kanthamani, Khilji, & Rustomji-Kerns, 1988 [Series 5b], and Kanthamani & Palmer, 1993, from Milton and Wiseman, 1999). These six were excluded from further analyses in this section.

The homogeneous databases consists of 102 studies: mean $z = 0.81$ ($SD = 1.23$; range: −2.30 to 4.32), mean ES $= 0.135$ ($SD = 0.20$; range: −0.44 to 0.65), and Stouffer $Z = 8.13$ ($p < 10^{-16}$). CIs (95%) are as follows: z scores, [0.56, 1.07]; ES values, [0.10, 0.17]. Note that neither of these includes MCE. Of the 102 studies, 74 (72.5%) had positive z scores. Twenty-seven (26.5%) of the 102 studies are independently significant ($\alpha \leq .05$).

Darlington and Hayes's (2000) online table[4] gives critical MeanZ = 1.46, where $s = 27$. In the Stouffer-max test, the mean z for this large database, at MeanZ = 2.32, is sufficiently higher than is required. With Rosenthal's (1995, p. 189) file-drawer formula, there would have to be approximately 2,414 unpublished and nonsignificant papers in existence to reduce our significant Stouffer Z to chance. Using Darlington and Hayes's (2000) table, for 27 studies with significant positive outcomes, pooled $p \leq .05$ if the fail-safe $N \leq 384$ studies, based

[4] For values of $k > 50$, Darlington and Hayes (2000, p. 503) refer researchers to their Cornell University online table (http://www.psych.cornell.edu/darlington/meta/S10.HTM).

on the 27 significant studies in our database (N = 102). It should be noted that 3,57 studies (i.e., 384 minus 27) are permitted to be psi-missing studies.

Free-response studies. In her original database of 78 free-response studies, Milton (1997a) found an ES of 0.16, mean z of 0.65, and Stouffer Z of 5.72 ($p = 5.40 \times 10^{-9}$). In her homogeneous database of 75 studies, having removed "three studies that contributed most to the heterogeneity" (p. 289), Milton found a mean ES of 0.17, mean z of 0.68, and Stouffer Z of 5.85 ($p = 2.46 \times 10^{-9}$).

For our homogeneous database of 14 free-response studies, we reported above a mean ES of −0.03, a mean z of −0.21, and a nonsignificant Stouffer Z of −2.29 (p = .989). Thus, Milton's reported ES (0.17) is considerably stronger than the ES reported in the present study. However, using mean z scores and applying Rosenthal and Rubin's (1979) Z_{diff} formula (see Appendix C) we see that the difference is not significant: $[0.68 - (-0.21)]/\sqrt{2} = 0.63$ (p = .264).

It is important to mention an anomaly in Milton's (1997a) database. There are 25 studies (44%) of the 57 studies marked with asterisks in her reference section (pp. 305–312) that are composed largely of designs using transcendental meditation, hypnosis, mental imagery training (or guided imagery), relaxation, and even ganzfeld. We are at a loss in understanding why these studies made it into her meta-analysis when her aim was to assess "free-response ESP studies without altered states of consciousness" (p. 279). Furthermore, two of these studies—Bierman et al. (1984) and Murre, van Dalen, Dias, and Schouten (1988)—are located (properly) in our Appendix A as ganzfeld studies because they *are* ganzfeld studies. Milton's database is a mixed bag of studies, and it certainly cannot be considered standard free response as we (and arguably Milton) categorize them. On this basis, we contend that nothing can be achieved in making any further comparisons of any databases, old or new, with Milton's database.

Decline Effects Across Time in the Ganzfeld Databases

Bierman, Bosga, Gerding, and Wezelman (1993) and Bierman (2001) were among the first to report decline effects in the ganzfeld domain, but others have drawn attention to the effect (see Milton & Wiseman, 1999). Such declines in the ganzfeld could indicate improvements in study quality over the years, which have therefore minimized or eliminated certain flaws. Consequently, psi may be nothing more than an artifact of such flaws. However, decline effects in psi research existed long before the Hyman–Honorton guidelines were implemented (see Hyman & Honorton, 1986), so it cannot be assumed that study quality is single-handedly responsible for declines in the ganzfeld domain. As Palmer (1986) has pointed out, it should not be assumed that failure to replicate (i.e., decline effects) results from the removal of flaws, just as it is presumptive to assume that past ganzfeld successes were due to the presence of flaws.

It is also possible, as Bierman (2001) has indicated, that rather than show declines, ganzfeld databases show so-called rebound effects. Indeed, we argue that ostensible declines could be explained by any number of maturational, historical, and/or environmental effects over the past 34 years, so that the declines themselves may be nothing but artifacts (see Bierman, 2001; Bierman et al., 1993; see also Storm & Ertel, 2001). In particular, we note Bierman's (2001) earlier finding of a ganzfeld rebound effect in the form of a flattened U-shaped (polynomial) curve for the period 1972–2001.

We assessed the evidence for an effect size decline in ganzfeld studies over a period of 34 years (1974 to 2008). Figure 2 shows F2 ES values plotted for 108 ganzfeld studies from five databases: (a) Honorton (1985; N = 28; period of analysis: 1974–1981), (b) Storm and Ertel (2001; N = 11; period of analysis: 1982–1989), (c) Bem and Honorton (1994; N = 10; period of analysis: 1983–1989), (d) Milton and Wiseman (1999; N = 30; period of analysis: 1989–1997), and (e) our ganzfeld database (N = 29; period of analysis: 1997–2008).

We note that the correlation between year of study and ES is negative and significant for the combined databases (N = 108), $r_s(106) = -.19$, p = .049 (two-tailed). This result indicates a linear decline in ESs over the 34-year period. However, a rebound effect is also indicated in the form of a significant quadratic polynomial curve: $\text{ES} = 0.0009 \times \text{YEAR}^2 + 3.4375 \times \text{YEAR} +$

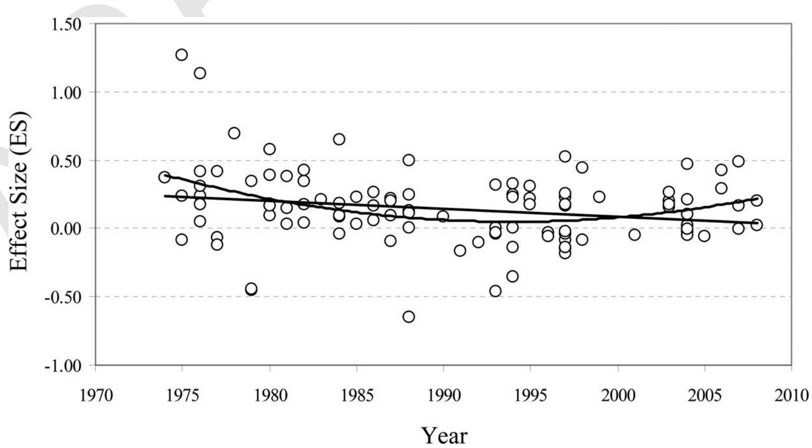

Figure 2. Scatterplot of ganzfeld studies over a period of 34 years (1974–2008). A slight but significant decline ($r = -.21$) is indicated (p = .029), although a significant rebound effect is also indicated in the form of a significant U-shaped curve (p = .002).

3427.2; $R^2 = .12$, $p = .004$ (see Figure 2). We tested a dual-line graphic representation of a 95% CI against our polynomial curve and found that our curve sits completely within that CI, so that we can be 95% confident that no effects reported during our 34-year period could be explained by chance.

Hyman (2008) relegates to a mere footnote his critique of a similar finding by Bierman (2001) of polynomial curves for two data sets, one of which is the ganzfeld database for the period 1972–2001:

> In two cases, Bierman [2001] suggests that after reaching zero, the effect size shows signs of increasing again. However, this is questionable and appears to be an artifact of fitting a polynomial to data where the zero effect size has existed for a while. Under such circumstances, a second-degree polynomial will better fit the data than will a linear regression line. (p. 43, Footnote 1)

However, we would argue that Hyman is attempting to predict a pattern in data that is 7 years old and not at all up-to-date at the time of his writing (i.e., 2008), and that much has changed for the ganzfeld since 2001. Bierman's earlier finding clearly anticipates the same effect we find in our more up-to-date analysis. When we tested for linear declines with the four outlier studies from the Honorton (1985) database removed, the association between study year and ES is negative, and only approaches significance, $r_s(102) = -0.15$, $p = .126$ (two-tailed), and thus is only suggestive of a linear decline. In other words, the decline may be attributable to these four outliers.

When the Milton and Wiseman (1999) database is removed as well, the relationship between study year and ES is relatively similar, and still not significant, $r_s(72) = -0.14$, $p = .242$ (two-tailed). The Milton and Wiseman database is not likely to have contributed to the significant linear decline in the larger database because the slope is about the same—the nonsignificant *p* value is more likely to be the result of a reduced (i.e., smaller) *N*. By way of a suggestive proof of this lack of contribution to the decline from the Milton and Wiseman database, when the four outliers are returned, but with the Milton and Wiseman database still excluded, the slope returns to that of the first correlation above for the full database ($N = 108$), and approaching significance, $r_s(76) = -0.19$, $p = .098$ (two-tailed).

We can assess the decline effect in another way by plotting the mean ES values for each of the five major databases. Figure 3 F3
shows the mean ES values and 95% CIs for the five databases. As can be seen from the figure, it is only one database out of five—that of Milton and Wiseman (1999)—for which the population estimate includes a zero ES. The upshot is that the Milton and Wiseman database may not be single-handedly responsible for a linear decline in the combined database of 109 studies, but it is an "outlier" database no less as stated by other parapsychologists (e.g., Bem et al., 2001; Schmeidler & Edge, 1999; Storm & Ertel, 2001). In contrast, our new database ($N = 30$), even with its relatively low mean ES, is still outside MCE.

We maintain that some caution is warranted in the interpretation of these results. Nevertheless, our findings indicate that the ganzfeld is one of the most consistent and reliable experimental paradigms in parapsychology.

Discussion

With respect to the hypotheses proposed in the present article, meta-analyses of the three categories—Category 1 (ganzfeld), Category 2 (non-Gz noise reduction), and Category 3 (standard

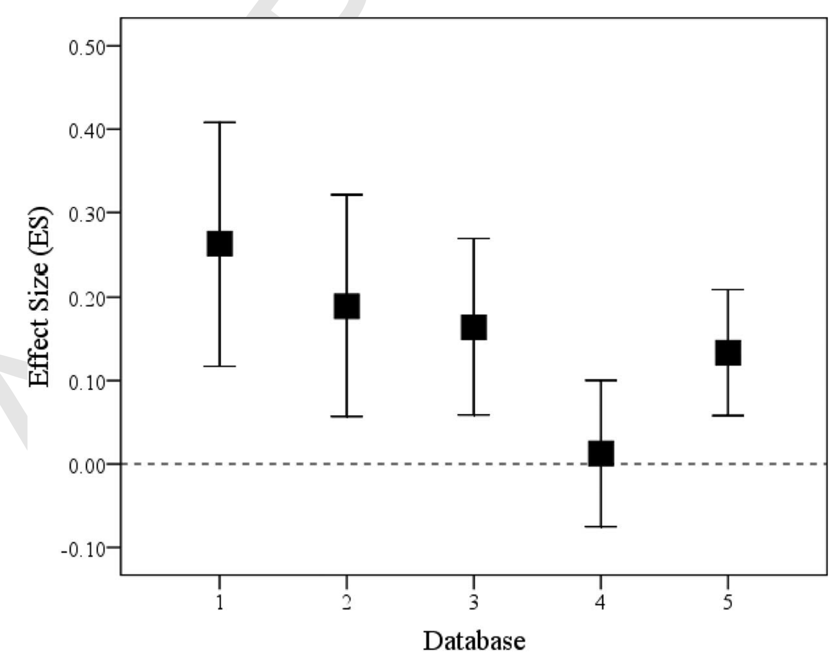

Figure 3. Comparison of five ganzfeld mean effect sizes (with 95% confidence interval): (1) Honorton (1985; $N = 28$), (2) Bem and Honorton (1994; $N = 10$), (3) Storm and Ertel (2001; $N = 11$), (4) Milton and Wiseman (1999; $N = 30$), and (5) Storm, Tressoldi, and Di Risio (present study; $N = 33$).

free response)—produced some fairly clear findings. The mean ES value for Category 1 (0.~~110~~2) was significantly higher than the mean ES for Category 2 (~~−0.110~~2) and Category 3 (−0.029). Our tests also showed that there was no performance difference (as measured by ES values) between unselected and selected participants in Categories 2 and 3, although the Category 1 studies (ganzfeld) did yield a significant difference between the two types. This effect is indicated by a significant interaction effect (see Figure 1). Thus, the ganzfeld seems to be particularly superior to the other two techniques, but only when selected participants are used. Interested psi researchers would do well to note these findings. Even though noise reduction techniques elicit statistical evidence for a communications anomaly, our findings suggest that researchers, if they plan to test selected participants, do so using pure ganzfeld techniques. Yet, even though psi research is nowadays largely process oriented, one still finds (a) recruitment of unselected participants to be about as frequent as recruitment of selected participants, and (b) target identification in nonganzfeld states is about as frequent as target identification in ganzfeld states. Note that for Conditions a and b, the ratio of selected to unselected participants, and ganzfeld to non-Gz studies, is about 1:1 in both cases.

For nearly two decades, ganzfeld pundits (e.g., Honorton & Ferrari, 1989; Morris, 1991) have been advising researchers to (a) select their participants through prior testing and/or training and (b) use a noise reduction technique. This combination still appears to be the best on offer as far as ES yields are concerned. It is probably for that reason that the ganzfeld paradigm has come to dominate the research agenda. However, other (simpler) treatments besides ganzfeld, such as meditation or relaxation, are still a reasonable option. We are aware that researchers' choices are often restricted by practical considerations (e.g., the necessity of recruiting a sufficient number of participants and/or running as many trials as possible, both of which are warranted under the argument from statistical power), so if practical considerations are a serious issue, researchers might be advised to run unselected (i.e., naive) participants through a nonganzfeld noise reduction treatment.

It may also be possible that the evidence obtained in the present study has a strong bearing on the debate about much-discussed declines in ganzfeld research, and the alleged unreliability of psi effects. Although we found evidence of a weak decline across five ganzfeld databases, we add that appearances can be deceiving: There is good evidence that the decline is "in decline," with effects showing an upward trend (see Figures 2 and 3). Furthermore, after we removed four outlier studies, there was only a marginally significant decline. In addition, with our study, the two significant databases (total $N = 45$), and only one small nonsignificant database ($N = 14$), as well as, more generally, only one dubious meta-analysis out of five (see Figure 3), the two negative assumptions made against ganzfeld research (i.e., that psi effects are in decline and are unreliable) are undermined. Whichever of our analyses is preferred, it appears that 34 years of ganzfeld research has more often than not produced a communications anomaly worth investigating further, as evidenced by the cumulative record: 74 (72.5%; $N = 102$) had positive z scores, even though only 27 (26.5%) of the 102 studies were independently significant ($\alpha \leq .05$). If 26.5% seems inconclusive or ambiguous, then in spite of, or due to, the statistical evidence, parapsychologists may still have some way to go to convince skeptics.

In summary, it appears that the noise reduction condition tends to produce stronger effects compared with standard free-response studies. This finding addresses a decade-old issue to do with the dubious efficacy of the ganzfeld (see W. G. Braud, 2002; Milton, 1997a). However, the only statistical evidence that selected participants have an advantage over unselected participants comes from the interaction effect that indicated performance was not the same across conditions, with ganzfeld seeming to be the condition that gave selected participants the advantage. A number of psychological variables, such as confidence, motivation, skill, and the like, may explain the difference.

The decline in ganzfeld effects, alleged to have been ongoing for a period of more than three decades, might possibly be considered a short-term (not a long-term) effect. In point of fact, we have also shown that the decline is an artifact of the presence of outliers. These findings directly address another old issue concerning skepticism that has emerged as a result of the ganzfeld decline, which some critics may have exclusively (and incorrectly) attributed to qualitative improvements and the elimination of flaws in the ganzfeld protocols.

In closing, we emphasize how important it is to free up this line of investigation from unwarranted skepticism and hasty judgments, so that these communication anomalies might be treated and investigated in like manner with other psychological functions.

References

References marked with a single asterisk indicate studies included in the meta-analyses for the present study (see Appendix A). References marked with a dagger indicate studies included in Appendix B from Storm and Ertel's (2001) study.

Akers, C. (1984). Methodological criticisms of parapsychology. In S. Krippner (Ed.), *Advances in parapsychological research* (Vol. 4, pp. 112–164). Jefferson, NC: McFarland.

Alcock, J. E. (1981). *Parapsychology: Science or magic?* New York, NY: Pergamon.

*Alexander, C. H., & Broughton, R. S. (1999). CL1-ganzfeld study: A look at brain hemisphere differences and scoring in the autoganzfeld. In *Proceedings of the 42nd Annual Convention of the Parapsychological Association* (pp. 3–18). Durham, NC: Parapsychological Association.

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin, 115*(1), 4–18. doi:10.1037/0033-2909.115.1.4

Bem, D. J., Palmer, J., & Broughton, R. S. (2001). Updating the ganzfeld database: A victim of its own success? *Journal of Parapsychology, 65,* 207–218.

Berger, R. E. (1989). A critical examination of the Blackmore psi experiments. *Journal of the American Society for Psychical Research, 83,* 123–144.

†Bierman, D. J. (1987). A test on possible implications of the OT's for Ganzfeld research. *European Journal of Parapsychology, 7,* 1–12.

Bierman, D. J. (2001). On the nature of anomalous phenomena: Another reality between the world of subjective consciousness and the objective world of physics? In P. Van Loocke (Ed.), *The physical nature of consciousness* (pp. 269–292). New York, NY: Benjamins.

†Bierman, D. J., Berendsen, J., Koenen, C., Kuipers, C., Louman, J., & Maisson, F. (1984). The effects of Ganzfeld stimulation and feedback in a clairvoyance task. In R. A. White & R. S. Broughton (Eds.), *Research in parapsychology 1983* (p. 14). Metuchen, NJ: Scarecrow.

Bierman, D. J., Bosga, D., Gerding, J. L. F., & Wezelman, R. (1993). Anomalous information access in the Ganzfeld: Utrecht—Novice series I and II. In *Proceedings of the 36th Annual Convention of the Parapsychological Association* (pp. 192–204). Durham, NC: Parapsychological Association.

Blackmore, S. J. (1985). The adventures of a psi-inhibitory experimenter. In P. Kurtz (Ed.), *A skeptic's handbook of parapsychology* (pp. 425–448). Buffalo, NY: Prometheus Books.

†Braud, L. W., Ackles, L., & Kyles, W. (1984). Free-response GESP performance during ganzfeld stimulation. In R. A. White & R. S. Broughton (Eds.), *Research in parapsychology 1983* (pp. 78–80). Metuchen, NJ: Scarecrow.

Braud, W. G. (2002). Psi-favorable conditions. In V. G. Rammohan (Ed.), *New frontiers of human science: A festschrift for K. Ramakrishna* (pp. 95–118). Jefferson, NC: McFarland.

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.

Collins, H. M., & Pinch, T. J. (1982). *Frames of meaning: The social construction of extraordinary science.* London, England: Routledge.

*Dalton, K. (1997). Exploring the links: Creativity and psi in the ganzfeld. In *Proceedings of the 40th Annual Convention of the Parapsychological Association* (pp. 119–134). Durham, NC: Parapsychological Association.

*Dalton, K., Steinkamp, F., & Sherwood, S. J. (1999). A dream GESP experiment using dynamic targets and consensus vote. *Journal of the American Society for Psychical Research, 96,* 145–166.

*Dalton, K., Utts, J., Novotny, G., Sickafoose, L., Burrone, J., & Phillips, C. (2000). Dream GESP and consensus vote: A replication. In *Proceedings of the 43rd Annual Convention of the Parapsychological Association* (pp. 74–85). Durham, NC: Parapsychological Association.

Darlington, R. B., & Hayes, A. F. (2000). Combining independent *p* values: Extensions of the Stouffer and binomial methods. *Psychological Methods, 5*(4), 496–515.

*da Silva, F. E., Pilato, S., & Hiraoka, R. (2003). Ganzfeld vs. no Ganzfeld: An exploratory study of the effects of Ganzfeld conditions on ESP. In *Proceedings of the 46th Annual Convention of the Parapsychological Association* (pp. 31–49). Durham, NC: Parapsychological Association.

*Del Prete, G., & Tressoldi, P. E. (2005). Anomalous cognition in hypnagogic state with OBE induction: An experimental study. *Journal of Parapsychology, 69,* 329–339.

~~*Goulding, A. (2005). Participant variables associated with psi Ganzfeld results. *European Journal of Parapsychology, 20,* 50–64.~~

*Goulding, A., Westerlund, J., Parker, A., & Wackermann, J. (2004). The first digital autoganzfeld study using a real-time judging procedure. *European Journal of Parapsychology, 19,* 66–97.

Hansel, C. E. M. (1966). *ESP: A scientific evaluation.* London, England: McGibbon & Kee.

Hansel, C. E. M. (1980). *ESP and parapsychology: A critical reevaluation.* Buffalo, NY: Prometheus Books.

†Haraldsson, E., & Gissurarson, L. R. (1985). Perceptual defensiveness, Ganzfeld and the percipient-order effect. *European Journal of Parapsychology, 6,* 1–17.

*Holt, N. J. (2007). Are artistic populations psi-conducive? Testing the relationship between creativity and psi with an experience-sampling protocol. In *Proceedings of the 50th Annual Convention of the Parapsychological Association* (pp. 31–47). Petaluma, CA: Parapsychological Association.

*Holt, N. J., & Roe, C. A. (2006). The sender as a PK agent in ESP studies: The effects of agent and target system lability upon performance at a novel PK task. *Journal of Parapsychology, 70,* 49–67.

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

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

Honorton, C. (1997). The ganzfeld novice: Four predictors of initial ESP performance. *Journal of Parapsychology, 61,* 143–158.

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., & 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.

†Houtkooper, J. M., Gissurarson, L. R., & Haraldsson, E. (1988–1989). Why the ganzfeld is conducive to ESP: A study of observational theory and the percipient-order effect. *European Journal of Parapsychology, 7,* 169–192.

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

Hyman, R. (1989). *The elusive quarry: A scientific appraisal of psychical research.* Buffalo, NY: Prometheus Books.

Hyman, R. (2008). Anomalous cognition? A second perspective. *Skeptical Inquirer, 32*(4), 40–43.

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

Kanthamani, H., Khilji, A., & Rustomji-Kerns, R. (1988). An experiment in ganzfeld and dreams with a clairvoyance technique. In *Proceedings of the 31st Annual Convention of the Parapsychological Association* (pp. 412–423). Durham, NC: Parapsychological Association.

Kanthamani, H., & Palmer, J. (1993). A ganzfeld experiment with "subliminal sending." *Journal of Parapsychology, 57,* 241–257.

*Lau, M. (2004). *The PSI phenomena: A Bayesian approach to the Ganzfeld procedure* (Unpublished master's thesis). University of Notre Dame, South Bend, IN.

Lawrence, T. R. (1993). Gathering in the sheep and goats: A meta-analysis of forced-choice sheep-goat ESP studies, 1947–1993. In *Proceedings of the 36th Annual Convention of the Parapsychological Association* (pp. 75–86). Durham, NC: Parapsychological Association.

*May, E. C. (2007). Advances in anomalous cognition analysis: A judge-free and accurate confidence-calling technique. In *Proceedings of the 50th Annual Convention of the Parapsychological Association* (pp. 57–63). Petaluma, CA: Parapsychological Association.

†Milton, J. (1987). Judging strategies to improve scoring in the ganzfeld. In D. H. Weiner & R. D. Nelson (Eds.), *Research in parapsychology 1986* (pp. 100–103). Metuchen, NJ: Scarecrow.

†Milton, J. (1988–1989). A possible "directive" role of the agent in the ganzfeld. *European Journal of Parapsychology, 7,* 193–214.

Milton, J. (1997a). Meta-analysis of free-response ESP studies without altered states of consciousness. *Journal of Parapsychology, 61,* 279–319.

Milton, J. (1997b). A meta-analytic comparison of the sensitivity of direct hits and sums of ranks as outcome measures for free-response studies. *Journal of Parapsychology, 61,* 227–241.

Milton, J., & Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin, 125*(4), 387–391. doi:10.1037/0033-2909.125.4.387

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

*Morris, R. L., Summers, J., & Yim, S. (2003). Evidence of anomalous information transfer with a creative population in ganzfeld stimulation. In *Proceedings of the 46th Annual Convention of the Parapsychological*

Association (pp. 116–131). Durham, NC: Parapsychological Association.
†Murre, J. M. J., van Dalen, A. C., Dias, L. R. B., & Schouten, S. A. (1988). A ganzfeld psi experiment with a control condition. *Journal of Parapsychology, 52,* 103–125.
Palmer, J. (1977). Attitudes and personality traits in experimental ESP research. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 175–201). New York, NY: Van Nostrand Reinhold.
Palmer, J. (1986). Comments on the "joint communiqué." *Journal of Parapsychology, 50,* 377–381.
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. (2000). A review of ganzfeld work at Gothenburg University. *Journal of the Society for Psychical Research, 64,* 1–15.~~
*Parker, A., & Sjödén, B. (2008). The subliminal priming of film clips used in the Ganzfeld. *Proceeding of the 51st Annual Convention of the Parapsychological Association* (pp. 356–359). Columbus, OH: Parapsychological Association.
*Parker A., & Westerlund, J. (1998). Current research in giving the ganzfeld an old and a new twist. In *Proceedings of the 41st Annual Convention of the Parapsychological Association* (pp. 135–142). Durham, NC: Parapsychological Association.
~~*Parra, A., & Villanueva, J. (2003). Personality factors and psi–ganzfeld sessions: A replication and extension. *Australian Journal of Parapsychology, 3,* 159–174.~~
*Parra, A., & Villanueva, J. (2004). Are musical themes better than visual images as ESP-targets? An experimental study using the ganzfeld technique. *Australian Journal of Parapsychology, 4,* 114–127.
*Parra, A., & Villanueva, J. (2006). ESP under the ganzfeld, in contrast with the induction of relaxation as a psi-conducive state. *Australian Journal of Parapsychology, 6,* 167–185.
*Pütz, P., Gäßler, M., & Wackermann, J. (2007). An experiment with covert ganzfeld telepathy. *European Journal of Parapsychology, 22,* 49–72.
Radin, D. I., & Ferrari, D. C. (1991). Effects of consciousness on the fall of dice: A meta-analysis. *Journal of Scientific Exploration, 5,* 61–83.
Rhine, J. B. (1954). *The reach of the mind.* Harmondsworth, England: Pelican/Penguin. (Original work published in 1948)
Rhine, J. B., Pratt, J. G., Stuart, C. E., Smith, B. M., & Greenwood, J. A. (1966). *Extra-sensory perception after sixty years.* Boston, MA: Humphries. (Original work published in 1940)
*Roe, C. A., & Flint, S. (2007). A remote viewing pilot study using a ganzfeld induction procedure. *Journal of the Society for Psychical Research, 71,* 230–234.
~~*Roe, C. A., & Holt, N. (2005). A further consideration of the sender as a PK agent in ganzfeld ESP studies. *Journal of Parapsychology, 69,* 113–127.~~
*Roe, C. A., & Holt, N. J. (2006). The effects of strategy ("willing" versus absorption) and feedback (intermediate versus delayed) on performance at a PK task. *Journal of Parapsychology, 70,* 69–90.
*Roe, C. A., Holt, N. J., & Simmonds, C. A. (2003). Considering the sender as a PK agent in ganzfeld ESP studies. *Journal of Parapsychology, 67,* 129–145.
*Roe, C., Jones, L., & Maddern, C. (2007). A preliminary test of the "four-factor model" using a dream ESP protocol. *Journal of the Society for Psychical Research, 71*(1), 35–42.
*Roe, C. A., McKenzie, E. A., & Ali, A. N. (2001). Sender and receiver creativity scores as predictors of performance at a ganzfeld ESP task. *Journal of the Society for Psychical Research, 65,* 107–121.
*Roe, C. A., Sherwood, S. J., Farrell, L., Savva, L., & Baker, I. S. (2007). Assessing the roles of the sender and experimenter in dream ESP research. *European Journal of Parapsychology, 22,* 175–192.
*Roe, C. A., Sherwood, S. J., & Holt, N. C. (2004). Interpersonal psi: Exploring the role of the sender in Ganzfeld GESP tasks. *Journal of Parapsychology, 68,* 361–380.
*Roe, C. A., Sherwood, S. J., Luke, D. P., & Farrell, L. M. (2002). An exploratory investigation of dream GESP using consensus judging and dynamic targets. *Journal of the Society for Psychical Research, 66,* 225–238.
*Roney-Dougal, S. M., & Solfvin, J. (2008). Exploring the relationship between two Tibetan meditation techniques, the Stroop effect and precognition. In *Proceedings of the 51st Annual Convention of the Parapsychological Association* (pp. 187–203). Columbus, OH: Parapsychological Association.
*Roney-Dougal, S. M., Solfvin, J., & Fox, J. (2008). An exploration of degree of meditation attainment in relation to psychic awareness with Tibetan Buddhists. *Journal of Scientific Exploration, 22,* 161–178.
Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*(2), 183–192. doi:10.1037/0033-2909.118.2.183
Rosenthal, R., & Rubin, D. B. (1979). Comparing significance levels of independent studies. *Psychological Bulletin, 86*(5), 1165–1168. doi:10.1037/0033-2909.86.5.1165
†Sargent, C. L. (1982). A ganzfeld GESP experiment with visiting subjects. *Journal of the Society for Psychical Research, 51,* 222–232.
†Sargent, C. L., & Harley, T. A. (1982). Precognition testing with free-response techniques in the ganzfeld and the dream state. *European Journal of Parapsychology, 4,* 243–256.
Schmeidler, G., & Edge, H. (1999). Should ganzfeld research continue to be crucial in the search for a replicable psi effect? Part II. Edited ganzfeld debate. *Journal of Parapsychology, 63,* 335–388.
*Sherwood, S. J., Dalton, K., Steinkamp, F., & Watt, C. (2000). Dream clairvoyance study II using dynamic video-clips: Investigation of consensus voting judging procedures and target emotionality. *Dreaming, 10,* 221–236.
*Sherwood, S. J., Roe, C. A., Holt, N. J., & Wilson, S. (2005). Interpersonal psi: Exploring the role of the experimenter and the experimental climate in a ganzfeld telepathy task. *European Journal of Parapsychology, 20,* 150–172.
*Sherwood, S. J., Roe, C. A., Simmonds, C. A., & Biles, C. (2002). An exploratory investigation of dream precognition using consensus judging and static targets. *Journal of the Society for Psychical Research, 66,* 22–28.
*Simmonds, C. A., & Fox, J. (2004). A pilot investigation into sensory noise, schizotypy, and extrasensory perception. *Journal of the Society for Psychical Research, 68,* 253–261.
*Simmonds-Moore, C., & Holt, N. J. (2007). Trait, state, and psi: A comparison of psi performance between clusters of scorers on schizotypy in a ganzfeld and waking control condition. *Journal of the Society for Psychical Research, 71,* 197–215.
*Smith, M. D., & Savva, L. (2008). Experimenter effects in the ganzfeld. In *Proceedings of the 51st Annual Convention of the Parapsychological Association* (pp. 238–249). Columbus, OH: Parapsychological Association.
†Sondow, N. (1987). Exploring hypnotizability, creativity, and psi: Conscious and unconscious components to psi success in the ganzfeld. In D. H. Weiner & R. D. Nelson (Eds.), *Research in Parapsychology 1986* (pp. 42–47). Metuchen, NJ: Scarecrow.
*Steinkamp, F. (2000). Does precognition foresee the future? A postal experiment to assess the possibility of true precognition. *Journal of Parapsychology, 64,* 3–18.
*Steinkamp, F. (2001). Does precognition foresee the future? Series 2, A laboratory replication and Series 3, A World Wide Web replication. *Journal of Parapsychology, 65,* 17–40.
*Steinkamp, F. (2005). Does precognition foresee the future? Series 4: A postal replication. *Journal of Parapsychology, 69,* 341–351.
*Stevens, P. (2004). Experimental evaluation of a feedback-reinforcement model for dyadic ESP. *Journal of Parapsychology, 68,* 65–92.

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

*Storm, L. (2003). Remote viewing by committee: RV using a multiple agent/multiple percipient design. *Journal of Parapsychology, 67,* 325–342.

*Storm, L., & Barrett-Woodbridge, M. (2007). Psi as compensation for modality impairment—A replication study using sighted and blind participants. *European Journal of Parapsychology, 22,* 73–89.

Storm, L., & Ertel, S. (2001). Does psi exist? Comments on Milton and Wiseman's (1999) meta-analysis of ganzfeld research. *Psychological Bulletin, 127*(3), 424–433. doi:10.1037/0033-2909.127.3.424

Storm, L., & Ertel, S. (2002). The ganzfeld debate continued: A response to Milton and Wiseman (2001). *Journal of Parapsychology, 66,* 73–82.

*Storm, L., & Thalbourne, M. A. (2001). Paranormal effects using sighted and vision-impaired participants in a quasi-ganzfeld task. *Australian Journal of Parapsychology, 1,* 133–170.

*Symmons, C., & Morris, R. L. (1997). Drumming at seven Hz and automated ganzfeld performance. In *Proceedings of the 40th Annual Convention of the Parapsychological Association* (pp. 441–454). Durham, NC: Parapsychological Association.

*Targ, R., & Katra, J. E. (2000). Remote viewing in a group setting. *Journal of Scientific Exploration, 14,* 107–114.

Thalbourne, M. A. (2003). *A glossary of terms used in parapsychology.* Charlottesville, VA: Puente.

*Tressoldi, P. E., & Del Prete, G. (2007). ESP under hypnosis: The role of induction instructions and personality characteristics. *Journal of Parapsychology, 71,* 125–137.

Utts, J. (1991). Replication and meta-analysis in parapsychology. *Statistical Science, 6*(4), 363–403. doi:10.1214/ss/1177011577

*Watt, C., & Wiseman, R. (2002). Experimenter differences in cognitive correlates of paranormal belief, and in psi. *Journal of Parapsychology, 66,* 371–385.

*Wezelman, R., & Bierman, D. J. (1997). Process oriented ganzfeld research in Amsterdam. In *Proceedings of the 40th Annual Convention of the Parapsychological Association* (pp. 477–491). Durham, NC: Parapsychological Association.

*Wezelman, R., Gerding, J. L. F., & Verhoeven, I. (1997). Eigensender ganzfeld psi: An experiment in practical philosophy. *European Journal of Parapsychology, 13,* 28–39.

*Wright, T., & Parker, A. (2003). An attempt to improve ESP scores using the real time digital ganzfeld technique. *European Journal of Parapsychology, 18,* 65–72.

Appendix A

Ganzfeld (Gz), Non-Gz (Noise Reduction), and Free-Response Studies by Category, Trials, *Z* Scores, and Effect Size Values

Study × Category	Trials	Hits	*Z*	*ES* ($z/\sqrt{n}$)
Category 1 (ganzfeld)				
1. Alexander and Broughton (1999)	50	18	1.60	0.23
2. Dalton (1997)	128	60	5.20	0.46
3. da Silva et al. (2003)	54	18	1.26	0.17
4. Goulding et al. (2004)	128	30	−0.31	−0.03
5. Lau (2004)	120	36	1.16	0.11
6. Morris et al. (2003)	40	15	1.64	0.26
7. Parker (2006)	20	8	1.29	0.34
8. Parker and Sjödén (2008)	29	8	0.11	0.02
9. Parker and Westerlund (1998), serial study	30	7	−0.49	−0.09
10. Parker and Westerlund (1998), Study 4	30	14	2.40	0.44
11. Parker and Westerlund (1998), Study 5	30	11	1.25	0.23
12. Parra and Villanueva (2004), picture	54	25	3.46	0.47
13. Parra and Villanueva (2004), musical clips	54	19	1.57	0.21
14. Parra and Villanueva (2006)	138	57	4.32	0.37
15. Pütz et al. (2007)	120	39	1.79	0.16
16. Roe and Flint (2007)	14	4	1.81	0.48
17. Roe et al. (2003)	40	14	1.28	0.20
18. Roe et al. (2001)	24	5	−0.24	−0.05
19. Roe et al. (2004), no sender	17	4	−0.19	−0.05
20. Roe et al. (2004), with sender	23	6	0.12	0.03
21. Sherwood et al. (2005)	38	8	−0.37	−0.06
22. Simmonds-Moore and Holt (2007)	26	6	−0.04	−0.01
23. Smith and Savva (2008)	114	39	2.16	0.20
24. Stevens (2004)	50	12	−0.03	−0.01
25. Symmons and Morris (1997)	51	23	2.97	0.42

(*Appendices continue*)

Appendix (*continued*)

Study × Category	Trials	Hits	*Z*	*ES* ($z/\sqrt{n}$)
26. Wezelman and Bierman (1997), Series IVB	32	5	−1.45	−0.26
27. Wezelman and Bierman (1997), Series V	40	8	−0.91	−0.14
28. Wezelman and Bierman (1997), Series VI	40	10	−0.15	−0.02
29. Wezelman et al. (1997)	32	14	2.15	0.38
30. Wright and Parker (2003)	74	24	1.34	0.16
Category 2 (nonganzfeld noise reduction)				
1. Dalton et al. (1999)	32	15	2.65	0.47
2. Dalton et al. (2000)	16	7	1.41	0.35
3. Del Prete and Tressoldi (2005)	120	45	3.06	0.28
4. Roe, Jones, and Maddern (2007)	15	2	−0.72	−0.19
5. Roe, Sherwood, et al. (2007), clairvoyance	40	14	1.28	0.20
6. Roe, Sherwood, et al. (2007), telepathy	40	12	0.55	0.09
7. Roney-Dougal and Solfvin (2008)	80	25	1.16	0.13
8. Roney-Dougal et al. (2008), clairvoyance	48	11	−0.17	−0.02
9. Roney-Dougal et al. (2008), precognition	48	7	−1.55	−0.22
~~10. Roney-Dougal & Solfvin (2008)~~	~~80~~	25	1.16	0.13
11. Sherwood et al. (2000)	28	11	1.53	0.29
12. Sherwood et al. (2002)	12	2	−0.28	−0.08
13. Steinkamp (2001), Series 2, clairvoyance	80	23	0.65	0.07
14. Steinkamp (2001), Series 2, precognition	80	26	1.42	0.16
15. Steinkamp (2005)	80	20	0.00	0.00
16. Tressoldi and Del Prete (2007)	120	40	2.00	0.18
Category 3 (free response)				
1. da Silva et al. (2003)	54	10	−0.94	−0.13
2. Holt (2007), artists	15	6	1.04	0.27
3. Holt (2007), artists	15	7	1.58	0.41
4. Holt and Roe (2006)	40	10	0.00	0.00
5. Lau (2004)	937	232	−0.12	−0.01
6. May (2007)	50	32	4.57	0.65
7. Parra and Villanueva (2006)	138	38	0.59	0.05
8. Roe and Holt (2006)	120	28	−0.32	−0.03
9. Roney-Dougal et al. (2008) clairvoyance	24	5	−0.24	−0.05
10. Roney-Dougal et al. (2008) precognition	24	4	−0.71	−0.15
11. Simmonds and Fox (2004), walking controls	20	2	−1.29	−0.29
12. Simmonds-Moore and Holt (2007)	26	8	0.45	0.09
13. Steinkamp (2000) clairvoyance	74	17	−0.27	−0.03
14. Steinkamp (2000) precognition	75	16	−0.60	−0.03
15. Steinkamp (2001) Series 3, precognition	100	21	−0.81	−0.08
16. Steinkamp (2001) Series 3, clairvoyance	100	28	0.58	0.06
17. Storm (2003)	10	5	1.84	0.58
18. Storm and Barrett-Woodbridge (2007)	76	16	−0.66	−0.08
19. Storm and Thalbourne (2001)	84	22	0.13	0.01
20. Targ and Katra (2000)	24	14	3.54	0.72
21. Watt and Wiseman (2002)	58	17	1.61	0.21

Note. Studies that gave different *z* scores in their original sources differ due to the fact that authors of those studies calculated *z* scores from sums-of-rank scores—not from direct hits, as we have done.

(*Appendices continue*)

Appendix B

Number of Trials, Adjusted *Z* Scores, and Effect Sizes From Storm and Ertel's (2001) Study

Study	Trials	Hits	*z* score	Effect size
Bierman (1987)	16	6	0.88[a]	0.22[a]
Bierman et al. (1984)	32	11	1.02	0.18
L. W. Braud et al. (1984)	10	6	2.06[a]	0.65[a]
Haraldsson and Gissurarson (1985)	70	19	0.28	0.03
Houtkooper et al. (1988–1989)	40	10	0.00	0.00
Milton (1987)	37	13	1.23	0.20
Milton (1988–1989)	35	13	1.46[a]	0.25[a]
Murre et al. (1988)	41	13	0.81	0.13
Sargent (1982)	20	7	0.79	0.18
Sargent and Harley (1982)	44	18	2.26	0.34
Sondow (1987)	60	12	−0.75	−0.10

[a] The *z* scores and effect sizes are adjusted from those given in Storm and Ertel (2001, p. 428, Table 1).

Appendix C

Formulae

File-Drawer Statistics

The formula given by Rosenthal (1995, p. 189), $X = [(\Sigma Z)^2/2.706] - k$, was used to calculate estimates of the number of studies averaging null results needed to reduce significant probability values to chance values (i.e., $p = .05$). The *k* value refers to the number of studies retrieved for the relevant meta-analysis.

Calculation of *Z* Scores

Rosenthal and Rubin's (1979) Z_{diff} formula is $Z_{\text{diff}} = (Z_1 - Z_2)/\sqrt{2}$. In regard to testing the free-response databases in the section Comparisons With Other Databases, we note that the Z_{diff} formula produces a conservative *z* difference that may overestimate the subsequent *p* value so that we may be making a Type II error in our calculations (i.e., we may not be finding statistical evidence of a difference when there is one). However, Milton (1997b) did not provide details of hit rates or *z* scores, so we would not be able to combine the two databases. We merely point out the possibility that the performance rate of the typical free-response experiment has not changed in over 45 years (i.e., since 1964).

Received September 17, 2009
Revision received February 17, 2010
Accepted February 24, 2010 ■