META-ANALYSIS OF FREE-RESPONSE STUDIES 2009–2018: ASSESSING THE NOISE-REDUCTION MODEL TEN YEARS ON

Lance Storm and Patrizio E. Tressoldi

ABSTRACT

We report the results of an update to the meta-analysis by Storm, Tressoldi, and Di Risio (2010). Three laboratory-based free-response designs were assessed: (i) ganzfeld (a technique proposed to enhance a communication anomaly referred to as extra-sensory perception; ESP); (ii) nonganzfeld noise reduction using alleged ESP-enhancing techniques (e.g., dreaming, hypnosis, relaxation, and meditation); and (iii) standard free-response (nonganzfeld; non-noise-reduction). For the period 2009 to 2018, a dataset of nine new ganzfeld studies (Category 1) yielded a mean effect size (*ES*) of 0.119 (where $ES = z/\sqrt{n}$); 19 new nonganzfeld noise-reduction studies (Category 2) yielded mean *ES* = 0.045; and 15 new standard free-response studies (Category 3) yielded mean *ES* = 0.050. Stouffer *Z* scores for all three databases were significant, but each new database was not significantly different from its respective database in Storm et al. (2010). The increased ganzfeld database ($N = 38$) yielded a mean *ES* = 0.133; the nonganzfeld noise-reduction database ($N = 37$) yielded a mean *ES* = 0.072; and the standard free-response database ($N = 33$) yielded mean *ES* = 0.027. Again, Stouffer *Z* scores were significant. We found Category 1 had a significantly higher mean effect than Category 3, and participants who were *selected* for the studies (they believed in psi, or practiced, say, meditation, etc.) produced a stronger mean effect than *unselected* (randomly assigned) participants, but there were no differences between type of modality tested (i.e., telepathy, clairvoyance, precognition), types of target (film clips, photographs, etc.), or experimenter/laboratory. There was also no evidence of a decline effect across a 44-year period. Finally, we conducted a Bayesian analysis and found that the case for a communication anomaly using free-response designs was upheld.

Introduction

The present study is a meta-analytic review of the free-response extrasensory perception (ESP) literature for the ten-year period 2009 to 2018. This study picks up from where Storm, Tressoldi, and Di Risio (2010) left off in 2008. The term *free response* "describes any test of ESP in which the range of possible targets is relatively unlimited and is unknown to the percipient" (Thalbourne, 2003, p. 44). One common form of free response design features an alleged ESP-enhancing, 'noise-reduction' technique known as *Ganzfeld*, which is "homogenous, unpatterned sensory stimulation" to the eyes and ears of the participant, who is usually in "a state of bodily comfort"

(Thalbourne, 2003, p. 45). Some free response studies feature any kind of noise-reduction technique (e.g., dreaming, hypnosis, relaxation, meditation), but not a ganzfeld condition. A third type of free response design is referred to as 'standard' because neither ganzfeld, nor noise-reduction techniques are used. Hence, Storm et al. (2010) categorized free-response studies in three ways: (i) ganzfeld; (ii) nonganzfeld noise-reduction; and (iii) standard free-response (i.e., nonganzfeld non-noise-reduction).

In this 'update' paper, we expected there would be sufficient studies accumulated in the ten years since 2008 to form three sizeable free-response databases in the same three categories, with the specific aim of determining if the newly compiled databases ('new' studies) would still demonstrate significant evidence of ESP, and we would combine these databases, where possible, with their respective earlier ('old') databases to cover a 27-year period (1992 to 2018).

The ganzfeld meta-analyses

The term 'free-response' implies a manifested 'mentation' that describes a real-time or future-generated target, both of which are outside normal sensory range. The rationale behind this protocol is that information about the target is embedded in the stream of images, thoughts, and ideas that arise in the mind of the perceiver. Mentation can either be used by the percipient to guide his/her target selection, or it can be used to assist in an independent judging process. There are three conjectured ESP modalities that are most commonly the subject of parapsychological research: (i) *telepathy* (mind-to-mind communication); (ii) *clairvoyance* (distant-seeing), and (iii) *precognition* (future telling), though a rarely featured fourth modality is also acknowledged, namely *retrocognition* (using ESP to target past events). Collectively, these anomalies are known as 'psi', a shorthand term for psychic functioning.[1] Storm et al. (2012) compared the three main ESP modalities (telepathy, clairvoyance, and precognition), but found no significant difference in effect size (no difference was also found by Storm et al., 2017).

The free-response domain has proved successful for parapsychologists, as shown by Milton (1997b) who conducted a meta-analysis of studies on participants who were in a normal waking state of consciousness during a free-response task. She found a weak but significant effect. The ganzfeld protocol has proven successful as well, with many studies having been subjected to meta-analytic treatment (e.g., Honorton, 1985; Honorton et al., 1990; Storm & Ertel, 2001; Storm et al., 2010).

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 mean chance expectation (MCE) was 25%. Various flaws in his approach were pointed out by Hyman (1985), but ultimately there was agreement that "there is an overall significant effect in this database that cannot reasonably be explained by selective reporting or multiple analysis" (Hyman & Honorton, 1986, p. 351). Hyman

---

[1] A fifth psi phenomenon, *psychokinesis* (PK; a.k.a. telekinesis), describes paranormal mental influence on matter.

and Honorton disagreed mainly on the "degree to which the effect constitutes evidence for psi" (p. 351).

A second major meta-analysis on a set of 'autoganzfeld' studies followed (Honorton et al., 1990). These studies adhered to the guidelines laid down by Hyman and Honorton (1986), but the autoganzfeld procedure avoids methodological flaws by using a computer-controlled target randomization, selection, and judging technique. They reported a hit rate of 32.2%. Milton and Wiseman (1999) followed with their meta-analysis of 30 studies collected for the period 1987 to 1997; they reported an overall nonsignificant *ES* of 0.013. Milton and Wiseman concluded that a significant communication anomaly for the ganzfeld had not been replicated by a "broader range of researchers" (p. 391). However, Jessica Utts (personal communication, December 11, 2009) using the exact binomial test on trial counts only in Milton and Wiseman's database ($N = 1198$; Hits = 327), found a significant hit rate of 27% ($p = .036$).

Storm and Ertel (2001) compared Milton and Wiseman's (1999) database with Bem and Honorton's (1994) database of ten studies and found the two did not differ significantly. A homogeneous database of 40 studies was formed; mean effect size (*ES*) = 0.05. Storm and Ertel went on to compile a 79-study database, which had a significant mean *ES* of 0.138. Studies published during the period 1997 to 2008 were then collected by Storm, Tressoldi, and Di Risio (2010), who formed a database of 29 ganzfeld studies yielding an *ES* of 0.14. They also formed a nonganzfeld noise-reduction database of 16 studies (*ES* = 0.11), and a third database of 14 standard free-response studies (*ES* = –0.03). All except this third database were significant overall, although we now consider that the removal of outliers from that nonsignificant database may have been excessive (we return to this problem in our *Post Hoc Analyses of Categories* section below).

Rouder, Morey, and Province (2013) reassessed Storm et al.'s (2010) meta-analysis, but they also conducted a Bayesian analysis on an entirely different dataset that was compiled with less than desirable precision. Storm, Tressoldi, and Utts (2013) noted that Rouder et al.'s Bayesian approach was not without merit *in principle*, and although Rouder et al. found evidence for the existence of psi *in the original dataset* by a factor of about 6 billion to 1, much of this effect was attributed to "difficulties in randomization" (p. 241). Rouder et al. argued that ganzfeld studies with computerized randomization supposedly yielded smaller psi effects than those with manual randomization — their revised Bayes factor evidence was a reasonable "330 to 1" (p. 241), which they regarded as "small and inconsequential" (p. 246). Nevertheless, Storm et al. (2013) showed that this conclusion was unconvincing as it was based on Rouder et al.'s faulty and inconsistent compilation methodology. Storm et al.'s (2013) own Bayesian analysis yielded contradictory evidence; "clear superiority of the combined ganzfeld and nonganzfeld noise-reduction studies emerges" (p. 252). Storm et al. suggested that the noise-reduction psi effect lies somewhere between 26% and 32% (where MCE = 25%).

Baptista, Derakhshani, and Tressoldi (2015) conducted a review of the ganzfeld databases and made various recommendations to researchers that might improve effect sizes. For example, they recommend the "exclusive use

of selected participants" (e.g., those who believe in the possibility of psi, or had prior spontaneous psi experiences, or had prior psychic training, or regularly practiced some type of mental discipline such as relaxation or meditation, etc.). Baptista et al. also advised experimenters to adhere to the standard ganzfeld methodology (pp. 198–199). Our evaluations in the present study are aimed at testing these claims.

Design of the present study

In this meta-analysis, we seek to test the claim of evidence for ESP effects in the databases of free-response studies dating from 2009 to 2018. We will check if the three 'new' databases (ganzfeld, nonganzfeld noise-reduction, and standard free-response) are independently significant, and if each is not significantly different from their corresponding 'old' database, as formed by Storm et al. (2010), we will combine them to form three larger datasets. Thus, we may be able to demonstrate the degree to which ESP effects differ according to the free-response methodologies used, with some methodologies possibly performing better than others. We make this point because of the alleged importance of noise reduction, which is considered 'psi conducive', thereby allegedly reducing irrelevant background noise, leaving mainly the psi signal.

We will assess the three major psi modalities (telepathy, clairvoyance, and precognition) to see if there is a difference in psi strength between them. This analysis was also conducted by Storm et al. (2010, p. 475). There were no significant differences between the three modalities in the free-response experimental domain, (Storm, Tressoldi, & Di Risio, 2012).[2] We expect the same will be found for free-response.

We will also test whether the ESP effect depends on (a) target type (pictures/photos, 3D objects, and video clips), and (b) participant type (selected vs. unselected) for reasons given above, since past research does suggest there are differences between selected and unselected participants (Storm et al., 2010).

Finally, we will scrutinize any suggested ESP effects in the context of their possible dependence on study quality, year of study, or a select group of researchers/laboratories. We will evaluate the file-drawer problem, and the possibility of decline/incline effects. We had aimed to test the *k*-choices hypothesis (where k = count of target + decoys in the array), based on Timm's (2000) claim that the number of target choices in a given trial is related to effect size (see Storm et al., 2012). However, there was not enough variability in k to run such a test. Nevertheless, we report statistics for studies grouped by k.

For convenience, the following hypotheses refer to the newly-found studies only. In the section *Post hoc hypotheses*, we will re-test the same hypotheses on the combined databases.

---

[2] The typical "forced-choice" experiment has a limited number of choices, and the participant is forced to guess the target from a limited set of symbols (e.g., one of five in the card-guessing design: square, cross, circle, star, and wavy lines).

Hypotheses

1. Free-response studies produce statistical evidence of a communication anomaly known as ESP.
2. There is a psi-scoring difference between the three categories: (i) ganzfeld, (ii) nonganzfeld noise-reduction, and (iii) standard free-response (i.e., nonganzfeld, non-noise-reduction).
3. There is a difference in psi strength between telepathy, clairvoyance, and precognition studies.
4. Effect size values vary in strength according to target types (pictures/photos, 3D objects, and video clips).
5. Selected participants (believers in the paranormal, meditators, etc.) have a performance advantage over unselected participants.

METHOD

Study retrieval

Suitable studies were sought from a range of peer-reviewed journals, including: *Australian Journal of Parapsychology*, *Consciousness and Cognition*, *Dreaming*, *European Journal of Parapsychology*, *Europe's Journal of Psychology*, *Journal of Cognitive Neuroscience*, *Journal of Parapsychology*, *Journal of Scientific Exploration*, *Journal of the Society for Psychical Research*, and *Proceedings of the Annual Convention of the Parapsychological Association*. The search period was between 2009 to 2018 inclusive, which continues from Storm et al. (2010) — their period of assessment was 1992 to 2008 inclusive.

Internet searches through EBSCOhost of the relevant databases, were conducted, including searches in PsycINFO, PsycARTICLES, and other relevant databases (e.g., Medline, Web of Science, Lexscien, and Informit). The following keywords and subject headings were entered in the search: 'anomalous cognition', 'clairvoyance', 'extrasensory perception', 'ESP', 'ganzfeld', 'paranormal', 'parapsychology', 'precognition', 'psi', and 'telepathy'.

Selection criteria

The following six selection criteria were adopted, and studies were excluded if any of these criteria were not met:

- Studies must use a free-response design;
- Studies must test ESP (i.e., telepathy, clairvoyance, and precognition) only (therefore excluding studies that expressly tested psychokinesis; no studies tested retrocognition);
- Studies must use human participants only (not animals);
- Number of participants must be in excess of two to avoid the problems and idiosyncrasies inherent in case studies;
- Target selection must be randomized by using a Random Number Generator (RNG) in a computer or similar electronic device, or a table of random numbers;
- Studies must provide sufficient information (e.g., number of trials and outcomes) for the authors to calculate the direct hit-rates and

effect size values, so that appropriate statistical tests can be conducted.

Procedure

- Databases (Excel and SPSS) were prepared that listed all the free-response studies found in our searches for the period 2009 to 2018;
- A table was derived from the database (see Table A1 in the Appendix). Data in the table includes authors and publication year, category (e.g., ganzfeld), trial counts, and hits. Despite Milton's (1997a) general finding of nonsignificant differences between direct hits and sum-of-ranks statistics, direct hits is our main measure as it provides a more "conservative" result (see Honorton, 1985, p. 54);
- With these data (see Table A1 in the Appendix), we derived proportion-of-hits values and compared these to the proportions expected by chance (i.e., mean chance expectation; or MCE). We also recorded the standard normal deviate z values and derived effect size (ES) values (where $ES = z/\sqrt{n}$). If z score data was not available, we drew on hits and trial counts to derive the Binomial Exact probabilities (http://www.vassarstats.net/binomialX.html)[3];
- We obtained quality ratings and checked inter-rater reliability;
- We calculated file-drawer statistics, and tested psi (ES) and its relationship with: (i) study quality; (ii) year of study; and (iii) research group (laboratory);
- Appropriate tables and figures were prepared to illustrate the various effects;
- Bayesian analysis was conducted;
- The database is to be made available on the Psi Open Data repository (https://open-data.spr.ac.uk).

For each study, we checked the following factors: (1) the criteria adopted for selecting participants, (2) number of participants, (3) number of trials, (4) mode of ESP tested (clairvoyance, telepathy, or precognition), (5) number of alternatives in the tasks, and (6) total number of hits. Studies included in previous meta-analyses that were then re-published after that meta-analysis, were not used in the present meta-analysis: for example, Holt (2007) = Holt (2013). If there was more than one version of the same study valid in the test period (using the same data), the more detailed study was used: thereby, Parra and Villanueva (2015) replaced two studies — i.e., Parra and Villanueva (2010, 2011); Parra and Argibay (2013b) replaced Parra and Argibay (2013c). One study ordinarily valid in the test period (i.e., Parker, 2010) happened to include more accurate data from a study already used in a previous meta-analysis (i.e., Parker, 2006, which was therefore excluded from the present analysis — for details, see Storm, Tressoldi, & Utts, 2013). One study outside the test period (Parra & Argibay, 2007) had to be included since it was overlooked by Storm et al. (2010).

---

[3] Each participant is usually given one trial, and each will either get a hit or not. 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 was 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 (i.e., homogeneous, unpatterned sensory stimulation to eyes and ears);
- Category 2 = Nonganzfeld noise-reduction, including (i) 'sensory attenuation' using only cotton mask + earplugs; but no red-light stimulation or white/pink noise as in ganzfeld (see Pérez-Navarro, Lawrence, & Hume, 2009a), and (ii) 'imagery cultivation' to induce a shamanic-like altered state of consciousness (see Storm & Rock, 2009);
- Category 3 = Standard free-response (i.e., no ganzfeld or similar noise-reduction techniques to alter the normal waking cognitive state through dreaming, hypnosis, relaxation, or meditation). We classified remote viewing (RV) studies as standard free-response studies in alignment with Milton's (1997b) claim that RV 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).

Selected studies were rated for quality based on seven criteria used by Honorton and Ferrari (1989), and Steinkamp, Milton, and Morris (1998). 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 pre-planned;
- sensory shielding from sender (agent) and receiver (perceiver);
- target independently checked by a second judge, and
- experimenters blind to target identity.

Studies were rated for quality by three judges; one was a graduate student of the second author; two others are long-time psi researchers. Judges were suitably 'blinded' to the identity of the assessable material. The judges answered 'Yes' or 'No' to each of the seven criteria. Each study score is simply the number of points awarded with respect to the seven items applicable (minimum score = 0; maximum score = 7). However, Cronbach's alpha of paired ratings was only acceptable between two judges, α = .73, suggesting a strong degree of inter-rater reliability. The following calculations were conducted on those two judges' ratings and mean ratings.

There were 43 studies in the total pool. Of those studies, 14 (33%) received a perfect score from at least one judge. This figure is higher than the 25% reported in Storm et al. (2010). Going by mean scores, most criteria (i.e., five or more out of seven) were met in 35 of 43 studies (i.e., 81%), which is just under the 87% reported in the Storm et al. study (p. 474). This is a 6% gap, but the difference is not significant (p = .396), and the 95%CI [–7.5%, 21.3%]

includes zero. We stress that failure to make explicit declarations of procedures does not mean that any given criterion was not incorporated into any given experiment (for further comments, see Storm et al., 2010, p. 475).

RESULTS

Descriptive statistics and quality ratings

Table A1 in the Appendix lists all studies in their respective categories. A total of 43 'studies' (i.e., various experimental treatments and/or control groups) were conducted by 30 experimenters who reported their results in a total of 21 articles (see References for articles by name; articles marked with a single asterisk indicate papers included in this meta-analysis). Thirty-four studies (79%) used a four-choice design (the remaining nine studies used two-, eight-, or ten-choice designs). Across all three categories combined, 3 studies (7%) tested telepathy, 28 studies (65%) tested clairvoyance, while 10 studies (23%) tested precognition (2 studies, or 5%, conflated clairvoyance and precognition — see Roney-Dougal et al., 2011, 2014).

Darlington and Hayes (2000) offer a conservative and reliable test that determines the number of unpublished nonsignificant studies needed to reduce a database to non-significance. They claim 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). We found that 11 of the 43 studies were independently significant ($\alpha \leq .05$). Using their Table 2 (p. 503), if 11 individual results are significant, then pooled *p* is less than or equal to 0.05 only if the total number of studies is 125 or less. In other words, we find a "fail-safe *N*" of up to 125 unpublished studies must exist in total to threaten the significance of the pooled *p* value (NB: we must subtract the 43 we found, bringing the missing *N* to 82). The hypothetical existence of 82 unpublished studies is unlikely.

The correlation between mean Quality and *ES* was weak and not significant, $r_s(41) = -0.21$, $p = .181$ (two-tailed), so we claim that effect size is not likely to be an artifact of poor experimental design. We note that mean quality ratings between the two judges did not correlate significantly with year of publication either, $r_s(41) = -0.19$, $p = .905$ (two-tailed). Thus, we claim that quality has not improved or worsened over the years, though there is in fact room for improvement, given the overall mean rating was 5 out of 7, indicating that studies on average failed to report at least two criteria in their procedures.

Z Statistics and Effect Sizes

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 nine ganzfeld studies (21% of all studies in the database). For all these nine studies, clairvoyance was the most often tested modality (7 times out of 9). Telepathy and precognition were each tested once only. Seven studies tested unselected participants only, and two tested selected participants only.

TABLE 1.

New data (2009–2018): Free-response studies by category: Mean Z, ES Values, SD, Skew, and SE, Stouffer Z, & p Values

Category	Mean Z (SD; Skew; SE)	Mean ES (SD; Skew; SE)	Sums of ΣZ	Stouffer Z	p (one-tailed)
1 ($N = 9$)*	0.854 (0.92; –0.59; 0.72)	0.119 (0.15; –0.11; 0.72)	7.69	2.56	5.23×10^{-3}
2 ($N = 19$)**	0.685 (1.62; 1.60; 0.52)	0.045 (0.10; 0.22; 0.52)	13.02	2.99	1.40×10^{-3}
3 ($N = 15$) †	0.850 (1.76; 0.58; 0.58)	0.050 (0.13; 0.45; 0.58)	12.75	3.29	5.01×10^{-4}

* Ganzfeld; ** Nonganzfeld Noise-Reduction; † Standard Free-Response

The skews of the *z*-scores and *ES* distributions were normal, as was kurtosis for each distribution.[4] Therefore the dataset was regarded as homogeneous. The nine ganzfeld studies yielded mean *z* = 0.85 (*SD* = 0.92), 95%CI [0.15 to 1.56]; mean *ES* = 0.119 (*SD* = 0.15), 95%CI [0.003 to 0.235]. Note that neither of the CIs (confidence intervals) includes MCE (i.e., zero). Of the nine studies, 8 (89%) had positive *z* scores. Two of the nine studies (22%) are independently significant (α ≤ .05).

We conducted a binomial exact test on trial and hit counts. All nine studies used a 4-choice design (i.e., *k* = 4), and we found a 31% hit-rate (146 hits out of 471 trials), where MCE = 25%; Binomial Exact *z* = 2.95, *p* = 1.59×10^{-3} (one tailed).

Although we report a significant Stouffer *Z* of 2.56 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). Their "Stouffer-max" test provides a "MeanZ(*s,k*)" value at α = .05, which is the "mean of the *s* highest of *k* mutually independent values of *z*" (p. 505), that is then compared to a critical MeanZ. Taking *s* = 4 (i.e., the four most positive values of *z* to yield a conservative calculation), and *k* = 9 (i.e., where *k* = *N* = 9), our MeanZ is 1.61. Table 3 in Darlington and Hayes (p. 506) gives critical MeanZ = 1.48. In other words, the mean *z* for the ganzfeld database is sufficiently higher than is required by the Stouffer-max test.

Category 2 (nonganzfeld noise-reduction): This database was composed of 19 studies (44.2% of all studies in our search). These 19 studies were classified as noise-reduction studies. Eleven studies used unselected participants, and eight studies used selected participants. Clairvoyance tasks outnumbered the other modalities (10 out of 19; 53%).

The distribution was normal. The homogeneous dataset yielded mean *z* = 0.69 (*SD* = 1.62), 95%CI [–0.095 to 1.47]; mean *ES* = 0.045 (*SD* = 0.10),

---

[4] Outlier studies that cause significant deviations in a distribution may deflate or inflate the mean *z* score and/or mean *ES*. Outliers were identified from SPSS Stem-and-Leaf and Box-and-Whiskers Plots. Only values more extreme than three IQRs (interquartile range) were excluded from analyses (see Hoaglin & Iglewitz, 1987).

95%CI [–0.003 to 0.093]. The confidence intervals include MCE (i.e., zero). Of the 19 studies, 13 (68%) had positive z scores. Four of the 19 studies (21%) are independently significant ($\alpha \le .05$).

We conducted a binomial exact test on trial and hit counts on those studies that used a '$k = 4$' design (three studies used a '$k = 2$' design). For 16 studies (trials = 1934, hits = 507), we found a 26.2% hit-rate, which is just above MCE = 25%, but was not significant; Binomial Exact $z = 1.21$, $p = .113$ (one tailed).

We applied Darlington and Hayes' (2000) "Stouffer-max" test once again. Given $s = 4$, and $k = 19$, Darlington and Hayes (2000, p. 506, Table 3) gives critical MeanZ = 1.88. Our MeanZ is higher at 3.00, so that the mean z for this database is sufficiently higher than is required in the Stouffer-max test.

Category 3 (standard free-response): This database consisted of 15 standard free-response studies (34.9% of all studies in our search). Eight of the 15 studies (53.3%) were composed completely of selected participants, and seven (46.7%) were composed of unselected participants. Eleven studies tested clairvoyance.

The distribution was normal, so the dataset was regarded as homogeneous. The 15 studies yielded mean $z = 0.85$ ($SD = 1.76$), 95%CI [–0.13 to 1.83]; mean $ES = 0.05$ ($SD = 0.13$), 95%CI [–0.019 to 0.120]. Both confidence intervals include zero.

Only nine studies used a '$k = 4$' design (five used a '$k = 10$' design; one used a '$k = 8$' design). For those nine studies (trials = 707, hits = 181), we found a non-significant hit-rate of 25.6% (Binomial Exact $z = 0.33$, $p > .05$, one tailed). However, the five studies that used a '$k = 10$' design (trials = 2560, hits = 344), yielded a hit-rate of 13.4% which was significant (Binomial Exact $z = 5.76$, $p = 4.21 \times 10^{-9}$, one-tailed). Of the 15 studies, 9 (60%) had positive z scores. Five of the 15 studies (33.3%) are independently significant ($\alpha \le .05$).

We applied Darlington and Hayes' (2000) "Stouffer-max" test once again. Given $s = 4$, and $k = 15$, Darlington and Hayes (2000, p. 506, Table 3) gives critical MeanZ = 1.76. Our MeanZ is higher at 3.08, so the mean z for this database is sufficiently higher than is required in the Stouffer-max test.

Planned analyses

H1a. Free-response studies produce statistical evidence of a communication anomaly known as ESP: As the 43 studies in the database have different Mean Chance Expectations (MCEs) due to the variance in k (the number of choices), it is not possible to make a single Binomial Exact z calculation for the combined databases. The best way to treat this general hypothesis, is to consider studies in terms of k:

- For the two-choice design ($n = 3$), there was a total of 1812 trials and 1021 hits, corresponding to a 56.3% hit rate where Mean Chance Expectation (MCE) = 50%. We calculated a significant Binomial Exact $z = 5.38$, $p = 3.72 \times 10^{-8}$ (one tailed).
- For the four-choice design ($n = 34$), there was a total of 3112 trials and 834 hits, corresponding to a 27% hit rate (MCE = 25%): Binomial Exact $z = 2.30$, $p = .012$ (one tailed).

- For the eight-choice design only ($n = 1$), there was a total of 448 trials and 43 hits, corresponding to a 9.6% hit rate (MCE = 12.5%): Binomial Exact $z = -1.79$, $p = .976$ (one tailed).
- For the ten-choice designs only ($n = 5$), there was a total of 2560 trials and 344 hits, corresponding to a 13.4% hit rate (MCE = 10%): Binomial Exact $z = 5.76$, $p = 4.21 \times 10^{-9}$ (one tailed).

Results were overall significant except for the eight-choice design, although that design was poorly represented by a single study. The hypothesis was generally supported.

H2a. There is a psi-scoring difference between the three categories: (i) ganzfeld, (ii) nonganzfeld noise-reduction, and (iii) standard free-response: From the above preliminary findings we can see that the ganzfeld studies produced the strongest mean *ES* values = 0.119, followed by standard free-response studies (mean *ES* = 0.050), followed closely by the nonganzfeld noise-reduction studies (mean *ES* = 0.045). We ran a Univariate ANOVA test, entering the variable 'category' as a fixed factor, as well as 'modality' (see *H3a* below), 'target' (see *H4a* below), and 'participant type' (see *H5a* below). This single ANOVA avoids capitalization on chance through multiple testing. The test showed no significant difference in mean *ES* values between categories, $F(2, 22) = 0.75$, $p = .487$ (two-tailed). The hypothesis was not supported.

H3a. There is a difference in psi strength between telepathy, clairvoyance, and precognition studies: Effect sizes for the three psi modalities were: telepathy, mean *ES* = .074; clairvoyance, mean *ES* = .038; precognition, mean *ES* = .079. Although precognition elicited the strongest effect, the Univariate ANOVA yielded no significant difference, $F(2, 22) = 1.74$, $p = .200$ (two-tailed). [5] The hypothesis was not supported.

H4a. Effect size values vary in strength according to target types: This hypothesis takes into account the possibility that target type may affect psi performance, which might then be reflected in *ES* value. We divided our data into three types of target: (a) pictures/photos (mean *ES* = .075), (b) objects (i.e., targets that occupy 3D-space; mean *ES* = .060), and (c) video clips (mean *ES* = .017). Although pictures/photos elicited the strongest effect, the Univariate ANOVA showed no significant difference between target types, $F(2, 22) = 0.15$, $p = .865$ (two-tailed). The hypothesis was not supported.

H5a. Selected participants have a performance advantage over unselected participants: Although selected participants produced stronger effects (mean *ES* = .078) than unselected participants (mean *ES* = .044), the same Univariate ANOVA test conducted previously showed no significant difference, $F(1, 22) = 0.47$, $p = .250$ (one-tailed). The hypothesis was not supported.

Post hoc analyses of categories

The following analyses are on the combined databases, which entail comparisons of the 'new' databases with the corresponding 'old' databases compiled by Storm et al. (2010).

---

[5] Roney-Dougal et al. (2011, 2014) were excluded as they conflated clairvoyance and precognition.

Category 1 (ganzfeld). Storm et al. (2010) constructed a homogeneous data set of 29 ganzfeld studies and reported mean $ES = 0.142$ ($SD = 0.20$). Storm et al. (2013) made some adjustments to the 2010 database. Specifically, Parker (2006) reported 20 trials and 8 hits, but there were actually 28 trials and 10 hits (see Parker, 2010); and Parker and Westerlund (1998, Study 5) reported 11 hits, whereas there were actually 12 hits (see Parker, 2000).[6] The revised mean $ES = 0.137$ ($SD = 0.20$).

When that database ($N = 29$) was tested against our new database ($N = 9$), we found no significant differences: z scores, $t(36) = 0.30$, $p = .768$ (two-tailed); ES values, $t(36) = 0.25$, $p = .801$ (two-tailed). We combined the two databases $(29 + 9 = 38)$, found no outliers, and thereby formed an updated homogeneous database of 38 studies collected for the period 1992 to 2018 — a 27-year span (actual publication dates were from 1997 to 2012):[7] mean $z = 0.97$ ($SD = 1.27$), 95%CI [0.549 to 1.381]; mean $ES = 0.13$ ($SD = 0.19$), 95%CI [0.071 to 0.194]. Neither of these CIs includes zero. Of the 38 studies, 27 (71%) had positive z scores. Ten (26%) of the 38 studies are independently significant ($\alpha \leq .05$). We also found that year of publication did not correlate with ES, $r_s(36) = -0.002$, $p = .988$ (two-tailed), indicating no decline or incline over the years.

Table 2 lists by category all the various statistics: mean z scores, mean ES values, Stouffer Z, and corresponding p values.

Category 2 (nonganzfeld noise-reduction). Storm et al. (2010) constructed a homogeneous data set of nonganzfeld noise-reduction studies ($N = 16$), yielding mean $ES = 0.11$ ($SD = 0.19$). In Storm et al. (2013) we pointed out that there were missing data from Tressoldi and Del Prete (2007), which were erroneously omitted in the original 2010 database. Also, Parra and Villanueva (2015), which was originally in Category 3, was transferred to Category 2 as there was a relaxation treatment. With a revised database now standing at $N = 18$, the revised mean $ES = 0.10$ ($SD = 0.18$).

TABLE 2.

Combined data (1992–2018): Free-response studies by category: Mean Z, ES Values, SD, Skew, and SE, Stouffer Z, & p Values

Category	Mean Z (SD; Skew; SE)	Mean ES (SD; Skew; SE)	Sums of ΣZ	Stouffer Z	p (one-tailed)
1 (N = 38)*	0.965 (1.27; 0.39; 0.38)	0.133 (0.19; 0.12; 0.38)	36.67	5.95	1.37×10^{-9}
2 (N = 37)**	0.724 (1.40; 1.17; 0.39)	0.072 (0.15; 0.47; 0.39)	26.78	4.40	5.00×10^{-6}
3 (N = 33) †	0.366 (1.39; 1.18; 0.41)	0.027 (0.14; 0.63; 0.41)	12.09	2.10	1.79×10^{-2}

[*] Ganzfeld; [**] Nonganzfeld Noise-Reduction; [†] Standard Free-Response

---

[6] We thank Bryan J. Williams for bringing these corrections to our attention (see Williams, 2011).

[7] Actually, four ganzfeld studies have been conducted since 2012, and before 2018, but described in abstracts only: Three featured in Roe et al. (2020); and the fourth by Watt, Dawson, Tullo, Pooley, and Rice (2020).

We tested this revised database ($N = 18$) against our new database ($N = 19$), and found no significant differences: z scores, $t(35) = 0.17$, $p = .866$ (two-tailed); ES values, $t(26.21) = 1.15$, $p = .262$ (two-tailed).

The two databases ($18 + 19 = 37$) were combined to yield a homogeneous database for the period 1992 to 2018 (actual publication dates were from 1999 to 2015): mean $z = 0.72$ ($SD = 1.40$), 95%CI [0.257 to 1.191]; mean $ES = 0.072$ ($SD = 0.15$), 95%CI [0.023 to 0.120]. Neither of these CIs includes zero. (See Table 2 for other statistics). Of the 37 studies, 24 (65%) had positive z scores. Seven (19%) of the 37 studies are independently significant ($\alpha \leq .05$). We also found that year of publication did not correlate with ES, $r_s(37) = -0.27$, $p = .102$ (two-tailed), indicating no decline or incline over the years.

Category 3 (standard free-response). Storm et al. (2010) constructed a database of standard free-response studies ($N = 21$), but Targ and Katra (2000) was later removed, whereas Del Prete and Tressoldi (2005) was added (for the explanation, see Storm et al., 2013). Also, Parra and Villanueva (2015) was removed as explained above (new $N = 20$). However, tests showed the database would not be homogeneous without the removal of the same six outliers listed in Storm et al. (2013, p. 251): Holt (2007; two datasets), May (2007), Simmonds and Fox (2004), Storm (2003), and Watt and Wiseman (2002). Though we had undertaken that step in Storm et al. (2013), we now deem the removal of six studies (30%) from a database of a mere 20 studies as extreme and potentially misleading. In particular, the degree to which heterogeneity accounts for any inconsistency in a database, with its subsequent effects on generalisability, should first be assessed lest the removal of outliers be unwarranted. One measure that assesses consistency of the results of studies in a meta-analysis, is Higgins' I^2 (Higgins, Thompson, Deeks, & Altman, 2003), which indicates the proportion of effect-size variance explained by heterogeneity across effect sizes.[8] Negative values of I^2 are set to zero so that I^2 lies between 0% and 100%. A value of 0% indicates no observed heterogeneity, and larger values show increasing heterogeneity. Our test for heterogeneity gave Cochran's $Q(df = 19) = 34.71$, $p = 0.015$, $I^2 = 48.10\%$. Very serious heterogeneity is indicated only when $I^2 > 75\%$, and serious heterogeneity is indicated when I^2 is $\geq 50\%$ (Higgins *et al.*, 2003), so we did not consider our database inconsistent (i.e., there is not enough inconsistency to warrant looking for explanations, so we did not label the above extreme studies as outliers).

The two databases ($20 + 15 = 35$) were not significantly different: z scores, $t(33) = -1.07$, $p = .291$ (two-tailed); ES values, $t(33) = 0.26$, $p = .793$ (two-tailed). However, two studies were outliers and their removal was not considered too severe: May (2007) and Storm (2003) — see Storm et al. (2010; Cases 6 and 17 in Category 3 of Table A1, p. 484). Our test for heterogeneity gave $Q(df = 32) = 51.28$, $p = 0.017$, $I^2 = 40.94\%$. Given the low I^2 value, we regarded as homogeneous the database of 33 studies for the period 1992 to

---

[8] Higgins' $I^2 = 100\% \times (Q - df)/Q$, where Q is Cochran's heterogeneity statistic and df is degrees of freedom. Heterogeneity benchmark values for I2 are 25% (low), 50% (moderate), and 75% (high). For details about Cochran's Q statistic, see Lipsey and Wilson (2001).

2018 (actual publication dates were from 2000 to 2015): mean z = 0.37 (*SD* = 1.39), 95%CI [–0.125 to 0.858]; mean *ES* = 0.027 (*SD* = 0.14), 95%CI [–0.024 to 0.078]. Both CIs include zero (See Table 2 for other statistics). Of the 33 studies, 15 (46%) had positive z scores. Only five (15%) of the 33 studies are independently significant ($\alpha \leq .05$). We also found that year of publication did not correlate with *ES*, $rs(31) = 0.24$, $p = .185$ (two-tailed), indicating no decline or incline over the years.

Post hoc hypotheses

H1b. Free-response studies produce statistical evidence of a communication anomaly known as ESP: As the 108 studies in the database have different MCEs due to the variance in k, it was again not possible to make a single Binomial Exact z calculation. The following statistics are given in terms of k:

- The two-choice design did not feature in the Storm et al. (2010) dataset, so the statistics are the same as reported above — i.e., three studies: 1812 trials, 1021 hits, 56.3% hit rate (MCE = 50%), $z = 5.38$, $p = 3.72 \times 10^{-8}$ (one tailed).
- For the four-choice design ($n = 970$), there was a total of 7650 trials and 2117 hits, corresponding to a 27.7% hit rate (MCE = 25%): Binomial Exact $z = 5.39$, $p < .001$ (one tailed).
- For the five-choice design ($n = 1$), there was a total of 58 trials and 17 hits, corresponding to a 29.3% hit rate (MCE = 20%): Binomial Exact $z = 1.61$, $p = .054$ (one tailed).
- For the eight-choice design only ($n = 2$), there were 462 trials and 47 hits, corresponding to a 10.2% hit rate (MCE = 12.5%): Binomial Exact $z = -1.44$, $p = .717$ (one tailed).
- The ten-choice design did not feature in the Storm et al. (2010) database, so the statistics are the same as reported above — i.e., five studies: 2560 trials, 344 hits, 13.4% hit rate (MCE = 10%), Binomial Exact $z = 5.76$, $p = 4.21 \times 10^{-9}$ (one tailed).

Results were overall significant except for the eight-choice design (poorly represented by a single study), and marginally significant for the five-choice design. The hypothesis was generally supported.

H2b. There is a psi-scoring difference between the three categories: (i) ganzfeld, (ii) nonganzfeld noise-reduction, and (iii) standard free-response: We ran a Univariate ANOVA on the three updated (enlarged) databases, entering the variable 'category' as a fixed factor, but again we also entered 'modality' (see *H3b* below), 'target' (see *H4b* below), and 'participant type' as fixed factors (see *H5b* below). Mean *ES* scores were in the direction hypothesized previously (see Table 2), and the test showed a marginally significant difference between categories, $F(2, 74) = 2.58$, $p = .082$ (two-tailed). The post hoc Tukey HSD test showed that only ganzfeld (Category 1) and standard free-response (Category 3) were significantly different ($p = .012$). The hypothesis was supported.

H3b. There is a difference in psi strength between telepathy, clairvoyance, and precognition studies: Effect sizes for the three psi modalities were: telepathy, mean *ES* = .154; clairvoyance, mean *ES* = .075; mean precognition,

$ES = .048.$[9] This time, telepathy (not precognition) elicited the strongest effect, but there was still no significant difference between the modalities, $F(2, 74) = 0.25$, $p = .777$ (two-tailed).

H4b. Effect size values vary in strength according to target types: Target types were the same as in H4a. *ES* values were: pictures/photos, mean $ES = .078$; objects, mean $ES = .087$; video clips, mean $ES = .100$. This time, video clips (not pictures/photos) elicited the strongest effect, but once again the Univariate ANOVA showed no significant difference between target types, $F(2, 74) = 0.35$, $p = .705$ (two-tailed). The hypothesis was not supported.

H5b. Selected participants have a performance advantage over unselected participants: Selected participants produced stronger effects (mean $ES = .120$) than unselected participants (mean $ES = .063$), but the ANOVA test result was not significant, $F(1, 74) = 0.00$, $p = .987$ (one-tailed). However, running the ANOVA again with only category and participants as IVs, a significant difference was found between categories, $F(2, 100) = 5.34$, $p = .003$; and between participants, $F(1, 100) = 6.70$, $p = .005$. As was the case with the Storm et al. (2010) study, the significant category difference was only between Category 1 and Category 3 ($p = .014$). Also, a marginally significant interaction effect was found, $F(2, 100) = 2.03$, $p = .069$. Figure 1 illustrates the differences (effects between participant groups vary disproportionately across category levels), and these are comparable to Figure 1 in Storm et al. (2010).

The ganzfeld and nonganzfeld noise-reduction databases

Effect size differences between experimenters/laboratories. In order to ascertain whether a limited pool of experimenters and/or laboratories were responsible for the significant effects, we conducted a one-way ANOVA on the pooled data of two databases (ganzfeld studies + nonganzfeld noise-reduction; $N = 75$), since these datasets were not significantly different (see *H2b*), and they both produce significant effects whereas the CIs for the *z* and *ES* values for the standard free-response database include zero. This test is a repeat of the test we conducted in Storm et al. (2010) on 45 studies, the test of which delivered a nonsignificant result. We were able to divide the database into experimenter/laboratory groups, but we could not test experimenter × laboratory interaction as we found a number of experimenters had worked in more than one specific laboratory.

Storm et al. (2010) used the following seven groups: 'Morris', 'Parker', 'Parra', 'Roe', 'Roney-Dougal', 'Tressoldi', and 'Wezelman'. These groups were used again, but we also formed six new groups with at least two studies in each: 'Luke', 'Pérez-Navarro', 'Rock/Storm', 'Steinkamp', 'Watt', and 'Other' (for single studies). Thus, there were a total of 13 experimenters/laboratories.

Effect size (*ES*) values were not significantly different between laboratory/ experimenter groups, $F(12, 62) = 1.11$, $p = .370$ (two-tailed). Re-running the test on the ganzfeld data only ($N = 38$) also gave a nonsignificant result, $F(6, 31) = 1.76$, $p = .139$ (two-tailed). For the period 1992 to 2018, there is no

---

[9] Once again, Roney-Dougal et al. (2011, 2014) were excluded as they conflated clairvoyance and precognition.
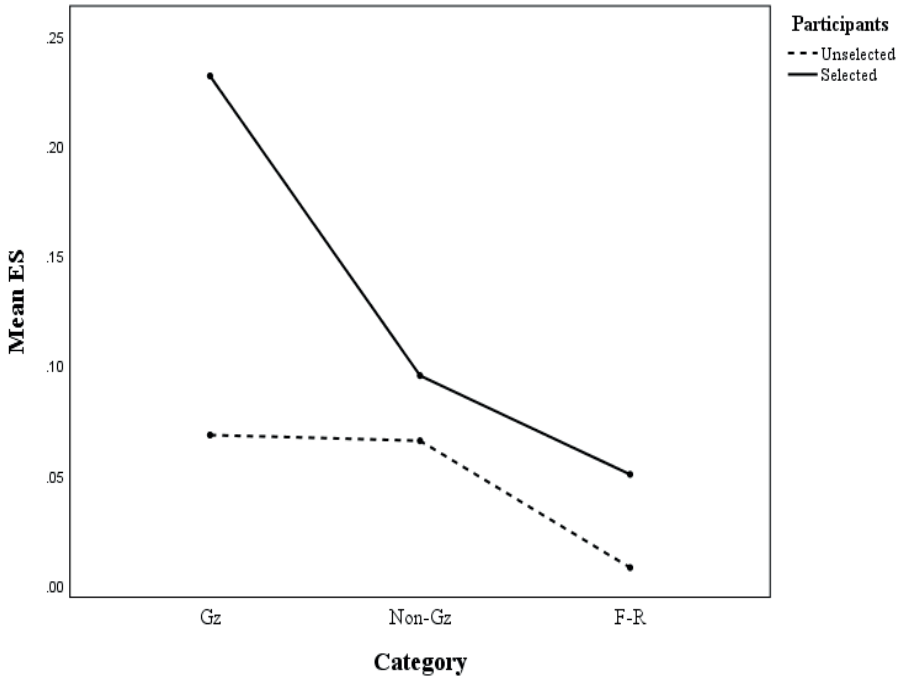
Figure 1. Effect size (ES) difference between two categories — Ganzfeld (Gz) and Standard Free-Response (F-R) — and an interaction effect between category and participant type

evidence of laboratory/experimenter differences in the mixed database (ganzfeld + nonganzfeld noise-reduction), or the ganzfeld-only database.

Decline/incline effects across time. Declines in ganzfeld performance over time, or suggestions of such, have been reported a few times (see Bierman, 2001; Bierman et al., 1993; Milton & Wiseman, 1999). It is noted that Bierman (2001) suggested there may also be a ganzfeld "rebound" effect in the form of a flattened U-shaped (polynomial) curve for the period 1972 to 2001. Storm et al. (2010) "assessed the evidence for an effect size decline [or incline] in ganzfeld studies" (p. 478), for a 34-year period (1974 to 2008), and there was either a significant linear decline if the four outlier studies from Honorton's (1985) database were removed, or a non-significant decline if they were retained. A significant polynomial (U-shaped) curve also fit the data.

In the present study, for an updated period (1974 to 2018), Figure 2 shows *ES* values for a total of 117 ganzfeld studies, from six databases, plotted over a 44-year period (1974 to 2018): (1) Honorton (1985; $N = 28$; period of analysis: 1974 to 1981); (2) Storm and Ertel (2001; $N = 11$; period of analysis: 1982 to 1989); (3) Bem and Honorton (1994; $N = 10$; period of analysis: 1983 to 1989); (4) Milton and Wiseman (1999; $N = 30$; period of analysis: 1989 to 1997); (5)
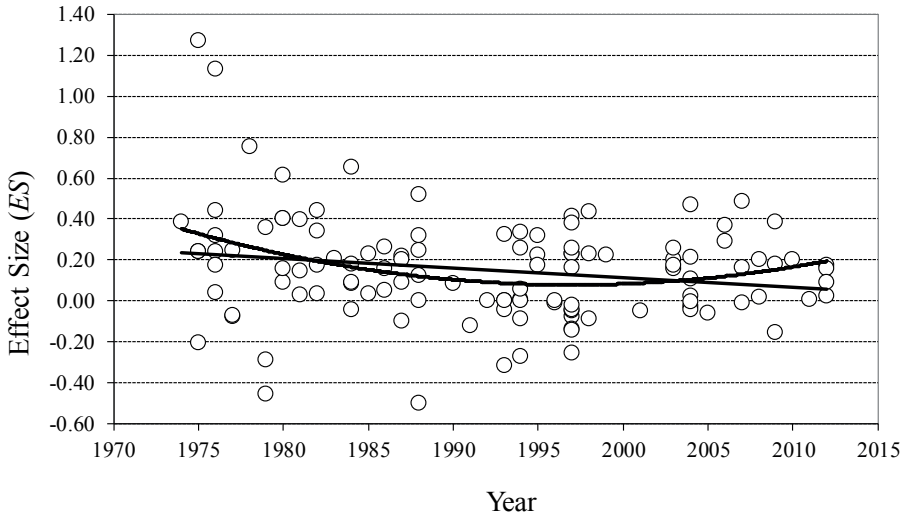
Figure 2. Scatter-plot of ganzfeld studies over a test period of 44 years (1974 to 2018). The decline (straight line) and rebound effect (U-shaped curve) were not significant

Storm et al. (2010; $N = 29$; period of analysis: 1997 to 2008), and (6) the present database ($N = 9$; period of analysis: 2009 to 2018). [10]

The combined ganzfeld database yields a mean $z = 0.85$ ($SD = 1.27$), 95%CI [0.612 to 1.079]; mean $ES = 0.15$ ($SD = 0.26$), 95%CI [0.105 to 0.198]. The CIs do not include zero. Of the 117 studies, 88 (75%) had positive z scores. For the four-choice design ($n = 114$), there was a total of 3885 trials and 1188 hits, corresponding to a 30.6% hit rate (MCE = 25%): Binomial Exact $z = 8.01$, $p = 5.55 \times 10^{-16}$ (one tailed). Thirty-three (28%) of the 117 studies are independently significant ($\alpha \leq .05$). A forest plot was generated from the data to show shifts in the cumulative weight of the evidence over time (Borenstein, Hedges, Higgins, & Rothstein, 2009). We added studies successively by their publication year and Figure 3 shows a tendency for ES values to converge towards a mean $ES = .15$.

Prior to our next analysis, it was discovered that two studies (Raburn, 1975; Honorton, 1976) were 'extreme' by our above definition (see footnote 4). Their removal rendered the database effectively homogeneous, and we proceeded with the decline/incline analysis. Running a hierarchical regression analysis on the dataset ($N = 115$), with year of study as the IV, and ES as the DV, Model 1 (testing for a linear effect) was not significant, $F(1, 113) = 1.06$, $p = .306$ (two-tailed), $R = .096$ ($R^2 = .009$). Model 2, which tested for a rebound effect in the form of a quadratic curve (year-squared), failed.

---

[10] We double-checked the z scores of the five previous databases to ensure that the values were exact binomials, and where necessary we recalculated ES values. A total of 46% were adjusted, but in most cases only to a very slight degree. For their database of 30 studies, Milton and Wiseman (1999) reported a Stouffer $Z = .70$, $p = .24$; our adjustment renders a Stouffer $Z = 1.54$, $p = .06$.
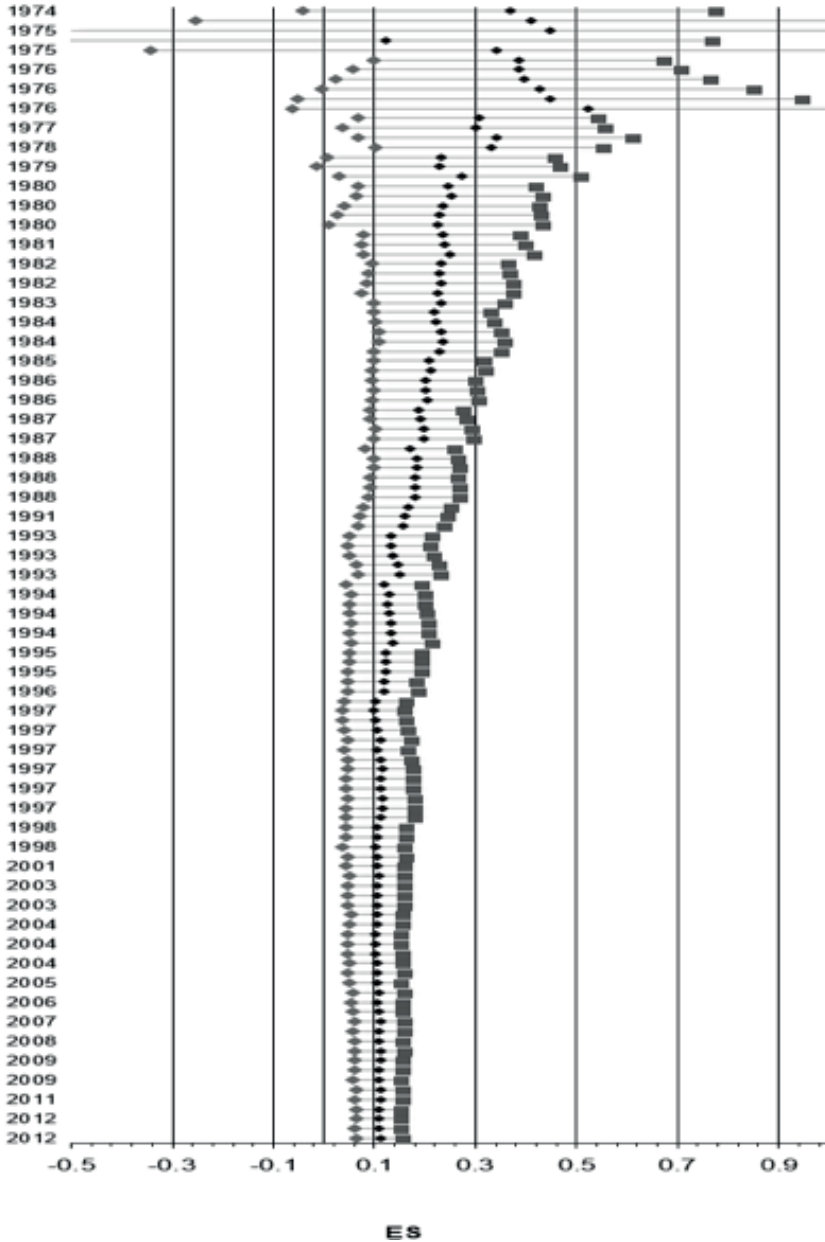
Figure 3. Forest plot showing that, accumulatively, effect size decreases on average across time 1974–2018 for the ganzfeld domain, averaging to a mean ES = 0.15 (N = 117). The central dots are estimated ES values, the diamond and square dots on both sides represent 95% CI minima and maxima, respectively.

Bayesian analysis

In the following analysis, we introduce a Bayesian meta-analysis parameter estimation to demonstrate the robustness of the evidence supporting the case for ESP. We analysed separately the three categories of studies. For all three categories we used the MetaBMA package (Heck, Gronau, & Wagenmakers, 2017) with the software R version 3.6.3. As priors for the overall random-model effect size we used a normal distribution with mean = 0.01 (SD = 0.03), constrained positive, lower bound = 0 (Rouder et al., 2019), given our expectation of positive values. For the parameter τ (tau) related to the between study heterogeneity, we used an inverse gamma distribution with shape = 1, and scale = 0.15. (A copy of the syntax is available in Appendix B.)

In the following Table 3 are reported the estimated effect sizes (ES) and corresponding 95% CIs for each of the three categories of studies. The ES values of all three categories of studies are compatible with those estimated using the classical procedure, thus supporting their robustness, notwithstanding the different statistical approach.

TABLE 3.

Bayesian Analysis: Estimated Effect Size (ES) and 95% CIs by Three Categories

Category	ES	95% Credible Interval
Ganzfeld (N = 38)	.12	0.06–0.18
Non-Ganzfeld, Noise-Reduction (N = 37)	.06	0.017–0.10
Standard Free-Response (N = 33)	.027	0.00–0.058

Discussion

Ten years ago, Storm et al. (2010) found that the free-response design generally produced statistical evidence of a communication anomaly known as ESP: For example, 63 studies using a four-choice design yielded a significant "29.9% hit rate where mean chance expectation (MCE) is equal to 25%" (p. 475). A corresponding hit rate for a new set of studies of the same design yielded a significant 27% hit rate (see *H1a*). In their original study, Storm et al. also constructed three categories of free-response studies: Category 1 (ganzfeld), Category 2 (nonganzfeld/noise-reduction), and Category 3 (standard free-response). The mean ES for Category 1 (ES = 0.142) was significantly higher than the mean ES for Category 2 (ES = 0.110) and Category 3 (ES = –0.029). In this present meta-analytic update, 43 new studies were found and assigned to the same three categories and analysed separately. Compared to the original statistics, two ES values had dropped slightly in strength (Category 1: ES = 0.119; Category 2: ES = 0.045), and the third had increased and was now *positive* (Category 3: ES = .050), so the downward trend of effects across Categories 1, 2, and 3 was not replicated (see *H2a*). Only Category 1 had an independently significant hit rate (31%,

where MCE = 25%). Also, the 95% CI for mean z and mean ES did not include zero (thus ruling out zero as the true mean of the population), but this was the case for Category 1 only, so we cannot be sure that the true means of the population are not zero for Categories 2 and 3. However, Stouffer Z values for all three categories were significant (see Table 1).

In the same Univariate ANOVA test on 'category' just mentioned (*H2a*), the other independent variables (IVs) — 'modality' (i.e., form of ESP), 'target type' (e.g., film clips, pictures), and 'participant type' (selected vs. unselected) — were entered into a single model so as not to capitalize on chance through multiple testing. For the new datasets, there were no differences in ES values within any of these IVs (see *H3a* to *H5a*).

As the paired comparisons of old and new datasets were not significant, we were able to increase the sizes of all three categories by combining corresponding old and new datasets. The free-response domain still produced a significant hit rate: For example, the hit rate of 27.7% for the set of studies with a four-choice design ($n = 97$) was significant (see *H1b*). On this occasion, the 95%CI for mean z and mean ES did not include zero for Category 1 *and* 2, but not Category 3. Again, Stouffer Z values for all three categories were significant (see Table 2).

Originally, Storm et al. (2010) found that "Category 1 and 3 differed significantly from each other" (p. 476). In fact, earlier findings presented in Storm et al., were generally replicated in the three enlarged databases (see *H1b, H2b, H3b*, and *Hb5*; target types [*H4b*] were not tested in Storm et al.) — there was again a significant difference between Category 1 and Category 3 (see *H2b*); no significant difference between the three modalities (see *H3b*), or target type (see *H4b*); but a significant main effect of 'participant type' (see *H5b*). Also, there was a near-significant 'category × participant type' interaction effect as illustrated in Figure 1. When it comes to selected participants, as Storm et al. had said before, "ganzfeld seems to be particularly superior to the other two techniques" (p. 480). We suggest our assessments in the present study support Baptista et al.'s (2015) recommendation to adhere to the standard ganzfeld methodology, *and* preferentially test selected participants. This advice is not new: very early in the history of free-response research, experimenters like Honorton and Ferrari (1989), and Morris (1991), advised researchers to (i) select their participants through prior testing and/or training, and (ii) use a noise-reduction technique. This combination still appears to be the best on offer as far as ES yields are concerned.

We presented free-response hit rates for '$k = 4$' (four-choice) design. Note, however, that over the decades, experimental designs have ranged in k (number of choices; target + decoys) from two choices (MCE = 50%) up to ten choices (MCE = 10%), with preferred k being four choices by an overwhelming margin (99 studies out of 110, or 90%). In all of six tests (new studies in each of the three categories and then the combined [old + new] studies for each of the three categories), the Stouffer Z statistics were all significant.

We were able to run two tests on experimenter/laboratory groups in this update: (i) a mixed database (ganzfeld + nonganzfeld noise-reduction) since there was no significant mean ES difference between the two databases, and (ii) the ganzfeld set of studies on their own. Neither test revealed significant

differences. For the period 1992 to 2018, there is no evidence of laboratory/ experimenter differences in the mixed set, or the ganzfeld-only database.

In our final set of analyses for this study, there was no statistical evidence for an effect-size decline in the ganzfeld domain (see Figure 2). For dubious claims made of the ganzfeld decline effects, see Bierman (2001), Hyman (2010), and Milton and Wiseman (1999). The average effect overall (i.e., in the long term) does appear to be reliable (consistent), albeit weak (as indicated in the ganzfeld forest plot; Figure 3). Inspection of Figure 3 indicates a relatively high *ES* in the mid-1970s going into a slow decline throughout the 1980s, to the mid-1990s, from which point (i.e., 1994) the effect gravitates towards 0.10, but rises slightly, levelling out to a mean of 0.15. We note there is no significant decline over the 44-year period, but we nevertheless suggest the initially large *ES* estimates were imprecise (see the large CIs in Figure 3), mainly because they are based on very low *N*s (low trial counts). Furthermore, in the cumulative meta-analysis, the mean *ES* does get more precise (with smaller CIs) because it takes into account all previous data. It is a different representation with respect to the simple raw representation (not cumulative) of *ES*s for each single study as shown in Figure 2 (for further discussion of this issue, see Borenstein et al., 2009, p. 371).

We close our discussion by emphasizing the fact that our Bayesian meta-analysis supports our findings — the *ES* values of all three categories of studies were found to be compatible with those estimates we found using the frequentist approach. Whichever of our analyses is preferred, it appears the cumulative record of 44 years of free-response research featuring noise-reduction procedures intended to be psi-conducive (especially ganzfeld) has more often than not produced a communication anomaly worthy of further investigation.

Conclusion

Given the findings for our updated databases are generally on par with those in our earlier paper (Storm et al., 2010), we find our comments of ten years ago are mostly supported by the replicated effects in this update paper. Specifically, it appears that the noise-reduction procedures (especially ganzfeld) tend to produce a consistent communication anomaly often described as 'extrasensory perception' (ESP), though the underlying process is not yet explained. Further to that, selected participants appear to have an advantage over unselected participants, but the near-significant interaction effect *suggests* no performance advantage of consequence for selected participants unless they are in the ganzfeld condition. Lastly, we found that (a) ESP effects tend to be replicated by independent researchers, though no specific laboratory or group of experimenters seems to have an advantage over any other, and (b) reports of a decades-long decline in the ganzfeld effect are not justified.

School of Psychology   LANCE STORM
University of Adelaide
Adelaide, South Australia 5005
Australia
lance.storm@adelaide.edu.au

REFERENCES

(References marked with a single asterisk indicate studies included in the meta-analyses for the present study; see Appendix A)

Baptista, J., Derakhshani, M., & Tressoldi, P. (2015). Explicit anomalous cognition: A review of the best evidence in ganzfeld, forced-choice, remote viewing and dream studies. In E. Cardeña, J. Palmer & D. Marcusson-Clavertz (Eds.), *Parapsychology. A handbook for the 21st century* (pp. 192–214). McFarland.

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.

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). Benjamins.

Bierman, D. J., Bosga, D., Gerding, J. L. F., & Wezelman, R. (1993). Anomalous information access in the Ganzfeld: Utrecht-Novice series I and II. *Proceedings of the PA Convention*, *36*, 192–204.

Borenstein, M., Hedges, L. V., Higgins, J. P. T, & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

* da Silva, F. E. (2011). Reporting an exploratory group experience in PSI training. Proceedings of the 54th Parapsychological Association Convention, Curitiba.

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

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

Heck, D. W., Gronau, Q. F., & Wagenmakers, E. (2017). metaBMA: Bayesian model averaging for random and fixed effects meta-analysis. https://doi.org/10.5281/zenodo.835494

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560.

Hoaglin, D. C., & Iglewicz, B. (1987). Fine tuning some resistant rules for outlier labeling. *Journal of American Statistical Association*, *82*, 1147–1149.

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. (2013). Comparing artists with controls on a free-response experience-sampling task: Creativity and anomalous cognition. *Journal of the Society for Psychical Research*, *77*, 1–18.

Honorton, C. (1976). Length of isolation and degree of arousal as probable factors influencing information retrieval in the ganzfeld. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology 1975* (pp. 184–186). Scarecrow Press.

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

Honorton, C., Berger, R. E., Varvoglis, M. P., Quant, M., Derr, P., Schechter, E. I., & Ferrari, D. C. (1990). Psi communication in the ganzfeld: Experiments with an automated testing system and a comparison with a meta-analysis of earlier studies. *Journal of Parapsychology*, *54*, 99–139.

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

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

Hyman, R. (2010). Meta-analysis that conceals more than it reveals: Comment on Storm et al. (2010). *Psychological Bulletin, 136*(4), 486–490. doi:10.1037/a0019676

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

Lipsey, M. W., & Wilson, D. (2001). *Practical meta-analysis*. Sage.

* Luke, D., & Zychowicz, K. (2014). Working the graveyard shift at the witching hour: Further exploration of dreams, psi and circadian rhythms. *International Journal of Dream Research*, *7*(2), 105–112.

* Luke, D., Zychowicz, K., Richterova, O., Tjurina, I., & Polonnikova, J. (2012). A sideways look at the neurobiology of psi: Precognition and circadian rhythms. *NeuroQuantology*, *10*(3), 580–590.

* Marcusson-Clavertz, D., & Cardeña, E. (2011). Hypnotizability, alterations in consciousness, and other variables as predictors of performance in a ganzfeld psi task. *Journal of Parapsychology*, *75*, 235–259.

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. (1997a). A meta-analytic comparison of the sensitivity of direct hits and sum of ranks as outcome measures for free-response studies. *Journal of Parapsychology*, *61*, 227.

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

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.

Parker, A. (2000). A review of the ganzfeld work at Gothenburg University. *Journal of the Society for Psychical Research, 64,* 1–15.

Parker, A. (2006). A ganzfeld study with identical twins. In *Proceedings of the 49th Annual Convention of the Parapsychological Association* (pp. 330–334). Parapsychological Association.

Parker, A. (2010). A ganzfeld study using identical twins. *Journal of the Society for Psychical Research*, *74*, 118–126.

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). Parapsychological Association.

* Parra, A., & Argibay J. C. (2007). Comparing a free-response psychometry test with a free-response visual imagery test for a non-psychic sample. *Journal of the Society for Psychical Research*, *71*, 91–99.

* Parra, A., & Argibay, J. C. (2013a). A free-response ESP test in two hypnotic susceptibility groups: A pilot study. *Australian Journal of Parapsychology*, *13*, 27–35.

* Parra, A., & Argibay, J. C. (2013b). Psi and death of the person-target: An experiment with highly emotional iconic representations. *NeuroQuantology*, *4*(1), 537–543.

Parra, A., & Argibay, J. C. (2013c). Psi and death of the person-target: An experiment with highly emotional iconic representations. Paper presented at the 56th Parapsychological Association Convention, Viterbo.

Parra, A., & Villanueva, J. (2010). Unusual perceptual experiences and ESP under psychomanteum stimulation: Imagery/Hallucination proneness and schizotypal personality measures. *Australian Journal of Parapsychology*, *10*, 120–138.

Parra, A., & Villanueva, J. (2011). Mirror-gazing facility and psi: Examining personality measures. *Journal of the Society for Psychical Research*, *75*, 177–190.

* Parra, A., & Villanueva, J. (2015). Exploring psychomanteum as a psi-conducive state of consciousness. *NeuroQuantology*, *13*(2), 154–159.

* Pérez-Navarro, J. M., & Cox, K. (2012). Context-dependence, visibility, and prediction using state and trait individual differences as moderators of ESP in a ganzfeld environment. *Europe's Journal of Psychology*, *8*(4), 559–572. doi:10.5964/ejop.v8i4.507

* Pérez-Navarro, J., & Guerra, X. M. (2012). An empirical evaluation of a set of recommendations for extrasensory perception experimental research. *Europe's Journal of Psychology*, *8*(1), 32–48. doi:10.5964/ejop.v8i1.297

* Pérez-Navarro, J. M., Lawrence, T., & Hume, I. (2009a). Personality, mental state, and procedure in the experimental replication of ESP: A preliminary study of new variables. *Journal of the Society for Psychical Research*, *73*, 17–32.

* Pérez-Navarro, J. M., Lawrence, T., & Hume, I. (2009b). Personality, mental state, and procedure in the experimental replication of ESP: A logistic regression analysis of a successful experimental condition. *European Journal of Parapsychology*, *24*, 68–92.

Raburn, L. (1975). Expectation and transmission factors in psychic functioning. Unpublished honors thesis, Tulane University, New Orleans, LA.

* Robinson, E. (2009). Student brief: Examining the case for dream precognition. *European Journal of Parapsychology*, *24*, 93–110.

* Rock, A. J., Storm, L. , Harris, K. & Friedman, H. (2012). Shamanic-like journeying and psi-signal detection: I. In search of the psi-conducive components of a novel experimental protocol. *Journal of Parapsychology*, *76*(2), 321–344.

Roe, C. A., Cooper, C. E., Hickinbotham, L., Hodrien, A., Kirkwood, L. & Martin, H. (2020). Performance at a precognitive remote viewing task, with and without ganzfeld stimulation: Three experiments. *Journal of Parapsychology*, *84*, 38–65.

* Roe, C., Cooper, C. & Martin, H. (2010). A comparison between remote viewing and ganzfeld conditions in a precognition task. Proceedings of the 53rd PA Conference, Paris, 21–22.

* Roney-Dougal, S. H., Ryan, A., & Luke, D. (2011). Preliminary study of the relationship between local geomagnetic activity, Tibetan Buddhist meditation and psychic awareness: Non-significant results. Proceedings of the 54th Parapsychological Association Convention, Curitiba, Brazil.

* Roney-Dougal, S. H., Ryan, A., & Luke, D. (2014). The relationship between local geo-magnetic activity and psychic awareness. *Journal of Parapsychology*, *78*(2), 235–254.

* Roney-Dougal, S. H., & Solfvin, J. (2011). Exploring the relationship between Tibetan meditation attainment and precognition. *Journal of Scientific Exploration*, *25*(1), 29–46.

Rouder, J. N., Haal, J. M., Davis-Stober, C., & Hilgard, J. (2019). Beyond overall effects: A Bayesian approach to finding constraints across a collection of studies in meta-analysis. *Psychological Methods*, *24*, 606–621.

Rouder, J. N., Morey, R. D., & Province, J. M. (2013). A Bayes factor meta-analysis of recent extrasensory perception experiments: Comment on Storm, Tressoldi, and Di Risio (2010). *Psychological Bulletin*, *139*(1), 241–247. doi: 10.1037/a0029008.

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.

Steinkamp, F., Milton, J., & Morris, R. L. (1998). A meta-analysis of forced-choice experiments comparing clairvoyance and precognition. *Journal of Parapsychology*, *62*, 193–218.

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

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., & Rock, A. J. (2009). Shamanic-like journeying and psi: I. Imagery cultivation, paranormal belief, and the picture-identification task. *Australian Journal of Parapsychology*, *9*, 165–192.

Storm, L., Sherwood, S. J., Roe, C. A., Tressoldi, P. E., Rock, A. J., & Di Risio, L. (2017). On the correspondence between dream content and target material under laboratory conditions: A meta-analysis of dream-ESP studies, 1966–2016. *International Journal of Dream Research*, *10*(2), 120–140.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2010). Meta-analyses of free-response studies 1992–2008: Assessing the noise reduction model in parapsychology. *Psychological Bulletin*, *136*(4), 471–485. doi:10.1037/a0019457.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2012). Meta-analyses of ESP studies 1987–2008: Assessing the success of the forced-choice design in parapsychology. *Journal of Parapsychology*, *76*, 243–273.

Storm, L., Tressoldi, P. E., & Utts, J. (2013). Testing the Storm et al. (2010) meta-analysis using Bayesian and frequentist approaches: Reply to Rouder et al. (2013). *Psychological Bulletin*, *139*, 248–254.

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*. Puente.

Timm, U. (2000). The problem of finding an optimal measure of effect size for psi experiments. *Journal of Parapsychology*, *64*, 253–254.

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

* Watt, C. (2014). Precognitive dreaming: investigating anomalous cognition and psychological factors. *Journal of Parapsychology*, *78*(1), 115–125.

Watt, C., Dawson, E., Tullo, A., Pooley, A., & Rice, H. (2020). Testing precognition and altered state of consciousness with selected participants in the ganzfeld: A pre-registered study. *Journal of Parapsychology, 84, 21–37*.

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

* Watt, C., Wiseman, R. & Vuillaume, L. (2015). Dream precognition and sensory incorporation: A controlled sleep laboratory study. *Journal of Consciousness Studies*, *22*(5–6), 172–190.

Williams, B. J. (2011). Revisiting the ganzfeld ESP debate: A basic review and assessment. *Journal of Scientific Exploration*, *25*, 639–661.

APPENDIX A

TABLE A1.

Ganzfeld, nonganzfeld / noise-reduction, and standard free-response studies (category, trials, Z score, and Effect Size).

Study	Trials	Hits	Z	ES (z/√n)
Category 1 (Ganzfeld)				
1. Marcusson-Clavertz & Cardeña (2011)	26	7	0.04	0.008
2. Pérez-Navarro & Cox (2012) — Exp1	60	20	1.34	0.173
3. Pérez-Navarro & Cox (2012) — Exp2	90	24	0.24	0.025
4. Pérez-Navarro & Cox (2012) — Exp3	55	18	1.17	0.158

Study	Trials	Hits	Z	ES $(z/\sqrt{n})$
5. Pérez-Navarro & Guerra (2012) — Expt condition	50	15	0.65	0.092
6. Pérez-Navarro, Lawrence, & Hume (2009a)	30	5	−0.84	−0.153
7. Pérez-Navarro, Lawrence, & Hume (2009a)	30	13	2.11	0.385
8. Pérez-Navarro, Lawrence, & Hume (2009b)	90	30	1.70	0.179
9. Roe, Cooper, & Martin (2010)	40	14	1.28	0.202
Totals (trials, hits, mean Z, and mean ES)	471	146	0.85	0.119

Study	Trials	Hits	Z	ES $(z/\sqrt{n})$
Category 2 (Nonganzfeld, Noise-Reduction)				
1. Luke, Zychowicz, et al. (2012)	143	33	−0.43	−0.036
2. Luke & Zychowicz (2014)	268	69	0.21	0.013
3. Parra & Argibay (2007) — psychometry	284	63	−1.03	−0.061
4. Parra & Argibay (2007) — non-psychometry	142	49	2.52	0.211
5. Parra & Argibay (2013b) — Living/Dead	856	511	5.64	0.193
6. Parra & Argibay (2013b) — Suicide/non-Suicide	856	458	2.02	0.069
7. Parra & Villanueva (2015) — psychomanteum	130	40	1.42	0.125
8. Pérez-Navarro, Lawrence, & Hume (2009a)	30	7	−0.36	−0.065
9. Pérez-Navarro, Lawrence, & Hume (2009a)	30	7	0.36	−0.065
10. Robinson (2009)	100	52	0.30	0.030
11. Rock, Storm, Harris, & Friedman (2012) — Drum	54	14	0.03	0.004
12. Rock, Storm, Harris, & Friedman (2012) — Voice	51	15	0.57	0.080
13. Rock, Storm, Harris, & Friedman (2012) — Voice/ Drum	52	14	0.16	0.022
14. Roney-Dougal, Ryan, & Luke (2011)	152	30	−1.40	−0.114
15. Roney-Dougal, Ryan, & Luke (2014)	224	49	−1.00	−0.067
16. Roney-Dougal, & Solfvin (2011)	80	25	1.16	0.130
17. Storm & Rock (2009)	55	19	1.48	0.200
18. Watt (2014)	219	67	1.83	0.124
19. Watt, Wiseman, & Vuillaume (2015)	20	6	0.26	0.058
Totals (trials, hits, mean Z, and mean ES)	3746	1528	0.69	0.045

Study	Trials	Hits	Z	ES $(z/\sqrt{n})$
Category 3 (Standard Free-Response)				
1. da Silva (2011)	448	43	−1.79	−0.085
2. da Silva (2011)	560	62	0.77	0.033

3. da Silva (2011)	560	70	1.90	0.080
4. da Silva (2011)	320	51	3.45	0.193
5. da Silva (2011)	520	84	4.60	0.202
6. da Silva (2011)	600	77	2.25	0.092
7. Luke, Zychowicz, et al. (2012)	57	10	−1.15	−0.152
8. Luke & Zychowicz (2014)	132	38	0.90	0.078
9. Parra & Argibay (2013a) — High Hypnotizability	40	16	2.01	0.318
10. Parra & Argibay (2013a) — Low Hypnotizability	162	34	−1.09	−0.086
11. Parra & Villanueva (2015) — nonpsychomanteum	130	38	1.01	0.089
12. Pérez-Navarro & Guerra (2012) — control	50	11	−0.33	−0.047
13. Rock, Storm, Harris, & Friedman (2012)	43	10	−0.09	−0.014
14. Roe, Cooper, & Martin (2010)	40	12	0.55	0.087
15. Storm & Rock (2009)	53	12	−0.24	−0.033
Totals (trials, hits, mean Z, and mean ES)	3715	568	0.85	0.050

APPENDIX B

Copy of the Syntax used for the Bayesian Meta-Analyses

```
library(metaBMA)
GZ<-subset(data,Category==1)
ASC<-subset(data,Category==2)
NO<-subset(data,Category==3)
#Ganzfeld category
GZ <- meta_random(GZ$ES, sqrt(GZ$VAR),data$Authors,
    tau = prior(family='invgamma', param = c(shape = 1, scale = 0.15)),
    d = prior(family='norm', param = c(mean=.01, sd=.3), lower = .0),
    iter = 10000, chains = 6, cores = 4)
GZ
#ASC category
ASC <- meta_random(ASC$ES, sqrt(ASC$VAR),
    tau = prior(family='invgamma', param = c(shape = 1, scale = 0.15)),
    d = prior(family='norm', param = c(mean=.01, sd=.3), lower = .0),
    iter = 10000, chains = 6, cores = 4)
ASC
#Normal category
NO <- meta_random(NO$ES, sqrt(NO$VAR),
    tau = prior(family='invgamma', param = c(shape = 1, scale = 0.15)),
    d = prior(family='norm', param = c(mean=.01, sd=.3), lower = .0),
    iter = 10000, chains = 6, control=list(adapt_delta=.97),cores = 4)
```