META-ANALYSIS OF ESP STUDIES, 1987–2010: ASSESSING THE SUCCESS OF THE FORCED-CHOICE DESIGN IN PARAPSYCHOLOGY

By Lance Storm, Patrizio E. Tressoldi*, and Lorenzo Di Risio*

ABSTRACT: We report the results of a meta-analysis on forced-choice ESP studies which used targets such as card symbols, numbers, letters, and so forth. For the period 1987 to 2010, a homogeneous dataset of 72 forced-choice studies yielded a weak but significant mean effect size (*ES*) of 0.01 (Stouffer $Z = 4.86$, $p = 5.90 \times 10^{-7}$). There was no evidence that these results were due to low-quality design or selective reporting. The clairvoyance studies did not produce a significantly higher mean *ES* than the precognition studies, and target type did not make a difference to effect size. We note that effects do not vary between investigators, but we did find suggestive evidence that the number of choices per trial is inversely related to the *p* value. We also found evidence of a linear incline in *ES* values indicating that effect sizes have increased over the period 1987 to 2010. Suggestions are made that might help facilitate further increases in effect sizes.

Keywords: ESP, forced-choice, meta-analysis, paranormal, parapsychology, psi

In the 1930s, J. B. Rhine became the first major laboratory-based empirical investigator of alleged paranormal phenomena. He is most famous for his card-guessing studies (Rhine et al., 1940/1966), first using the standard 52-card deck of playing cards, and then switching to so-called Zener cards (consisting of five symbols: star, wavy lines, square, circle, and cross) named after perceptual psychologist Karl Zener (see Figure 1).

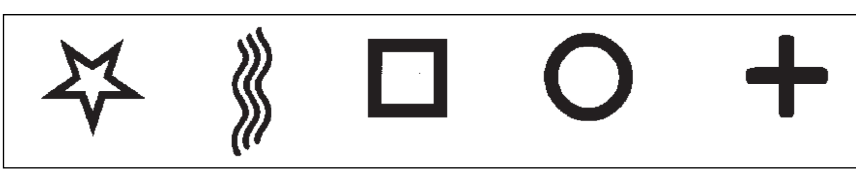

Figure 1. The Zener-card symbols: star, wavy lines, square, circle, and cross.

timuli in these early parapsychological studies were not only card symbols, but were also numbers, alphabet letters, words, shapes, and so forth. Collectively, tests of extrasensory perception (ESP) using these kinds of stimuli are known as “forced-choice” because the target-guess is “one of a limited range of possibilities which are known to [the participant] in advance” (Thalbourne, 2003, p. 44).

Decades later, meta-analytic studies of forced-choice studies confirmed that the collective efforts of many investigators, in a number of laboratories worldwide, over many years, had been fruitful, as indicated by accumulative effects that were statistically significant, though effects were variable in strength and tended to be weak (see Honorton & Ferrari, 1989; Steinkamp, Milton, & Morris, 1998; Tart, 1983). However, since more than a decade has passed since the last meta-analysis of the forced-choice domain (i.e., Steinkamp, Milton, & Morris, 1998), one of the major aims of the present article is to evaluate the performance of forced-choice studies conducted in the period 1987 to 2010. This evaluation will serve to determine if there is still statistical support for the hypothesis that human beings have a sensory modality that functions in an anomalous manner, often referred to as "psi." We also assess whether there is a difference between two types of psi function known as "clairvoyance" and "precognition." Clairvoyance is "paranormal acquisition of information concerning an object or contemporary physical event" (Thalbourne, 2003, p. 18), whereas precognition is a form of ESP "in which the target is some future event that cannot be deduced from normally known data in the present" (Thalbourne, 2003, p. 90). The following section is a review of ESP forced-choice studies.

Telepathy, Precognition, and Clairvoyance

In a typical forced-choice study, participants attempt to identify a concealed target such as a symbol, letter, or number. Forced-choice designs can test telepathy (the "paranormal acquisition of information concerning the thoughts, feelings or activity of another conscious being," Thalbourne, 2003, p. 125), but they can also test clairvoyance (distant seeing), or precognition (future-seeing), depending on the experimental design.

Tart (1983) conducted one of the first meta-analyses on the forced-choice domain. He found 85 ESP studies, but discarded studies if they did not reach independent significance nominally set at a critical $\alpha = .05$. He found that real-time ESP (i.e., clairvoyance/telepathy) out-performed precognition ($p < 5.00 \times 10^{-4}$). Steinkamp, Milton, and Morris (1998) argued that Tart's selection criteria introduced a bias which may have put the precognition studies at a disadvantage.

Forced-choice precognition experiments from as early as 1935 up to 1987 were meta-analyzed by Honorton and Ferrari (1989). They used only the studies where participants had to "predict the identity of target stimuli selected randomly over intervals ranging from several hundred milliseconds to 1 year following the subjects' responses" (p. 281). A total of 309 studies (62 of which were from "senior authors") were analyzed, amassing 50,000 participants and approximately two million individual trials. The effect size (*ES*), calculated using the formula $\Sigma z/\sqrt{N}$ (where z = z score for each study, and N = number of studies), for all studies was 0.02 (mean z = 0.65). Ninety-two studies (30%) showed significant hitting at the 5% level. When outlier studies contributing to the heterogeneity of the database were removed (reduced N = 248) the *ES* fell to 0.01; mean z = 0.38 (Stouffer Z = 6.02, $p = 1.10 \times 10^{-9}$, one-tailed). Also of note is the fact that experienced participants performed better than naïve participants, *ES* = 0.05 versus *ES* = 0.01, respectively.

Honorton and Ferrari also found that precognition forced-choice experiments, although demonstrating a weak effect, produced consistent ("very robust") and highly significant results across a time span of more than 50 years. While effects have remained stable, the quality of studies has improved, indicating that the effects could not have been artifacts of poor design. Unfortunately, Honorton and Ferrari (1989) felt it was "bad news" that the effect sizes had not increased over this time since the stable *ES* meant investigators had not "develop[ed] sufficient understanding of the conditions underlying the occurrence (or detection) of these effects to reliably increase their magnitude" (p. 294). However, Honorton and Ferrari's meta-analysis revealed that the largest effect sizes were found in experiments using (a) experienced participants, (b) independent testing (one participant at a time) as opposed to group testing, and (c) trial-by-trial feedback—important factors for future researchers to consider.

A later study by Steinkamp et al. (1998) meta-analyzed forced-choice studies for the period 1935–1997 (an additional 10 years compared to Honorton and Ferrari's meta-analysis). At the same time, Steinkamp et al. compared matched clairvoyance and precognition studies in order to ascertain statistical evidence of a phenomenological difference between the two. They hypothesized that clairvoyance studies would have a higher effect size because precognition had an extra "calculational step," involving "real-time ESP" (telepathy/clairvoyance) and then extrapolation from that information "to make an informed prediction about future events" (p. 193).

Steinkamp et al. (1998) used a total of 22 comparable study-pairs in their meta-analysis, where procedures were effectively the same in both types of studies. Effect sizes for precognition and clairvoyance were almost identical. Because the sample was so small (*N* = 22 study-pairs), they calculated *N*-weighted effect sizes, again with essentially no difference in outcome (precognition: *ES* = 0.034; clairvoyance: *ES* = 0.030). Steinkamp et al. felt that their coding method may have been responsible for this nonsignificant result, and that a different method for coding study comparability might yield different results. They concluded that the burden of proof rested with those "who argue for a difference between effect sizes under real-time and future ESP" (p. 209). Based on the reasonable argument put forward by Steinkamp et al., we also assessed whether clairvoyance studies have a higher mean effect size than precognition studies.

Steinkamp's (2005) comprehensive review of forced-choice studies from 1880 to 1989 considered various predictors and variables alleged to be conducive of paranormal effects. She noted that "there are few variables that have correlated clearly with success" (p. 155) and she was rather critical of the variations in study designs because these made it difficult to ascertain clear patterns due to conflicting outcomes. Nevertheless, Steinkamp found that "reviews consistently suggested that low neuroticism, extraversion, and good social adjustment may be positively related to forced-choice ESP scoring" (p. 158).

Steinkamp also found "promising evidence" that selected participants on the basis of prior testing performed better in precognition tests, and nonbelievers in psi scored lower than believers (see also the meta-analysis on paranormal belief by Lawrence, 1993). She also found "partial evidence" that meditation and hypnosis susceptibility were conducive of paranormal effects.

Target Types

In the forced-choice experimental domain, target types have varied considerably over the decades. Examples include face and symbol cards from the 52-card playing deck and Zener cards (Rhine et al., 1940/1966; see Figure 1); pictures (Watt, 1996); drawings (Simmonds-Moore & Moore, 2009); and even letters (Vaughan & Houck, 1993, 2000) and numbers (Palmer, 2009). Other experimenters have opted for more ecologically valid targets based on the premise that these are more in keeping with real world scenarios where psi might therefore be better elicited (e.g., Roe, Davey, & Stevens, 2003, with their horse racing design, and Storm & Thalbourne, 1998–1999, with their *I Ching* design). More recently, Ertel (2005) used physical targets (ping-pong balls and beads) under the same premise.

It is a moot point whether any given target type is uninteresting and/or meaningless (e.g., Zener cards, numbers, letters) compared to others that may be emotionally stimulating and/or meaningful (e.g., divinatory readings, real pictures, video clips). Indeed, shifts from significant to nonsignificant effects can be shown to occur within designs (see Storm's, 2009, *I Ching* series of studies) so that even if there is some degree of participant interest (i.e., emotional stimulation) in divinatory readings, interest does not necessarily lead to a significant psi effect. However, Storm (2008) *did* find that the mean meaningfulness rating was higher for psi hitters (those who successfully predicted their hexagram readings) than for psi missers (those who were not successful in their predictions). In the present study, we considered the issue of target type and tested the hypothesis that target type has influence on effect size.

Experimenter Effect

Rosenthal (1974, 1976) was one of the first psychologists to draw attention to the so-called experimenter effect. He found that the experimenter's expectations, biases, and beliefs can have an independent causal influence on the phenomenon under investigation. The experimenter effect has long been a serious issue in the sciences, so much so that the problem has come to the attention of the popular media (Lehrer, 2010).

Indeed, the experimenter effect threatens to undermine the notion of objective verification of treatment effects through replication.

In parapsychology, the problem is even more pertinent, given the psi hypothesis, because there is virtually no limit to the extent the experimenter effect may have in the experimental setting. Most researchers are familiar with the terms "psi-conducive" and "psi-inhibitive," but these terms are applied equally to both participants *and* experimenters. Note that psi-conducive means "favourable to, or facilitative of, the occurrence of psi" (Thalbourne, 2003, p. 93), whereas psi-inhibitive or "psi-inhibitory" mean "unfavourable to, or suppressive of, the occurrence of psi" (p. 94). Therefore, the experimenter effect becomes a problem when parapsychologists need to know that psi is an ability found in participants, not an effect independent of participants, limited to a handful of experimenters. The effect has been comprehensively reviewed by White (1977). The effect has been observed, or is alleged to be an influence, in all experimental domains in parapsychology, including free-response (e.g., ganzfeld), in relation to which the effect is often mentioned and tested. However, the effect has never been found in the meta-analyses across a broad range of investigators (see Bem & Honorton, 1994; Honorton et al., 1990), which reflects admirably on the domain and the field. More recently, Storm, Tressoldi, and Di Risio (2010) did not find a significant experimenter/laboratory difference for their new ganzfeld database, which they subdivided into seven mutually exclusive experimenter/laboratory groups. We note the distinction that may be drawn between experimenter psychology and experimenter psi, and in this case we are arguing that there may be a conflation of both. Only further investigation can help tease out the difference (see Tart, 2011 for a discussion of these issues). In the forced-choice domain, Honorton and Ferrari (1989) showed no significant difference across 57 investigators in their homogeneous database, thus suggesting that the effect was consistent across many experimenters. Effect size differences between groups of experimenters were investigated in the present study.

Decline Effects and High *k*-Choice Designs

Time-dependent effect size declines have been reported in the parapsychological literature. However, even though Honorton and Ferrari (1989) reported a performance effect that showed no signs of improvement over a 50-year period (even with improvements in experimental designs), they did not report a decline per se. Bierman (2001) too only intimated a decline in the forced-choice domain: "Based on observations in the other databases, we would expect that this [card-guessing] database too would exhibit decline effects" (p. 279). On this assumption, he stated: "It appears these sequential effects are the rule in studies of psi phenomena" (p. 279).

Of course, "sequential effects" are not necessarily problematic if they appear only as declines in the short term but show inclines over the long term. On the basis of Bierman's assumption, we assessed the evidence for an effect-size decline in our forced-choice database of studies that span a 24-year period (1987 to 2010).

Along a similar line of interest for psi researchers is the claim that declines in effect size might be related to the number of choices *k* (i.e., the number of target choices) in a target set. Timm (2000) argued that effect size measures have limited use if they do not adequately account for *k*. He argues that "the significance of ESP experiments must increase not only with *n* but also with decreasing hit probability *p* (or with increasing number of target alternatives $k = 1/p$)" (p. 253). Thus far, there is no empirical support for this claim in the forced-choice databases. However, forced-choice designs lend themselves quite readily to this kind of investigation because the values of *k* have been quite variable over the decades, ranging in the case of our database from 2 up to as many as 26. If Timm's claim is supported statistically, then researchers should be using *k*-choice designs where *k* is not low but high (within reasonable limits) in order to increase significance. In the present study, we tested whether number of target choices is related to z.

General Aims of the Present Study

We note the importance of the forced-choice meta-analyses by Honorton and Ferrari (1989) and Tart (1983), but as of 2010 it has been more than 12 years since forced-choice studies were last meta-analysed, dating from the conclusion date of the most recent meta-analysis by Steinkamp et al. (1998), which we also regard as a pivotal study. We point out, though, that Steinkamp et al. used only matching studies, so that a number of precognition and clairvoyance studies from their search period were overlooked. The general aim of the present study was to conduct an exhaustive meta-analytic review of the forced-choice literature from 1987 to 2010 to determine if a comprehensive up-to-date database is still significant, as has been the case with the previous databases. Apart from testing the *k*-choices hypothesis, we also investigated the differences between databases (precognition and clairvoyance), target types, and experimenters. Our last analysis was on effect size values over time to determine if these increased over the years. Based on the above reviews and claims, the following hypotheses were proposed:

1. Forced-choice studies produce statistical evidence of a communications anomaly known as ESP.
2. Clairvoyance studies have a higher mean effect size (*ES*) than precognition studies.
3. *ES* values vary in strength according to target types.
4. *ES* values vary between experimenters.
5. Number of choices (*k*) per trial is positively related to z.
6. *ES* values increased over the period of analysis (i.e., 1987–2010).

Method

Study Retrieval

The following major peer-reviewed journals were accessed for studies: *Australian Journal of Parapsychology, British Journal of Psychology, European Journal of Parapsychology, Explore, International Journal of Parapsychology, Journal of Cognitive Neuroscience, Journal of Parapsychology, Journal of Scientific Exploration, Journal of the American Society for Psychical Research, Journal of the Society for Psychical Research, Perceptual and Motor Skills, Proceedings of the Annual Convention of the Parapsychological Association, Psychological Reports,* and *The Open Psychology Journal.*

The search period for our meta-analysis was 1987 to 2010, which continues from Honorton and Ferrari (1989). Their period of assessment was 1935 to 1987. We note one problem with the Honorton and Ferrari study—it includes precognition studies only. In that regard, the Steinkamp et al. (1998) study is more comprehensive as it includes precognition and clairvoyance and covers a longer period, 1935–1997. Although the conclusion year of Steinkamp et al.'s (1998) study was actually 1997, we did not start our search from 1998 because Steinkamp et al. sought only matching precognition/clairvoyance studies, of which there were 22, although 25 studies came under their consideration (pp. 210-212). We found an additional 14 papers (21 studies) within the period 1987 to 1997 that Steinkamp et al. did not include due to their matching criterion. We include these 21 studies in our database.

We also realize that part of our period of analysis (i.e., 1987 to 1997) overlaps with that of Steinkamp et al., but there is only one study (i.e., Haraldsson, 1993) from Steinkamp et al., that is of interest to us. Two case studies (Don, McDonough, & Warren, 1992, and McDonough, Warren, & Don, 1990) did not meet one of our criteria (see Selection Criteria below in regard to case studies). Thus, the component of our meta-analysis that covers precognition studies is a direct continuation from Honorton and Ferrari, so that a 75-year period of precognition studies is comprehensively represented. The component of our meta-analysis that deals with clairvoyance studies covers the period 1987 to 2010. Though not comprehensive, it is more representative of the forced-choice literature by including telepathy studies as well.

To find appropriate papers online we conducted exhaustive Internet searches through EBSCOhost of the most relevant databases, including PsycINFO, PsycARTICLES, and CINAHL, as well as other relevant databases: Medline, Web of Science, LexScien, Informit, Google Scholar, Scopus, WorldCat, and UMI (dissertations). The following keywords and subject headings were entered in the search: "extrasensory perception," "ESP," "telepathy," "clairvoyance," "precognition," "anomalous cognition," "parapsychology," "paranormal," and "psi" (the term "psi" is short-hand for a psychic or paranormal effect). Most of these Internet searches yielded studies already found in the above-listed journals.

Selection Criteria

We adopted the following selection criteria, and if any of these criteria were not met, we excluded the study:

1. Studies had to use a forced-choice design;
2. Studies had to test ESP only (therefore excluding studies that expressly tested psychokinesis);
3. Studies had to use human participants only (not animals);
4. Number of participants had to be in excess of two to avoid the inherent problems that are typical in case studies;
5. Randomization procedures for selection of targets could not be manipulated by the experimenter or participant;
6. Studies had to provide sufficient information (e.g., number of trials and outcomes) for the authors to calculate the direct hit rates and apply appropriate statistical tests and calculate effect size (*ES*) as $z/\sqrt{n}$.

Procedure

The Appendix lists a total of 91 forced-choice studies drawn from 65 papers that satisfied the above criteria for inclusion in the meta-analysis (see References for articles by name—articles marked with an asterisk indicate papers included in the meta-analysis).

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 (telepathy, clairvoyance, or precognition), (e) number of alternatives in the tasks, (f) total number of hits. We preferred the direct hits 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 proportion of hits and compared these to the proportions expected by chance (i.e., mean chance expectation, or MCE). When available, we collected the corresponding standard normal deviate z value and effect size (where $ES = z/\sqrt{n}$). When missing, we calculated these data using the binomial exact probability (http://faculty.vassar.edu/lowry/binomialX.html).

Forced-choice studies were rated for quality based on criteria used by Honorton and Ferrari (1989), and Steinkamp et al. (1998). These criteria are:

1. Number of trials preplanned;
2. Appropriate randomisation (using electronic apparatuses or random tables);
3. Random target positioning;
4. Automated scoring;
5. Safeguard(s) against data manipulation by participants;
6. Experimenter(s) blind to target identity.

Two judges were used to rate the studies for their quality. One (PET) is a full time researcher with 8 years of experimental experience in parapsychological investigations, and the other (LDR) is a graduate student with 3 years of experience as research assistant in parapsychological investigations. They 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 1/6 = 0.17; maximum score is 6/6 = 1.00). Forty-one studies out of 91 (45%) received a perfect score from at least one judge. Most criteria (i.e., 4 or more out of 6) were met in 71 of the 91 studies (i.e., 78%).

Results

Descriptive Statistics

Ninety-one studies were reported in 65 papers conducted by 96 investigators (see Appendix). There were a total of 812,626 trials and 221,034 hits. Forty-four studies (48%) were clairvoyance, 33 studies (36%) were precognition, and 14 studies (16%) were telepathy. Four-choice and 5-choice designs were the most frequent (57 studies in total, or 63%); 36 studies used a 4-choice design, accounting for 40% of all 91 studies; and 21 studies used a 5-choice design, accounting for 23% of the studies. Twenty-two studies (24%) used a 2-choice design, six studies (7%) used a 6-choice design, and the remaining six studies used 3- or 26-choice designs (6%).

A forest plot, which is a cumulative representation of studies illustrating possible time trends in effect sizes as new studies are added, was generated to show shifts in the cumulative weight of the evidence over time (Rothstein, Sutton, & Borenstein, 2005). We added studies successively by their publication year and Figure 2 shows a tendency for *ES* values to increase, thus giving evidence that the accumulation of effects over time is not attributable to older *ES* values.

Quality Ratings

Forced-choice. Cronbach's alpha for the two judges' ratings was 0.94, indicating very high inter-rater reliability. The correlation between

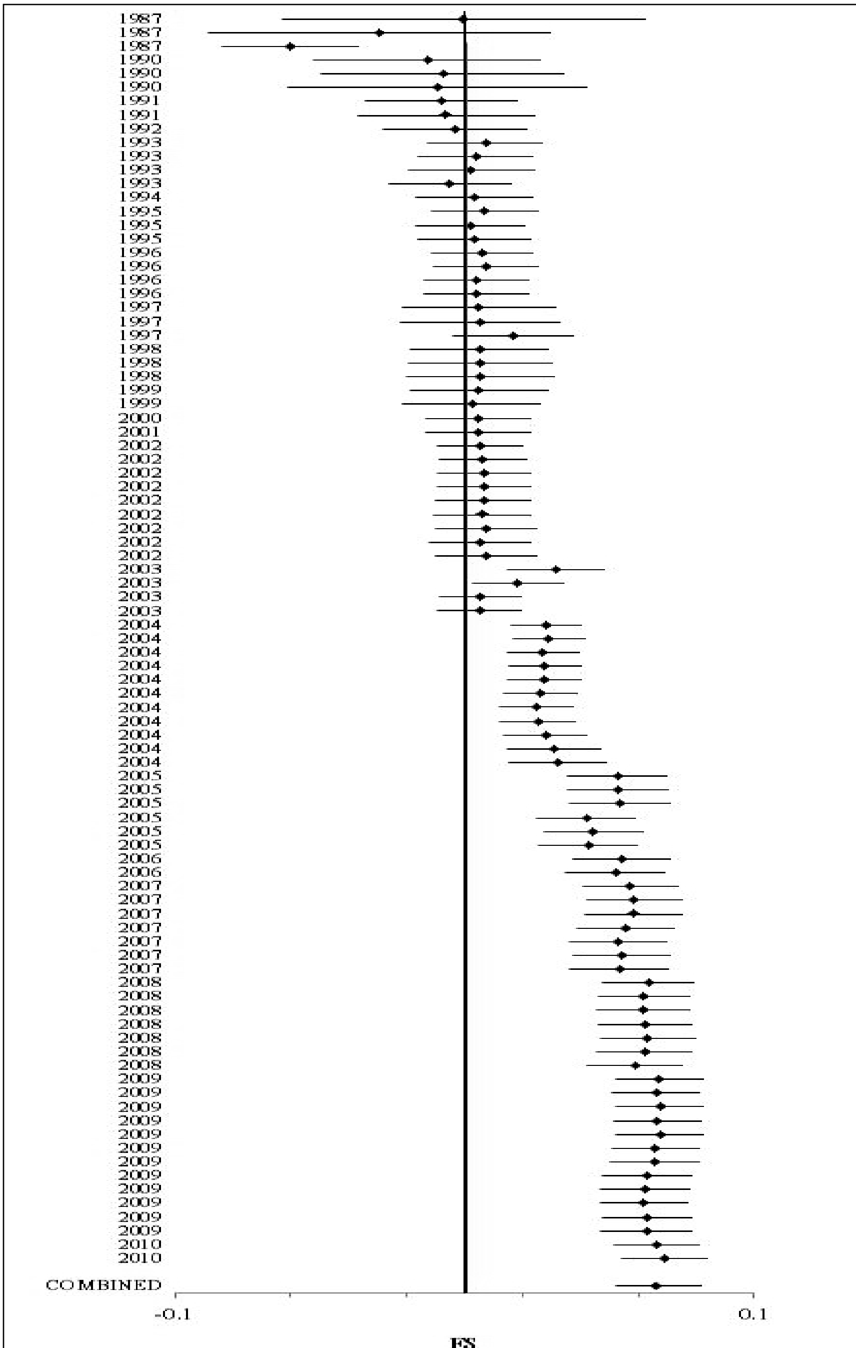

Figure 2. Forest plot showing that, accumulatively, effect size increases on average across time 1987–2010 ($N = 91$).

mean Quality Scores and *ES* values (for the latter, see Appendix) was very weak, negative, and not significant, $r_s(89) = -0.08$ ($p = .45$, two-tailed), suggesting that effect size was not an artifact of poor experimental design. We note that the relationship between quality ratings and year of publication is positive and significant, $r_s(89) = 0.25$ ($p = .016$, two-tailed). In other words, by our criteria, study quality has improved across the span of our period of analysis. (For the test on the relationship between *ES* values and year of publication, see the section Testing for an Effect Size Incline.)

Planned Analyses

H1: Z statistics and effect sizes. We hypothesized that the database would yield a significant effect size (*ES*). The forced-choice database of 91 studies yielded a mean *ES* of 0.04 ($SD = 0.09$), and a mean *z* score of 1.13 ($SD = 3.06$; Stouffer $Z = 10.82$, $p < 10^{-16}$, one-tailed). A single-sample *t* test revealed that *ES* values significantly deviated from chance—specifically mean chance expectation (MCE)—where the test statistic is zero (i.e., MCE = 0.00), $t(90) = 4.02$, $p < .001$, two-tailed. Although the database is significant, the skew of the *ES* distribution was not normal. Outlier studies that cause significant deviations in the distribution may deflate or inflate mean values. Outliers were identified from SPSS Stem-and-Leaf and Box-and-Whiskers Plots as significantly deviant ["extreme"] cases. Such indicators of heterogeneity are standard practice in meta-analysis. Normality was achieved after the removal of 19 studies (study #7, #8, #12, #25, #36, #48, #50, #51, #54, #55, #56, #66, #67, #73, #74, #75, #76, #77, and #80—see Appendix). These 19 studies had *ES* values as low as –0.15 and as high as 0.46. Removal of the 19 studies reduced the database to a homogeneous set of 72 studies (55 papers). There were a total of 790,465 trials and 214,513 hits. Once again, a *t* test revealed that *ES* values significantly deviated from MCE, $t(71) = 3.49$, $p < .001$, two-tailed. Thirty-eight studies (53% of 72 studies) were clairvoyance (Mean $ES = 0.01$, $SD = 0.03$), 25 studies (35% of 72 studies) were precognition (Mean $ES = 0.01$, $SD = 0.04$), and 9 studies (12% of 72 studies) were telepathy, which produced the strongest effect (Mean $ES = 0.04$, $SD = 0.05$).

The reduced database yielded a now weak mean *z* score of 0.57 ($SD = 1.58$; Range: –3.63 to 7.10; 95% CI = 0.20 to 0.94), with a correspondingly weak *ES* of 0.01 ($SD = 0.03$; Range: –0.06 to 0.10; 95% CI = 0.01 to 0.02); and Stouffer $Z = 4.86$ ($p = 5.90 \times 10^{-7}$, one-tailed). The confidence intervals do not include MCE. Of the 72 studies, 44 (61%) had positive *z* scores. Sixteen (22%) of the 72 studies are independently significant ($\alpha = .05$). The database of 72 studies has a quality-weighted $Z = 4.36$ ($p = 6.50 \times 10^{-6}$). (The quality-weighted *z* was calculated using Rosenthal's, 1984, p. 89, Formula 4.31: quality-weighted $z = \Sigma w_j z_j / [\Sigma w_j^2]^{1/2}$, where *w* is the weight, and j ranges from 1 to *k*.) When the two *z* scores were tested using the $Z_{diff.}$ formula, $Z_{diff.} = [Z_1 - Z_2]/\sqrt{2}$, we found no significant difference between them (4.86

– 4.36]/ /√2 = 0.35, p = .36). This result means there are only chance differences between the weighted and unweighted values, so that we can assume that the overall performance of the database is not an artifact of poor quality (i.e., ESP is not an acronym for "Error Some Place," as some skeptics like to claim).

Although we report a significant Stouffer Z of 5.90 for this database, a more stringent approach to testing the significance of 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, which is the "mean of the s highest of k mutually independent values of z" (p. 505), which is then compared to a critical MeanZ. Taking s = 10 (i.e., 10 studies with the highest z scores), and k = 72 (i.e., where k = N = 72, as given above), our MeanZ is 3.28. Darlington and Hayes's online table gives critical MeanZ = 1.91 (For values of k > 50, Darlington and Hayes, 2000, p. 503, refer researchers to their Cornell University online table at www.psych.cornell.edu/darlington/meta/S10.HTM). In other words, the mean z for the forced-choice database is considerably higher than the critical value. Thus, for our first hypothesis, where we proposed that forced-choice studies produce significant effects, we found a 22% success rate for independent studies, a highly significant Stouffer Z score, and a mean z score above the critical value set by Darlington and Hayes (2000).

Using the "file-drawer" formula given by Rosenthal (1995, p. 189), there would have to be no fewer than 557 unpublished papers with overall nonsignificant results to reduce our significant finding to a chance result. 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. Darlington and Hayes (2000) offer a more conservative and reliable test, and they point out that: "If publication bias existed, the average z of the unpublished studies would typically be below zero" (p. 498). The "flaw" in Rosenthal's (1995) formula is that it only calculates the number of studies with an average z value *at* zero. Darlington and Hayes's test determines the number of unpublished nonsignificant studies needed to reduce a database to nonsignificance, but these may also be "highly negative" (i.e., "psi-missing" in the case of parapsychology). Darlington and Hayes (2000) 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). Using Darlington and Hayes's (2000, p. 503) Table 2, if 16 individual results are significant with critical alpha = .05, then the pooled p is less than or equal to .05 only if the total number of studies (the "fail-safe N") is up to 203 studies, with 187 of these being unpublished (i.e., 203 minus 16 studies). Note that these could all be negative (psi-missing) studies, yet our 16 significant studies in this database constitute proof against the null hypothesis. The hypothesized existence of such a large number of (as yet) unpublished null or even psi-missing studies (i.e., up to 187) is unlikely.

H2: Modality comparisons—clairvoyance, precognition, and telepathy. Like Steinkamp et al. (1998), we might have compared the clairvoyance studies with the precognition studies only, but since we have telepathy studies in our database, we considered it useful to compare all three databases simultaneously. We note, however, that our clairvoyance and precognition studies are not matched as was the case with Steinkamp et al. (1998). We nevertheless present the results of our test to give some idea as to the possible quantitative differences that may exist between these psi modalities.

We used a Univariate ANOVA test, entering the variable Psi Modality, but we also entered the variable Target Type (see *H3* below). Our test showed that there was a significant difference between the three psi modalities, $F(2, 44) = 3.29$, $p = .047$, two-tailed, but the post hoc Tukey test showed no significant difference between Clairvoyance and Precognition (Mean Diff. = 0.005 $p = .88$), and no significant difference between Clairvoyance and Telepathy (Mean Diff. = –0.03, $p = .14$). The only difference, which is the apparent source of the significant main effect, is between Precognition and Telepathy, but even so, the difference only approached significance (Mean Diff. = –0.03, $p = .09$).

H3: Effect size and target types. Our third hypothesis takes into account the possibility that target types may affect participants' performances, which might be reflected in *ES* values. We divided our data into five types of target: (a) Pictures/Drawings, (b) Symbols (e.g., trigrams), (c) Numbers, (d) Words/Letters, and (e) Objects (i.e., targets that occupy 3-D physical space). Studies using symbols and numbers had to be excluded as their corresponding cells were empty (i.e., there were no telepathy studies that used symbols and only precognition studies used numbers).

The same ANOVA test conducted to test Hypothesis 2 showed that there was no significant difference between the three groups defined by target type, $F(2, 44) = 1.91$, $p = .16$, two-tailed. Target type did not make a difference for effect size. However, there was a significant interaction effect between Modality and Target Type, $F(4, 44) = 3.24$, $p = .02$, two-tailed; partial eta squared = .23. Figure 3 shows the interaction effect. Telepathy produced the strongest effects but only for Words/Letters (*ES* = 0.07) and Objects (*ES* = 0.08), but there were only five studies with these two target types combined. All other effects ranged between –0.001 and 0.02.

H4: Effect size differences between experimenters. For our fourth hypothesis, in order to ascertain whether our database was the result of extremely positive *ES* values for a limited pool of experimenters, we conducted a Kruskal-Wallis ANOVA on the pooled data after dividing them into experimenter groups. We formed 16 mutually exclusive experimenter groups with at least two studies in each: "Auriol," "Crandall," "Dalkvist," "Don," "Haraldsson," "Palmer," "Pitman," "Rammohan," "Rao," "Roe," "Sheldrake," "Steinkamp," "Storm," "Vaughan," "Watt," and "Wiseman."

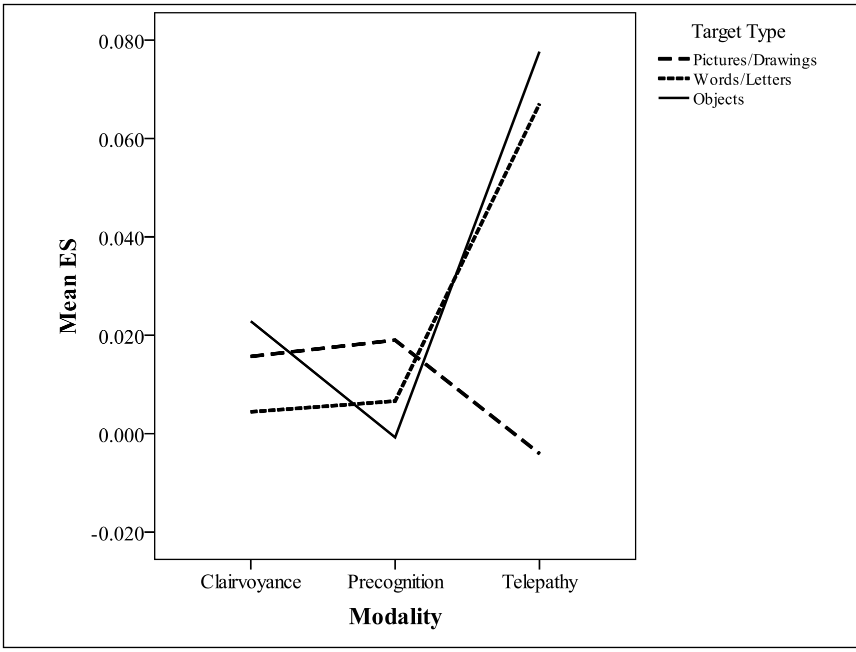

Figure 3. Interaction effects. Effect sizes are not the same across levels of psi modality.

Effect size (*ES*) values were not significantly different between experimenter groups, $\chi^2(15, N = 57) = 18.10$, $p = .36$ (two-tailed). The effects cannot be said to be due to a few outstanding investigators, which contradicts Akers's (1987) claim that psi is an effect not replicated amongst the majority of investigators.

H5: The advantage in using a high *k*-choices design. For our fifth hypothesis, we proposed that the number of choices *k* per trial is positively related to z. There are six values of *k*: 2, 3, 4, 5, 6, and 26. The number of trials and hits are presented, with hit-rates and mean *z* scores, in Table 1.

In accordance with Timm's (2000) conjecture, increasing the number of target alternatives *k* must lead to increased levels of significance in ESP esperiments, which can be discerned from *z* scores. A Pearson's *r* test on *k* values and *z* scores was significant, $r(5) = 0.79$, $p = .03$ (one-tailed). Thus, *z* scores tended to increase as *k* increased. Researchers may wish to consider using high *k*-choice designs if this finding indicates a valid effect.

H6: Testing for an effect size incline. For our sixth and final hypothesis, we proposed that *ES* values have increased over the period 1987 to 2010. We pointed out above that the relationship between quality ratings and year of publication was positive and significant. We note that the correlation between year of study and *ES* is positive and significant for the database ($N = 72$), $r_s(70) = 0.31$, $p = .01$; linear $r^2 = 0.10$), indicating an

Table 1
Hit-Rates for Six Forced-Choice Designs by Number of Choices (k)

k-Choices	*N*	Trials	Hits	Hit-rate(%)	P_{MCE} (%)	Mean *Z*
2-choice	16	195,060	97,430	49.95	50.00	0.11
3-choice	3	113,520	37,961	33.44	33.33	1.46
4-choice	26	113,999	32,333	28.36	25.00	0.49
5-choice	19	190,686	38,886	20.39	20.00	0.99
6-choice	6	6,200	1,008	16.26	16.67	–0.38
26-choice	2	171,000	6,895	4.03	3.85	2.83

incline (see Figure 4), meaning that *ES* values increased over the 24-year period. The linear trend line formula is $ES = (0.002 \times \text{YEAR}) - 3.34$. Based on our other findings presented in the section Quality Ratings above, these results show that improvements in quality over time do not meet with reduced *ES* values, but on the contrary, *ES* values increased. These findings bode well for the forced-choice paradigm.

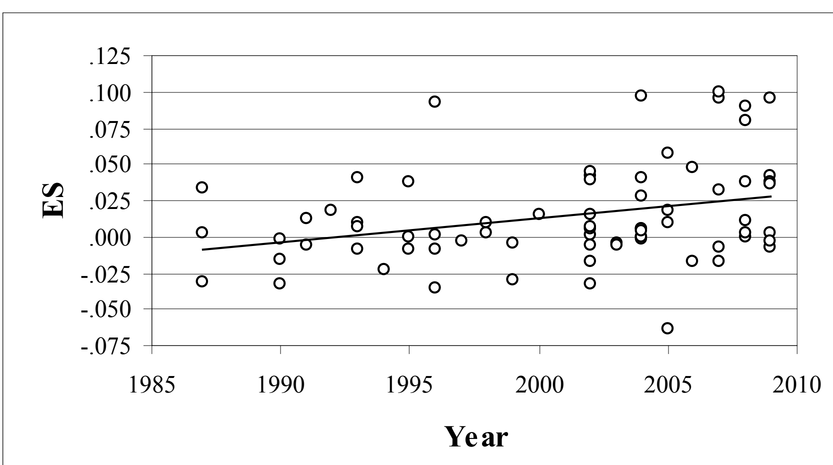

Figure 4. Scatter-plot of forced-choice studies showing a significant incline over 24 years ($r = .31$, $p = .007$).

Honorton and Ferrari (1989) reported no improvement in *ES* values in their homogeneous precognition dataset of 248 studies dating from 1935 to 1987. This finding does not imply a decline, and as we said above, Bierman (2001) only had "expectations" of declines in the forced-choice domain. We, therefore, tested for an incline in our precognition data ($N = 25$) to track the progress of that modality over the period 1987 to 2010. We found the correlation between year of study and *ES* was positive, but not significant, $r_s(23) = .14$, $p = .26$ (one-tailed). Given the small *N* for precognition, the statistic failed to reach significance, so we can only tentatively claim that precognition *ES* values increased over the 24-year period. We note that the precognition incline is not as strong as it is for the three ESP categories combined, but using the Fisher *r*-to-*z* transformation, it is not significantly weaker, $p = .23$ (one-tailed). Because our dataset suggests an incline, we tentatively claim that Honorton and Ferrari's (1989) finding of no improvement of *ES* values for precognition may not be true.

Post Hoc Analyses

Comparisons with two earlier databases. Compared to our database, there are clear limitations in the meta-analyses by Steinkamp et al. (1998) and Honorton and Ferrari (1989). Both meta-analyses included case studies, which we avoided because they are of limited generalizability. Steinkamp et al. restricted their analysis to precognition, although these were matched with clairvoyance studies so their dataset did not comprehensively and independently represent clairvoyance or precognition for their period of analysis. Honorton and Ferrari only considered precognition studies, but they also included self-studies by experimenters, and "an-psi" studies (i.e., ESP studies that tested animals such as rats and gerbils). These studies contributed to the heterogeneity of Honorton and Ferrari's database, and even their trim of almost 20% of outliers (pp. 287–288) produced a spuriously "homogeneous" dataset, because their trim rule was essentially arbitrary. And finally, there is considerable overlap in both studies. Nevertheless, our comparisons give reasonable guidelines as to the progress of precognition and clairvoyance over many decades. Also, we did not compare our precognition database with the precognition database of Steinkamp et al. (1998) because Honorton and Ferrari's (1987) is a far more comprehensive picture of the precognition literature, given that Steinkamp et al. limited their selection to matched precognition/clairvoyance studies. Furthermore, most of Steinkamp et al.'s precognition studies are included in Honorton and Ferrari's (1987) meta-analysis.

First, for clairvoyance, Steinkamp et al. (1998) reported a mean *ES* of 0.009, and a Stouffer $Z = 2.81$ ($p = .002$, one-tailed). For our clairvoyance database ($n = 38$), we report a slightly stronger mean *ES* of 0.013, and a higher Stouffer *Z* of 3.07 ($p = 1.07 \times 10^{-3}$), although our Stouffer *Z* is not quality-weighted. The mean *z* scores for the two databases are 0.57 and 0.50, respectively. Using mean *z* scores and applying Rosenthal and Rubin's (1979) $Z_{diff.}$ formula, $Z_{diff.} = [Z_1 - Z_2]/\sqrt{2}$, we see that the difference is not significant, $Z_{diff} = [0.57 - 0.50]/\sqrt{2} = 0.05$ ($p = .48$). The results indicate no substantial differences between databases.

Second, for precognition, Honorton and Ferrari (1989) reported a mean *ES* of 0.012, mean $z = 0.38$ (Stouffer $Z = 6.02$, $p = 1.10 \times 10^{-9}$, one-tailed). For our precognition database ($n = 25$), we report a slightly lower mean $ES = 0.010$, and a lower Stouffer $Z = 1.92$ ($p = 2.74 \times 10^{-2}$). Interestingly, the mean *z* scores for both databases are 0.38. Again, the results indicate no substantial differences between databases.

Discussion

Our meta-analysis on the forced-choice database of studies from 1987 to 2010 and subsequent tests on possible alternative sources for the statistical anomaly indicate that the forced-choice domain generally produces significant psi effects above mean chance expectation. The results of two analyses show that this significant mean effect is not an artifact of poor-quality experimental design. We stress, however, that effects on average are very weak for precognition, clairvoyance, and even telepathy, which was the strongest effect of the three, albeit marginally significantly stronger than precognition. We add that *ES* values were not dependent on target type, but the interaction effect (see Figure 3) does show that telepathy studies produced the strongest effects if experimenters used words, letters, and objects as targets, though this effect is not reliable given that the size of that group was small ($n = 5$).

Overall, we point out that strength of effect does not undermine the fact that there was a tendency for the elicitation of some kind of anomalous effect in these studies. Furthermore, weak but important effects are often reported in other scientific fields such as medicine. A weak effect ($r = .03$) was reported in the aspirin/heart-attack study by the Steering Committee of the Physicians' Health Study Research Group (1988), but the study was halted on ethical grounds because 45% fewer heart attacks were reported in the experimental group compared to the control group. Likewise, on ethical grounds, the National Heart, Lung, and Blood Institute (DeMets, Hardy, Friedman, & Lan, 1984; Kolata, 1981) discontinued a study 9 months ahead of schedule because of the clear benefits of propranolol on patients with a recent myocardial infarction, even though the effects were also reported to be weak ($r = .04$). These are but two examples (see also Spencer, 1995).

Having divided our database into seven mutually exclusive investigator/experimenter groups, we found no single group that produced effects significantly different from any other group. This finding bodes well for experimenters as they need not fear that psi can be elicited only by a select group of psi-conducive researchers in certain parts of the world.

We considered an issue raised by Timm (2000) that number of choices (*k*) per trial may be positively related to *z* score, since he argued that increasing *k* would mean subsequent increases in significance in ESP experiments. We found evidence of a trend indicating that higher *k* values tend to be associated with higher *z* scores. If this relationship is real, researchers would do well to consider it in future forced-choice designs. Researchers should be especially mindful of this point, since replication is a particularly salient issue in regard to forced-choice studies, the effects of which have been shown to be amongst the weakest, in the order of 1/10 compared to other paranormal domains (see Storm, 2006b; Storm & Thalbourne, 2000; Storm et al., 2010). Among the possible reasons for such low effects are (a) the use of uninteresting and/or meaningless targets (i.e., Zener cards, numbers, letters, etc.) instead of emotionally stimulating and/or meaningful targets (i.e., divinatory readings, real pictures, video clips, etc.), (b) the recruitment of convenience samples instead of selected participants, (c) the use of normal instead of reduced cognitive noise conditions (i.e., ganzfeld, relaxation, meditation), and (d) artificial instead of ecologically valid tasks (such as e-mail or phone call predictions as featured in the Sheldrake & Smart, 2003a, 2003b, studies). We draw the reader's attention to the fact that we did find that target type may make a difference to effect size, but this finding refers only to an interaction effect (see again, Figure 3) whereby the strength of effect was not the same across psi modalities, with telepathy eliciting the greatest differences between target types.

In addition, we found evidence of a time-dependent increase in *ES* values for the period of analysis 1987 to 2010. Other researchers have not found declines in various forced-choice databases (see Bierman, 2001; Honorton & Ferrari, 1989). We add that an increase in *ES* values in our precognition database is merely *suggested* since we had to make do with a low *N* of only 25 studies. Nevertheless, the general trend is that *ES* values do appear to be increasing in a linear fashion. These findings are particularly important given the improvements in design quality over recent decades (i.e., since 1987), yet we did not find the corresponding decreases in *ES* values that might be expected due to design flaws being eliminated. This good news counteracts the "bad news" that Honorton and Ferrari (1989) reported 20 years prior when they bewailed the fact that they saw no indication of effects increasing for forced-choice precognition designs over the span of more than five decades, even with improvements in design quality over that period (i.e., 1935-1987).

Finally, from having compared past databases with our own, we point out that there is suggestive evidence that the performance rate of the typical forced-choice study has been stable over the last 70 years (i.e., since 1935). Some researchers (e.g., Honorton & Ferrari, 1989) might argue that a consistent mean effect size since 1935 indicates insufficient understanding of the conditions underlying the effect. But Honorton and Ferrari were referring only to precognition studies. More generally, it might equally be argued from Steinkamp's (2005) findings for forced-choice studies from 1880 to 1989 that effect sizes do vary when independent variables such as neuroticism, extraversion, and social adjustment are manipulated. With greater focus these days on the processes underlying the hypothesized psi function, we might expect increases in psi effects, even though the database comparisons above do not give statistical evidence of such trends.

Notwithstanding the above findings and considerations, it might still be argued that the forced-choice domain has some distance yet to cover if the scientific community is to be convinced that psi is worthy of investigation when effect sizes become stronger and more consistent. We can only state once again that the significant incline effect reported above may help ameliorate that skeptical attitude.

In closing, we suggest that, as a communications anomaly, there is much to be understood and learned about psi from further research. Nevertheless, as research currently stands, and from our contributions to the meta-analytic body of evidence (see, for example, Storm, Tressoldi, & Di Risio, 2010), this anomaly appears to constitute a major theoretical and practical deviation from mainstream understandings about the nature of information, how it is transmitted and received, and how it might be used. Skeptics taking a pragmatic position might argue that a weak effect might prove difficult to deploy efficaciously in the communications disciplines and industries, as it needs to be sufficiently reliable. Only further research can ascertain whether psi research has reached its limits, or whether the signal-to-noise ratio can be improved to satisfy the needs of disciplines and industry.

We add, as a rebuttal to that skeptical (if not conservative) view, that parapsychologists working in public institutions (e.g., U.S. Department of Defense—see McMoneagle, 1997, 2000), or gainfully employed in other psi-practices, have produced useful and/or profitable results (see also studies by Schouten, 1993; and Schwartz, 1983, 2000, 2005). In his comprehensive review, Schouten (1993) remarked that "complementary medicine deserves a place in the health care system" (p. 399). Schwartz has gone to great lengths to demonstrate the uses of psi in archaeology and anthropology, and this approach extends to crime solving and detection—areas which have already seen applications (Schouten, 1994). Psi may already have been proved to be a human faculty that can make a worthwhile contribution to the services of mankind. Even so, the limits of that faculty are still undetermined.

References

(References marked with an asterisk indicate studies included in the meta-analysis.)

Akers, C. (1987). Parapsychology is science, but its findings are inconclusive. *Behavioral and Brain Sciences, 10,* 566–568.

* Auriol, B. M., Garcia, F., Puech, L., Lagrange, S., Morer, C., Campardon, M., . . . & Leconte, E. (2004). Agapè: Group telepathy. A long-

term experimental series. *Proceedings of Presented Papers: The Parapsychological Association 47th Annual Convention*, 325–246.

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin, 115*, 4–18.

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: Benjamins.

* Crandall, J. E. (1987). Effects of cognitive style and type of target on displacements. *Journal of Parapsychology*, *51*, 191–216.

* Curtis, J. T., & Wilson, J. P. (1997). Sensation seeking and ESP test performance: A preliminary investigation. *Journal of the Society of Psychical Research, 62*, 1–21.

* Dalkvist, J., & Westerlund, J. (1998). Five experiments on telepathic communication of emotions. *Journal of Parapsychology, 62,* 219–253.

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

* De Graaf, T. K., & Houtkooper, J. M. (2004). Anticipatory awareness of emotionally charged targets by individuals with histories of emotional trauma. *Journal of Parapsychology, 68,* 93–127.

DeMets, D. L., Hardy, R., Friedman, L. M., & Lan, K. K. G. (1984). Statistical aspects of early termination in the Beta-Blocker heart attack trial. *Control Clinical Trials*, *5*, 362–372.

Don, N. S., McDonough, B. E., & Warren, C. A. (1992). Psi testing of a controversial psychic under controlled conditions. *Journal of Parapsychology*, *56*, 87–96.

* Don, N. S., McDonough, B. E., & Warren, C. A. (1998). Event-related potential (ERP) indicators of unconscious psi: A replication using subjects unselected for psi. *Journal of Parapsychology, 62,* 127–145.

* Ertel, S. (2005). Psi test feats achieved alone at home: Do they disappear under lab control? *Australian Journal of Parapsychology*, *5*, 149–164.

* Haraldsson, E. (1993). Are religiosity and belief in an afterlife better predictors of ESP performance than belief in psychic phenomena? *Journal of Parapsychology*, *57*, 259–273.

* Haraldsson, E., Houtkooper, J. M., Schneider, R., & Bäckström, M. (2002). Perceptual defensiveness and ESP performance: Reconstructed DMT ratings and psychological correlates in the first German DMT-ESP experiment. *Journal of Parapsychology, 66*, 249–270.

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.

* Houtkooper, J. M. (2003). An ESP experiment with natural and simulated sferics: Displacement scores and psychological variables. *European Journal of Parapsychology, 18,* 49–64.

* Houtkooper, J. M., Schienle, A., Stark, R., & Vaitl, D. (1999). Atmospheric electromagnetism: A possible disturbing influence of natural sferics on ESP. *Perceptual and Motor Skills, 89,* 1179–1192.

Kolata, G. B. (1981). Drug found to help heart attack survivors. *Science, 214,* 774–775.

* Krishna, S. R., & Rao, K. (1991). Effect of ESP feedback on subjects' responses to a personality questionnaire. *Journal of Parapsychology, 55,* 147–158.

* Kugel, W. (1999). Amplifying precognition: Four experiments with roulette. *Proceedings of Presented Papers: The Parapsychological Association 42nd Annual Convention,* 136–145.

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

Lehrer, J. (2010, December 13). The truth wears off. *The New Yorker,* pp. 52–57.

* Lobach, E., & Bierman, D. (2004). Who's calling at this hour? Local sidereal time and telephone telepathy. *Proceedings of Presented Papers: The Parapsychological Association 47th Annual Convention,* 91–97.

* Luke, D. P., Delanoy, D., & Sherwood, S. J. (2008). Psi may look like luck: Perceived luckiness and beliefs about luck in relation to precognition. *Journal of the Society for Psychical Research, 72,* 193–207.

* Luke, D. P., Roe, C. A., & Davison, J. (2008). Testing for forced-choice precognition using a hidden task: Two replications. *Journal of Parapsychology, 72,* 133–154.

* McDonough, B. E., Don, N. S., & Warren, C. A. (2002). Differential event-related potentials to target and decoys in a guessing task. *Journal of Scientific Exploration, 16,* 187–206.

McDonough, B. E., Warren, C. A., & Don, N. S. (1990). Precognition and real-time ESP performance with an exceptional subject: A replication attempt. *Journal of Parapsychology, 54,* 141–150.

McMoneagle, J. (1997). *Mind trek: Exploring consciousness, time, and space through remote viewing.* Charlottesville, VA: Hampton Roads.

McMoneagle, J. (2000). *Remote viewing secrets: A handbook.* Charlottesville, VA: Hampton Roads.

* Milton, J. (1994). Guessing strategies and confidence-call criteria of uninstructed participants in a forced-choice ESP experiment. *Journal of the Society of Psychical Research, 60*, 65–77.

* Mossbridge, J. A., Grabowecky, M., & Suzuki, S. (2009). Evidence for subconscious but not conscious psi in remote stare detection and precognition tasks. *Proceedings of Presented Papers: The Parapsychological Association 52nd Annual Convention*, 12–13.

* Moulton, S. T., & Kosslyn, S. M. (2008). Using neuroimaging to resolve the psi debate. *Journal of Cognitive Neuroscience, 1*, 182–192.

* Murray, C. D., Howard, T., Wilde, D. J., Fox, J. & Simmonds-Moore, C. (2007). Testing telepathy using an immersive virtual environment. *Journal of Parapsychology, 71*, 105–124.

Palmer, J. (1977). Attitudes and personality traits in experimental ESP research. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 175–201). New York: Van Nostrand Reinhold.

* Palmer, J. (1992). Effect of a threatening subliminal stimulus on the perceptual ESP test: A partial replication. *Journal of Parapsychology, 56*, 189–204.

* Palmer, J. (2009). Decision augmentation in a computer guessing task. *Journal of Parapsychology, 73*, 119–136.

* Parra, A., & Argibay, J. C. (2007). Comparing psychics and non-psychics through a "token-object" forced-choice ESP test. *Journal of the Society for Psychical Research, 71*, 80–90.

* Pitman, J. A., & Owens, N. E. (2004). The effect of manipulating expectations both before and during a test of ESP. *Journal of Parapsychology, 68*, 45–63.

* Rammohan, V. G. (1990). Psi in the academic context: Two further studies. *Journal of Parapsychology, 54*, 229–244.

* Rao, R. K., Kanthamani, H., & Palmer, J. (1990). Exploring normal-paranormal interaction within a memory-ESP testing paradigm. *Journal of Parapsychology, 54*, 245–260.

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

* Roe, C. A., Davey, R., & Stevens, P. (2003). Are ESP and PK aspects of a unitary phenomenon? A preliminary test of the relationship between ESP and PK. *Journal of Parapsychology, 67*, 343–366.

* Roe, C. A., Davey, R., & Stevens, P. (2004). Arousal and performance at ESP and PK tasks using a common protocol. *European Journal of Parapsychology, 19*, 29–43.

* Roe, C. A., Davey, R., & Stevens, P. (2005). Are ESP and PK aspects of a unitary phenomenon? The effects of deception when testing the relationship between ESP and PK. *Journal of the Society for Psychical Research, 69*, 18–32.

* Roe, C. A., Davey, R., & Stevens, P. (2006). Experimenter effects in laboratory tests of ESP and PK using a common protocol. *Journal of Scientific Exploration*, *20*, 239–253.

* Roe, C. A., Henderson, S. J., & and Matthews, J. (2008). Extraversion and performance at a forced-choice ESP task with verbal stimuli: Two studies. *Journal of the Society for Psychical Research*, *72*, 208–221.

Rosenthal, R. (1974). *On the social psychology of the self-fulfilling prophecy: Further evidence for pygmalion effects and their mediating mechanisms.* New York: MSS Modular.

Rosenthal, R. (1976). *Experimenter effects in behavioral research.* New York: Irvington.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, *118*, 183–192.

Rosenthal, R., & Rubin, D. B. (1979). Comparing significance levels of independent studies. *Psychological Bulletin*, *86*, 1165–1168.

* Sartori, L., Massacessi, S., Martinelli, M., & Tressoldi, P. E. (2004). Physiological correlates of ESP: Heart rate differences between targets and nontargets. *Journal of Parapsychology*, *68*, 351–360.

* Schmidt, S., Erath, D., Ivanova, V., & Walach, H. (2009). Do you know who is calling? Experiments on anomalous cognition in phone call receivers. *Open Psychology Journal*, *2*, 12–18.

Schouten, S. A. (1993). Applied parapsychology studies of psychics and healers. *Journal of Scientific Exploration*, *7*, 375–401.

Schouten, S. A. (1994). An overview of quantitatively evaluated studies with mediums and psychics. *Journal of the American Society for Psychical Research*, *88*, 221–254.

Schwartz, S. (1983). Preliminary report on a prototype applied parapsychological methodology for utilization in archaeology, with a case report [Abstract]. In W. G. Roll, R. L Morris, & R. White (Eds.), *Research in parapsychology 1981* (pp. 25–27). Metuchen, NJ: Scarecrow Press.

Schwartz, S. (2000). *The Alexandria project.* New York: Author's Guild.

Schwartz, S. (2005). *The secret vaults of time.* Charlottesville, VA: Hampton Roads.

* Sheldrake, R., & Avraamides, L. (2009). An automated test for telepathy in connection with emails. *Journal of Scientific Exploration*, *23*, 29–36.

* Sheldrake, R., Avraamides, L., & Novák, M. (2009). Sensing the sending of SMS messages: An automated test. *Explore*, *5*, 272–276.

* Sheldrake, R., & Beharee, A. (2009). A rapid online telepathy test. *Psychological Reports*, *104*, 957–970.

* Sheldrake, R., & Lambert, M. (2007). An automated online telepathy test. *Journal of Scientific Exploration*, *21*, 511–522.

* Sheldrake, R., & Smart, P. (2003a). Experimental tests for telephone telepathy. *Journal of the Society for Psychical Research, 68*, 184–199.

* Sheldrake, R., & Smart, P. (2003b). Videotaped experiments on telephone telepathy. *Journal of Parapsychology, 67*, 187–206.

* Sheldrake, R. & Smart, P. (2005). Testing for telepathy in connection with e-mails. *Perceptual and Motor Skills, 101*, 771–786.

* Simmonds-Moore, C. A., & Moore, S. L. (2009). Exploring how gender role and boundary thinness relate to paranormal experiences, beliefs and performance on a forced-choice clairvoyance task. *Journal of the Society for Psychical Research, 73*, 129–143.

* Smith, M. D., Wiseman, R., Machin, D., Harris, P., & Joiner, R. (1997). Luckiness, competition and performance on a psi task. *Journal of Parapsychology, 61*, 33–43.

Spencer, B. (1995). Correlations, sample size, and practical significance: A comparison of selected psychological and medical investigations. *Journal of Psychology, 129*, 469–475.

* Steinkamp, F. (2002). Psi on the web: Two replication studies. *Journal of Parapsychology, 66*, 169–176.

Steinkamp, F. (2005). Forced-choice ESP experiments: Their past and their future. In M. A. Thalbourne & L. Storm (Eds.), *Parapsychology in the 21st century: Essays on the future of psychical research* (pp. 124–163). Jefferson, NC: McFarland.

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.

Steering Committee of the Physicians Health Study Research Group (1988). Preliminary report: Findings from the aspirin component of the ongoing physicians' health study. *New England Journal of Medicine, 318*, 261–264.

* Storm, L. (2002). A parapsychological investigation of the *I Ching*: Seeking psi in an ancient Chinese system of divination. *Australian Journal of Parapsychology, 2*, 44–62.

* Storm, L. (2006a). A parapsychological investigation of the *I Ching*: The relationships between psi, intuition, and time perspective. *Journal of Parapsychology, 70*, 121–141.

Storm, L. (2006b). Meta-analysis in parapsychology: II. Psi domains other than ganzfeld. *Australian Journal of Parapsychology, 6*, 135–155.

* Storm, L. (2008). Investigations of the *I Ching*: I. Relationships between psi, time perspective, paranormal belief and meaningfulness. *Australian Journal of Parapsychology, 8*, 103–127.

Storm, L. (2009). Investigations of the *I Ching*: II. Reliability and validity studies. *Australian Journal of Parapsychology, 9*, 111–142.

* Storm, L., & Thalbourne, M. A. (1998–1999). The transliminal connection between personality and paranormal effects in an experiment with the *I Ching*. *European Journal of Parapsychology, 14*, 100–124.

Storm, L., & Thalbourne, M. A. (2000). A paradigm shift away from the ESP-PK dichotomy: The theory of psychopraxia. *Journal of Parapsychology, 64*, 279–300.

* Storm, L., & Thalbourne, M. A. (2001). Studies of the *I Ching*: I. A replication. *Journal of Parapsychology, 65*, 105–124.

* Storm, L., & Thalbourne, M. A. (2005a). The effect of a change in pro attitude on paranormal performance: A pilot study using naïve and sophisticated skeptics. *Journal of Scientific Exploration, 19*, 11–29.

* Storm, L., & Thalbourne, M. A. (2005b). The effects of intuition and attitudes towards gambling on ESP performance during a gambling task. *European Journal of Parapsychology, 20*, 22–49.

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*, 491–494. doi:10.1037/a0019840.

Tart, C. T. (1983). Information acquisition rates in forced-choice ESP experiments: Precognition does not work as well as present-time ESP. *Journal of the American Society for Psychical Research, 77*, 293–310.

* Thalbourne, M. (1996). An attempt to predict precognition scores using transliminality-relevant variables. *Journal of the Society of Psychical Research, 62*, 129–140.

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

* Thorisson, K. R., Skulason, F., & Haraldsson, E. (1991). Effects of belief in ESP and distorted feedback on a computerized clairvoyance task. *Journal of Parapsychology, 55*, 45–58.

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., Martinelli, M., Zaccaria, E., & Massacessi, S. (2009). Implicit intution: How heart rate can contribute to prediction of future events. *Journal of the Society for Psychical Research, 73*, 1–16.

* Tressoldi, P. E., Martinelli, M., Scartezzini, L., & Massacessi, S. (2010). Further evidence of the possibility of exploiting anticipatory physiological signals to assist implicit intuition of random events. *Journal of Scientific Exploration, 24*, 411–424.

* Vaughan, A., & Houck, J. (1993). A "success" test of precognition and attitude toward the future. *Journal of the Society for Psychical Research, 59*, 259–267.

* Vaughan, A., & Houck, J. (2000). Intuition-training software: A second pilot study. *Journal of the Society for Psychical Research, 63*, 177–184.

* Walsh, K., & Moddel, G. (2007). Effect of belief on psi performance on a card guessing task. *Journal of Scientific Exploration, 21*, 501–510.

* Watt, C. (1996). What makes a good PSI target? Three studies of forced-choice ESP varying target emotionality and complexity. *Journal of Parapsychology, 60*, 25–41.

* Watt, C., & Morris, R. (1995). The relationships among performance on a prototype indicator of perceptual defence/vigilance, personality and extrasensory perception. *Personality and Individual Differences, 19*, 635–648.

* Westerlund, J., & Dalkvist, J. (2004). A test of predictions from five studies on telepathic group communication of emotions. *Proceedings of Presented Papers: The Parapsychological Convention 47th Annual Convention*, 268–277.

* Wilson, S., & Hamlin, I. (2007). Implicit learning in a card prediction task. *European Journal of Parapsychology*, 22, 3–29.

* Wiseman, R., & Greening, E. (2002). The mind machine: A mass participation experiment into the possible existence of extrasensory perception. *British Journal of Psychology, 93*, 487–499.

White, R. A. (1977). The influence of experimenter motivation, attitudes, and methods of handling subjects on psi test results. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 273–301). New York: Van Nostrand Reinhold.

* Zilberman, M. (1995). On the training of precognitive ability. *Journal of the Society for Psychical Research, 60*, 289–292.

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

** Dipartimento di Psicologia Generale*
Università di Padova
Padua, Italy

Acknowledgments

We thank Jerry Solvfin, Serena Roney-Dougal, Chris Roe, Dean Radin and Simon Sherwood for their collaboration in data retrieval. We thank Dean Radin, Dick Bierman, and Eva Lobach for their helpful comments.

Correspondence concerning this article should be addressed to Lance Storm, School of Psychology, University of Adelaide, Adelaide, South Australia 5005, Australia. Electronic mail may be sent to lance.storm@adelaide.edu.au.

Abstracts in Other Languages

German

EINE METAANALYSE VON ASW-STUDIEN 1987–2010:
ZUR BEURTEILUNG DES ERFOLGS DER VERSUCHSPLANUNG
MIT BEGRENZTER WAHL IN DER PARAPSYCHOLOGIE

ZUSAMMENFASSUNG: Wir berichten über die Ergebnisse einer Metaanalyse von ASW-Studien mit begrenzter Wahl, bei denen Kartensymbole, Zahlen, Buchstaben und so weiter als Zielobjekte verwendet wurden. Für den Zeitraum 1987 bis 2010 ergab ein homogener Datensatz bei 72 Studien mit begrenzter Wahl eine schwache, aber signifikante mittlere Effektstärke (ES) von 0.01 (Stouffer Z = 4.86, p = 5.90 × 10–7). Keine Hinweise fanden sich, dass diese Ergebnisse mit geringerer Studienqualität oder selektiver Publikation zusammenhingen. Wir stellen fest, dass die Effekte nicht unter den Untersuchern variierten, aber es zeigten sich Hinweise darauf, dass die Anzahl der Wahlmöglichkeiten pro Versuchsdurchgang eine inverse Beziehung mit dem p-Wert aufwies. Auch fanden wir Hinweise auf einen linearen Anstieg der ES-Werte, der darauf hindeutet, dass die Effektstärken von 1987 bis 2010 zugenommen haben. Es werden Vorschläge gemacht, wie sich die Größen von Effektstärken weiter steigern lassen.

Spanish

META-ANÁLISIS DE ESTUDIOS PES, 1987-2010: EVALUACIÓN
DEL DISEÑO DE ELECCIóN FORZADA EN PARAPSICOLOGíA

RESUMEN: Presentamos los resultados de un meta-análisis de los estudios de elección forzada PES que utilizaron como objetivos símbolos en cartas, números, letras, etc. Durante el período de 1987 a 2010, un conjunto de datos homogéneos de 72 estudios de elección forzada produjeron un tamaño promedio de efecto (ES) débil pero significativo de 0.01 (Stouffer Z = 4.86, p = 5.90 × 10–7). No hubo evidencia de que estos resultados se debieran a mala calidad de diseño o información selectiva. Los estudios de clarividencia no produjeron un ES significativamente mayor que los estudios de precognición, y el tipo de objetivo no produjo ninguna diferencia en el ES. Tomamos nota de que los efectos no variaron entre los investigadores, pero encontramos indicios de que el número de opciones por prueba tuvo una relación inversamente proporcional al valor de p. También encontramos evidencia de una pendiente lineal en los valores de ES que indican que los tamaños del efecto se incrementaron durante el período de 1987 a 2010. Damos algunas sugerencias que pueden facilitar incrementos en el tamaño del efecto.

French

META-ANALYSE DES ETUDES DE PES, 1987–2010: EVALUER LE SUCCES DU PARADIGME A CHOIX FORCE EN PARAPSYCHOLOGIE

RESUME: Nous présentons les résultats d'une méta-analyse des recherches à choix forcé sur la PES, qui emploient des cibles telles que des symboles de cartes, des chiffres, des lettres, etc. Pour la période de 1987 à 2010, une base de données homogènes de 72 études à choix forcé donne une taille d'effet faible mais significative de 0.01 (Stouffer Z = 4.86, p = 5.90 × 10–7). Il n'y a pas de preuve que ces résultats soient dus à une qualité plus faible des expérimentations ou à une publication sélective des résultats. Les études sur la clairvoyance n'ont pas produit une taille d'effet significativement plus forte que les études sur la précognition, et le type de cibles ne fait pas de différence par rapport à la taille d'effet. Nous remarquons que les effets ne varient pas suivant les chercheurs, mais nous avons trouvé des éléments tendant à montrer que le nombre de choix par essai est inversement relié à la valeur p. Nous avons également trouvé des preuves d'une inclinaison linéaire des valeurs de taille d'effet indiquant que les tailles d'effets ont augmenté sur la période allant de 1987 à 2010. Des suggestions sont faites pour tenter d'améliorer encore davantage les tailles d'effet.

APPENDIX

List of Studies in the Meta-analysis and Their Results

Study	Trials	Hits	Z	ES ($z/\sqrt{n}$)	Task
1. Auriol, Garcia, Puech, et al. [7 others] (2004) (Two Pictures)	27081	13524	–0.19	–0.001	CL
2. Auriol, Garcia, Puech, et al. [7 others] (2004) (Three Words)	102634	34223	0.07	0.000	CL
3. Auriol, Garcia, Puech, et al. [7 others] (2004) (Five Words)	120347	24226	1.12	0.003	CL
4. Crandall (1987) Study 1	4750	891	–2.12	–0.030	CL
5. Crandall (1987) Study 2	5900	1187	0.21	0.003	CL
6. Crandall (1987) Study 3	3000	640	1.80	0.033	CL
7. Curtis & Wilson (1997)	3520	1051	6.63	0.112	CL
8. Curtis & Wilson (1997)	4440	872	–7.90	–0.119	CL
9. Dalkvist & Westerlund (1998)	9960	5030	0.99	0.010	TE
10. De Graaf & Houtkooper (2004)	4800	970	0.34	0.005	CL
11. Don, McDonough, & Warren (1998)	1760	442	0.08	0.002	PR
12. Ertel (2005) (Lab Expt 1)	9360	2628	19.52	0.202	CL
13. Ertel (2005) (Lab Expt 2)	15244	3400	7.10	0.058	CL
14. Haraldsson (1993) (Pictures)	20291	5001	–1.16	–0.008	CL
15. Haraldsson (1993) (Pictures)	2160	468	1.91	0.041	CL
16. Haraldsson (1993) (Letters)	14611	3695	0.80	0.007	PR
17. Haraldsson, Houtkooper, Schneider, & Bäckström (2002)	2680	719	2.16	0.042	CL
18. Haraldsson, Houtkooper, Schneider, & Bäckström (2002)	2680	632	–1.67	–0.032	PR
19. Houtkooper (2003)	4560	1130	–0.32	–0.005	CL
20. Houtkooper, Schienle, Stark, & Vaitl (1999)	4000	991	–0.31	–0.005	CL
21. Krishna & Rao (1991)	22500	5744	1.82	0.012	CL
22. Lobach & Bierman (2004)	214	63	1.42	0.097	TE
23. Kugel (1999)	14975	7265	–3.63	–0.03	PR
24. Luke, Delanoy, & Sherwood (2008)	1000	285	2.52	0.080	PR
25. Luke, Roe, & Davison (2008) Study 1 (General Public)	250	85	3.21	0.203	PR
26. Luke, Roe, & Davison (2008) Study 2 (Students)	320	93	1.61	0.090	PR
27. McDonough, Don, & Warren (2002)	755	204	1.24	0.045	PR
28. Milton (1994)	10300	5033	–2.30	–0.023	CL
29. Mossbridge, Grabowecky, & Suzuki (2009)	500	123	–0.15	–0.007	PR
30. Moulton & Kosslyn (2008)	3687	1842	–0.03	0.000	CL
31. Murray, Howard, Wilde, Fox, & Simmonds-Moore (2007)	200	48	–0.24	–0.017	TE
32. Palmer (1992)	1600	413	0.72	0.018	CL
33. Palmer (2009) Study 1 (Believers)	3200	695	2.41	0.043	PR
34. Palmer (2009) Study 2 (Nonbelievers)	6300	1250	–0.30	–0.004	PR
35. Parra & Argibay (2007) (Psychics)	264	145	1.54	0.095	CL
36. Parra & Argibay (2007) (Non-Psychics)	180	83	–0.97	–0.072	CL
37. Pitman & Owens (2004) Expt. 1	775	164	0.76	0.027	CL
38. Pitman & Owens (2004) Expt. 2	1775	384	1.69	0.040	CL

Continued

Study—*Cont'd*	Trials	Hits	*Z*	*ES* ($z/\sqrt{n}$)	Task
39. Rammohan (1990) Study 1	1356	262	–0.59	–0.016	CL
40. Rammohan (1990) Study 2	1919	382	–0.07	0.002	CL
41. Rao, Kanthamani & Palmer (1990)	520	251	–0.75	–0.033	CL
42. Roe, Henderson, Matthews (2008) Study 1	3600	737	0.69	0.012	CL
43. Roe, Henderson, Matthews (2008) Study 2	1960	422	1.67	0.038	CL
44. Roe, Davey & Stevens (2003)	240	39	–0.09	–0.006	CL
45. Roe, Davey & Stevens (2004)	240	41	0.09	0.007	CL
46. Roe, Davey & Stevens (2005)	480	68	–1.41	–0.064	PR
47. Roe, Davey & Stevens (2006)	240	38	–0.26	–0.017	CL
48. Sartori, Massacessi, Martinelli, & Tressoldi (2004)	240	73	1.86	0.120	CL
49. Schmidt, Erath, Ivanova, & Walach (2009) Study 1 (Phone)	397	106	0.72	0.036	TE
50. Schmidt, Erath, Ivanova, & Walach (2009) Study 2 (Email)	100	30	1.04	0.104	TE
51. Sheldrake & Avraamides (2009)	419	175	3.61	0.176	TE
52. Sheldrake, Avraamides, & Novák (2009)	886	336	2.86	0.096	TE
53. Sheldrake, R. & Beharee, A. (2009)	6000	1599	2.94	0.038	TE
54. Sheldrake & Smart (2003a)	571	231	8.48	0.355	TE
55. Sheldrake & Smart (2003b)	271	122	7.54	0.458	TE
56. Sheldrake & Smart (2005)	552	235	9.49	0.404	TE
57. Sheldrake & Lambert (2007)	1980	581	4.44	0.100	TE
58. Simmonds-Moore & Moore (2009)	5250	1095	0.15	0.002	CL
59. Smith, Wiseman, Machin, Harris, & Joiner (1997)	2800	1395	–0.17	–0.003	PR
60. Steinkamp (2002) Study 2	10000	3402	1.45	0.008	PR
61. Steinkamp (2002) Number Questions	2500	400	–0.87	–0.017	PR
62. Steinkamp (2002) Verbal Questions	2500	422	0.26	0.005	PR
63. Storm (2002)	78	20	0.06	0.007	PR
64. Storm (2006)	353	95	0.90	0.048	PR
65. Storm (2008)	272	68	0.04	0.002	PR
66. Storm & Thalbourne (1998-99)	172	54	1.95	0.149	PR
67. Storm & Thalbourne (2001)	186	58	1.97	0.144	PR
68. Storm & Thalbourne (2005a)	6750	1377	0.81	0.010	CL
69. Storm & Thalbourne (2005b)	1000	208	0.59	0.019	CL
70. Thalbourne (1996)	990	232	–1.10	–0.035	PR
71. Thalbourne & Storm (2002-2005)	363	83	0.75	0.039	PR
72. Thorisson, Skulason, & Haraldsson (1991)	9160	2267	–0.54	–0.006	CL
73. Tressoldi, Martinelli, Zaccaria, & Massaccesi (2009)	200	111	1.48	0.105	PR
74. Tressoldi, Martinelli, Zaccaria, & Massaccesi (2009)	240	111	–1.10	–0.071	PR
75. Tressoldi, Martinelli, Zaccaria, & Massaccesi (2009)	180	103	1.86	0.139	PR
76. Tressoldi, Martinelli, Scartezzi, & Massaccesi (2010)	440	246	2.43	0.116	PR
77. Tressoldi, Martinelli, Scartezzi, & Massaccesi (2010)	240	102	–2.26	–0.146	PR
78. Vaughan & Houck (1993)	149400	6000	3.41	0.009	PR
79. Vaughan & Houck (2000)	21600	895	2.25	0.015	PR

Study—*Cont'd*	Trials	Hits	*Z*	*ES* ($z/\sqrt{n}$)	Task
80. Walsh & Moddel (2007)	600	151	3.11	0.127	CL
81. Walsh & Moddel (2007)	600	128	0.77	0.031	CL
82. Watt (1996) Exp 1	1152	577	0.03	0.001	CL
83. Watt (1996) Exp 2	336	184	1.69	0.092	CL
84. Watt (1996) Exp 3	1800	892	–0.35	–0.008	CL
85. Watt & Morris (1995) Exp 1	1152	576	–0.03	–0.001	TE
86. Watt & Morris (1995) Exp 2	1800	892	–0.35	–0.008	TE
87. Westerlund & Dalkvist (2004)	18150	9069	–0.08	–0.001	CL
88. Wilson & Hamlin (2007)	1760	434	-0.30	-0.007	PR
89. Wiseman & Greening (2002)	55479	27784	0.37	0.002	CL
90. Wiseman & Greening (2002)	55479	27590	–1.27	–0.005	PR
91. Zilberman (1995)	5100	2646	2.67	0.037	PR

Note: CL = Clairvoyance; TE = Telepathy, PR = Precognition