Journal of Scientific Exploration
Anomalistics and Frontier Science

SPECIAL SUBSECTION

Assessing 36 Years of the Forced Choice Design in Extra Sensory Perception Research: A Meta-Analysis, 1987 to 2022

Lance Storm
lance.storm@adelaide.edu.au

Patrizio Tressoldi
patrizio.tressoldi@unipd.it

SUBMITTED April 12, 2023
ACCEPTED June 29, 2023
PUBLISHED October 31, 2023

https://doi.org/10.31275/20232967

HIGHLIGHTS

Statistical analysis shows that forced-choice experiments successfully demonstrate potential extrasensory perception, yet only to a small degree.

ABSTRACT

This meta-analysis is an update of Storm, Tressoldi, and Di Risio (2012); a meta-analysis on forced-choice ESP studies (1987 to 2010), which use targets such as card symbols, pictures, and letters. We formed two datasets: New Studies #1 (studies that included actual hit rates) and New Studies #2 (Reaction Time [‘RT’] Studies; which are studies that measured only reaction time, not hits, as indicators of psi responses). New Studies #1: For the period 2011 to 2022, a homogeneous dataset of 38 studies yielded a mean effect size (*ES*) of 0.02 (Stouffer $Z = 5.55$, $p = 1.43 \times 10^{-8}$). New Studies #2 (‘RT’ Studies): For the same period, a homogeneous dataset of 23 studies yielded a weaker mean *ES* of 0.01 (Stouffer $Z = 5.50$, $p = 1.90 \times 10^{-8}$). The two databases were combined. In this dataset, telepathy, clairvoyance, and precognition studies were not significantly different from each other. Nor were target types. We updated the forced-choice database by combining our revised original database with the new studies to form a homogeneous database ($N = 141$): mean effect size (*ES*) of 0.02 (Stouffer $Z = 8.52$, $p < 10^{-16}$). Effects did not vary between investigators or laboratories, and we found a near-significant incline in *ES* values over a 36-year period (i.e., no evidence of a decline). These results confirm that the forced-choice design adequately tests extra-sensory perception (ESP). We compare the overall results with those obtained in other domains, focusing on ‘selected’ participants (meditators, psychics, psi-test experienced) and ‘unselected’ (i.e., untrained, naïve) participants.

KEYWORDS

Consciousness, ESP, extra-sensory perception, forced-choice, meta-analysis, psi.

INTRODUCTION

Biologist J. B. Rhine was one of the first empirical investigators 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). The problem with the 52-card deck is that extra-sensory perception (ESP) effects can be determined statistically in a number of ways (i.e., by suit, by face value, by color), giving skeptics the opportunity to accuse experimenters of “hypothesis saving” (Irwin & Watt, 2007, p. 52).

Rhine’s various kinds of card-guessing experiments

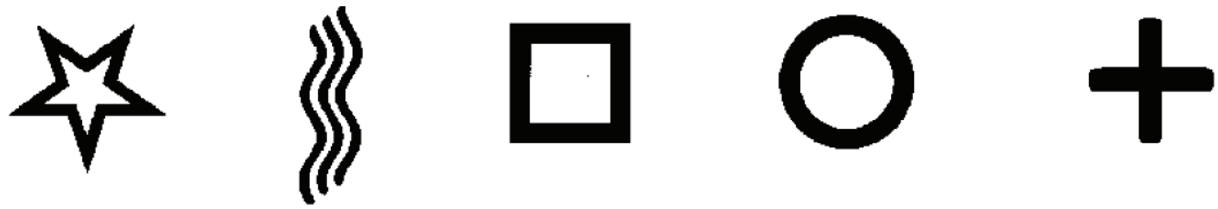


Figure 1. The Zener card symbols: star, wavy lines, square, circle, and cross. The deck of 25 cards has five of each symbol.

came to be known generally as the 'forced-choice' design, which soon produced impressive findings (Rhine, 1934). In the typical forced-choice ESP design, the target to be 'guessed' (i.e., identified without sensory clues) is "one of a limited range of possibilities which are known to [the participant] in advance" (Thalbourne, 2003, p. 44). The advantage of card-guessing is that outcomes are unambiguous, so independent judging is not required, but the long runs of multiple trials were tedious and boring for participants, which saw declines in hit rates. Rhine's forced-choice design also came under criticism for its methodological flaws, such as sensory leakage due, for example, to subtle tell-tale folds on cards or poor-quality printing that leaked information through to the reverse sides of the cards. Although these problems were rectified by Rhine, card testing ultimately fell out of vogue over the decades with the introduction of computers and the adoption of more interesting, ecologically valid targets. These target types still include card symbols, but also used nowadays are pictures, alphabet letters, words, shapes, and so forth. Studies in recent decades now include fractal images, SMS messages, and Chinese characters.

As the number of forced-choice studies accumulated from research conducted around the world and the technique of meta-analysis came to the fore, it became possible to evaluate large datasets of studies. A number of forced-choice meta-analytic studies are extant in the literature, and these have generally confirmed a consistent so-called 'psi' (i.e., paranormal) effect independent of experimenters/laboratories worldwide. The forced-choice psi effect, however, has varied in strength and tends to be weak (see Honorton & Ferrari, 1989; Steinkamp et al., 1998; Storm, Tressoldi, & Risio, 2012; Tart, 1983).

A major aim of the present meta-analysis is to evaluate the performance of forced-choice studies conducted since our last meta-analysis, and thus, we planned to cover the period 2011 to 2022. This evaluation may provide statistical evidence that there is an anomalous sensory modality that can manifest as either 'telepathy', 'clairvoyance', or 'precognition'. Telepathy refers to the "paranormal acquisition of information concerning the thoughts, feelings or activity of another conscious being" (Thalbourne, 2003, p. 125), but in the case of the forced-choice design, the agent (sender) is sensorially shielded from the percipient (receiver), and the agent's thoughts are confined to ESP targets. Clairvoyance is defined as "paranormal acquisition of information concerning an object [or ESP target suitable for a forced-choice design] or contemporary physical event" (Thalbourne, 2003, p. 18). Precognition is "a form of extra-sensory perception in which the target is some future event that cannot be deduced from normally known data in the present" (Thalbourne, 2003, p. 90). The 'future event' may include an ESP target suitable for a forced-choice study. The following section is a review of ESP forced-choice studies with a focus on modalities.

Forced-Choice Telepathy, Clairvoyance, and Precognition

Tart (1983) found 85 forced-choice 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) outperformed precognition. Steinkamp et al. (1998) argued that Tart's selection criteria introduced a bias which may have put the precognition studies at a disadvantage.

Honorton and Ferrari (1989) only looked at precognition forced-choice experiments from 1935 to 1987. 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). Out of 309 studies, 92 (30%) showed significant hit-rates at the 5% level. The authors noted that experienced participants performed better than naïve participants. They also found that precognition, although weak, produced "very robust" and highly significant results across a time span of more than 50 years, with the quality of studies remaining stable, or even improving, in that time. Honorton and Ferrari's meta-analysis revealed that the largest effect sizes were found in experiments using (a) experienced (i.e., selected) participants and (b) trial-by-trial feedback. These are important factors for future psi researchers to consider in their designs, and we will assess the influence of both in our post hoc analyses.

The meta-analysis by Steinkamp et al. (1998), for the period 1935 to 1997, was a comparison of 'matched' (procedurally similar) clairvoyance and precognition studies in order to test for a phenomenological difference between the two modalities. They found 22 comparable study-pairs, but tests showed no difference. 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).

For the very comprehensive expansive period 1880 to 1989, Steinkamp's (2005) review of forced-choice studies showed that there were few variables that correlate with psi success, partly because the variability in study designs made it difficult to discern clear patterns due to conflicting outcomes. Nevertheless, she found "low neuroticism, extraversion, and good social adjustment may be positively related to forced-choice ESP scoring" (p. 158). She also found some support for the notion that 'selected' (i.e., meditators, psychics, previously psi tested) participants performed better than unselected participants in precognition tests, whereas those who did not believe in psi scored lower than those who did (see also the meta-analysis on paranormal belief by Storm & Tressoldi, 2017).

Storm et al. (2012) reported the following in their meta-analysis covering the period 1987 to 2010: For a homogeneous dataset of 72 studies, there was a very weak but significant mean effect size (*ES*), calculated from the formula $z/\sqrt{n}$ (where z is a standardized score with a mean of 0.00 and *SD* of 1.00; and n is the number of studies). *ES* did not correlate with study quality, and there was no evidence of selective reporting. Clairvoyance and precognition studies were not significantly different, as was found previously by Steinkamp et al. (1998). Also, *ES* did not vary between investigators. Storm et al. also found that target type did not make a difference to effect size, but they did find suggestive evidence that the number of choices per trial was inversely related to the p-value. Evidence of a linear incline in *ES* values was also found over the period 1987 to 2010.

The most recent meta-analytic treatment of forced-choice studies was by Zdrenka and Wilson (2017). They meta-analyzed 55 studies dating from 1945 to 2016 but evaluated precognition studies only. They found that psi performance correlated significantly with six individual difference variables: "luck belief (the belief that luck is primarily controllable), perceptual defensiveness, openness to experience, belief in psi, extraversion, and time belief as dynamic" (p. 9). They did not present an overall effect size value for the 55 studies. However, of the 23 measures in their Table 2, effect-size values (r) ranged between -0.006 and 0.14, with 17 (74%) of them positive.

In conclusion, the general findings are that (i) forced-choice effects are very weak (i.e., small) but consistent and suggestive of psi; (ii) extroversion is a significant correlate of psi (found twice); (iii) selected participants perform better than unselected participants (found twice), and (iv) there is some evidence that no psi modality (telepathy, clairvoyance, or precognition) is superior to any other. However, with the noted exclusion of the study by Storm et al. (2012), little attention has been given in these meta-analyses to the influence on psi of target types, the experimenter effect, the number of choices in the target set (i.e., k-choices), and the decline effect. These factors are considered next. However, our main aim in the present study is to see if we can replicate the results of our original six hypotheses (see Storm et al., 2012, p. 248). Tests on other variables are relegated to the section, *Post Hoc Analyses*.

Target Types

As mentioned above, target types are now quite varied. Most trials are conducted on-screen using computer monitors rather than with decks of cards or other physical objects. Apart from conventional images such as photographs of faces, Zener cards, pictures, drawings, letters, and numbers, ecologically valid targets are also being used, such as racing-horse images as targets on simulated racetracks (e.g., Roe, Davey & Stevens, 2003), fractal images (Luke, Delanoy, & Sherwood, 2008), and Chinese characters (Vernon, Hitchman, & Roe, 2021).

Previously, we argued that there is no consensus on whether some target types are "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)" (Storm et al., 2012, p. 246). In that paper, it was shown that target type did not make a difference to effect size, although there were significantly stronger mean effects for word/letter targets *and* for objects in the telepathy condition compared to the other two modalities (clairvoyance and precognition). In the present study, we once again test whether target type has an influence on effect size.

Experimenter Effect

Experimenter-psi (or E-psi) becomes a problem when parapsychologists want to know that participants are exclusively responsible for psi; not the experimenter. However, E-psi has not been found across a broad range of investigators in various meta-analyses (Bem & Honorton, 1994; Honorton et al., 1990; Storm, Tressoldi, & Di Risio, 2010). Indeed, pertinent to the present study on forced-choice studies only, Storm et al. (2012) found "no single group that produced effects significantly different from any other group" (p. 259; see also Honorton & Ferrari, 1989). The same effect will be tested in the present study.

High k-Choice Designs

Parapsychologists are also interested in whether or not declines in effect size might be due 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). The values of *k* have been quite variable over the decades in the forced-choice domain, ranging from 2 to as many as 26. Empirical support for Timm's claim is minimal for forced-choice, but worthy of note. Specifically, Storm et al. (2012) found a significant positive correlation between *k* values and *z* scores. That is, *z* scores tended to increase as *k* increased. If this finding indicates a valid effect, researchers may wish to consider using high *k*-choice designs. In the present study, we will again test whether the number of target choices per trial is related to *z*.

Decline Effects

It is thought that significant declines in effects over the long-term (i.e., decades) indicate some kind of artifact due to, say, improvements in study quality over the years or deliberate changes in experimental design (from simple and fun to complex and tedious). These changes are made because theoretically oriented experimenters want to understand the *psi process* rather than merely *prove* the presence of a so-called communication anomaly. While lengthy (chronological) declines have been noted on occasion in the parapsychological literature, generally, these declines are spurious, as shown in a large collection of meta-analyses reviewed by Storm (in press). For forced-choice studies in particular, researchers have not found declines (Bierman, 2001; Honorton & Ferrari, 1989). And for the period 1987 to 2010, Storm et al. (2012) noted that *ES* values in the forced-choice domain actually increased significantly in their database of 72 studies.

General Aims of the Present Study

As of 2022, ten years have passed since forced-choice studies were last meta-analysed (see Storm et al., 2012). The general aim of the present study was to conduct a meta-analytic review of the new forced-choice studies (dating from 2011 to 2022) to determine if a comprehensive, up-to-date database is still significant, as was the case with the Storm et al. meta-analysis. Further to that, we aim to see if we can replicate the results of our original six hypotheses (see Storm et al., 2012, p. 248), in addition to testing the difference between the 'old' (1987-2010) and 'new' (2011-2022) datasets. Tests on other variables (feedback, participant type) are relegated to the section, *Post Hoc Analyses*. The following hypotheses were thus proposed:

1. Forced-choice studies produce statistical evidence of a communications anomaly known as ESP;
2. The mean *ES* values for telepathy, clairvoyance, and precognition are different;
3. *ES* values vary in strength according to target types;
4. *ES* values vary between experimenters/laboratories;
5. Number of choices (*k*) per trial is positively related to *z*;
6. *ES* values increased over the period of analysis (i.e., 2011–2022);
7. The original database (1987 to 2010), and the new database (2011 to 2022), are not different and can be combined to form a larger dataset.

Should Hypothesis 7 be supported, we will conduct additional (post hoc) analyses on a single combined database to further test the validity of the forced-choice paradigm.

METHOD

Meta-Analysis Reporting Guideline

We followed the APA meta-analysis reporting standards guideline (MARS, Appelbaum et al., 2018).

Study Retrieval

- The period of interest was January 2011 to December 2022. Full-text studies were retrieved from the following sources:
- The meta-analyses by Bem, Tressoldi, Rabeyron, and Duggan (2016); Google Scholar, PubMed, Scopus databases by using the keywords 'forced-choice AND extra-sensory perception';
- Specialized scientific (peer-reviewed) journals, including the *Australian Journal of Parapsychology*, *Consciousness, and Cognition*, *Explore: The Journal of Science and Healing*, *Heliyon*, *Journal of Anomalous Experience and Cognition*, *Journal of Consciousness Studies*, *Journal of Parapsychology*, *Journal of Scientific Exploration*, *Journal of Personality and Social Psychology*, *Journal of the Soci-*

ety for Psychical Research, and *NeuroQuantology*;
- Proceedings of the Annual Convention of the Parapsychological Association and the Society for Psychical Research.

The meta-analysis by Storm et al. (2012) provided the original dataset of forced-choice publications dating from January 1987 to December 2010 (see the Appendix in Storm et al., 2012, p. 271). That dataset contains 91 studies, which were reported in 65 papers conducted by 96 investigators.

Inclusion Criteria

To be included in the present update, studies had to:

1. Be published between January 2011 and December 2022;
2. Have adopted a forced-choice procedure for the ESP identification of targets (therefore excluding studies that expressly tested free-response and psychokinesis);[1]
3. Have used human participants only (not animals);
4. Have been carried out with groups of participants and not single cases;
5. Have incorporated randomization procedures for selection of targets which could not be manipulated by the experimenter or participant;
6. Have been peer-reviewed;[2]
7. Have 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 $z/\sqrt{n}$ (where n = number of trials). Studies with only reaction times were also included separately as these did not have hits data but did have t scores by which *ES* values could be calculated.

Procedure

We present a PRISMA flowchart in Appendix A (Page et al., 2021). This figure details the counts of papers across four stages: identification, screening, eligibility, and inclusion. For each study, we checked the following factors: (a) the criteria adopted for selecting participants; (b) authors' names; (c) year of publication; (d) whether participants were selected or unselected; (e) type of ESP task (telepathy, clairvoyance, or precognition); (f) number of participants; (g) number of trials; (h) number of alternatives k per trial in the tasks; and (i) total number of hits (we preferred the direct hits measure as it provides a more "conservative" result—see Honorton, 1985, p. 54).

With these data, we derived the 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 *ES* (where $ES = z/\sqrt{n}$). These values were double-checked, and in some cases (e.g., Hitchman, Sherwood, & Roe, 2015), we found discrepancies between published z scores and our calculations using the Vassarstats Exact Binomial calculator (http://www.vassarstats.net/binomialX.html), which requires only hits, trial counts, and MCE.

Variables Considered

Each study included in the database was classified with the following variables: Authors, Year of Publication, Selected or Unselected Participants, Task Type (Telepathy, Clairvoyance, or Precognition); Target Type (pictures, letters, symbols, objects, etc.); Sample Size; Trials; Hits; Hits Percentage; Number of Choices (k); Type of Feedback (none = 0; trial-by-trial = 1; end of trials = 2), Peer Reviewed ('Yes' or 'No').

Meta-Analysis Design

We applied a random-effects model by using the metafor package v. 3.8 (Viechtbauer, 2010), adopting the restricted maximum likelihood (REML) to estimate the heterogeneity variance (Langan et al., 2019) and the Hartung method to control effect-size non-normality (Rubio-Aparicio et al., 2018) and corresponding confidence intervals estimation. We identified effect size outliers by using SPSS Stem-and-Leaf and Box-and-Whiskers Plots.

The whole database and the code used for all the statistical analyses are available for open access on:(https://figshare.com/articles/dataset/Forced-Choice_1987_-_2022_meta-analysis/22574218/5),for independent reproducibility, controls, and analyses.

RESULTS

After all inclusion and exclusion criteria had been met, 27 articles *with hits data* were included for analysis. There were 52 individual experiments/treatments/conditions ('studies' as distinct from 'articles') that had sufficient data for the calculation of specific z scores and corresponding effect sizes (*ES*). (For a list of these 52, see Appendix B.)

There were 17 articles *with no hit data* (these studies measured Reaction Time; 'RT'). In these articles were 25 individual experiments with t scores allowing for the calculation of corresponding *ES*s. (For a list of these 25, see Appendix C.) The References section lists all articles meta-analyzed (marked by asterisks). The total number of articles is 43 (not to count Bem, 2011, twice), that included a total of 77 experiments/treatments/conditions ('studies') with sufficient information to calculate *ES*s.

Two articles in our original database (Storm et al., 2012) were expanded to include various conditions/ treatments (Luke, Roe, & Davison, 2008; Pitman & Owen, 2004); Thalbourne and Storm (2002-2005) remains in the 2012 meta-analysis even though it was later published as Thalbourne and Storm (2014) and is therefore not counted in the update.

In the original database (Storm et al., 2012) were two articles now excluded due to their datasets being incomplete (Vernon, Sandford, & Moyo, 2019; Zilberman, 1995). A further two articles by Tressoldi et al. (2009, 2010) were removed because it was decided that they did not use typical behavioral forced-choice procedures (i.e..., they used sound targets).

Also, regarding the original database (Storm et al., 2012), six articles previously overlooked from the initial period were added in. We will correct the original statistics and re-do relevant analyses—these articles are: Batthyány, Kranz, and Erber (2009); Ertel (2010); Hadlaczky (2005); Savva, Child, and Smith (2004); Savva, Roe, and Smith (2005); Watt and Nagtegaal (2000).

Descriptive Statistics

We compiled two databases: New Studies #1, the set of articles *with hits data*; and (2) New Studies #2, the set of articles *with no hits data* ('RT' studies). In New Studies #1, there are a total of 162,989 trials and 71,678 hits. In New Studies #2, there are a total of 207,019 trials (no hits).

H1: Z statistics and effect sizes (ES). It was hypothesized that the new databases would yield statistical evidence of a communications anomaly known as ESP.

New Studies #1. For a heterogeneous dataset of 52 studies, mean *ES* = -0.004 (*SD* = 0.20), and mean *z* = -0.84 (*SD* = 8.03). These negative values are largely attributable to extreme scoring in two studies: Escolà-Gascón (2022), and Escolà-Gascón et al. (2022). It is noted that the skew of the *ES* distribution was not normal. Outliers were identified from SPSS Stem-and-Leaf and Box-and-Whiskers Plots as significantly deviant ("extreme") cases. Seven of nine extremely low-scoring values were found in the two Escolà-Gascón studies just mentioned (see Appendix B: #45, #47, #48, #49, #50, #51, & #52), and a further seven outliers removed (see Appendix B: #9, #12, #13, #14, #16, #24, & #32). Normality (homogeneity) was achieved after the removal of these 14 studies. A dataset of 38 studies is now homogeneous with mean *z* = 0.90 (*SD* = 1.74); mean *ES* = 0.02 (*SD* = 0.04). Ninety-five percent confidence intervals (CIs) are as follows: *z* scores, [0.33, 1.47]; *ES* values, [0.006, 0.03]. Note that neither of these 95% CIs includes values of MCE (i.e., zero). Stouffer *Z* = 5.55, $p = 1.43 \times 10^{-8}$ (one-tailed). A single-sample *t*-test revealed that *ES* values significantly deviated from mean chance expectation (MCE), where the test statistic is zero (i.e., MCE = 0.00), $t(37) = 2.94$, $p = .006$ (two-tailed). Eleven studies (29% of 38 studies) were independently significant at $\alpha < .05$ level.

We considered it necessary to assess homogeneity in a different context. Higgins' I^2 (Higgins, Thompson, Deeks, & Altman, 2003) indicates the proportion of effect-size variance explained by heterogeneity across effect sizes. I^2 lies between 0% and 100%. A value of 0% indicates no observed heterogeneity, and larger values show increasing heterogeneity. We found significant heterogeneity in the dataset, $Q(df = 37) = 436.3$, $p < .001$, and $I^2 = 93.04\%$ (very high heterogeneity). It is important to regard heterogeneity not as a measure of the quality of the studies, but as a measure of between-studies differences. We point out that experimental designs of forced-choice studies vary to a far greater extent than, say, free-response (especially Ganzfeld designs) in terms of type of task, targets used, number of trials, and so on.

New Studies #2 ('RT' Studies). For a heterogeneous dataset of 25 studies, mean *ES* = 0.006 (*SD* = 0.020), and mean *t* = 0.54 (*SD* = 1.21). The skew of the *ES* distribution was normal. However, two outliers were identified: Rabeyon (2014), and Wittmann et al. (2021, expt. 2). (See Appendix C: #13 & #23.) The homogeneous dataset of 23 studies has a mean *t* = 0.74 (*SD* = 1.01); mean *ES* = 0.01 (*SD* = 0.01). Ninety-five percent CIs are as follows: *t* scores, [0.31, 1.18]; *ES* values, [0.004, 0.02]. Note that neither of these 95% CIs includes values of MCE (i.e., zero). For comparative purposes, we calculated a Stouffer *Z* statistic since *z* approximates *t* when samples have 30+ trials, which is the case in this homogeneous dataset. For the 23 studies, Stouffer *Z* = 5.50, $p = 1.90 \times 10^{-8}$ (one-tailed). A single-sample *t*-test revealed that *ES* values significantly deviated from chance, $t(22) = 3.34$, $p = .003$ (two-tailed). Four studies (17% of 23 studies) were independently significant at $\alpha < .05$ level. Again, there was significant heterogeneity, $Q(22) = 104.07$; $p < .001$; $I^2 = 81.02\%$.

New Studies #1 and New Studies #2 ('RT' Studies) combined. The two databases (New Studies #1 & New Studies #2), totaling 61 studies, were not significantly different on *ES* values, $t(50.43) = 1.28$, $p = .207$ (two-tailed). The skew was normal, but two outliers were removed: Luke et al. (2012), and Sheldrake (2015). (See Appendix B: #10, & #33). The homogeneous dataset of 59 studies has a mean *ES* = 0.016 (*SD* = 0.03), and mean *z* = 0.84 (*SD* = 1.47). Stouffer *Z* = 6.42, $p = 6.81 \times 10^{-11}$ (one-tailed). Table 1 lists statistics for the combined database of new studies only (*N* = 59), as well as subgroups of experimental conditions after outliers and other data exclusions. Subsequent analyses in this

Table 1. Combined Database of New Studies (2011 to 2022): Effect Sizes, and 95% Confidence Intervals, *p*-value and I^2

Datasets and Subsets	*n*	*ES (SD)*	*95% CI*	*p*	I^2
Combined New Studies (*N*)	59	0.02 (0.03)	0.01, 0.02	3.5×10^{-6}	92.6
Telepathy	4	0.025 (0.03)	-0.03, 0.08	2.3×10^{-1}	97.5
Clairvoyance	14	0.007 (0.04)	-0.01, 0.03	4.7×10^{-1}	88.2
Precognition	41	0.017 (0.03)	0.01, 0.03	2.3×10^{-6}	90.9
Selected participants	9	0.02 (0.04)	-0.01, 0.05	2.0×10^{-1}	83.3
Unselected participants	50	0.02 (0.03)	0.01, 0.02	4.5×10^{-6}	92.3
With feedback	47	0.02 (0.03)	0.01, 0.03	1.1×10^{-8}	91.1
No feedback	12	0.002 (0.04)	-0.02, 0.03	8.6×10^{-1}	87.2

paper are conducted on this larger database of 59 studies. Telepathy seems an outstanding performer as a modality, but with only six studies in the subset, it is difficult to gauge its pertinence (see *H2* below).

H2: Effect size differences for telepathy, clairvoyance, and precognition. It was hypothesized that the mean *ES* values for telepathy, clairvoyance, and precognition are different. Table 1 lists the ES values for the three modalities. Four studies (7% of 59 studies) tested telepathy; 14 studies (24%) tested clairvoyance; 41 studies (69%) tested precognition. Telepathy produced the strongest effect. A Univariate ANOVA test was conducted, entering the variable Psi Modality, as well as the variable Target Type (see *H3* below). There was a just significant *ES* difference between the three modalities, $F(2, 51) = 3.16$, $p = .051$ (two-tailed). The greatest difference was between telepathy and clairvoyance (0.018), but a Tukey's HSD test showed no significant difference.

H3: Target types. Target types may affect participants' performances. As was done by Storm et al. (2012), data was divided into five types of targets: (1) Pictures/drawings/faces, (2) Symbols/fractals/*I Ching* hexagrams, (3) Numbers, (4) Letters/words/messages, and (5) Objects (i.e., targets that occupy 3-D physical space). The new set of telepathy studies used only 'Pictures'; clairvoyance studies used all types except numbers; and precognition studies did not use numbers or objects. The same ANOVA from above (see *H2*) showed no significant *ES* difference between target types, $F(3, 51) = 0.61$, $p = .612$ (two-tailed). Nevertheless, the preferred targets were 'Symbols' and 'Letters' with *ES* about equal ($ES = 0.04$).

H4. ES differences between experimenters/laboratories. In order to ascertain whether our database was the result of extremely positive *ES* values for a limited pool of laboratories/experimenters, we conducted a Kruskal-Wallis ANOVA on the pooled data after dividing them into laboratory/experimenter groups. We formed 11 mutually exclusive experimenter groups with at least two studies in each. Some of these groupings were used in Storm et al. (2012): 'Argibay', 'Bem', 'Bierman', 'Luke', 'Roe', 'Schlitz', 'Schönwetter', 'Simmonds-Moore', 'Sheldrake', 'Storm', and 'Watt'. Table 2 lists the mean *ES* values for each group. The same combined dataset was used as was tested in *H2* & *H3* (for the full-database tests, see *Post Hoc Analyses*). Fourteen single studies (24%) did not qualify, so these were excluded, but it would be misleading to categorize these as 'Other' since the variability of laboratory/author is too great. Mean *ES* values varied from -0.021 ('Schönwetter') to 0.049 ('Luke'). Using a Kruskal-Wallis ANOVA, a marginally significant difference was found between experimenter groups, $\chi^2(10) = 18.13$, $p = .053$ (two-sided). However, no two groups were significantly extreme

Table 2. Experimenter/Laboratories: Effect Sizes, *SD*, and 95% Confidence Intervals

Grp.	Expt./Lab.	n	*ES* (*SD*)	95% CI
#1	'Bem'	8	0.040 (0.02)	0.02, 0.06
#2	'Schlitz'	3	0.004 (0.01)	-0.02, 0.02
#6	'Argibay'	4	0.042 (0.04)	-0.02, 0.10
#9	'Simmonds-Moore'	2	-0.018 (0.02)	-0.17, 0.14
#10	'Schönwetter'	2	-0.021 (0.03)	-0.33, 0.29
#12	'Bierman'	5	0.018 (0.02)	-0.005, 0.04
#14	'Luke'	3	0.049 (0.03)	-0.04, 0.14
#16	'Roe'	10	0.014 (0.03)	-0.004, 0.03
#17	'Sheldrake'	3	0.040 (0.02)	-.0006, 0.08
#19	'Storm'	2	0.010 (0.02)	-0.20, 0.22
#21	'Watt'	2	0.007 (0.02)	-0.14, 0.16

n = number of studies

in scoring to reach significance. The effects cannot be said to be due to a few outstanding investigators.

H5. The advantage in using a high k-choices design. We proposed that the number of choices *k* per trial is positively related to *z* (or *t*). In the combined dataset ($N = 59$), there are only three values for *k* choices: $k = 2$ ($n = 47$; mean $z = 0.95$); $k = 4$ ($n = 9$; mean $z = 0.23$); and $k = 5$ ($n = 3$; mean z = 0.87). Visual inspection shows no clear trend. A Pearson's *r* test on these values (grouped) was not significant, $r(1) = -0.29$, $p = .409$ (one-tailed). Correlating *z* scores for individual studies with their *k* values, was also not significant, $r(57) = -0.13$, $p = .158$ (one-tailed). Given these analyses, *z* (or *t*) scores tend not to increase with *k*.

H6. Change in ES values over the period of analysis (2011–2022). We propose that *ES* values have *increased* over this period based on a similar finding by Storm et al. (2012). However, we find the correlation between year of study and *ES* is negative and significant for the combined database ($N = 59$), $r_s(57) = -0.34$, $p = .004$ (one-tailed). Our hypothesis failed as this statistic indicates a *decline* in *ES* values over the 12-year period.

H7. The original database (1987 to 2010), and the new database (2011 to 2022), are not different and can be combined to form a larger dataset. The old database in Storm et al. (2012) identified 19 outlier studies. The database was reduced from 91 studies to 72 studies. Storm et al. reported the following: mean *z* score = 0.57 (SD = 1.58); $ES = 0.01$ ($SD = 0.03$); Stouffer $Z = 4.86$, $p = 5.90 \times 10^{-7}$ (p. 253). As explained above, we deleted some old studies that did not fit our criteria, but found new ones, warranting a re-assessment of the forced-choice database of studies dating 1987 to 2010.

There were 68 articles for this period, with 102 experiments/treatments/conditions (i.e., studies) with associated *ES* values. The heterogeneous database has a mean *z* score = 1.34 ($SD = 3.33$); 95%CI [0.69, 1.99]; ES = 0.04 ($SD = 0.09$); 95%CI [0.03, 0.06].

Once again, 19 outliers were removed to render the database homogeneous ($N = 83$).[3] The database has mean *z* score = 0.68 ($SD = 1.70$); 95%CI [0.31, 1.05]; $ES = 0.02$ (SD = 0.04); 95%CI [0.01, 0.02]; Stouffer $Z = 6.18$ ($p = 3.21 \times 10^{-10}$). Note that these 95% CIs do not include values of MCE (i.e., zero). A single-sample *t*-test revealed that *ES* values significantly deviated from chance, $t(82) = 4.26$, $p < .001$ (two-tailed). Of 83 studies, 16 (19%) are independently significant ($\alpha = .05$).

The revised original database (1987 to 2010; $N = 83$) and the database of new studies (2011 to 2022; $N = 59$) were not significantly different on *ES* values, $t(140) = 0.08$, $p = .940$ (two-tailed). The two databases were combined. One outlier was removed.[3] Table 3 lists statistics for the combined homogeneous database of studies from 1987 to 2022 ($N = 141$), as well as subgroups of experimental conditions. Statistics not given in Table 3, include mean *z* score = 0.72 ($SD = 1.58$); 95%CI [0.45, 1.00]; Note that most of the 95% CIs do not include values of MCE (i.e., zero). Stouffer $Z = 8.52$ ($p < 10^{-16}$). A single-sample *t*-test revealed that *ES* values deviated significantly from chance, $t(140) = 5.78$, $p < .001$ (two-tailed). Of 141 studies, 29 (21%) are independently significant ($\alpha = .05$).

Post Hoc Analyses

We re-assessed hypotheses *H2* (modality) and *H3* (target type), this time applying our tests to the largest database of forced-choice studies assembled to date ($N = 141$). As telepathy was not represented in two types of targets (Symbols and Numbers), and clairvoyance was not represented in one type of target (Numbers), relevant data were removed for this analysis only, reducing the database to $N = 103$. Again, we conducted a Univariate ANOVA. The following results were obtained:

(i) Psi modality, $F(2, 94) = 0.30$, $p = .739$ (two-tailed)

Table 3. Combined Database of Studies (1987 to 2022): Effect Sizes, and 95% Confidence Intervals, *p*-value and I^2

Datasets and Subsets	*n*	*ES (SD)*	*95% CI*	*p*	I^2
Combined Studies (*N*)	141	0.02 (0.03)	0.009, 0.02	5.9×10^{-9}	97.1
Telepathy	12	0.03 (0.04)	0.004, 0.05	2.0×10^{-1}	95.6
Clairvoyance	58	0.01 (0.03)	0.001, 0.02	1.0×10^{-1}	96.4
Precognition	71	0.02 (0.03)	0.01, 0.02	1.1×10^{-6}	96.7
Selected participants	15	0.03 (0.04)	0.005, 0.05	2.0×10^{-2}	91.3
Unselected participants	126	0.01 (0.03)	0.007, 0.02	2.2×10^{-7}	97.0
With feedback	88	0.02 (0.03)	0.01, 0.02	1.0×10^{-9}	97.3
No feedback	53	0.01 (0.03)	-0.001, 0.02	1.8×10^{-2}	94.4

(ii) Target type, $F(2, 94) = 3.79$, $p = .026$ (two-tailed)

However, there were no significant differences between any two specific target types. There was a significant modality/target-type interaction effect, $F(4, 94) = 3.18$, $p = .017$ (two-tailed). Figure 2 illustrates the interactions between modality and target type. Letters were the most successful of the three target types, but only for telepathy.

Experimenter/laboratory differences (*H4*) were assessed separately since so much data was lost in this analysis due to the large number of single studies; 41 (29%) were excluded for this analysis. Also, we now had 19 mutually exclusive experimenter groups as a number of them from the old period (1987-2010) did not conduct studies in the new period (2011 to 2022). All 19 groups are: 'Argibay', 'Bem', 'Bierman', 'Dalkvist', 'Don', 'Ertel', 'Haraldsson', 'Luke', 'Palmer', 'Rao', 'Roe', 'Schlitz', 'Schönwetter', 'Simmonds-Moore', 'Sheldrake', 'Storm', 'Vaughan', 'Watt', and 'Wiseman'. There was a significant difference between groups of experimenters, $\chi^2(18) = 34.58$, $p = .011$ (two-sided). However, when mean *ES* values by group were checked against each other, there were no significant differences.

Finally, *H5* concerning the relationship between *z* scores and *k*-choices was re-tested on the large database ($N = 141$). Previously, Storm et al. (2012) reported a significant trend ($r = 0.79$). On this occasion, the effect was moderate in strength but not significant, $r(5) = 0.48$, $p = .139$ (one-tailed). Correlating *z* scores for individual studies with their *k* values produced a significant outcome, $r(139) = 0.20$, $p = .008$ (one-tailed). The strongest *z* scores tended to correlate with higher *k* values.

For the final database ($N = 141$), Table 3 shows differences between (i) the three modalities; (ii) selected and unselected participants; and (iii) feedback/no-feedback conditions. A Univariate ANOVA was conducted to test these differences. There were significant *ES* differences between participant type and feedback condition:

(i) Psi modality, $F(2, 131) = 0.18$, $p = .834$ (two-tailed)

(ii) Selected vs. unselected participants, $F(1, 131) = 4.39$, $p = .038$ (two-tailed)

(iii) Feedback vs. no-feedback condition, $F(1, 131) = 5.43$, $p = .021$ (two-tailed)

Selected participants (mean $ES = .03$) were superior in ESP performance compared to unselected participants (mean $ES = .01$). Studies that gave feedback to participants (mean $ES = .02$) produced superior ESP performances than studies that did not give feedback (mean $ES = .01$). There were no significant interaction effects.

These findings prompted a comparison of selected participants who received feedback with selected participants who did not receive feedback. The former were very few in number ($n = 5$) compared to the latter ($n = 136$), but we can assume equal variance ($p = .205$), and the difference was significant, $t(139) = 1.87$, $p = .032$ (one-tailed). 'Selected with feedback' (mean $ES = 0.04$; $SD = 0.04$) outperformed 'selected with no feedback' (mean $ES = 0.02$; $SD = 0.03$).

Decline vs. incline effect. In this research field, some authors suggest *episodic* (within study) declines in effects

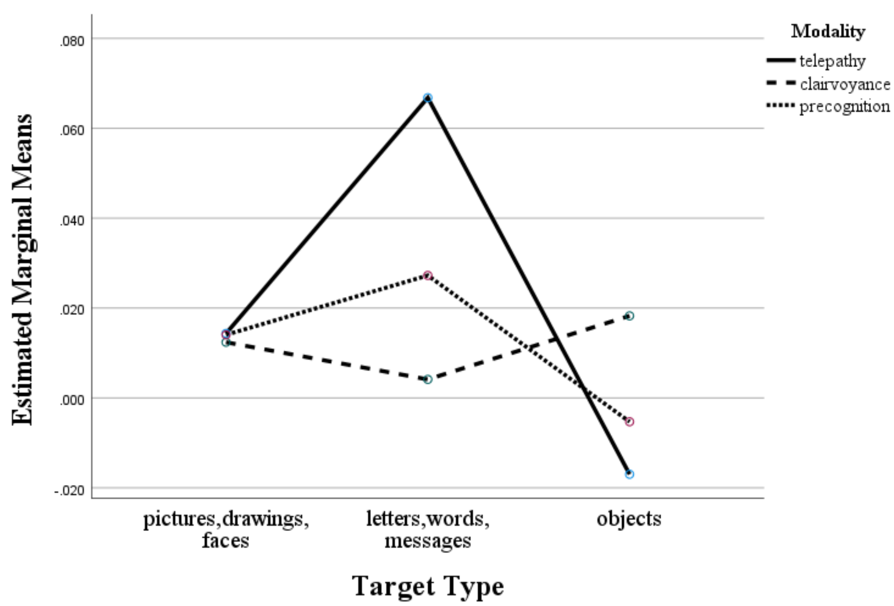


Figure 2. The target-type advantage: ESP differences on psi modality are not the same across levels of target type.

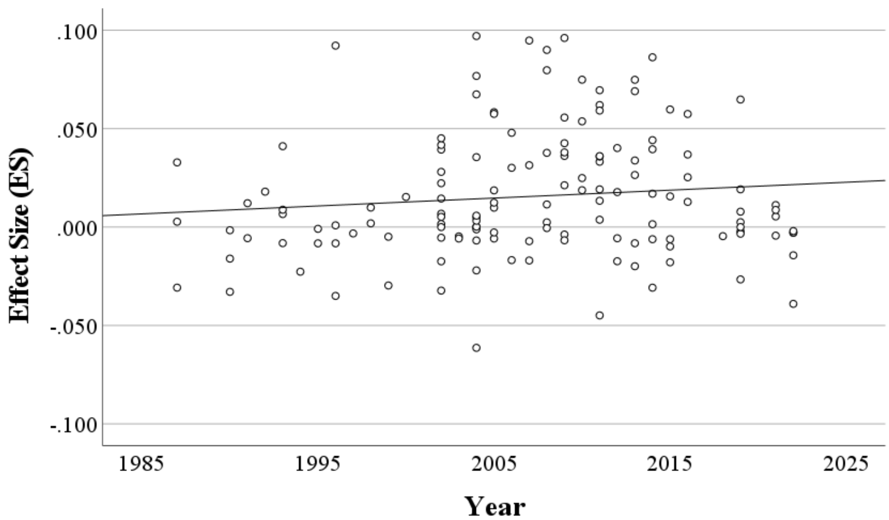


Figure 3. Scatter-plot of effect sizes (*ES*) for forced-choice studies showing a slight incline over a 36-year period.

are inevitable and appear as slow but constant decreases in the strength of effects due to repetition of similar tasks (Bierman, 2001, Kennedy, 2003). However, as Storm (in press) has pointed out, there is little evidence for *chronological* (between-study) declines. We tested the assumption of a chronological decline in two ways, where the hypothesis is that effect size (*ES*) covaries with year of publication and with a non-parametrical correlation between year of study and *ES*. The meta-regression coefficient is zero, with a *p*-value = .09. The Spearman's *rho* correlation was positive and approached significance, $r(139) = 0.11$, $p = .098$ (one-tailed), suggesting an incline. The linear trend line formula is $ES = [0.0004 \times \text{YEAR}] - 0.79$; linear $R^2 = .011$ (see Figure 3). Our hypothesis (*H6*) of an incline was partially supported.

Publication bias. As a means of avoiding publication bias, it has been the policy of the *Journal of Parapsychology* since 1975 to publish all papers passing peer review whether the reported results are significant or not. Other parapsychology journals have adopted the same policy. Grimes, Bauch, and Ioannidis (2017) note the "top-tier journals possess a limited number of publication slots and are thus overwhelmingly weighted towards publishing only novel or significant results" (p. 2). Parapsychological journals are free of that pressure due to the limited number of researchers in the field. We note that the majority of studies in this meta-analysis (91%) were either published in journals specializing in parapsychology or journals known to be favorable to parapsychological research. Clearly, the other 9% of journals expressed no bias at that time.

One empirical method to test if authors disseminated only experiments with positive statistical results is to count how many of them reached the statistical threshold of $p \leq .05$. In our database, we counted the number of experiments obtaining a *z* or *t* value equal or higher than 1.65, corresponding to a one-tailed *p*-value of .05. For the heterogeneous database (i.e., before we removed outliers), we counted 52 (29%) out of 179 effect sizes that are independently significant. As reported above, even 29 (21%) of 141 studies in the homogeneous database is a small fraction of the total. This result supports the hypothesis that our database is not likely to have been contaminated by publication bias, as we should expect a much larger percentage of successful (significant) studies as a result of authors withholding (not publishing) unsuccessful studies.

Participant comparisons across experimental designs. We compared selected and unselected participants across a range of experimental designs. Table 3 and our *Post Hoc* tests show that selected participants outperform unselected participants. Figures *4* and *5* show comparisons with forced-choice and various other meta-analyses—namely, free-response remote viewing (Tressoldi & Katz, 2023), free-response in a Ganzfeld environment (Tressoldi & Storm, 2023), and presentiment design (Duggan & Tressoldi, 2018). Figure 4 compares unselected participants; Figure 5 compares selected participants. Comparing Figure 4 with Figure 5, we see marked selection differences in other designs.

In reference to the two figures, we note that selected participants in 'Forced-Choice' are nearing *unselected* participants' in 'Free-Response Ganzfeld' in terms of mean *ES*. However, forced-choice continues to deliver smaller effects compared to the other three designs.

DISCUSSION

The above two-stage forced-choice meta-analysis on (a) two newly-formed databases, 2011 to 2022, and (b)

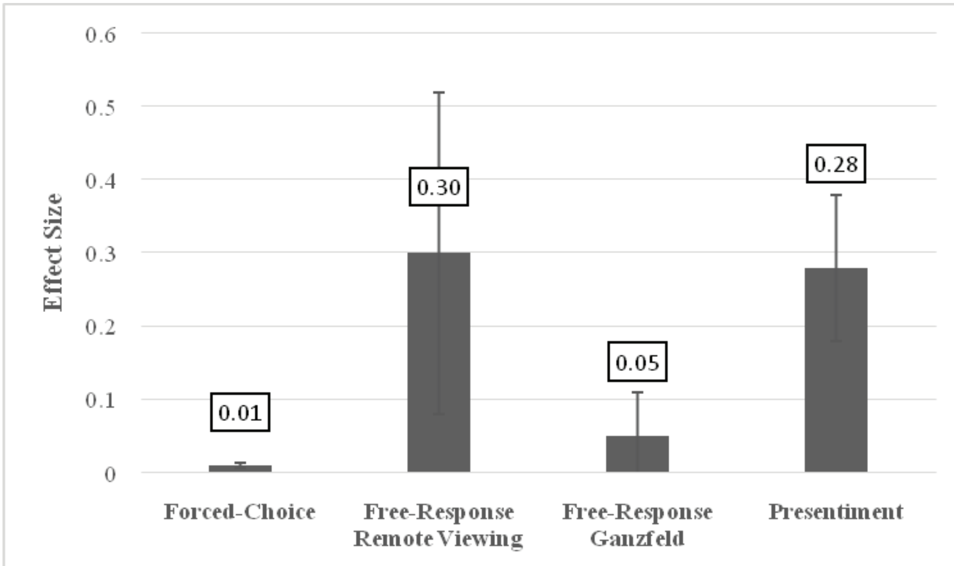


Figure 4. Unselected participants: Effect size with corresponding 95%CI, observed in the meta-analyses related to forced-choice, free-response remote viewing, free-response in a Ganzfeld environment, and presentiment designs.

the enlarged complete database 1987 to 2022 (*N* = 141), indicate that the forced-choice domain generally produces significant psi effects above MCE. Up until the present study, we had considered forced-choice effects to be very weak, and our position has not changed. The new findings are nevertheless encouraging, even across modalities, target types, and participant types (i.e., selected vs. unselected). We formed two databases—a dataset of New Studies #1 (studies reporting hit rates), and a dataset of New Studies #2 ('RT' Studies; i.e., measures not recording hits)—and although mean *ES* values were different (0.02 vs. 0.01, respectively), the difference was not statistically significant, so we combined the two. Also, New Studies #1 was more successful than New Studies #2 ('RT' Studies), with 29% of studies independently significant in the former compared to 17% in the latter. Generally, however, our results are very similar to those observed by Honorton and Ferrari (1989), Steinkamp et al. (1998), and Storm et al. (2012). The following is a break-down of our findings across a series of hypotheses.

We proposed six hypotheses and tested *H2* to *H6* using the combined dataset (*H1* was tested separately for both smaller datasets #1 and #2). Interestingly, Storm et al. (2012) reported that effects by modality were "very weak for precognition, clairvoyance, and even telepathy, which was the strongest effect of the three" (p. 259)—this finding

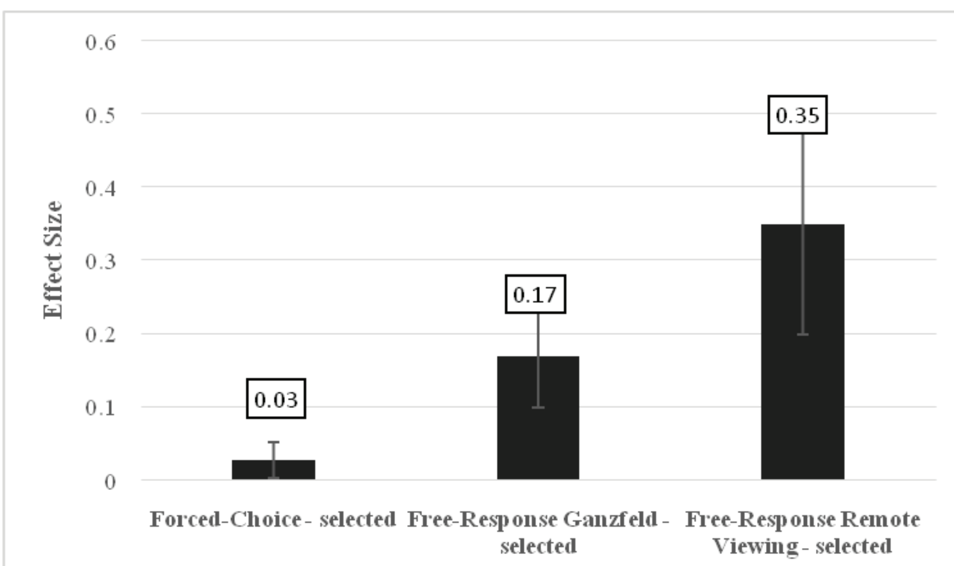


Figure 5. Selected participants: Effect size with corresponding 95%CI, observed in the meta-analyses related to forced-choice, free-response in a Ganzfeld environment, and free-response remote viewing designs (no data for presentiment).

was repeated with telepathy still the strongest effect, though not significantly (*H2*).

We found target type made no difference to effects in our test on the new studies (*N* = 59), but the test on the large database (*N* = 141) was significant (*H3*). However, there were no apparent differences between specific target types in the post hoc analysis. As we had already pointed out (see Storm et al., 2012, p. 260), possible reasons for these failures are to do mainly with participants' attitudes and reactions to targets—specifically, targets may be uninteresting and/or meaningless (i.e., Zener cards, numbers, letters, etc.) whereas experimenters should aim to offer emotionally stimulating and/or meaningful targets (i.e., divinatory readings, real pictures, video clips, etc.). Storm et al. also argued that cognitive noise may be higher in forced-choice studies compared to Ganzfeld and other designs where relaxation is offered. It is also thought that ecologically valid tasks are preferred by participants, such as e-mail or phone call predictions, and these are always telepathy tests (see Sheldrake et al., 2015). Preferred targets included various symbols (such as fractals) and text (i.e., words, letters, and messages).

Following Storm et al. (2012), we divided our database of new studies into mutually exclusive investigator/experimenter groups. Again, we found groups tended not to outperform each other (*H4*). No significant differences were found in the larger database.

Timm (2000) claimed that number of choices per trial (*k*) may be positively related to *z* score. However, when we tested the new set of studies, no proposed advantage based on *k* was found (see *H5*). A re-test on the larger database was not significant when *z* scores were grouped by *k* to form mean scores for each *k*-group, but correlating *z* scores for individual studies with *k* also produced a significant effect. The strongest *z* scores tended to correlate with higher *k* values.

In testing *H6*, we found a significant time-dependent *decline* in *ES* values in the new studies (*N* = 59), which covers a 12-year period, but for the longer period of analysis (1987 to 2022), there was a marginally significant incline (see Figure 3). We noted earlier that the correlation between year of study and *ES* indicated an incline, "meaning that *ES* values increased over the 24-year period" (Storm et al., 2012, p. 257). That effect was significant, and after 36 years, there is now good evidence the incline has been maintained.

Further tests on the largest ever database of forced-choice studies produced some additional findings that warrant mentioning:

First, we did not find a performance difference between modalities, but an interaction effect showed that telepathy was a 'show-case' modality, provided targets were letters, words, or messages (see Figure 2).

Second, testing selected participants, compared to unselected participants, showed a significant mean *ES* difference in ESP performance. This finding shores up the general understanding that some gifted participants tend to obtain high scores in ESP tasks in forced-choice designs (for examples, see Honorton, 1987; Kanthamani & Kelly, 1974; Steinkamp, 2005).

Third, in testing studies with a feedback condition, we found that giving participants feedback (either after trials or after runs) also gives an advantage. Given this finding, and the previous one concerning selection, it followed that we should test the 'selected with feedback' condition against the 'selected with no feedback' condition. We found a significant difference between the two, with selected participants who received feedback producing a stronger mean *ES* than selected participants who did not receive feedback.

Fourth, we found no evidence of publication bias; otherwise, we would expect a much higher rate of independently significant studies than a mere 21% (29 out of 141 studies).

Our fifth and final point is that the forced-choice domain, with *ES* = 0.02, has confirmed previously reported low *ES* values (see Honorton & Ferrari, 1989; Steinkamp et al., 1998; Storm et al., 2012). The effect is enduring and consistent across time and other variables, albeit weak (which is to say 'small'), and it might be argued that there is room for improvement. We note that the effect is weak because it is estimated by considering the number of *trials* and not the number of *participants*, and number of trials in forced-choice protocols are much larger than those used in typical free-response protocols. For all studies from 1987 to 2022 (*N* = 179) before outliers were removed, we estimate the typical participant performs 14 trials on average, whereas most participants in Ganzfeld experiments seldom do more than one trial each. Nevertheless, the number of forced-choice studies that are independently significant (21%) is better than that for the free-response domain, such as non-ganzfeld noise-reduction and standard free-response (19% and 15%, respectively; see Storm & Tressoldi, 2020, pp. 205-206), and not so far from Ganzfeld (26%). Also, for experimenters to get some kind of additional advantage in forced-choice studies, they are advised to test selected participants and certainly to offer feedback. They may then find that their participants (selected with feedback; mean *ES* = .04) perform possibly even better on average than participants in Bem-type precognition studies (mean *ES* = .03)—see Cardeña (2018, p. 667).

Given these findings, skeptics, critics, and even psi advocates may need to reconsider their current positions on the efficacy of forced-choice designs insofar as their views may be negative. Yet, we too, suggested that the forced-choice "weak effect might prove difficult to deploy efficaciously" (Storm et al., 2012, p. 261). We have seen that there are steps that can be taken to improve that outlook.

ENDNOTES

1. For a review of the meta-analyses of these and other experimental domains, see Cardeña (2018).
2. Although Delorme et al. (2016) was initially peer-reviewed, it was later retracted by the hosting journal and is, therefore, officially unpublished. Hence, we excluded it from this meta-analysis on the grounds that it must again go through peer review prior to publication.
3. 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 I^2 are 25% (low), 50% (moderate), and 75% (high). For details about Cochran's Q statistic, see Lipsey and Wilson (2001).

REFERENCES

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

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

* Arora, S., Schmidt, M., Boylan, J., & Pantazatos, S. P. (2022). Attempt to replicate Bem's precognitive avoidance task and detect relationships with trait anxiety. *Journal of Consciousness Studies*, *29*(5-6), 8–20. doi.org/10.53765/20512201.29.5.008

Batthyány, A., Kranza, G. S., & Erber, A. (2009). Moderating factors in precognitive habituation: The roles of situational vigilance, emotional reactivity, and affect regulation. *Journal of Society for Psychical Research*, *73*, 65–82.

* Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407–425. doi:10.1037/a0021524

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

Bem, D., Tressoldi, P., Rabeyron, T., & Duggan, M. (2016). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events [version 2; referees: 2 approved]. *F1000Rsearch*, https://doi.org/10.12688/f1000research.7177.2

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. https://doi.org/10.1075/aicr.29.12bie

* Bierman, D. J. (2011). Anomalous switching of the bi-stable percept of a Necker Cube: A preliminary study. *Journal of Scientific Exploration*, *25*(4), 721–733.

* Bierman, D., & Bijl, A. (2014). Anomalous 'retrocausal' effects on performance in a Go/NoGo Task. *Journal of Scientific Exploration*, *28*(3), 437-452.

* Billows, H., & Storm, L. (2015). Believe it or not: A confirmatory study on predictors of paranormal belief, and a psi test. *Australian Journal of Parapsychology*, *15*, 7-35.

* Boer, De R., & Bierman, D. (2006). The roots of paranormal belief: divergent associations or real paranormal experiences? *Proceedings of the 49th Parapsychological Association Annual Convention*, 283–298.

Cardeña, E. (2018). The experimental evidence for parapsychological phenomena: A review. *American Psychologist*, *73*(5), 663–677. https://doi.org/10.1037/amp0000236

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. (2013). Performance in group telepathy experiments as a function of target picture characteristics. *Journal of Parapsychology*, *77*(1), 79–106.

Delorme, A., Pierce, A., Michel, L., & Radin, D. (2016, May 17). Prediction of mortality based on facial characteristics. *Frontiers in Human Neuroscience*, *10*. doi.org/10.3389/fnhum.2016.00173

Duggan, M., & Tressoldi, P. (2018). Predictive physiological anticipatory activity preceding seemingly unpredictable stimuli: An update of Mossbridge et al.'s meta-analysis. *F1000Research*, *7*, 407. https://doi.org/10.12688/f1000research.14330.2

Ertel, S. (2010). Psi in a skeptic's lab: A successful replication of Ertel's Ball Selection Test. *Journal of Scientific Exploration*, *24*(4), 581–598.

* Ertel, S. (2013). Psi effect or sensory leakage: Scrutinizing the Ball Selection Test. *Journal of Scientific Exploration*, *27*(3), 387–391.

* Escolà-Gascón, Á. (2022). Forced-choice experiment on anomalous information reception and correlations with states of consciousness using the Multivariable Multiaxial Suggestibility Inventory-2 (MMSI-2). *Ex-*

plore: The Journal of Science and Healing, *18*, 170–178. https://doi.org/10.1016/j.explore.2020.11.009

* Escolà-Gascón, Á., Wright, A. C., & Houran, J. (2022). 'Feeling' or 'sensing' the future? Testing for anomalous cognitions in clinical versus healthy populations. *Heliyon*, e11303. https://doi.org/10.1016/j.heliyon.2022.e11303

Grimes, D. R., Bauch, C. T., & Ioannidis, J. P. A. (2017). Modelling science trustworthiness under publish or perish pressure. *Royal Society Open Science*, *5*, 171511. http://dx.doi.org/10.1098/rsos.171511

Hadlaczky, G. (2005). Precognitive habituation. Accessed on http://www.diva-portal.org/smash/get/diva2:189200/FULLTEXT01.pdf

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. https://doi.org/10.1136/bmj.327.7414.557

* Hitchman, G. A. M., Roe, C. A., & Sherwood, S. J. (2012). A re-examination of non-intentional precognition with openness to experience, creativity, psi beliefs and luck beliefs as predictors of success. *Journal of Parapsychology*, *76*(1), 109–145.

* Hitchman, G. A. M., Roe, C. A., & Sherwood, S. J. (2015). The relationship between lability and performance at intentional and nonintentional versions of an implicit PMIR-type psi task. *Journal of Parapsychology*, *79*(1), 65–86.

* Hitchman, G. A. M., Sherwood, S. J. & Roe, C. A. (2015). The relationship between latent inhibition and performance at a non-intentional precognition task. *Explore: The Journal of Science and Healing*, *11*(2), 118–126. DOI: 10.1016/j.explore.2014.12.004

* Hitchman, G. A. M., Pfeuffer, C. U., Roe, C. A., & Sherwood, S. J. (2016). The effects of experimenter-participant interaction qualities in a goal-oriented nonintentional precognition task. *Journal of Parapsychology*, *80*(1), 45–69.

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

Honorton, C. (1987). Precognition and real-time ESP performance in a computer task with an exceptional subject. *Journal of Parapsychology*, *51*(4), 291–320.

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.

* Houran, J., & Lange, R. (2012). *I Ching* outcomes from experimental manipulations of transliminality paranormal belief. *Australian Journal of Parapsychology*, *12*, 39–58.

* Houran, J., & Lange, R. (2014). Applying the theory of reasoned action to a computerized test of intuition, part II: Decision-making in a hidden test of psi. *Journal of the Society for Psychical Research*, *77*(4), 236–251.

Irwin, H. J., & Watt, C. A. (2007). *An introduction to parapsychology* (5th ed.). McFarland.

* Jolij, J., & Bierman, D. (2019). Two attempted retro-priming replications show theory-relevant anomalous connectivity. *Journal of Scientific Exploration*, *33*(1), 43–60. https://doi.org/10.31275/2019.1262

Kanthamani, H., & Kelly, E. F. (1974). Awareness of success in an exceptional subject. *Journal of Parapsychology*, *38*(4), 355–382.

* Kekecs, Z., Palfi, B., Szaszi, B., Szecsi, P., Zrubka, M., Kovacs, M., ... Aczel, B. (2019, August 1). Raising the value of research studies in psychological science by increasing the credibility of research reports: The transparent psi project. *PsyArxiv* https://doi.org/10.31234/osf.io/uwk7y For the final published versions see also https://royalsocietypublishing.org/doi/pdf/10.1098/rsos.191375

Langan, D., Higgins, J. P. T., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., ... Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*, *10*(1), 83–98. https://doi.org/10.1002/jrsm.1316

* Lange, R. & Houran, J. (2013). Towards a replicable formula for significant *I Ching* outcomes. *Australian Journal of Parapsychology*,*13*, 9–25.

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

* Luke, D., & Morin, S. (2014). Exploration of the validity and utility of a reward contingency in a non-intentional forced-choice precognition task. *Journal of the Society for Psychical Research*, *78*(917), 207–218.

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 Society for Psychical Research*, *72*(4),193–207.

* 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. https://doi.org/10.14704/nq.2012.10.3.614

* Luke, D., & Zychowicz, K. (2014). Comparison of out-

comes with nonintentional and intentional precognition tasks. *Journal of Parapsychology*, *78*(2), 223–234.

* Muhmenthaler, M. C., Dubravac, M., & Meier, B. (2022). The future failed: No evidence for precognition in a large scale replication attempt of Bem (2011). *Psychology of Consciousness: Theory, Research, and Practice*. Advance online publication. https://doi.org/10.1037/cns0000342

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n71

* Parra, A., & Argibay, J. C. (2013a). Anomalous remote diagnosis: Mental and motor psi impressions under iconic representation of the person-target. *Journal of Parapsychology*, *77*(1),123–130.

* Parra, A., & Argibay, J.C. (2013b). Psi and death of the person-target: An experiment with highly emotional iconic representations. *NeuroQuantology*, *11*(4), 537–543. https://doi.org/10.14704/nq.2013.11.4.704

* Rabeyron, T. (2014). Retro-priming, priming, and double testing: psi and replication in a test-retest design. *Frontiers in Human Neuroscience*, *8*, 1–7. doi:10.3389/fnhum.2014.00154

* Rabeyron, T., Roe, C., Mousseau, M. C., & Deledalle, A. (2018). Anomalous experiences, mental health, and creativity: Is psi the missing link? *Journal of Consciousness Studies*, *25*(3-4), 207–232.

* Rabeyron, T. & Watt, C. (2010). Paranormal experiences, mental health and mental boundaries, and psi. *Personality and Individual Differences*, *48*, 487–492. doi: 10.1016/j.paid.2009.11.029

Rhine, J. B. (1934). *Extra-sensory perception.* Boston Society for Psychical Research. https://doi.org/10.1037/13314-000

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

* Roe, C. A., Grierson, S., & Lomas, A. (2012). Feeling the future: Two independent replication attempts. In *Abstracts of Presented Papers Parapsychological Association 55th Annual Convention* (pp. 52–53). Parapsychological Association.

Rubio-Aparicio, M., López-López, J. A., Sánchez-Meca, J., Marín-Martínez, F., Viechtbauer, W., & Van den Noortgate, W. (2018). Estimation of an overall standardized mean difference in random-effects meta-analysis if the distribution of random effects departs from normal. *Research Synthesis Methods*, *9*(3), 489–503. https://doi.org/10.1002/jrsm.1312

Savva, L., Child, R. & Smith, M. D. (2004). The precognitive habituation effect: An adaptation using spider stimuli. *Paper presented at the 47th Annual Convention of the Parapsychological Association*, Vienna, Austria.

* Savva, L., & French, C. C. (2002). Is there time-reversed interference in Stroop- based tasks. The Parapsychological Association 45th Annual Convention, Proceedings of the Presented Papers, 194-205.

Savva, L., Roe, C., & Smith, M. D. (2005). Further testing of the precognitive habituation effect using spider stimuli. *Proceedings of Presented Papers: The Parapsychological Association 48th Annual Convention* (pp. 163–170), Parapsychological Association.

* Schlitz, M., Bem, D., Marcusson-Clavertz, D., Cardeña, E., Lyke, J. Grover, R., Blackmore, S., Tressoldi, P., Roney-Dougal, S., Bierman, D., Jolij, J., Lobach, E., Hartelius, G., & Delorme, A. (2021). Two replication studies of a time-reversed (psi) priming task and the role of expectancy in reaction times. *Journal of Scientific Exploration*, *35*(1), 69–94. https://doi.org/10.31275/20211903

* Schlitz, M. & Delorme, A. (2021). Examining implicit beliefs in a replication attempt of a time-reversed priming task [version 2; peer review: 2 approved]. *F1000Research*, *10*(5) https://doi.org/10.12688/f1000research.27169.2

* Schönwetter, T., Ambach, W., & Vaitl, D. (2011a). Does autonomic nervous system activity correlate with events conventionally considered as unperceivable? Using a guessing task with physiological measurement. *Journal of Parapsychology*, *75*(2), 327–348.

* Schönwetter, T., Ambach, W., & Vaitl, D. (2011b). Does a modified guilty knowledge test reveal anomalous interactions within pairs of participants? *Journal of Parapsychology*, *75*(1), 93–118.

* Sheldrake, R., & Beeharee, A. (2016). Is joint attention detectable at a distance? Three automated, internet-based tests. *Explore: The Journal of Science and Healing*, *12*(1), 34–41. https://doi.org/10.1016/j.explore.2015.10.006

* Sheldrake, R., Smart, P., & Avraamides, L. (2015). Automated tests for telephone telepathy using mobile phones. *Explore: The Journal of Science and Healing*, *11*, 310–319.

* Simmonds-Moore, C. (2014). Exploring the perceptual biases associated with believing and disbelieving in paranormal phenomena. *Consciousness and Cognition*, *28*, 30–46. https://doi.org/10.1016/j.concog.2014.06.004

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).

McFarland.

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

* Storm, L., Ertel, S. & Rock, A. (2013). The sheep-goat effect as a matter of compliance vs. noncompliance: The effect of reactance in a forced-choice ball selection test. *Journal of Scientific Exploration*, *27*(3), 93–413.

Storm, L. (in press). The dark spirit of the trickster archetype in parapsychology. *Journal of Scientific Exploration*.

Storm, L., & Tressoldi, P. E. (2017). Gathering in more sheep and goats: A meta-analysis of forced-choice sheep-goat studies, 1994-2015. *Journal of the Society for Psychical Research*, *81*(2), 79-107.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2010). Meta-analyses of freeresponse studies 1992–2008: Assessing the noise reduction model in parapsychology. *Psychological Bulletin*, *136*, 491–494. doi:10.1037/ a0019840

Storm, L., Tressoldi, P. E., & Di Risio, L. (2012). Meta-analysis of ESP studies, 1987-2010: Assessing the success of the forced- choice design in parapsychology. *Journal of Parapsychology*, *76*(2), 243–273.

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. A., & Storm, L. (2014). A further study of psychopraxia using the *I Ching*. *Australian Journal of Parapsychology*, *14*, 115–142.

Tressoldi, P., & Katz, D. (2023). Remote Viewing: a 1974-2022 systematic review and meta-analysis. *PsyArXiv*, https://doi.org/10.31234/osf.io/r9yw5

Tressoldi, P. E., Martinelli, M., Zaccaria, E., & Massacessi, S. (2009). Implicit intuition: 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.

Tressoldi, P. E., & Storm, L. (2023). Stage 2 Registered Report: Anomalous perception in a Ganzfeld condition - A meta-analysis of more than 40 years investigation [version 1; peer review: awaiting peer review]. *F1000Research*, *10*, 234. https://doi.org/10.12688/f1000research.51746.2

* Varvoglis, M., Bancel, P. A., Bailly, J.-P., Boban, J., & Ahmed, D. s. (2019). The Selfield: Optimizing precognition research. *Journal of Parapsychology, 83*(1), 13–24. https://doi.org/10.30891/jopar.2019.01.02

* Vernon, D. (2015). Exploring precognition using a repetition priming paradigm. *Journal of the Society for Psychical Research*, *79*, 65–79.

Vernon, D., Sandford, T., & Moyo, E. (2019). A test of telepathy using immersive virtual reality. Paper presented at the 62nd Annual Convention of the Parapsychological Association, Paris, 2019.

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor package. Journal of Statistical Software, 36(3), 148. https://doi.org/10.18637/jss.v036.i03

* Wagenmakers, E. J., Wetzels, R., Borsboom, D., et al. (2012). An agenda for purely confirmatory research. *Perspective in Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Watt, C. & Nagtegaal, M. (2000). Luck in action? belief in good luck, Psi-mediated instrumental response, and games of chance. *Journal of Parapsychology*, *64*(1), 33–52.

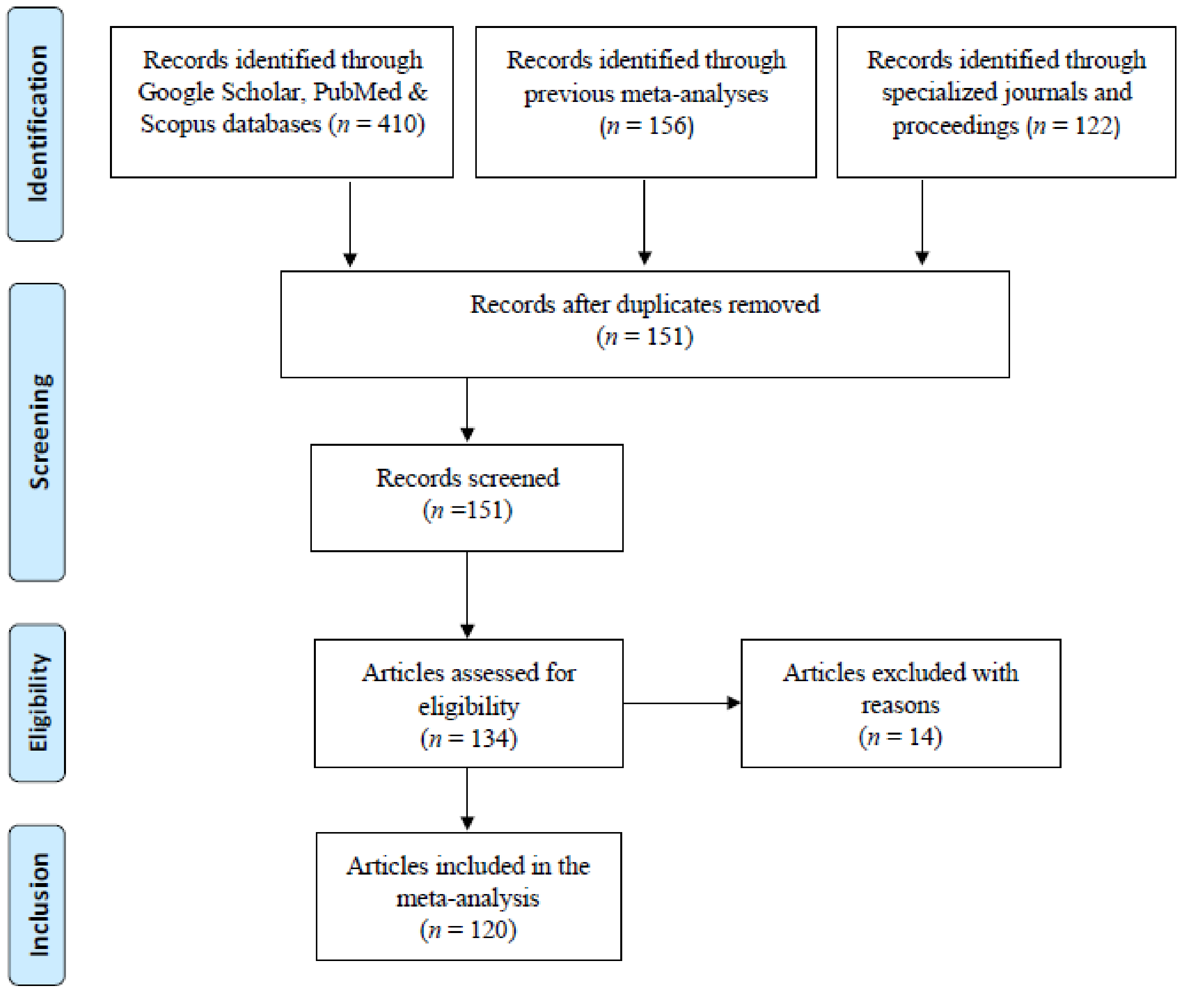
* Wehrman, J. J. (2019). No evidence for retrocausation in two classic cuing paradigms. *SAGE Open*, *9*(2), 2158244019855852

* Wittmann, M., Scheck, F., Feldmann, J., Glaesmann, A., Mossbridge, J., & Bem, D. (2021). The German version of a retroactive priming task shows mixed effects. *Psychology of Consciousness: Theory, Research, and Practice*. Advance online publication. https://doi.org/10.1037/cns0000296

Zdrenka, M., & Wilson, M. S. (2017). Individual difference correlates of psi performance in forced-choice precognition experiments: A meta-analysis (1945-2016). *Journal of Parapsychology*, 81(1), 9–32.

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

APPENDIX A: Prisma Flowchart



APPENDIX B: List of New Studies with Hits Data in the Meta-Analysis and Their Results

	Author(s)	Year	Trials	Hits	Z	ES (z/√n)	Task
1	Bem (expt. 1)	2011	3600	1912	3.72	0.062	PR
2	Bem (expt. 2)	2011	5400	2790	2.44	0.033	PR
3	Bem (expt. 3)	2011	1800	963	2.95	0.070	PR
4	Bem (expt. 5)	2011	2400	1274	2.90	0.059	PR
5	Bem (expt. 6)	2011	3600	1865	2.15	0.036	PR
6	Schönwetter et al.	2011a	657	151	-1.15	-0.045	CL
7	Schönwetter et al.	2011b	350	71	0.07	0.004	CL
8	Hitchman et al. (nonintentional)	2012	750	201	1.10	0.040	PR
9	Houran & Lange	2012	60	27	3.43	0.443	PR
10	Luke et al.	2012	200	43	-1.06	-0.075	PR
11	Dalkvist	2013	19560	9585	-2.78	-0.020	TE
12	Ertel	2013	7740	2293	21.16	0.241	CL
13	Lange & Houran	2013	60	34	5.52	0.713	PR
14	Parra & Argibay (believers; mental)	2013a	896	520	4.78	0.160	CL
15	Parra & Argibay (believers; motor)	2013a	896	482	2.24	0.075	CL
16	Parra & Argibay (live/nolive)	2013b	856	511	5.64	0.193	CL
17	Parra & Argibay (suicide/nosuicide)	2013b	856	458	2.02	0.069	CL
18	Parra & Argibay (motor live/nolive)	2013b	856	424	-0.24	-0.008	CL
19	Parra & Argibay (motor suicide/nosuicide)	2013b	856	443	0.99	0.034	CL
20	Storm et al.	2013	12016	2531	2.90	0.026	CL
21	Houran & Lange	2014	744	187	0.04	0.001	CL
22	Luke & Morin (contingent)	2014	210	57	0.64	0.044	PR
23	Luke & Morin (no-contingent)	2014	200	58	1.22	0.086	PR
24	Luke & Zychowicz (intentional; intuition)	2014	200	40	-1.55	-0.110	PR
25	Luke & Zychowicz (nonintentional; preference)	2014	200	52	0.24	0.017	PR
26	Simmonds-Moore	2014	95	22	-0.30	-0.031	CL
27	Simmonds-Moore	2014	95	24	-0.06	-0.006	CL
28	Billows & Storm	2015	3725	736	-0.35	-0.006	CL
29	Hitchman et al. (intentional)	2015	500	245	-0.40	-0.018	PR
30	Hitchman et al. (nonintentional)	2015	500	247	-0.22	-0.010	PR
31	Hitchman et al. (nonintentional)	2015	735	390	1.62	0.060	PR
32	Sheldrake et al. (three callers)	2015	1728	718	7.22	0.174	TE
33	Sheldrake et al. (two callers)	2015	660	370	3.08	0.120	TE
34	Hitchman et al. (nonintentional)	2016	624	160	0.32	0.013	PR
35	Sheldrake & Beeharee (study 1)	2016	11160	5901	6.07	0.057	TE
36	Sheldrake & Beeharee (study 2)	2016	2720	1395	1.32	0.025	TE
37	Sheldrake & Beeharee (study 3)	2016	8860	4594	3.47	0.037	TE
38	Kekecs et al.	2019	37836	18876	-0.43	-0.002	PR
39	Varvoglis et al. (all no feed)	2019	1172	570	-0.91	-0.027	PR
40	Varvoglis et al. (all with feed)	2019	1828	932	0.82	0.019	PR
41	Varvoglis et al. (selected no feed)	2019	358	179	0.00	0.000	PR
42	Varvoglis et al. (selected with feed)	2019	602	321	1.59	0.065	PR
43	Arora et al.	2022	5148	2566	-0.21	-0.003	PR
44	Escolà-Gascón (mediums; positive)	2022	900	432	-1.17	-0.039	CL
45	Escolà-Gascón (mediums; neutral)	2022	900	287	-10.83	-0.361	CL
46	Escolà-Gascón (mediums; haunted)	2022	900	443	-0.43	-0.014	CL
47	Escolà-Gascón (non-believers; positive)	2022	900	238	-14.10	-0.470	CL
48	Escolà-Gascón (non-believers; neutral)	2022	900	252	-13.17	-0.439	CL
49	Escolà-Gascón (non-believers; haunted)	2022	900	248	-13.43	-0.448	CL
50	Escolà-Gascón et al. (study 1)	2022	7110	2133	-33.72	-0.400	PR
51	Escolà-Gascón et al. (study 2; healthy)	2022	3630	1117	-23.15	-0.384	PR
52	Escolà-Gascón et al. (study 2; psychiatric)	2022	3540	1310	-15.45	-0.260	PR

APPENDIX C: List of New Studies with No Hits Data (RT Studies) in the Meta-Analysis and Their Results

	Author(s)	Year	Trials	*t*	ES (z/√n)	Task
1	Savva & French (expt. 1)	2002	1371	0.83	0.022	PR
2	Savva & French (expt. 2)	2002	1800	1.19	0.028	PR
3	Savva & French (expt. 3)	2002	4320	0.00	0.000	PR
4	Boer & Bierman	2006	9792	2.98	0.030	PR
5	Rabeyron & Watt	2010	4960	1.32	0.019	PR
6	Bem (expt. 4)	2011	3168	2.03	0.036	PR
7	Bem (expt. 7)	2011	9600	1.31	0.013	PR
8	Bierman	2011	5408	1.41	0.019	PR
9	Roe et al. (study 1)	2012	1504	0.69	0.018	PR
10	Roe et al. (study 2)	2012	1344	-0.64	-0.017	PR
11	Wagenmakers et al.	2012	1500	-0.22	-0.006	PR
12	Bierman & Bijl	2014	4288	2.59	0.040	PR
13	Rabeyron	2014	896	-1.35	-0.045	PR
14	Vernon	2015	9792	1.55	0.016	PR
15	Rabeyron et al.	2018	2944	-0.25	-0.005	PR
16	Jolij & Bierman (expt. 1)	2019	7104	0.21	0.002	PR
17	Jolij & Bierman (expt. 2)	2019	1984	-0.15	-0.003	PR
18	Wehrman	2019	4080	0.50	0.008	CL
19	Schlitz & Delorme	2021	4449	-0.29	-0.004	PR
20	Schlitz et al. (expt. 1)	2021	19720	0.76	0.005	PR
21	Schlitz et al. (expt. 2)	2021	22560	1.68	0.011	PR
22	Wittmann et al. (expt. 1)	2021	3640	0.53	0.009	PR
23	Wittmann et al. (expt. 2)	2021	3720	-2.34	-0.038	PR
24	Muhmenthaler et al. (expt. 1)	2022	26171	-0.45	-0.003	PR
25	Muhmenthaler et al. (expt. 2)	2022	50904	-0.46	-0.002	PR