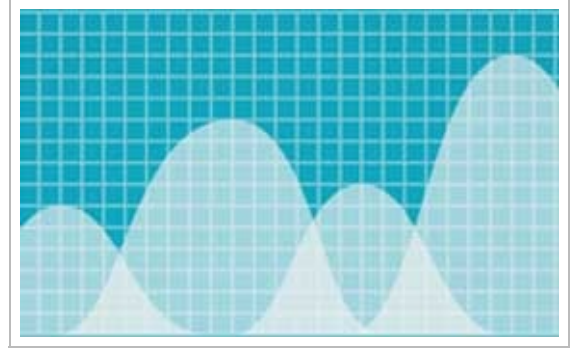# Meta-Analysis in Parapsychology

Meta-analysis is a statistical procedure that combines the results of a number of studies in a particular area of research in order to provide a more robust finding. The method has been embraced by parapsychologists since the 1980s, since the results of meta-analyses tend overwhelmingly to confirm the statistically significant findings of individual psi studies, underscoring the existence of psi as a genuine phenomenon. However, this apparent success has contributed to doubts about the value of meta-analysis among those who question the reality of psi, fuelling controversy about its true worth.

Outline of Meta-Analysis

For parapsychology, meta-analysis is a major form of scientific evidence that psi is real, or as stated by statistician Jessica Utts, an 'anomalous effect in need of an explanation'.[1]

Twelve major experimental domains feature in this review:

- ganzfeld
- autoganzfeld
- free-response
- remote viewing
- forced-choice (includes card-guessing)
- biological systems
- dice-throwing
- micro-PK (random number generators)
- dream-ESP
- sheep-goat effect
- presentiment effect
- hypnosis/comparison condition ESP

The meta-analytic results are an encouraging step towards establishing the replicability of psi effects, since significant effects have been found across the full range of domains. These results suggest a very real (albeit statistical) anomaly worthy of continued investigation, especially in domains such as ganzfeld, autoganzfeld, remote viewing, biological systems, micro-PK, and dream-ESP.

Criticisms and Responses

Apart from parapsychology, the use of meta-analysis has provided favourable evidence in other controversial fields, such as psychotherapy. This success has increasingly raised questions about the worth of the method among critics opposed to 'fringe' science.[2] Some have argued that results are corrupted if data from methodologically flawed experiments are included with data from better-designed experiments.[3] Variations in the way supposedly identical experiments are carried out may also yield a tainted result.

To meet these criticisms parapsychologists have made improvements aimed at controlling confounding factors. Rosenthal advocates differential weighting as an effective way of dealing with 'variation in the quality of research'.[4] Hence the 'blocking' procedure is used to code experiments according to their 'quality' and type (methodology, hypothesis, etc.). Sample size and the population from which the sample is drawn are other critical considerations. Credit is given to studies if sample size is specified in advance, as well as the nature of the analyses — pre-planned or *post hoc*. Acceptable randomisation methods are also credited, and even the date of the experiment and the identity of the investigator are now important criteria in meta-analytic studies.[5]

Although these procedures may be seen as subjective, some degree of qualitative assessment can be made about studies, and these assessments are converted to numerical values to arrive at a more objective, albeit pseudo-precise numerical result that is still seen as a gain over previous methods that did not consider study quality.

File Drawer Problem

Another major criticism of meta-analysis is that a significant outcome is inevitable as long as the majority of studies included in the analysis were individually significant. If, as critics tend to assume, many studies have nonsignificant outcomes but are rarely published and lie unnoticed in a file-drawer, none of these can be included in the meta-analysis, therefore artificially skewing the outcome.[6] This problem is often referred to as 'selective reporting' or the 'file drawer problem'.

Parapsychologists offer three responses to this charge. First, parapsychology journals go to great lengths to publish studies with nonsignificant results precisely to avoid such studies ending up in the file drawer. In 1975, the Parapsychological Association adopted a policy of opposing the exclusive publication of studies with positive outcomes. As a result, negative findings have been routinely reported at the association's meetings and in its affiliated publications since that date.[7]

Second, estimates can be made that account for unpublished studies. As is described in more detail below, the number of nonsignificant studies that would be needed to reduce a significant meta-analytic result to a chance outcome is often far in excess of that which would be possible for the few researchers in the field of parapsychology.[8]

Third, the 'funnel-plot' technique gives meta-analysts the means by which all the studies used in the meta-analysis can be distributed and presented on a two-axis array (effect size on the *x*-axis, and number of studies, *N*, on the *y*-axis). This usually appears as a scatter of data points that look like an inverted funnel-shape evenly distributed around a mean effect size value. The funnel shape results from the general rule that effect sizes tend to approach zero as *N* increases. If the funnel-plot is asymmetrical, the researcher can determine how many studies (and their effect size values) are *theoretically* missing in order to produce a symmetrical plot.

Empirical Findings

The Ganzfeld Procedure

The ganzfeld is a form of free-response test — 'free response' being a term that 'describes any test of ESP in which the range of possible targets is relatively unlimited and is unknown to the percipient'.[9] In a free-response test, the target is not restricted to a few choices, but can be almost anything, thus hopefully reducing the risk of boredom so common in forced-choice experiments because free responses 'more nearly resemble the conditions of spontaneous psi occurrences'.[10]

The ganzfeld is a 'special type of environment (or the technique for producing it) consisting of homogeneous, un-patterned sensory stimulation' to the eyes and ears of the participant who is usually in 'a state of bodily comfort'.[11] A number of investigators pioneered the technique in the 1970s.[12] The technique allegedly minimizes mental 'noise' and ambient noise in the laboratory, thus allowing optimal opportunity for the psi 'signal' to be perceived.

Procedurally, the eyes of the participant are covered with halved ping-pong balls illuminated by a uniform source of light (usually of a single wavelength, such as red light). A uniform auditory signal of 'white' noise (full-range audio signal),[13] or 'pink' noise (high-frequency filtered sound),[14] is channelled through headphones to the ears. The participant reclines on a chair or lies on a bed. This technique has remained essentially the same since the 1970s.

Ganzfeld Meta-Analyses

The first major meta-analytic study in parapsychology started in 1981 when Ray Hyman, an American psychology professor and noted skeptic of parapsychology, began evaluating 42 ganzfeld psi studies conducted during the period 1974 to 1982.[15] Hyman initially chose the ganzfeld studies because they supposedly held a 'high level of research sophistication and rigor' — a claim that Hyman was to criticise heavily.[16]

A public debate ensued between Hyman and parapsychologist Charles Honorton, a pioneer of the ganzfeld method, since they arrived at conflicting conclusions from the same data set. Hyman first argued that the 'alleged' 55% success rate of 42 studies determined from a vote-count made by Honorton was inflated due to the fact that many of the studies were not independent, but rather subsets of ongoing experiments.[17]

Hyman also cited evidence that suggested there was bias in how the studies were reported. For example, some studies were not planned as such, but were 'given this status retrospectively just because they yielded significant results'.[18] Hyman

reduced the number of successful studies to 31%. He further criticised many of the studies for their multiple analyses (for example, use of differing measures of ESP) that he argued gave increased opportunity for a good result, especially since investigators were not adjusting their criterion significance levels according to the number of statistical tests they performed.

Honorton accepted the criticism of multiple analysis, and to correct it applied a Bonferroni correction across all studies.[19] He now found that only 45% of the 42 studies were significant — not 55% which he originally claimed, although still a higher proportion than Hyman's 31%. Honorton arrived at a final total of 28 studies that used direct hits alone: 12 (43%) were significant at $p \leqslant .05$, and 23 (82%) had positive z scores. Honorton reported a composite Stouffer Z score of 6.60 across the 28 studies. The suggestion of a file-drawer problem was also rendered less plausible by the fact that 15 nonsignificant and unknown studies would have to exist for every one of the 28 direct hit studies to reduce the result to a chance outcome.

Hyman had more to say about Honorton's 28 studies.[20] He claimed to identify 12 major flaws, such as inadequate randomization of targets and failure to use a duplicate set of targets for judges. Rather than continue the debate, Hyman and Honorton produced a 'Joint Communiqué' addressing fundamental issues in parapsychological experimentation.[21] The communiqué recommended that 'more stringent standards' be implemented in experiments, which should also be conducted by a 'broader range of investigators'.[22] Utts listed these standards as including

- controls against any kind of sensory leakage
- thorough testing and documentation of randomization methods used
- better reporting of judging and feedback protocols
- control for multiple analyses [and statistics]
- and advance specification of number of trials and type of experiment.[23]

Autoganzfeld Meta-Analyses

The automated ganzfeld ('autoganzfeld') procedure was adopted as a more rigorous approach to psi testing, while still maintaining the ganzfeld paradigm. It came into being as a proactive response to the recommendations in the 'Joint Communiqué' that studies be computer-controlled, and targets be randomly-selected, presented, and scored. As in the ordinary ganzfeld, targets can be 'dynamic' (short scenes from movies, cartoons, documentaries), or 'static' (photographs, art prints, advertisements).

A series of 11 autoganzfeld experiments was conducted by eight experimenters during the period 1983-1989.[24] As reported in Bem and Honorton there was a significant 32% hit-rate when 25% would be expected by chance.[25] The ordinary ganzfeld and the autoganzfeld appeared to be equally effective, since they produced similar effect sizes. Given that Honorton's database of 28 studies,[26] and the new Honorton *et al.* database of 11 autoganzfeld studies were not significantly different on mean effect sizes and mean z scores,[27] Honorton *et al.* combined the two into a 39-study database that was highly significant, Stouffer $Z = 7.53$ ($p = 9.00 \times 10^{-14}$).[28]

Milton and Wiseman followed up with a meta-analysis of 30 new ganzfeld studies dating from 1987 to 1997.[29] Studies prior to 1987 were not used because it was assumed that investigators needed time to familiarise themselves with Hyman and Honorton's guidelines so that earlier studies would be too flawed for serious consideration in a meta-analysis.[30] Milton and Wiseman deemed suitable for analysis 30 studies by '10 different principal authors from 7 laboratories'.[31] They calculated a Stouffer Z of 0.70, $p = .24$ ($ES = 0.013$),[32] and concluded that 'the autoganzfeld results have not been replicated by a "broader range of researchers".'[33]

Taking issue with the Milton-Wiseman meta-analysis, Storm and Ertel argued that the pair had failed to demonstrate a thorough meta-analysis of the available literature.[34] Storm and Ertel found an additional 11 pre-communiqué studies not previously meta-analysed, and after step-by-step performance comparisons, combined them with the three ganzfeld databases currently extant: Honorton's database of 28 studies,[35] Bem and Honorton's databases of 10 studies (having removed one outlier study),[36] and Milton and Wiseman's database of 30 studies.[37] The resulting 79-study database had a significant mean ES of 0.14 ($Z = 5.66$, $p = 7.78 \times 10^{-9}$).

Milton and Wiseman argued that the 11 pre-Communiqué studies used in Storm and Ertel's meta-analysis should not have been used at all because they were poor in quality due to their ostensible 'methodological problems'.[38] However, Milton and Wiseman overlooked Storm and Ertel's performance comparisons of (a) pre-communiqué authors with post-communiqué authors, and (b) pre-communiqué studies with post-communiqué studies, both of which yielded no statistical evidence that the guidelines in the communiqué had any 'influence on effect size outcomes'.[39] Logically, there

was no indication that the mean effect size of the pre-communiqué database was 'inflated' (i.e., an artifact of flaws) because it compared favourably with the allegedly 'flawless' post-communiqué studies. It follows that there was no evidence that the mean effect size of the post-communiqué database was 'deflated' due to the removal of these flaws.

That same year in 2001, Bem, Palmer, and Broughton reported a further ten ganzfeld studies that were conducted after Milton and Wiseman's analysis.[40] When these studies were combined with the 30 studies in the Milton-Wiseman database, a significant hit rate of 30.1% was obtained, where 25% would be expected by chance.

Some years later, Storm, Tressoldi, and Di Risio,[41] covering the period 1997 to 2008, reported statistics for a homogeneous dataset of 29 ganzfeld studies. They calculated a mean effect size of 0.14 (Stouffer $Z = 5.48$, $p = 2.13 \times 10^{-8}$). This effect was significantly higher than the mean effect size of a combined set of standard (nonganzfeld) free-response studies. Also, so-called 'selected' participants (believers in the paranormal, meditators, etc.) had a performance advantage over 'unselected' participants, but only if they were in the ganzfeld condition. They also found no convincing evidence of a decline effect in the ganzfeld database over a period spanning four decades.

Hyman disputed these meta-analytic results, and he critiqued the rationale for the study, but only gave the usual responses that such meta-analytic findings elicit.[42] For example, he cloaked conventional meta-analytic practice in the guise of some kind of untoward procedure when he accused the authors of manufacturing 'homogeneity and consistency by eliminating many outliers and combining databases whose combined effect sizes are not significantly different'.[43] Storm, Tressoldi, and Di Risio pointed out that they had not broken any rules, and they argued that Hyman did not tell the full story about the ganzfeld meta-analytic findings — in fact, he focused on studies that apparently failed to showed replication, but did not mention those that did not fail.[44]

The following year, Williams revisited the ganzfeld database, with a focus on post-communiqué (autoganzfeld) studies only, arguing that this database would be the least likely to contain 'serious flaws' that would 'inflate or otherwise confound the overall results'.[45] He also endeavoured to show that the ganzfeld effect had been replicated independently of the highly successful series of 10 studies reported by Bem and Honorton.[46] Indeed, Williams was able to demonstrate replication — his database of 59 studies yielded a hit rate of 31% ($Z = 7.37$, $p = 8.59 \times 10^{-14}$).

The ganzfeld meta-analyses have attracted considerable attention, with critics conducting alternative investigations of the databases. An update by Storm, Tressoldi, and Utts[47] might never have been written had it not been for a critical paper by Rouder, Morey, and Province[48] which attempted to undermine the initial findings of Storm *et al.*[49] Rouder *et al.* reassessed Storm et al.'s meta-analysis, but they then conducted a Bayesian analysis on an entirely different database, one which was compiled with less than desirable precision.

Rouder et al.'s Bayesian approach was not without merit *in principle* for 'Bayes factors allow the analyst to state evidence for the no-psi effect null as well as for a psi-effect alternative'.[50] Although they found evidence for the existence of psi by a factor of about 6 billion to 1, much of this effect was attributed to 'difficulties in randomization',[51] for ganzfeld studies with *computerized randomization* supposedly had smaller psi effects than those with *manual randomization*. Storm *et al.* showed that this conclusion was unconvincing as it was based on Rouder et al.'s above-mentioned faulty and inconsistent compilation methodology. In addition, Storm et al.'s own Bayesian analysis yielded contradictory evidence to that of Rouder et al., where 'clear superiority of the combined ganzfeld and nonganzfeld noise reduction studies emerges, with an HDI (high density interval that indicates the most plausible 95% of the values in the distribution) ranging from 0.26 to 0.32'.[52] In other words, the authors found that the noise-reduction psi effect lies somewhere between 26% and 32%, well outside the 25% rate expected by chance.

Free-Response and Remote Viewing

Milton meta-analysed all available free-response studies for the period 1964 to 1992.[53] Her analysis putatively excluded studies using altered states of consciousness (ASCs), but included so-called 'remote viewing' studies in which the percipient 'attempts to describe the surroundings of a geographically distant agent'.[54] Milton found 78 studies with a mean effect size of 0.16 ($Z = 5.72$, $p = 5.40 \times 10^{-9}$). A file-drawer of 866 studies would be necessary to reduce this significant result to a chance outcome. An homogenised database of 75 studies had a slightly higher *ES* of 0.17 ($Z = 5.85$, $p = 2.46 \times 10^{-9}$). However, the homogeneity of the database may be questioned from another perspective because many of these studies featured 'meditation, hypnosis, mental imagery training (or guided imagery), relaxation, and even ganzfeld'.[55] Storm *et al.* formed a homogeneous database of 14 standard free-response studies free of ASCs that yielded a weak nonsignificant negative mean effect size of -0.03.[56] It is important to note, though, that this analysis was for the short period 1992 to 2008, and a 14-study dataset is far from representative.

The first major experimental research in remote viewing (RV) was by Targ and Puthoff (who coined the term).[57] They were particularly successful with Pat Price, a retired police commissioner, who was able to see objects large and small, such as furniture, and buildings, and concealed text.[58] Neither distance, nor size of target, seemed to influence RV outcomes. RV has been investigated systematically by a number of researchers.[59] The RV protocol is described as a 'non-altered state free response protocol', though the claim that there is no ASC is disputed — for example, well-known RV'ers like Joe McMoneagle make reference to their own altered states during RV sessions.[60]

Apart from Milton's (1998) review, which does not draw out the strength of the RV effect *per se*, the review that comes closest to a meta-analysis of sorts is by Baptista, Derakhshani, and Tressoldi.[61] They present a summary of results which show that effects range from as low as 0.16 (for studies conducted between 1964 to 1992),[62] to as high as 0.39 (for studies conducted between 1994 and 2014), depending on which laboratory conducted the studies.

Looking at individual laboratories, the Stanford Research Institute (SRI) produced a mean RV effect of 0.20, and its later incarnation, the Science Applications International Corporation (SAIC), produced a mean RV effect of 0.23. Princeton Engineering Anomalies Research (PEAR), run by Robert Jahn and Brenda Dunne, produced a mean RV effect of 0.21.[63] These effects are comparable and demonstrate a consistent effect over time and across laboratories.

Forced-Choice Meta-Analyses

Forced-choice experiments require that the participant 'guess a target that is one of a limited range of possibilities which are known to them in advance [such as in the card-guessing experiment]'.[64] Forced-choice precognition experiments from as early as 1935 up to 1987 were meta-analysed by Honorton and Ferrari.[65] They used only the studies where the procedure was to select the target 'randomly after the subject had attempted to predict what it would be'.[66] A total of 309 studies (62 of which were from 'senior authors') were analysed, amassing a phenomenal 50,000 participants and approximately two million individual trials. The effect size was 0.02 (mean $z = 0.65$, all studies).

Ninety-two studies (30%) showed significant hitting at the 5% level. A homogeneous database yielded a lower effect size of 0.012 ($Z = 6.02$, $p = 1.10 \times 10^{-9}$). The 'fail-safe N' was 14,268 studies,[67] which would be needed in order to reduce the significant effect to a chance outcome (requiring 46 unreported and unsuccessful studies for every successful study). Honorton and Ferrari concluded that precognition forced-choice experiments, although demonstrating a weak effect, produced consistent ('robust') and highly significant results across a time span of more than 50 years. Their meta-analysis also 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.

A later study by Steinkamp, Milton, and Morris meta-analysed forced-choice studies for the period 1935-1997,[68] and simultaneously compared clairvoyance with precognition in order to ascertain statistical evidence of a phenomenological difference between the two. They hypothesised that clairvoyance studies have a significantly higher effect size because precognition had an extra 'calculational step', involving 'real-time ESP' (clairvoyance) and then extrapolation from that information 'to make an informed prediction about future events'.[69]

Steinkamp *et al.* used a total of 22 comparable study-pairs in their meta-analysis, where procedures were effectively the same in both types of studies.[70] Effect sizes for precognition and clairvoyance were almost identical. Being such a small sample ($N = 22$ study-pairs) N-weighted effect sizes were calculated, again with essentially no difference in outcome (precognition: 0.034; clairvoyance: 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'.[71] It should be noted that Storm *et al.* have also not found a significant difference between psi modalities (telepathy, clairvoyance, and precognition) in the ganzfeld condition.[72] These results suggest that psi may not be compromised by apparent complexity of process.

Steinkamp's comprehensive review of forced-choice studies from 1880 to 1989 considered various predictors and psi-conducive variables.[73] She noted that 'there are few variables that have correlated clearly with success' and she was rather critical of the variations in study designs because these made it difficult to ascertain clear patterns due to conflicting outcomes.[74] However, she showed that there has been strong evidence in the past that low-neuroticism, high extraversion, prior testing (pre-selection of participants), and trial-by-trial feedback, are the most 'promising' and relevant variables in terms of yielding evidence for psi.

Finally, Storm, Tressoldi, and Di Risio looked at the forced-choice database for the period 1987 to 2010.[75] They formed a homogeneous dataset of 72 studies with an extremely weak but significant mean effect size of 0.01 (Stouffer $Z = 4.86$, $p =$

5.90×10^{-7}). There was no evidence that these results were due to low-quality design or selective reporting. They noted that effects did not vary between investigators, and there was no evidence of a decline effect over the period 1987 to 2010.

Forced-Choice Biological Systems

Braud and Schlitz conducted a 13-year-long series of studies that looked at eight living target systems:[76]

- electrodermal activity (EDA, skin resistance), participant's *influence* on a target system's skin resistance
- electrodermal activity (EDA) of a participant's *attention* away from target system
- ideomotor reactions (reactions associated with thought)
- muscular tremor (measured by the movement of a hand-held metal stylus in a small aperture)
- blood pressure
- fish orientation
- mammal locomotion (gerbil activity in a wheel)
- rate of hemolysis of human red blood cells

The goal in each case was to influence these systems to bring about 'increments' or 'decrements' in the activities of the monitored systems.[77] Experiments in this domain were described as testing participants' 'direct mental influence on living systems' or DMILS.[78]

A number of 30-second 'influence epochs' in a session were reduced to a single score (the unit of analysis). A 'percent influence score' was then calculated, a percent measure of the 'total activity that occurred in the prescribed direction during the entire set of influence (decremental or incremental aim) periods'.[79] A score of 50% (the result expected by chance) set the baseline for influence outcomes (no effect), and the *t* test was used to compare actual percent influence scores against this baseline.

Influence on remote biological systems was generally found to be significantly above chance on all target systems except muscular tremor, although a total of only 19 sessions were run for that system, whereas the next lowest was 40 sessions, and the average number of sessions was over 65. All eight systems produced significant effect sizes ranging from 0.095 for EDA (attention) to 0.300 for mammal location. Meta-analysis of the eight studies showed a significant mean effect size of 0.178 ($Z = 3.79$, $p = 1.00 \times 10^{-4}$). Braud and Schlitz discussed rival hypotheses that might explain these successful results, such as external stimuli, common internal rhythms, recording errors and biased misreading of records, participants' prior knowledge of when influence was to take place followed by appropriate responses, and even fraud.[80] All these 'explanations' were inapplicable. The overall conclusion was that 'effect[s] appear to occur in a "goal-directed" manner' because influencers were able to bring about effects without a specific understanding or awareness of how the physical or physiological processes brought about the desired outcomes.[81]

More recently, Schmidt, Schneider, Utts, and Wallach's meta-analysis of a pool of 36 DMILS studies (EDA only), comprised of 1,015 single sessions, conducted between 1977 and 2000, produced a highly significant effect size, Cohen's *d* = .106 ($p = .001$).[82] This meta-analysis included a subset of 15 remote-staring studies. Remote staring is another form of DMILS. The standard design requires two participants — one sees the other on live video at randomly-selected times. During those times, the viewer stares at the participant on the screen aiming to activate the target-person's nervous system. During times when the screen is blank, the staring participant rests. The participant being stared at has his/her EDA monitored the whole time. In 15 remote staring studies (379 trials), the mean effect size was significant, Cohen's *d* = 0.128 ($p = .013$).

The Schmidt *et al.* meta-analysis also included results from an analysis of studies on Attention Focusing Facilitation, whereby receivers ('helpees') focus their minds on a candle placed directly in front of them; they press a button every time their attention drifts, while a remote helper in another room helps, or does not help, the helpee concentrate according to a signal to do either.[83] The frequencies of button presses during helping and non-helping (control) periods are compared. Eleven studies produced a total of 576 trials; the mean effect size was significant, Cohen's *d* = 0.114 ($p = .029$).

Dice-Throwing

The dice-throwing experiment is one of a number of experiments designed to test whether consciousness can influence physical systems at the 'macro' ('greater than molecular') level. Radin and Ferrari examined dice-throwing studies spanning more than 50 years (1935 to 1987).[84] There were 148 experimental studies and 31 control studies considered. A total of 2,500 participants attempted to influence 2.6 million dice throws.

Forty-four percent of the 148 experimental studies gave results significant at the 5% level. The weighted mean effect size for the experimental studies was 0.012, which was '19 standard errors from chance'.[85] The control studies' weighted mean, however, was a low 0.00093, which was within one standard error from chance.[86] The combined Stouffer Z for the experimental studies was 18.20, but the control studies gave a low 0.18. The fail-safe N was 17,974 (121 nonsignificant studies to every one significant study).

Given that die faces are rarely equal in mass due to scooping out of the die face to mark the numbers, biases would have existed in many of the 148 studies. Radin and Ferrari took into consideration the fact that only 69 studies used protocols where targets were evenly balanced among all six die-faces.[87] A conservative quality-weighted effect size of 0.007 was calculated ($Z = 7.62$, $p = 1.30 \times 10^{-14}$). Eliminating the outlier studies that contributed to the heterogeneity of the database resulted in a database of 59 studies with an even more conservative, but still significant, quality-weighted effect size of .003 ($Z = 3.19$, $p = 7.16 \times 10^{-4}$).

Radin and Ferrari found no evidence that the overall effect size was due to a 'few exceptional investigators'.[88] Of note was their finding that methodological quality improved over time, but they also found, in a first analysis, that quality-rating correlated negatively and significantly with effect size, suggesting that design flaws present in low quality studies were contributing to the success of earlier experiments. However, analysis of a homogeneous subset of the original database (from which outliers were removed) found no suggestive evidence for a possible 'regression to the mean effect' in the 'perfect' dice-experiment. The general conclusion, based on the 'homogeneous subset of balanced protocol studies', was that, if not strong, the mean effect size for the dice-throwing experiments was still significant and consistent over time, indicating a 'genuine mental . . . intention effect on dice'.[89]

Micro-PK (Random Number Generators)

Paranormal influence on physical systems at the 'micro' level can be tested experimentally using random number generators (RNGs), or random event generators (REGs). These machines are like electronic 'coin-flippers.' RNG experiments are designed to test the hypothesis that 'the statistical output of an electronic RNG is correlated with observer intention in accordance with prespecified instructions'.[90] Radin and Nelson conducted an initial meta-analysis with an accumulated database of 591 studies, the mean effect size of which proved significant. They also found a significant mean effect ($\sim 3.00 \times 10^{-4}$) for a homogeneous set of 490 quality-weighted experiments.

Fourteen years later, Radin and Nelson updated that meta-analysis, accumulating a total of 515 RNG studies (423 published up to and including 1987; 92 published after 1987).[91] Overall mean z score was small, at 0.17. Given that result, and noting that the Stouffer Z value is extremely large (3.81) with a correspondingly small p value (6.94×10^{-5}), the meta-analytic evidence for mind-matter interaction using RNGs has been consistent over this 44-year period (1959 to 2003).

It should be pointed out that Bösch, Steinkamp and Boller undermined the claim of an anomalous effect in the RNG paradigm, claiming that it was an artifact of publication bias.[92] But Bösch *et al.* assumed that effect size is entirely independent of sample size. Thus, Radin, Nelson, Dobyns, and Houtkooper argued that effect size is not entirely independent of sample size,[93] indicating there was no likely evidence of selective reporting.[94]

Dream-ESP

Dream-ESP is paranormal communication in an altered state of consciousness (ASC) commonly known as dreaming. This state is considered particularly conducive to psi because consciousness is reduced — in a strong sense it resembles the state elicited in the ganzfeld condition because stimulation from all the sensory modalities is considerably reduced, or even blocked completely, due to decreased activity of the reticular formation. The dream state thus may enable the psi signal the best possible chance of being detected above sensory noise.

In 1960, Montague Ullman was one of the first to conduct serious dream-ESP research using medium Eileen Garrett.[95] Pictures were used as target sets, from which one picture was telepathically sent. After some successful trials using this method, Ullman set up a sleep laboratory at Maimonides Medical Center in New York.

The Maimonides lab was particularly productive with 379 dream-ESP sessions conducted between 1966 and 1973 inclusive. [96] Independent judges were used to evaluate the reports from dreamers so that judges' ratings could be used to ascertain degree of correspondence with the target material, but participants also rated their attempts to identify targets.

Results from this period were mixed and complex, involving a number of researchers with different methodologies, different statistical testing procedures, and different goals. Some researchers tested general extrasensory perception

(where it was not possible to discern the psi modality), while others tested precognitive dreaming or clairvoyant dreaming.

Sherwood and Roe published a meta-analytic review of Maimonides dream-ESP studies conducted at the Maimonides Dream Lab (MDL). They found 15 MDL studies.[97] Effect sizes ranged from -0.22 to 1.10, with a mean effect size of 0.33.[98] They also published a meta-analytic review of dream-ESP studies conducted since the closure of the Maimonides lab (post-MDL). They found 21 post-MDL studies.[99] Effect sizes ranged from -0.49 to 0.80, with a mean effect size of 0.14.[100]

Results bode well for this second set of dream-psi studies, though the mean effect size for the typical post-Maimonides dream-psi study is not as strong compared to the typical MDL study. Sherwood and Roe argued that this performance difference may be due to 'procedural differences, including that post-Maimonides receivers tended to sleep at home and were generally not deliberately awakened from REM sleep'.[101]

Other possible weaknesses in the post-Maimonides studies included a tendency to use participant judging whereas the Maimonides series mostly used independent judging. Sherwood and Roe are of the opinion that independent judges are probably better skilled at judging dream material purely by 'aptitude or through experience',[102] whereas participants are usually naïve. Also, the majority of Maimonides studies investigated telepathy whereas the majority of post-Maimonides studies investigated clairvoyance. Finally, the Maimonides studies featured in their procedures targets that had emotional themes, and were noted for their 'vividness, colour, and simplicity',[103] whereas post-Maimonides studies used neutral targets. In an updated review, Sherwood and Roe found another seven more non-MDL studies bringing the total number of studies to 28, but with a slightly lower mean effect size of 0.11.[104]

Storm, Sherwood, Roe, Tressoldi, Rock, and Di Risio reported meta-analytic results on experimental dream-ESP studies for the period 1966 to 2014.[105] Studies fell into two categories: the MDL studies (n = 14), and independent (non-MDL) studies (n = 36). Though the databases were constructed with more critical selection, inclusion, and exclusion criteria prevailing, the effect size was the same for the MDL dataset (mean effect size = 0.33); and the non-MDL studies yielded a mean effect size of 0.14 (matching Sherwood and Roe's earlier finding, and similar to Sherwood and Roe's updated post-MDL database effect size of 0.11). The difference between the two mean values was not significant. A homogeneous dataset (N = 50) yielded a mean z of 0.75 (mean effect size of 0.20), with corresponding significant Stouffer Z = 5.32 (p = 5.19×10^{-8}), suggesting that dream content can be used to identify target materials correctly and more often than would be expected by chance.

Storm et al., found that significant improvements in the quality of the studies were not related to effect size, but effect size did decline over the 49-year period. Bayesian analysis of the same homogeneous dataset yielded results supporting the 'frequentist' finding that the null hypothesis should be rejected. Storm *et al.* conclude that the dream-ESP paradigm in parapsychology is worthy of continued investigation.

Sheep-Goat Effect

Schmeidler introduced the term 'sheep' to describe a person who believes in the possibility of ESP under given experimental conditions, and 'goats' as those who reject this possibility.[106] Lawrence conducted a meta-analysis on studies that included measures of the sheep-goat effect (SGE), which refers to the consistent finding that sheep score significantly better than goats on paranormal tasks.[107] Lawrence confined his search to forced-choice ESP studies, and his search covered the period 1947 to 1993.

From a total pool of 73 studies (4,500 participants, 685,000 guesses), Lawrence calculated an effect size of 0.03, with a highly significant Stouffer Z = 8.17 (p = 1.33×10^{-16}). The mean effect size per investigator was also 0.03. Eighteen studies (24%) showed a significant SGE. Lawrence also found that study quality and effect size had not changed in 46 years. The file-drawer estimate was 1,726 (23 unreported, nonsignificant studies for every one successful study).

Presentiment Effect

The presentiment effect is essentially a reversal of a standard psychological effect — where we may usually expect a response to *follow* a stimulus, the presentiment literature suggests a psychophysiological response may *precede* a stimulus. Thus, the presentiment effect can be regarded as a form of predicting the future. Mossbridge, Tressoldi, and Utts describe two paradigms:[108]

(1) randomly ordered presentations of arousing vs. neutral stimuli... participants are shown, for example, a randomly inter-mixed series of violent and emotionally neutral photographs on each trial, and there is no *a priori* way to predict which type of stimulus will be viewed in the upcoming trial.

(2) guessing tasks for which the stimulus is the feedback about the participant's guess (correct vs. incorrect)... on each trial participants are asked to predict randomly selected future stimuli (such as which of four cards will appear on the screen) and once they have made their prediction, they then view the target stimulus, which becomes feedback for the participant.

In their meta-analysis, Mossbridge *et al.* looked at 26 reports published between 1978 and 2010 that tested different randomly presented stimuli that produced different 'post-stimulus physiological activity'.[109] Variables of interest were EDA, heart rate, blood volume, pupil dilation, electroencephalographic activity, and blood oxygenation level dependent (BOLD) activity.

Analyses yielded a significant but small effect of 0.21 calculated as both a 'fixed effect' (where it is assumed that effect size is the same for all studies), and a 'random effect' (where it is assumed that effect sizes differ across studies, and are sampled from a distribution of differing effect sizes). The number of contrary unpublished reports that would be necessary to reduce the level of significance to chance was conservatively calculated to be 87.

Mossbridge *et al.* explored alternative (non-paranormal) explanations for the effects, such as sensory cueing and expectation bias, but these were ruled out. They advised that similar studies must be conducted across a number of laboratories and experimenters before replicability can be generalized.

Hypnosis/Comparison Condition ESP

Hypnosis has long been associated with psi, and the review of the literature by Dingwall showed an association between hypnosis and paranormal events, including ESP performance.[110] Twenty-five studies that tested hypnosis and comparative conditions (controls) for their effects on ESP performance, were meta-analysed by Stanford and Stein. [111] A significant unweighted Stouffer Z score of 8.77 ($p < 10^{-16}$) was found for the 25 studies, whereas for comparison conditions the effect was not significant, Stouffer $Z = 0.34$ ($p = .367$). Effect size was rather small for hypnosis ($\pi = 0.52$, where MCE = 0.50), but was essentially at chance for comparative conditions ($\pi = 0.51$).

While the hypnotic state appears conducive to psi performance (judging from the cumulative Z values), further statistical analysis showed a tendency toward psi-hitting among both hypnotic participants and the comparative condition participants when consideration was given to the chief investigator. Some investigators were better than others at inducing an effective hypnotic state. Nevertheless, Stanford and Stein reached the conclusion that hypnosis, generally speaking, may still enhance psi performance, as long as the expectations of the investigator, and the skill and personal attributes necessary in the participants and the investigator, are present or can be implemented in the experimental situation.[112]

Conclusion

The above reviews are not comprehensive but can be regarded as better than representative. While the major domains have been featured a few domains, such as remote viewing and the Global Consciousness Project, have not been thoroughly meta-analysed — these domains (perhaps up-and-coming in some cases) may as yet be represented by too few studies to warrant meta-analysis, or they have been subsumed by other domains (for instance, RV in free-response), or they do not lend themselves to meta-analytic treatment (for instance, GCP which is a single albeit worldwide project).[113] For most domains, experimentation continues, with experimental designs becoming increasingly sophisticated and innovative.

Lance Storm

Literature

Bandura, A. (1978). On paradigms and recycled ideologies. *Cognitive Therapy and Research*, *2*, 79-103.

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

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

Bem, D. J., Palmer, J., & Broughton, R. S. (2001). Updating the ganzfeld database: A victim of its own success? Journal of Parapsychology, 65(3), 207-218.

Bösch, H., Steinkamp, F., & Boller, E. (2006). Examining psychokinesis: The interaction of human intention with random number generators. *Psychological Bulletin*, *132*, 497-523.

Braud, W. G., & Schlitz, M. A. (1991). Consciousness interactions with remote biological systems: Anomalous intentionality effects. *Subtle Energies*, *2*, 1-46.

Braud, W. G., Wood, R., & Braud, L. W. (1975). Free-response GESP performance during an experimental hypnagogic state induced by visual and acoustic ganzfeld techniques: A replication and extension. *Journal of the American Society for Psychical Research*, *69*, 105-113.

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

Burdick, D. S., & Kelly, E. F. (1977). Statistical methods in parapsychological research. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 81-130). New York: Van Nostrand Reinhold.

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist*, *33*, 517.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. London: Sage.

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

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

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

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

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

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

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

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

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

Milton, J. (1998). A meta-analysis of waking state of consciousness, free response ESP studies. In N. L. Zingrone, & M. J. Schlitz (Eds.). *Research in parapsychology 1993* (pp. 31-34). Lanham, MD: Scarecrow Press.

Milton, J., & Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin*, *125*, 387-391.

Milton, J., & Wiseman, R. (2001). Does psi exist? Reply to Storm and Ertel (2001). *Psychological Bulletin*, *127*, 434-438.

Mossbridg, J., Tressoldi, P. & Utts, J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: a meta-analysis. Frontiers in Psychology, 3, 1-18. doi: 10.3389/fpsyg.2012.00390

Nelson, R. D. (2015). Implicit physical psi: The global consciousness project. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Eds.). *Parapsychology: A handbook for the 21st century* (pp. 282-292). Jefferson, NC: McFarland.

Oakes, M. (1986). *Statistical inference: A commentary for the social sciences*. Chichester, UK: John Wiley.

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

Radin, D. I., & Nelson, R. D. (1989). Evidence for consciousness-related anomalies in random physical systems. *Foundations of Physics*, *19*,

1499-1514.

Radin, D. I., & Nelson, R. D. (2003). Research on mind-matter interaction (MMI): Individual intention. In W. B. Jones & C. C. Crawford (Eds.), *Healing, intention, and energy medicine: Research and clinical implications* (pp. 39-48). Edinburgh, Scotland: Churchill Livingstone.

Radin, D. I., & Ferrari, D. C. (1991). Effects of consciousness on the fall of dice: A meta-analysis. *Journal of Scientific Exploration*, *5*, 61-83.

Radin, D. I., Nelson, R. D., Dobyns, Y., & Houtkooper, J. (2006). Reexamining psychokinesis: Comment on Bösch, Steinkamp & Boller (2006). *Psychological Bulletin*, *132*, 529-532.

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

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

Schmeidler, G. R. (1943). "Predicting Good and Bad Scores in a Clairvoyance Experiment: A Preliminary Report." *Journal of the American Society for Psychical Research*, 37: 103-110.

Schmidt, S., Schneider, R., Utts, J., & Wallach, H. (2004). Distant intentionality and the feeling of being stared at: The meta-analyses. *British Journal of Psychology*, *95*, 235-247.

Shapiro, D. A., & Shapiro, D. (1977). The 'double standard' in evaluation of psychotherapies. *Bulletin of the British Psychological Society*, *30*, 209-210.

Sherwood, S. J., & Roe, C. A. (2003). A review of dream ESP studies conducted since the Maimonides dream ESP studies. In J. Alcock, J. Burns, & A. Freeman (Eds.), *Psi wars: Getting to grips with the paranormal* (pp. 85-109). Thorverton, UK: Imprint Academic.

Sherwood, S. J., & Roe, C. A. (2013). An updated review of dream ESP studies conducted since the Maimonides dream ESP program. In S. Krippner, A. J. Rock, J. Beischel, & H. Friedman (Eds.), *Advances in parapsychological research 9* (pp. 38-81). Jefferson, NC: McFarland.

Smith, M., & Glass, G. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*, 752-760.

Stanford, R. G. (1979). The influence of auditory ganzfeld characteristics upon free-response ESP performance. *Journal of the American Society for Psychical Research*, *73*, 253-272.

Stanford, R. G., & Stein, A. G. (1994). A meta-analysis of ESP studies contrasting hypnosis and a comparison condition. *Journal of Parapsychology*, *58*, 235-269.

Steinkamp, F. (2005). "Forced-Choice ESP Experiments: Their Past and Their Future. In M. A. Thalbourne & L. Storm (Eds.), *Parapsychology in the Twenty-First 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.

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

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

International Journal of Dream Research: Psychological Aspects of Sleep and Dreaming, 10(2), 120-140.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2010a). A meta-analysis with nothing to hide: Reply to Hyman (2010). *Psychological Bulletin*, *136*(4), 491-494. doi:10.1037/a0019840.

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

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

Storm, L., Tressoldi, P. E., & Utts, J. (2013). Testing the Storm *et al.* (2010) meta-analysis using Bayesian and Frequentist approaches: Reply to Rouder *et al.* (2013). *Psychological Bulletin*, *139*(1), 248-254. doi:10.1037/a0029506.

Targ, R. (1996). Remote viewing at Stanford Research Institute in the 1970s: A memoir. *Journal of Scientific Exploration*, *10*, 77-88.

Targ, R., & Puthoff, H. (1974). Information transmission under conditions of sensory shielding. *Nature*, *251*, 602-607.

Thalbourne, M. A. (2003). *A glossary of terms used in parapsychology* (2nd Ed.). Charlottesville, VA: Puente.

Ullman, M., & Krippner, S. with Vaughan, A. (1973). *Dream telepathy: Experiments in nocturnal ESP*. Jefferson, NC: McFarland.

Ullman, M., Krippner, S., & Vaughan, A. (1974). *Dream telepathy*. Baltimore, MD: Penguin Books.

Utts, J. (1991). Replication and meta-analysis in parapsychology. *Statistical Science*, *6*, 363-378.

Varvoglis, M., & Bancel, P. A. (2015). Micro-psychokinesis. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Eds.). *Parapsychology: A handbook for the 21st century* (pp. 266-281). Jefferson, NC:

Williams, B. (2011). Revisiting the Ganzfeld ESP Debate: A Basic Review and Assessment. *Journal of Scientific Exploration*, 25(4), pp. 639-661.

References

Footnotes

1.ˆ Utts, 1991, p. 363.
2.ˆ Smith & Glass, 1977.
3.ˆ Bandura, 1978; Eysenck, 1978; Oakes, 1986; Shapiro & Shapiro, 1977.
4.ˆ Rosenthal, 1984, p. 127
5.ˆ Broughton, 1991, p. 283.
6.ˆ Hyman, 1985.
7.ˆ Bem & Honorton, 1994, p. 6; see also Honorton, 1985, p. 66.
8.ˆ Broughton, 1991, p. 286; Utts, 1991, pp. 370, 372, 375-376.
9.ˆ Thalbourne, 2003, p. 44.
10.ˆ Burdick & Kelly, 1977, p. 109.
11.ˆ Thalbourne, 2003, p. 45.
12.ˆ Braud, Wood, & Braud, 1975; Honorton & Harper, 1974; Parker, 1975.
13.ˆ see Thalbourne, 2003, p. 45; Utts, 1991, p. 369.
14.ˆ see Stanford, 1979, p. 253.
15.ˆ Hyman, 1985.
16.ˆ Hyman, 1985, p. 4.
17.ˆ Hyman, 1985, p. 5.
18.ˆ Hyman, 1985, p. 16.
19.ˆ Honorton, 1985.
20.ˆ Hyman, 1985, pp. 30-35.
21.ˆ Hyman & Honorton, 1986.
22.ˆ Hyman & Honorton, 1986, p. 351.

23.ˆ Utts, 1991, p. 371.
24.ˆ see Honorton, Berger, Varvoglis, Quant, Derr, Schechter, & Ferrari, 1990.
25.ˆ Bem & Honorton, 1994.
26.ˆ Honorton, 1985.
27.ˆ Honorton Honorton, Berger, Varvoglis, Quant, Derr, Schechter, & Ferrari, 1990.
28.ˆ Honorton Honorton, Berger, Varvoglis, Quant, Derr, Schechter, & Ferrari, 1990, p. 99.
29.ˆ Milton & Wiseman, 1999.
30.ˆ Milton & Wiseman, 1999, p. 388.
31.ˆ Milton & Wiseman, 1999, p. 388.
32.ˆ Unless otherwise stated, all effects sizes are calculated using the mean z score and the sample size: $z/\sqrt{n}$.
33.ˆ Milton & Wiseman, 1999, p. 391.
34.ˆ Storm & Ertel, 2001.
35.ˆ Honorton, 1985.
36.ˆ Bem & Honorton, 1994.
37.ˆ Milton & Wiseman, 1999.
38.ˆ Milton & Wiseman, 2001, p. 434.
39.ˆ Storm & Ertel, 2001, p. 430.
40.ˆ Bem, Palmer, & Broughton, 2001.
41.ˆ Storm, Tressoldi, & Di Risio, 2010b.
42.ˆ Hyman, 2010.
43.ˆ Hyman, 2010, p. 486.
44.ˆ Storm, Tressoldi, & Di Risio, 2010a.
45.ˆ Williams, 2011, p. 648.
46.ˆ Bem & Honorton, 1994.
47.ˆ Storm, Tressoldi, & Utts, 2013.
48.ˆ Rouder, Morey, & Province, 2013.
49.ˆ Storm, Tressoldi, & Di Risio, 2010b.
50.ˆ Rouder, Morey, & Province, 2013, p. 241.
51.ˆ Rouder, Morey, & Province, 2013, p. 241.
52.ˆ Storm, Tressoldi, & Utts, 2013, p. 252.
53.ˆ Milton, 1998.
54.ˆ Thalbourne, 2003, p. 107.
55.ˆ Storm, Tressoldi, & Di Risio, 2010b, p. 478.
56.ˆ Storm, Tressoldi, & Di Risio, 2010b, p. 476.
57.ˆ Targ & Puthoff, 1974.
58.ˆ Targ, 1996.
59.ˆ For example, see Dunne, Dobyns, & Intner, 1989; Dunne, Jahn, & Nelson, 1983; Jahn & Dunne, 1987; Schlitz & Gruber, 1981; Schlitz & Haight, 1984.
60.ˆ Baptista, Derakhshani, & Tressoldi, 2015, p. 202.
61.ˆ Baptista, Derakhshani, & Tressoldi, 2015.
62.ˆ Milton, 1997.
63.ˆ Baptista, Derakhshani, & Tressoldi, 2015, p. 203.
64.ˆ Thalbourne, 2003, p. 44.
65.ˆ Honorton & Ferrari, 1989.
66.ˆ Utts, 1991, p. 374.
67.ˆ Rosenthal, 1984.
68.ˆ Steinkamp, Milton, & Morris, 1998.
69.ˆ Steinkamp, Milton, & Morris, 1998, p. 193.
70.ˆ Steinkamp, Milton, & Morris, 1998.
71.ˆ Steinkamp, Milton, & Morris, 1998, p. 209.
72.ˆ Storm, Tressoldi, & Di Risio, 2010b, p. 475.
73.ˆ Steinkamp, 2005.
74.ˆ Steinkamp, 2005, p. 155.
75.ˆ Storm, Tressoldi, &Di Risio, 2012.
76.ˆ Braud & Schlitz, 1991.
77.ˆ Braud & Schlitz, 1991, p. 2.
78.ˆ Braud & Schlitz, 1991, p. 3.

79.ˇ Braud & Schlitz, 1991, p. 5.

80.ˇ Braud & Schlitz, 1991, pp. 31-34.

81.ˇ Braud & Schlitz, 1991, p. 41.

82.ˇ Schmidt, Schneider, Utts, & Wallach, 2004.

83.ˇ Schmidt, Schneider, Utts, & Wallach, 2004.

84.ˇ Radin & Ferrari, 1991.

85.ˇ Radin & Ferrari, 1991, p. 79.

86.ˇ Radin & Ferrari, 1991, p. 79.

87.ˇ Radin & Ferrari, 1991, pp. 74-76.

88.ˇ Radin & Ferrari, 1991, p. 68.

89.ˇ Radin & Ferrari, 1991, pp. 79-80.

90.ˇ Radin & Nelson, 1989, p. 1502.

91.ˇ Radin & Nelson, 2003.

92.ˇ Bösch, Steinkamp & Boller, 2006.

93.ˇ Radin, Nelson, Dobyns, & Houtkooper, 2006.

94.ˇ see also, Varvoglis & Bancel, 2015, for a supporting argument.

95.ˇ Ullman, Krippner, & Vaughan, 1974

96.ˇ Ullman, Krippner with Vaughan, 1973.

97.ˇ See Table 1 in Sherwood & Roe, 2003, p. 89.

98.ˇ Sherwood & Roe, 2003, pp. 88, 104.

99.ˇ See Table 2 in Sherwood & Roe, 2003, p. 94.

100.ˇ Sherwood & Roe, 2003, pp. 103, 104.

101.ˇ Sherwood & Roe, 2003, p. 85.

102.ˇ Sherwood & Roe, 2003, p. 105.

103.ˇ Sherwood & Roe, 2003, p. 106.

104.ˇ Sherwood & Roe, 2013.

105.ˇ Storm, Sherwood, Roe, Tressoldi, Rock, & Di Risio, 2017.

106.ˇ Schmeidler, 1943.

107.ˇ Lawrence, 1993.

108.ˇ Mossbridge, Tressoldi, & Utts, 2012, p. 1.

109.ˇ Mossbridge, Tressoldi, & Utts, 2012, p. 1.

110.ˇ Dingwall, 1967; cited in Stanford & Stein, 1994, p. 235.

111.ˇ Stanford & Stein, 1994.

112.ˇ Stanford & Stein, 1994, pp. 260-261.

113.ˇ Nelson (2015) gives a very impressive example of the GCP effect with a database of 461 events/happenings that have produced significant cumulative deviations from random behaviour in RNGs, but there are no other independent teams running similar projects on a global scale.