Australian Journal of Parapsychology
2006, Volume 6, Number 2, pp. 135–155

Meta-Analysis in Parapsychology: II. Psi Domains other than Ganzfeld

LANCE STORM[1]

ABSTRACT: The present article completes the two-part review on meta-analyses in parapsychology (for Part I, see L. Storm, 2006). The reviewed literature other than ganzfeld/autoganzfeld studies, includes meta-analyses on: (i) biological systems (DMILS), (ii) forced-choice ESP, (iii) free-response ESP, (iv) dice-throwing, (v) micro-PK (RNG), and (vi) dream-psi. Meta-analyses by T. R. Lawrence (1993), E. Haraldsson (1993), and R. G. Stanford and A. G. Stein (1994) are also reviewed. Results indicate that these meta-analyses provide considerable evidence that there is an anomalous effect in the field of parapsychology in need of an explanation. It is concluded that these and other meta-analyses in parapsychology have revealed significant non-zero effects across studies, although these tend to be rather small, but process-oriented research may further our understanding of these anomalies.

REVIEW OF META-ANALYSES OTHER THAN GANZFELD

The present article is a review of the meta-analytic literature for: (i) biological systems, (ii) forced-choice ESP, (iii) free-response ESP, (iv) dice-throwing, (v) micro-PK (RNG), and (vi) dream-psi. Also reviewed are meta-analyses by Lawrence (1993), Haraldsson (1993), and Stanford and Stein (1994). The article concludes with various comments on the meta-analyses by parapsychologists and researchers from other fields.

[1] The author thanks the Bial Foundation for support in the preparation of this article.

1. Biological Systems

Braud and Schlitz (1991) conducted a 13-year-long series of studies that looked at eight living target systems: (1) electrodermal activity, participant's *influence* on a target system's skin resistance, (2) electrodermal activity of a participant's *attention* away from target system, (3) ideomotor reactions (reactions associated with thought), (4) muscular tremor (measured by the movement of a hand-held metal stylus in a small aperture), (5) blood pressure, (6) fish orientation, (7) mammal locomotion (gerbil activity in a wheel), and (8) 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 "monitored systems' activities" (Braud & Schlitz, 1991, p. 2). Experiments in this domain were described as testing participants' "direct mental influence on [or interactions with] living systems" or DMILS (Braud & Schlitz, 1991, p. 3).

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, which was a percent measure of the "total activity that occurred in the prescribed direction during the entire set of influence (decremental or incremental aim) periods" (Braud & Schlitz, 1991, p. 5). 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, whereas the next lowest was 40 sessions, and the average number of sessions was over 65 (see Table 1 for results of individual studies). Meta-analysis of the eight studies showed a significant mean effect size (*ES*) of 0.178 ($Z = 3.79$, $p = 1.00 \times 10^{-4}$). Braud and Schlitz (1991, pp. 31-34) 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. All these 'explanations' were inapplicable.

The overall conclusion was that "effect[s] appear to occur in a 'goal-directed' manner" (p. 41) 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. Braud and Schlitz (1991) argued that participants' "intentionality" was the "key factor" that brought about the significant results":

> Maintaining a strong intention of a desired goal event, focusing attention upon the relevant aspect of the target system, and filling oneself with strong imagery of the desired biological activity are, under certain conditions, accompanied by a shift in the target system's activity in the intended direction. (p. 41)

Table 1
Summary of Results of Direct Mental Influence Experiments

Living target system	No. of sessions	Mean *z* score	Mean *ES*	Stouffer *Z*	*p*
Electrodermal activity (influence)	323	1.05	0.058	4.08	2.25×10^{-5}
Electrodermal activity (attention)	78	0.84	0.095	1.68	4.65×10^{-2}
Ideomotor reactions	40	1.72	0.272	2.98	1.44×10^{-3}
Muscular tremor	19	–0.42	–0.096	–0.59	7.22×10^{-1}
Blood pressure	41	1.35	0.210	1.91	2.81×10^{-2}
Fish orientation	40	1.88	0.297	3.78	7.84×10^{-5}
Mammal locomotion	40	1.90	0.300	3.81	6.92×10^{-5}
Rate of hemolysis	74	2.43	0.282	4.20	1.33×10^{-5}
All systems combined	655	1.34	0.178	7.72	2.58×10^{-14}

Note: Table adapted from Braud and Schlitz (1991, p. 30, *Table IX*).

More recently, Schmidt, Schneider, Utts, and Wallach's (2004) meta-analysis of a pool of 36 DMILS studies, comprised of 1,055 sessions, conducted between 1977 and 2000, produced a highly significant effect size, Cohen's $d = .11$ ($p < .001$).[2] 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 skin resistance (i.e., Electro-Dermal Activity, or EDA) monitored the whole time. In 15 remote staring

[2] Cohen's *d* is the standardised mean difference, which is the difference in the treatment and control group means divided by either the control group standard deviation (*SD*) or pooled *SD*.

studies, the mean effect size was Cohen's $d = 0.13$ ($p = .01$). Although the effect is consistent, the overall pool of studies is still small, as is the effect size, thus warranting further research in this area aimed at (a) reaching an understanding of the nature of the effect, and (b) increasing its size.

2. Forced-Choice (Precognition and Clairvoyance)

Forced-choice experiments require that the participant "guess a target which is one of a limited range of possibilities which are known to them in advance [such as in the card-guessing experiment]" (Thalbourne, 2003, p. 44). Forced-choice precognition[3] experiments from as early as 1935 up to 1987 were meta-analysed by Honorton and Ferrari (1989).

Honorton and Ferrari (1989) used only the studies where the procedure was to select the target "randomly after the subject [*sic*] had attempted to predict what it would be" (Utts, 1991, p. 374). 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 $z/\sqrt{n}$ formula was used as the measure of effect size (*ES*). The *ES* was 0.02 (mean $z = 0.65$, all studies).

Ninety-two studies (30%) showed significant hitting at the 5% level. When outlier studies contributing to the heterogeneity (inconsistency) of the database were removed the *ES* fell to 0.012 ($Z = 6.02$, $p = 1.10 \times 10^{-9}$). (See Table 2 for other results.) The "fail-safe *N*" (Rosenthal, 1984) was 14,268 studies, 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). Naïve participants, where Cohen's $h = 0.008$, did not perform as well as experienced participants, Cohen's $h = 0.051$.

Honorton and Ferrari also found 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. Unfortunately, Honorton and Ferrari (1989) felt it was "bad news" that the effect sizes had not increased over this time since the stable *ES* meant investigators had not "develop[ed] sufficient understanding of the conditions underlying the occurrence (or detection) of these effects to reliably increase their magnitude" (p. 295). However, Honorton and Ferrari's meta-analysis revealed that the largest effect sizes were found in experiments using (a) experienced participants, (b) independent testing (one

[3] Precognition is a form of ESP "in which the target is some future event that cannot be deduced from normally known data in the present" (Thalbourne, 2003, p. 90).

participant at a time) as opposed to group testing, and (c) trial-by-trial feedback—important factors for future researchers to consider.

Table 2
Meta-Analyses of Seven Psi Domains: DMILS, Forced-Choice, Free-Response, Dice-Throwing, RNG, Clairvoyance, Precognition

Domain meta-analysed and author(s)	No. studies (k) and period of analysis	Mean z score ($\Sigma z/k$)	Mean ES ($\Sigma[z/\sqrt{n}]/k$)	Stouffer Z ($\Sigma z/\sqrt{k}$)	p
DMILS (Braud & Schlitz, 1991)	8 (1979-1991)	1.34	0.178	3.79	7.53×10^{-5}
Forced-Choice (Honorton & Ferrari, 1989)	248 (1935-1987)	0.38	0.012[a]	5.98	1.12×10^{-9}
Free-Response (GESP/Remote Viewing) (Milton, 1998)	75 (1964-1993)	0.68	0.170[a]	5.89	1.93×10^{-9}
Dice-Throwing (Radin & Ferrari, 1991)	59 (1935-1987)	0.42	0.003[b]	3.23	6.19×10^{-4}
RNG (Radin & Nelson, 2003)	515[c] 1959-2003)	0.17	n/a	3.81	6.94×10^{-5}
Clairvoyance (Steinkamp et al., 1998)	22 (1935-1997)	0.60	0.009	2.81	2.48×10^{-3}
Precognition (Steinkamp et al., 1998)	22 (1935-1997)	1.02	0.010	4.78	8.76×10^{-7}

Note: The most conservative values were used in the table. DMILS = Direct Mental Influence on Living Systems; *ES* = Effect Size; GESP = General Extra-Sensory Perception; RNG = Random Number Generator (note that *ES* values are not given due to extremely wide ranging number of event-counts from a mere $N = 312$ events to a staggering $N = 3.90 \times 10^8$).

[a] Calculated from a homogeneous (consistent) data set.

[b] Calculated from a quality-weighted, homogeneous dataset of 'balanced' studies (fair dice only).

[c] PEAR lab data collapsed into one data-point; 176 additional studies included since Radin and Nelson's (1989) meta-analysis.

A later study by Steinkamp, Milton, and Morris (1998) meta-analysed forced-choice studies for the period 1935-1997, while at the same

time comparing clairvoyance[4] 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" (p. 193).

Steinkamp et al. (1998) used a total of 22 comparable study-pairs in their meta-analysis, where procedures were effectively the same in both types of studies. Effect sizes for precognition and clairvoyance were almost identical (see Table 2). 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" (p. 209).

Steinkamp's (2005) comprehensive review of forced-choice studies from 1880 to 1989 considered various predictors and psi-conducive variables. She noted that "there are few variables that have correlated clearly with success" (p. 155) and she was rather critical of the variations in study designs because these made it difficult to ascertain clear patterns due to conflicting outcomes. However, her Table 6.7 (Steinkamp, 2005, p. 156) shows that there has been strong evidence in the past that low-neuroticism, high extraversion, prior testing (i.e., pre-selection of participants), and trial-by-trial feedback, are the most "promising" and relevant variables in terms of yielding evidence for psi.

3. Free-Response (GESP/remote-viewing)

The free-response[5] experiment tests participants in a "normal, waking state of consciousness," thus differentiating it from the ganzfeld experiment (Milton, 1998, p. 31). Milton (1998) meta-analysed all available free-response studies published during the period 1964-1993 (these studies included remote viewing studies, in which the percipient "attempts to

[4] Clairvoyance is "paranormal acquisition of information concerning an object or contemporary physical event" (Thalbourne, 2003, p. 18).

[5] The free-response method describes any test of ESP using a relatively unlimited range of possible targets, thus permitting the participant to "respond freely with whatever impressions come to mind" (Thalbourne, 2003, p. 44). The participant may, for example, respond by drawing a pictorial representation of the target.

describe the surroundings of a geographically distant agent," Thalbourne, 2003, p. 107).

Milton found a mean effect size for 78 studies of .16 ($Z = 5.72$, $p = 5.33 \times 10^{-9}$) (see Table 2 for other results). 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 an *ES* of .17 ($Z = 5.85$, $p = 2.46 \times 10^{-9}$). Milton also wondered if the large effect sizes in the ganzfeld domain were due to the ganzfeld technique or to the free-response methodology itself as used in ganzfeld research. When she compared mean effect sizes between her database and Honorton's (1985a) ganzfeld database, there was no significant difference, $t(104) = 1.49$, $p > .05$. This result suggests that there is no basis as yet for assuming that the ganzfeld condition provided a superior method of eliciting psi compared with the standard free-response protocol.

4. 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 (1991) examined dice-throwing studies spanning more than 50 years (1935 to 1987). 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 *ES* for the experimental studies was 0.012, which was "19 standard errors from chance" (Radin & Ferrari, 1991, p. 79). The control studies' weighted mean, however, was a low 0.00093, which was within one standard error from chance (p. 79). 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). (See Table 2 for other results.)

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 (1991, pp. 74-76) took into consideration the fact that only 69 studies used protocols where targets were evenly balanced among all six die-faces. A conservative quality-weighted *ES* 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 (see Table 2) with an even more conservative, but still significant, quality-weighted *ES* of .003 ($Z = 3.19$, $p = 7.16 \times 10^{-4}$).

Radin and Ferrari (1991, p. 68) found no evidence that the overall effect size was due to a "few exceptional investigators." Of note was their finding that methodological quality improved over time, but they also found, in a first analysis, that quality 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" (Radin & Ferrari, 1991, pp. 79-80).

It is important to note that, after RNG experiments (reviewed next), dice-throwing experiments produce the weakest effect sizes on average of all domains. This fact must be seriously considered by investigators wishing to conduct research in the areas of micro-PK since the probability of a 'pseudo-failure to replicate' (i.e., Type II error) would be high.

5. Micro-PK (Random Number Generators)

Paranormal influence on physical systems at the 'micro' level can be tested experimentally using random number generators (RNGs), which are also referred to as 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" (Radin & Nelson, 1989, p. 1502). Radin and Nelson (1989) conducted an initial meta-analysis with an accumulated database of 490 studies.

Fourteen years later, Radin and Nelson (2003) updated that meta-analysis, and consequently they accumulated 515 RNG studies—423 published up to and including 1987, and 92 published after 1987. Overall mean z score was small (0.17; see Table 2). Given that result, and noting that the Stouffer Z value is extremely large with a correspondingly small p-value, 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 a recent finding in *Psychological Bulletin* by Bösch, Steinkamp and Boller (2006) undermined the claim of an anomalous effect in the RNG paradigm, claiming that it was an artifact of publication bias. But Bösch et al. assumed that effect size is entirely independent of sample size. Thus, in the same issue of *Psychological*

Bulletin, Radin, Nelson, Dobyns, and Houtkooper (2006) argued that effect size is not entirely independent of sample size, indicating there was no likely evidence of selective reporting.

Although replicability and robustness have been demonstrated in these meta-analyses, statistical power must be considered in RNG experiments since effect sizes are shown to be very small. In fact, RNG experiments to date represent the domain that delivers the lowest mean *z* score of all domains meta-analysed (again see Table 2). However, at an earlier date, Broughton (1991, p. 290) stressed the point that results such as these were still dramatic evidence that consciousness could have an effect on micro-systems, and his words are just as true today. In conclusion, the extreme odds against the RNG effect sizes being the result of chance represent the most convincing evidence so far (after dice-throwing) for an anomalous effect *of the psychokinetic kind*.

6. Dream-Psi

Dream-psi 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-psi research using medium Eileen Garrett. Ullman set up a dream laboratory with Karlis Osis and Douglas Dean. Three pictures from *Life* magazine were used as a target set, from which one picture was telepathically sent to Garrett while she slept in the lab. 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-psi sessions conducted between 1966 and 1973 inclusive (Ullman, Krippner with Vaughan, 1973). 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 GESP, while others tested precognitive dreaming or clairvoyant dreaming. Results for independent judges' ratings and rankings are given in Table 3.

Table 3
Summary of Maimonides Dream Studies[6]

Series	Raw Score		*z* score
	Hit	Miss	
GESP (dreams monitored and recorded throughout night; agent "transmitting" during each REM period):			
A. 1st screening—Ullman et al. (1966)	7	5	0.71[a]
B. 1st Erwin—Ullman et al. (1966)	5	2	2.53[a]
C. 2nd screening—Ullman (1969)	4	8	–0.25[a]
D. Posin—Ullman (1969)	6	2	1.05[b]
E. Grayeb—Ullman et al. (1973)	3	5	–0.63[b]
F. 2nd Erwin—Ullman & Krippner (1969)	8	0	2.49[c]
G. Van de Castle—Krippner & Ullman (1969)	6	2	2.06[c]
H. Pilot sessions	53	14	4.20[a]
Precognition (Dreams monitored and recorded throughout night; target experience next day):			
I. 1st Bessent—Krippner et al. (1971)	7	1	2.06[c]
J. 2nd Bessent—Krippner et al. (1972)	7	1	1.84[c]
K. Pilot sessions	2	0	0.67[b]
GESP (Dreams monitored and recorded throughout night; agent active only at beginning or sporadically):			
L. Sensory bombardment—Krippner et al. (1971)	6	0	3.11[a]
M. Grateful Dead—Krippner et al. (1973)	7	5	0.61[b]
Clairvoyance (Dreams monitored and recorded throughout night: concealed target known to no one):			
N. Pilot sessions	5	3	0.98[a]
GESP (Single dreams):			
O. Vaughan, Harris, Parise—Honorton et al. (1972)	105	98	0.63[b]
Total	231[d]	146	22.03[e]

Note: GESP = General Extra-Sensory Perception.
[a] Rankings; [b] Score (count of hits and misses); [c] Original ratings were *t* scores—these were converted to *z* scores from *r* values calculated by Sherwood and Roe, 2003; [d] Total hits = 61%; [e] Mean $z = 1.47$ (Stouffer $Z = 5.69$).

[6] Table 3 is adapted from the table in Child (1985).

The *Z* scores for 13 of 15 series deviated positively (87%), and 7 of these (47%) were significant at $p < .05$ (Series B, F, G, H, I, J, and L). For convenience, *z* scores for all 15 series were combined so that a single mean *z* score could be calculated. The result was a marginally significant mean $z = 1.47$ ($p = .071$); Stouffer $Z = 5.69$ ($p = 6.35 \times 10^{-9}$). Overall, Child (1985) concluded that chance could not explain the results, and he argued that there was "some systematic—that is, nonrandom—source of anomalous resemblance of dreams to targets" (Child, 1985, p. 122).

Table 4
Summary of Post-Maimonides Dream Studies[7]

Study	Type of ESP	Trials	*z* score
1. Child et al. (1977) Experiment 1 2. Child et al. (1977) Experiment 2	} Telepathy	13	2.55[a]
3. Kanthamani et al. (1988) Pilot	Clairvoyance	10	0.76[a]
4. Kanthamani & Khilji (1990)	Clairvoyance	20	1.70[a]
5. Kanthamani & Broughton (1992)	Clairvoyance	20	2.82[a]
6. Braud (1977) Pilot	Telepathy	50	–1.90
7. Braud (1977) Experiment 1	Telepathy	30	1.29
8. Sargent & Harley (1982)	Precognition	20	0.30
9. Harley (1989)	Clairvoyance	20	–2.19[a]
10. Markwick & Beloff (1983)	Clairvoyance/Telepathy	100	1.87
11. Markwick & Beloff (1988)	Clairvoyance/Precognition	100	–0.39
12. Hearne (1981)	Telepathy	—	0.00
13. Hearne (1987)	Telepathy	8	–0.39
14. Hearne (1989)	Telepathy	10	0.31
15. Dalton et al. (1999)	Clairvoyance	32	3.58
16. Sherwood et al. (2000)	Clairvoyance	28	1.44
17. Dalton et al. (2000)	Clairvoyance	16	2.35
18. Eppinger (2001)	Clairvoyance	50	–0.07
19. Roe et al. (2002)	Clairvoyance	31	0.80
20. Sherwood et al. (2002)	Precognition	12	–1.16
21. Weiner & McCain (1981)	Clairvoyance	12	1.97[a]
Total		582	15.64[b]

[a] Original *r* values are converted to *z* scores; [b] Mean z = 0.78 (Stouffer Z = 3.50).

[7] Table 4 adapted from Sherwood and Roe (2003, Table 2).

Sherwood and Roe (2003) published a meta-analytic review of dream-psi studies conducted since the closure of the Maimonides lab. They found 21 studies with a total of 582 trials (see Table 4).

The *Z* scores for 14 of 21 series deviated positively (67%), and 7 of these were significant (33%) at $p < .05$ (Studies [1 + 2], 4, 5, 10, 15, 17 and 21). The combined mean *z* score = 1.47 ($p = .218$); Stouffer $Z = 3.50$ ($p = 2.33 \times 10^{-4}$). While this Stouffer *Z* score bodes well for this second set of dream-psi studies, the mean *z* score suggests the typical post-Maimonides dream-psi study does not elicit psi. 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" (p. 85).

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 (2003) are of the opinion that independent judges are probably better skilled at judging dream material purely by "aptitude or through experience" (p. 105), 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" (p. 106), whereas post-Maimonides studies used neutral targets.

However, in spite of these procedural differences, the difference *statistically* between the two databases (i.e., in terms of mean effect sizes) is not significant, $t(33) = 1.69$, $p = .100$ (two-tailed). This result means that, in spite of the procedural differences between the two databases, we can assume that they belong to a homogeneous set, so that combining the two is justified. In other words, the slightly poorer performance of the post-Maimonides database can be attributed to chance fluctuation. The combined mean *z* score = 1.08 ($p = .140$); Stouffer $Z = 6.37$ ($p = 9.45 \times 10^{-11}$).

Again, these findings indicate that the 'average' dream-psi study is not likely to elicit evidence of psi, but when the collection is looked at over a 40-year period, there is an extremely strong indication of a correspondence between dream motifs and target material under laboratory conditions.

The Sheep-Goat Effect

Schmeidler (1943) 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. Lawrence (1993) looked at

the sheep-goat effect (SGE), which refers to the consistent finding that sheep score significantly better than goats on paranormal tasks (see also the reviews by Palmer, 1971, 1977). From the 73 forced-choice studies analysed (4,500 participants, 685,000 guesses), Lawrence calculated an *ES* of 0.03, with a highly significant Stouffer $Z = 8.17$ ($p = 1.33 \times 10^{-16}$; 18 studies [24%] showed a significant SGE at $p \leq .05$). The mean *z* was 0.96, and mean *ES* per investigator was 0.03. 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).

Lawrence (1993, p. 82) determined that, of the 65 studies that reported the order of administration of the belief scale and ESP task, there were 15 studies where the sheep-goat scale was given after the ESP task (*ES* = 0.04), and 50 studies where the scale was given before the ESP task (*ES* = 0.03). He thus found suggestive evidence of an artifact (an order effect), though the difference between effect sizes was not significant. Although the number of studies stating the kind of feedback (i.e., 'no feedback' and 'trial-by-trial feedback') given was "insufficient" (p. 81), Lawrence nevertheless advised that measures of belief be given before the ESP task, since outcome feedback from the ESP task could have a biasing effect on stated belief.

Religiosity, Belief in an Afterlife, Belief in Psychic Phenomena, and ESP

Haraldsson (1993) conducted a meta-analysis of eight forced-choice ESP experiments to determine correlates, if any, of three belief variables with ESP performance. The Icelandic Sheep-Goat Scale (Thalbourne & Haraldsson, 1980) was administered in six of the experiments. Seven of the experiments were tests of clairvoyance and precognition, and one was a test of precognition only.

The three belief variables were religiosity, belief in an afterlife (both of which correlated significantly with ESP performance), and belief in psychic phenomena (which failed to correlate significantly with ESP performance). The *z* value was highest for religiosity, second highest for belief in an afterlife, and lowest for belief in psychic phenomena (see Table 5).

A regression analysis revealed that religiosity was the only significant predictor of ESP performance. Haraldsson (1993) concluded: "religiosity might be a more efficient predictor of ESP performance than the traditional sheep-goat variable" (p. 270). However, Haraldsson added that inferences from these findings might be applicable only to Iceland's population. Religiosity has "differences in content" in Iceland compared to the USA, "which may be the reason for the less consistent relationship in

the US between religiosity and belief and reporting of psychic experiences than in Iceland" (Haraldsson, 1993, p. 270).

Table 5
Possible Predictors of ESP Performance (Haraldsson, 1993)

Predictor variable	*z*	*ES*	*p*
Religiosity (*n* = 383)	2.48	0.13	.007
Belief in an Afterlife (*n* = 382)	1.92	0.10	.028
Belief in Psychic Phenomena (*n* = 448)	0.73	0.04	.232

Note: *p* values are one-tailed

Hypnosis/Comparison Condition ESP

Hypnosis has long been associated with psi, and the review of the literature by Dingwall (1967; cited in Stanford & Stein, 1994, p. 235) showed an association between hypnosis and paranormal events, including ESP performance. Twenty-five studies that tested hypnosis and comparative conditions (controls) for their effects on ESP performance, were meta-analysed by Stanford and Stein (1994). An unweighted Stouffer *Z* score of 8.77 ($p < 10^{-16}$) was found for the 25 studies, whereas for comparison conditions, Stouffer $Z = 0.34$ ($p = .367$). Effect size was rather small for hypnosis ($\pi = 0.52$), 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 (1994, pp. 260-261) 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.

COMMENTS ON THE META-ANALYSES

Skeptics will continue to criticise the validity of the significant nonzero effect-sizes found in the meta-analyses, whilst many proponents of

psi may base their conviction that psi is real exclusively on the statistical evidence. Even so, the very low effect sizes may raise serious doubts from both skeptics and believers about the ultimate utility of psi. As it happens, the problem of low effect sizes is not unique to parapsychology. This problem is quite common in other fields (see the examples from the field of medicine in Greenhouse & Greenhouse, 1988, and Rosenthal, 1990). However, Greenhouse (1991) does not believe that "anyone's views regarding the existence of paranormal phenomena" will change (p. 386). He makes it clear that only until certain prejudices are dissolved can new paradigms be adopted. Greenhouse (1991, p. 367) cites the aspirin/heart-attack study used by Utts (1991, p. 367) to clarify this point. In spite of the very low effect size reported in this study, the relationship was easily accepted because the biological mechanism that explained this relationship is a causal process that is easily understood, even by the layperson. The point to be noted here is that parapsychology must focus on process studies that aim to explain the mechanism of psi, rather than just demonstrate its presence.

Hyman (1991, p. 392) yields to Utts's conclusion that there is an "anomaly in the parapsychological findings." His only reservation in this regard is that the effect sizes vary so much, both within experimental domains ("the effect sizes obtained by [Robert G.] Jahn are much smaller than those obtained by [Helmut] Schmidt with similar experiments on random number generators," Hyman, 1991, p. 392), and between domains (for example, meta-analysis shows that the RNG domain yields the lowest mean effect size, while the DMILS and ganzfeld domains yield the highest—see Storm, 2006). Hyman asks what it is we are supposed to account for when there are "nonchance departures from a statistical model" (p. 392) concomitant with such varied effect sizes. He uses an anecdotal and misleading astronomical example to illustrate his point (Uranus's anomalous orbit is explained as being due to another planet, which led to the discovery of Neptune), which falsely implies that non-parapsychological science can always give account of its anomalies. However, unaccounted anomalies abound in orthodox science where, in many cases, it is "empirical departures from expectation [that have] led to important findings or theoretical models" (Utts, 1991, p. 400). Parapsychologists have merely followed the example of the other scientific disciplines. Utts's (1991) response to Hyman is to propose "intensive investigation" as the means by which we may arrive at explanations for psi effects (p. 400).

Finally, Morris (1991) notes that meta-analysis allows the parapsychologist the opportunity of identifying the moderator variables that need to be incorporated into experiments in order to yield the strongest effects. For example, it appears to be better to test individuals rather than

groups, better to use dynamic rather than static targets, and better to use experienced participants rather than novices. Currently, though, there are doubts about the last two of these three recommendations (Milton & Wiseman, 1999).

CONCLUSION

Though meta-analysis is a generally accepted methodology, it does have its critics. Those critics of meta-analysis, and statistics in general, will assure us that statistical procedures can "mislead and be misused" (Diaconis, 1991, p. 386). There seems little doubt that such an assurance is based on fact, but to avoid such pitfalls a basic tenet of science has always been to proceed with caution. Nevertheless, the evidence to date from the meta-analyses reviewed in this article (and its companion article—see Storm, 2006) indicates that there are significant effect sizes for all the major experimental domains in parapsychology. These results suggest a very real (albeit statistical) anomaly worthy of continued investigation. Thus the field of parapsychology justifies itself as a discipline, and in the spirit of Rosenthal's (1986, p. 333) words—"at any point in time some judgment can be made"—it is hoped that this pair of review articles fosters a judgement in scientists, researchers, academics, and laypersons alike, that might help draw positive attention to the field, thereby allowing the opportunity of reaching a deeper understanding of the nature of psi.

REFERENCES

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. (1977). Long-distance dream and pre-sleep telepathy. In J. D. Morris, W. G. Roll, and R. L. Morris (Eds.). *Research in parapsychology 1976* (pp. 154-155). Metuchen, NJ: Scarecrow Press.

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

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

Child, I. L. (1985). Psychology and anomalous observations: The question of ESP in dreams. *American Psychologist*, *40*, 1219-1230.

Child, I. L., Kanthamani, H., & Sweeney, V. M. (1977). A simplified experiment in dream telepathy. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology 1976* (pp. 91-93). Metuchen, NJ: Scarecrow.

Dalton, K., Steinkamp, F., & Sherwood, S. J. (1999). A dream GESP experiment using dynamic targets and consensus vote. *Journal of the Society for Psychical Research*, *93*, 145-166.

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

Diaconis, P. (1991). Comment. *Statistical Science*, *6*, 386.

Eppinger, R. (2001). *A comparative examination of ganzfeld and dream reports in free response ESP studies.* Unpublished PhD thesis, University of Edinburgh.

Greenhouse, J. B. (1991). Comment: Parapsychology—On the margins of science? *Statistical Science*, *6* (4), 386-389.

Greenhouse, J. B., & Greenhouse, S. W. (1988). An aspirin a day . . . ? *Chance*, *1*, 24-31.

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

Harley, T. A. (1989). Psi missing in a dream clairvoyance experiment. *Journal of the Society for Psychical Research*, *56*, 1-7.

Hearne, K. M. T. (1981). The effect on the subject (in waking, SWS and REM states) of electric shocks to the agent: An 'ESP" experiment. *Journal of the Society for Psychical Research*, *51*, 87-92.

Hearne, K. M. T. (1987). A dream-telepathy experiment using a home "dream machine." *Journal of the Society for Psychical Research*, *53*, 38-40.

Hearne, K. M. T. (1989). A nationwide mass dream-telepathy experiment. *Journal of the Society for Psychical Research*, *55*, 271-274.

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., Krippner, S., & Ullman, M. (1972). Telepathic perception of art prints under two conditions. *Proceedings of the 80th Annual Convention of the American Psychological Association*, *7*, 319-320.

Hyman, R. (1991). Comment. *Statistical Science*, *6*, 389-392.

Kanthamani, H., & Broughton, R. (1992). An experiment in ganzfeld and dreams: A further confirmation. *Proceedings of the 35th Annual*

Convention of the Parapsychological Association, Las Vegas, NV, 59-73.

Kanthamani, H., & Khilji, A. (1990). An experiment in ganzfeld and dreams: A confirmatory study. *Proceedings of the 33rd Annual Convention of the Parapsychological Association, Chevy Chase, MD*, 126-137.

Kanthamani, H., Khilji, A., & Rustomji-Kerns, R. (1988). An experiment in ganzfeld and dreams with a clairvoyance technique. *Proceedings of the 31st Annual Convention of the Parapsychological Association, Montreal, Quebec*, 413-423.

Krippner, S., Honorton, C., Ullman, M., Masters, R., & Houston, J. (1971). A long-distance "sensory bombardment" study of ESP in dreams. *Journal of the American Society for Psychical Research*, *65*, 468-475.

Krippner, S., Honorton, C., & Ullman, M. (1972). A second precognitive dream study with Malcolm Bessent. *Journal of the American Society for Psychical Research*, *66*, 269-279.

Krippner, S., Honorton, C., & Ullman, M. (1973). A long-distance ESP dream study with the "Grateful Dead." *Journal of the American Society of Psychosomatic Dentistry and Medicine*, *20*, 9-17.

Krippner, S., & Ullman, M. (1970). Telepathy and dreams: Controlled experiment with electroencephalogram-electro-oculogram monitoring. *Journal of Nervous and Mental Disease*, *151*, 394-403.

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

Markwick, B., & Beloff, J. (1983). Dream states and ESP: A distance experiment with a single subject. In W. G. Roll, J. Beloff, & R. A. White (Eds.). *Research in parapsychology 1982* (pp. 228-230). New York: Scarecrow Press.

Markwick, B., & Beloff, J. (1988). Dream states and ESP: A distance experiment featuring a pure clairvoyance, free-response design. In D. H. Weiner, & R. L. Morris (Eds.). *Research in parapsychology 1987* (pp. 77-81). London: Scarecrow Press.

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.

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

Palmer, J. (1971). Scoring in ESP tests as a function of belief in ESP: Part I. The sheep-goat effect. *Journal of the American Society for Psychical Research*, *65*, 373-408.

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

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

Roe, C. A., Sherwood, S. J., Luke, D., & Farrell, L. M. (2002). An exploratory investigation of dream GESP using consensus judging and dynamic targets. *Journal of the Society for Psychical Research*, *66*, 225-238.

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

Rosenthal, R. (1986). Meta-analytic procedures and the nature of replication: The ganzfeld debate. *Journal of Parapsychology*, *50*, 315-336.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist*, *45*, 775-777.

Sargent, C. L., & Harley, T. A. (1982). Precognition testing with free-response techniques in the ganzfeld and the dream state. *European Journal of Parapsychology*, *4*, 243-256.

Schmeidler, G. 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.

Sherwood, S. J., Dalton, K., Steinkamp, F., & Watt, C. (2000). Dream clairvoyance study II using dynamic video clips: Investigation of consensus voting judging procedures and target emotionality. *Dreaming*, *10*, 221-236.

Sherwood, S. J., Roe, C. A., Simmonds, C. A., & Biles, C. (2002). An exploratory investigation of dream precognition using consensus judging and static targets. *Journal of the Society for Psychical Research*, *66*, 22-28.

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.

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

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.

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

Storm, L. (2006). Meta-analysis in parapsychology: I. The ganzfeld domain. *Australian Journal of Parapsychology*, *6*, 35-52.

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

Thalbourne, M. A., & Haraldsson, E. (1980). Personality characteristics of sheep and goats. *Personality and Individual Differences*, *1*, 180-185.

Ullman, M. (1969). Telepathy and dreams. *Experimental Medicine and Surgery*, *27*, 19-38.

Ullman, M., & Krippner, S. (1969). A laboratory approach to the nocturnal dimension of paranormal experience: Report of a confirmatory study using the REM monitoring technique. *Biological Psychiatry*, *1*, 259-270.

Ullman, M., Krippner, S., & Feldstein, S. (1966). Experimentally induced telepathic dreams: Two studies using EEG-REM monitoring technique. *International Journal of Neuropsychiatry*, *24*, 420-437.

Ullman, M., Krippner, S., with Vaughan, A. (1973). *Dream telepathy: Experiments in nocturnal ESP* (2nd ed.). Jefferson, NC: McFarland.

Weiner, D. H., & McCain, D. L. (1981). The use of the majority-vote technique in dream ESP: A pilot study. *Journal of Parapsychology*, *45*, 156-157.

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

Anomalistic Psychology Research Unit
School of Psychology
University of Adelaide
Adelaide SA 5005
AUSTRALIA

E-mail: lance.storm@adelaide.edu.au