Experimental Evidence Suggestive of Anomalous Consciousness Interactions

Deborah L. Delanoy,
Department of Psychology, University of Edinburgh, Edinburgh, Scotland, U.K.

1 Introduction

The so-called "mind-body problem" is arguably humankind's most enduring question. The crux of this question is whether mind can exist independently of the body. Or, to re-phrase it, is mind an epiphenomenon of brain functioning, or is it, to some degree, independent of the mechanistic properties of our physical brains? Throughout written history, the greatest philosophical thinkers have pondered this matter. However, outside of the field of parapsychology, there has been very little experimental research exploring whether consciousness can interact with its environment independently of the physical body. This paper will address these issues by exploring patterns found in experimental parapsychological research which suggest that mind or consciousness can interact directly with its environment without mediation by known physical mechanisms, e.g. senses, motor activity, physiological output. If the patterns emerging from this experimental work are as they appear, they may help shed some light on the ability of consciousness to act independently of the physical body.

Parapsychological research can broadly be conceptualised as addressing two main areas. The first of these, extrasensory perception (ESP), refers to the apparent obtaining of information by the mind without recourse to currently understood sensory means of gaining such information. The second area, psychokinesis (PK), refers to changes in physical systems apparently brought about by an act of conscious intention, without recourse to currently understood means of effecting such changes. Both ESP and PK can be conceptualised as anomalous interactions between mind and its environment, apparently not mediated by any currently understood physical, sensory means. Psi is a term used to refer to both ESP and PK phenomena.

This paper will present seven major *meta-analyses* carried out on various parapsychological databases. These seven were chosen as they demonstrate both ESP and PK research, and highlight the wide scope of psi experimentation which has been conducted over the last 60 years. Meta-analysis is a term which refers to a group of statistical procedures that are used to summarise and describe bodies of research. They provide a systematic means of combining results from groups of related individual studies to assess overall consistency of results, and can assist in identifying variables within the database that appear to affect outcomes, known as "moderating variables". Meta-analytic techniques provide *quantitative,* as opposed to qualitative, reviews of bodies of research. The term "meta-analysis" was first coined by Glass in 1976[1], although the basic procedures had been known for several decades (Snedecor [2]; Mosteller and Bush [3]). More recently, many books have been published detailing methods, procedures and theoretical considerations for conducting meta-analyses (e.g. Glass, McGaw and Smith [4]; Hedges and Olkin [5]; Wolf [6]; Hunter and Schmidt [7]; and Rosenthal [8]); these references will provide further details of the procedures and statistical formulae described generally below.

For readers who are unfamiliar with meta-analytic techniques, a brief summary of the basic components of meta-analysis will be given. After identifying a domain of study, all relevant studies are gathered together. The characteristics of those which are of interest are then coded, e.g. procedural variables and constants, study quality, etc. Ideally, this coding should be performed by one or more individuals who are not closely involved with the research topic, to avoid investigators' biases influencing any coding decisions. The statistical measures generated for each study (commonly referred to as "test statistics", e.g. z, t, chi-square, etc.) are converted into effect sizes. An effect size is a measure of the degree to which a phenomenon is present in the population (i.e., of how large the effect is). As noted by Rosenthal [8], commonly used statistical measures are usually a product how large the effect is and some function of the size of the study, often the square root of the number of trials or individuals. He expresses this (p. 20) as:

Test of significance = Size of effect × Size of Study

Most effect size measures are defined so that they are zero when the null hypothesis is true. Unlike "test" statistical measures, effect sizes do not grow in magnitude with the size of the study. Thus they provide a more accurate picture of replication across studies than can be provided by the standard statistical measures alone. This can be especially important when dealing with small effects and thus with studies with relatively low power, such as those commonly found in parapsychological research. Effect sizes allow studies to be assessed with a continuous measure, rather than the dichotomous measure to which test statistics are often reduced (i.e., statistically significant or not).

Replication across different studies is measured in terms of the consistency, or the "homogeneity" of the magnitude of the observed effect sizes. Again, this differs from the more traditional approach using test statistics, in which replication is defined by whether or not the null hypothesis is rejected in each study. Of course, when evaluating replication across a group of studies, confidence in any estimate of overall effect will be increased as the amount of confirming data increases. Using test statistics, outcomes from a group of studies can be combined and/or summarised to give an overall outcome for the database, weighting each study according to its size. In the following meta-analyses of parapsychological studies, the overall likelihood of observing the results if the null hypothesis is always true can be assessed by finding a combined z-score for all studies. This is simply a weighted average of the number of standard deviations the results deviated from chance, and its likelihood can be assessed using the standard statistical tables. One method of combining studies in this way is with a "Stouffer z" [3, 6, 9], and if the null hypotheses are always true, this statistic follows a standard normal curve. Stouffer's z provides a measure of how many standard deviations from chance the combined results of all of the studies fell. Using Stouffer's z, we can compute a "p-value" which gives us the probability of observing such extreme results if chance alone is the explanation. As we will see, the p-values for the meta-analyses in parapsychology are extremely low, thus effectively ruling out chance as an explanation for the data.

Using meta-analytic techniques, the impact of flaws upon study outcome and of various moderating variables can be quantitatively assessed, leading to improvements in study design and identification of factors associated with optimal outcomes. Possible relationships between variables can be recognised and tested in future experiments.

One problem that plagues all literature reviews is the tendency to report/publish only significant findings, commonly referred to as the "file drawer problem". However, there are a variety of methods available to estimate the size of the file drawer effect (i.e., the number of non-significant studies which would be required to nullify the outcome of a meta-analysis). For example, Rosenthal [8] provides a statistical measure, referred to as the "Fail-Safe N", to determine how many unpublished, null studies would be needed to negate an observed effect in any size of database, with the general guideline that a 5:1 ratio of null, unpublished studies to each published study should be obtained before the possibility of a negating file drawer effect can be safely eliminated.

In the following sections the findings of the selected meta-analyses will be presented. Consideration of possible interpretations, explanations and implications of this work will be found in the concluding Discussion section.

2 The ganzfeld debate

One group of parapsychological studies, the ganzfeld studies, have received more recent publicity, in terms of published articles examining the overall effect of the database, than has any other area of psi research. This attention is the result of detailed meta-analyses of the ganzfeld studies by a leading ganzfeld researcher, Honorton [9], and a critic of this work, Hyman [10].

Following the publication of these two meta-analyses, many other "pro" and "con" evaluations and commentaries have been published [11, 12, 13, 14]. The ganzfeld debate, often referred to as the "Honorton/Hyman debate", will be summarised below, but first a brief description of a ganzfeld study will be presented.

The ganzfeld technique consists of presenting a relaxed percipient with homogenous, unpatterned visual and auditory stimuli, which assists in increasing the mental imagery experienced by the percipient. While receiving this stimulus, the percipients verbalise all their experiences, their goal being to gain impressions which will relate to a sensorially isolated and remote target picture or short video clip. The "target" is being watched frequently by another person (a "sender" or "agent") who is attempting mentally to convey impressions of the target to the percipient or "receiver". These studies utilise a "free-response" methodology, in which the contents of the target material are unknown to the receiver (i.e., the percipient is "free" to respond with whatever impressions they generate, as he or she has no information regarding the specific contents of the possible target). The most common method of analysis used in ganzfeld studies is for the percipient or an independent judge(s) to compare the obtained impressions to four different target pictures/video clips, one of which is a duplicate of the actual target, looking for similarities. Using blind procedures, the judge has a one in four chance of correctly identifying the actual target (i.e., mean chance expectancy = 25 per cent "hit" rate). Study outcome is based upon whether similarities between the percipient's impressions and the actual target enabled the target to be correctly identified significantly more often than chance would allow. For further information regarding this experimental technique, procedural details and methods of analysis, see Honorton [9], and Honorton et al. [15].

A meta-analysis of twenty-eight ganzfeld studies was performed by Honorton (1985) [9], in response to a flaw analysis of the ganzfeld database conducted by Hyman (1985) [10]. Hyman found a highly significant overall effect in the database, but concluded that this effect was negated as he found a significant relationship between the study outcomes and procedural and statistical flaws contained in the studies. However, Hyman's flaw categorisations were severely criticised by Honorton, and a psychometrician, Saunders [16], found faults in Hyman's statistical analyses.

Honorton's meta-analysis found there were no significant relationships between study outcomes and quality. The overall composite (Stouffer) z score for the 28 ganzfeld studies included in the Honorton meta-analysis was highly significant ($z = 6.6$, $p < 10^{-9}$, two-tailed). The effect sizes were homogeneous, overall and across experimenters. The discrepancy between the Honorton and Hyman analyses of the ganzfeld studies prompted a further meta-analysis by Rosenthal [17], an independent specialist in meta-analysis. Like Honorton, Rosenthal found an overall composite z score of 6.60 for the twenty-eight ganzfeld studies. His file drawer estimate agreed with that of Honorton, requiring 423 unreported, null studies to negate the significance of the database. Here it is worth noting that another critic, Blackmore [18] conducted a survey to discover the number of unreported ganzfeld studies in 1980, prior to the Honorton/Hyman debate. Her survey found 32 unreported studies, of which 12 were never completed, and one could not be analysed. Of the remaining 19 studies, 14 were judged by Blackmore to have adequate methodology, with 5 of these (36 percent) reporting significant results. She concluded that "the bias introduced by selective reporting of ESP ganzfeld studies is not a major contributor to the overall proportion of significant results" (p. 217). Rosenthal, after considering the possible influence of various flaws upon study outcome, concluded that the overall hit rate of the studies could be estimated to be 33 percent, whereas chance expectancy was 25 percent.

In 1986 Honorton and Hyman published a "Joint Communiqué" [19] in which they agreed that there was an overall effect in the database, but continued to disagree as to what extent this effect may have been influenced by methodological flaws. In their communiqué they outlined the necessary methodological precautions that should be taken to avoid the possibility of future studies giving rise to the same level of debate that had surrounded the previous ones. They concluded that more studies needed to be conducted, using the controls they had documented, before any final verdict about the database could be reached.

Honorton and his research team proceeded to design a new ganzfeld system which met the criteria he and Hyman had specified in their communiqué. This system, and studies using it, are referred to as "autoganzfeld studies", as much of the procedure is under automated computer control in order to avoid the problems found in some of the earlier studies. Before Honorton's lab closed in 1989, 11 experimental series, representing 355 sessions, conducted by eight experimenters, had been collected using the autoganzfeld. Honorton et al. [15] published a summary of the autoganzfeld studies and compared them with his earlier meta-analysis. The autoganzfeld sessions yielded overall significant results ($z = 3.89$, $p = 0.00005$), with an obtained hit rate of 34.4 percent (with 25 percent being chance expectancy). The effect sizes by series and by experimenter were both homogeneous. Comparing the autoganzfeld outcomes to those of the 28 studies of the earlier meta-analysis revealed very similar outcomes, with the autoganzfeld showing slightly better ESP scoring than that obtained in the earlier studies (autoganzfeld results by series: effect size or *es* = .29, earlier 28 meta-analysis studies by experiment: *es* = .28).

Hyman, in 1991 [20] commenting upon a presentation of these results by the statistician, Utts [12], concluded that "Honorton's experiments have produced intriguing results. If, as Utts suggests, independent laboratories can produce similar results with the same relationships and with the same attention to rigorous methodology, then parapsychology may indeed have finally captured its elusive quarry." (p. 392). Replications are currently being undertaken at various labs; the only replication using a full autoganzfeld environment which has been reported to date was conducted at the University of Edinburgh [21], where the obtained significant, overall hit rate was 33 percent ($z = 1.67$, $p < 0.05$). This outcome is consistent with Honorton's autoganzfeld scoring rate of 34.4 percent, and replicates Rosenthal's hit rate estimate based on the earlier ganzfeld studies. The procedure for the Edinburgh study incorporated additional safeguards against subject and experimenter fraud.

3 Looking into the future: Meta-analysis of precognition ESP studies

Folklore and many anecdotal stories have relayed how some individuals have claimed to be able to "foretell" the future, or have experienced premonitions of events before they actually occurred. While much of this information is likely due to misinterpretation, misrepresentation or other flaws of human perception, memory and reasoning, there are experimental findings which suggest that precognition may occur (see Wiseman and Morris [22] for an overview of ways we can be deceived, or can deceive ourselves into interpreting a normal incident as being paranormal).

Honorton and Ferrari [23] conducted a meta-analysis of 309 precognition studies conducted between 1935 and 1987. These studies all used a "forced-choice" methodology, in which the subject is aware of the possible target choices, and is asked to choose one of them as his answer (as opposed to "free-response" methodologies, such as ganzfeld studies). In all of these studies, the subject made their choice as to the target identity prior to the target identity actually being randomly generated. Thus the subjects' responses were to targets which did not exist at the time of their response. These studies are thought by some to be methodologically superior to other ESP studies as there is little possibility of the subject "cheating", or receiving any subtle cues about the target identity, as the target does not exist when their response is made.

The studies included in this meta-analysis were conducted by 62 different senior investigators, and included nearly two million individual trials contributed by over 50,000 subjects. While the mean effect size per trial is small (*es* = .02), it is sufficiently consistent for the overall effect from these studies to be highly significant (combined $z = 11.41$, $p = 6.3 \times 10^{-25}$). Using eight different measures of study quality, no systematic relationship was found between study outcome and study quality. A "fail-safe N" estimate would require 14,268 unreported, null studies to reduce the significance of the database to chance levels. Given the wide diversity of study methods and procedures found in this database, it is not surprising that the study outcomes were extremely heterogeneous. The authors eliminated outliers by discarding those studies with z scores falling within the top and bottom 10 percent of the distribution, leaving 248 studies. It should be noted that the elimination of outlier studies to obtain homogeneity is a common practice, and in other, non-parapsychological reviews "it is sometimes necessary to discard as many as 45% of the studies to achieve a homogeneous effect size distribution" (p. 1507) [24]. The resulting mean trial effect size was .012, and the combined z still highly significant ($z = 6.06$, $p = 1.1 \times 10^{-9}$). While it was found that study quality improved significantly over the 55 year period during which these studies were conducted (correlation coefficient *r*[246 degrees of freedom] = .282, $p = 2 \times 10^{-7}$), study effect sizes did not significantly co-vary with the year of publication. Study effect sizes are homogenous across the 57 investigators contributing to the trimmed database. The rest of the analyses conducted were all performed upon this smaller database.

The authors identified four "moderating" variables that appeared to relate systematically to study outcome. The first variable involved the subject population. It was found that studies using subjects who were selected on the basis of good ESP performance in previous experimental sessions obtained significantly better ESP effects than those studies using unselected subjects (a *t* test with 246 degrees of freedom [df] giving $t = 3.16$, $p = 0.001$). Another variable which covaried with study effect size was whether the subjects were tested individually or in groups, with individual testing studies obtaining significantly higher outcomes than those using group testing methods (*t*[200 df] = 1.89, $p = 0.03$).

A further moderating variable involved the type of feedback subjects received about the accuracy of their responses. There were four feedback categories, including no feedback, delayed feedback (usually via mail), feedback given after a sequence of responses (often after 25 responses), and feedback given after each response. Of the 104 studies which supplied the necessary information, there was a linear and significant correlation between the precognition effect and feedback level (*r*[102 df] = .231, $p = 0.009$), with effect sizes increasing with level of feedback. A related finding involves the time interval between the subject's responses and the target selection. This finding is confounded by the feedback level, as time duration between the response and target generation may co-vary with feedback level (i.e., when feedback was given after every response, the time interval between response and target selection would have to be shorter than was necessarily the case when feedback was given after a sequence of calls, or a month after the responses had been made). There were seven different time interval categories, varying from a millisecond to months. There was found to be a significant decline in precognition effect sizes as the time interval between response and target selection increased (*r*[142 df] = –.199, $p = 0.017$). The significant temporal decline/study effect size relationship is due entirely to those studies which used unselected subjects, with the studies that tested selected subjects showing a small, non-significant increase in precognition scoring as the time interval increased (the difference between these groups was not significant).

It should be noted that there was no significant difference in quality between studies using selected and unselected subjects. Also, studies which tested subjects individually did show significantly higher study quality than those utilising group testing procedures (*t*[137 df] = 3.08, $p = 0.003$). A correlation between feedback level and research quality was positive, but not significant (*r* [103] = .173, $p = 0.82$).

In summarising the precognition findings, Honorton and Ferrari concluded "the forced-choice precognition experiments confirm the existence of a small but highly significant precognition effect." (p. 300). Furthermore, they concluded that the most important outcome of the meta-analysis was the identification of moderating variables, which not only provides guidelines for future research, but may also help expand our understanding of the phenomena.

4 Influencing randomness in physical systems: Two meta-analyses

"Mind over matter" is a frequently used phrase, but is there any evidence suggesting that mind can exert some influence over the behaviour of physical, material systems? Two meta-analyses dealing with such effects will be reviewed, both of which suggest that mind can directly interact with matter. Both these databases involve participants attempting to make a random system behave in a non-random manner.

The first of these databases involves studies in which people tried to influence the outcome of falling dice. This work was initially suggested by claims of gamblers that they were able to influence the outcome in dice throwing situations in gaming casinos. Radin and Ferrari [25] conducted a meta-analysis of 148 dice studies conducted between 1935 and 1987. This database also included 31 control studies in which no conscious influence of outcome was attempted. The results showed a significant overall effect for the experimental influence studies ($es = .012$, Stouffer $z = 18.2$, $p < 10^{-70}$), and chance results for the control studies (Stouffer $z = 0.18$). To obtain a homogeneous distribution of effect sizes, 53 studies (35 per cent) of the database had to be deleted. Of these deleted studies, 33 had positive and 19 had negative effect sizes. Eleven study quality measures were considered. While the relationship was not significant, the authors did find that effect size decreased as study quality increased.

Another methodological problem affecting this database is that the probability of obtaining a specific outcome is not necessarily equally distributed across all the die faces (e.g., if using pipped dice, the six typically has the least mass and is thus most likely to come up). To examine the possible influence of this "non-random" aspect of dice throwing, the results for a subset of 69 studies, in which targets were balanced equally across the six die faces, were examined. A significant overall effect was still obtained (Stouffer $z = 7.617$, $p < 10^{-11}$). For these 69 studies, the effect size was relatively constant across the different measures of study quality, and a file drawer analysis revealed that a 20:1 ratio of unreported, nonsignificant studies for each reported study would be required to reduce the database to chance expectations.

The second "mind over matter" meta-analysis involves studies in which a person attempts to influence a microelectronic random number generator (RNG) to behave in a non-random manner. This meta-analysis, conducted by Radin and Nelson [24], involves the largest parapsychological database to date, with 832 series, of which 597 were experimental series and 235 control series. The general protocol of these studies involves having a RNG drive a visual display, which an observer tries to influence, by means of mental intention, in accordance with prespecified instructions. The randomness of the RNG is usually provided by radioactive decay, electronic noise or pseudorandom number sequence seeded with true random sources; the RNG's are frequently monitored to ensure true random output in these studies. The observer initiates a "trial" by means of a button push, which starts the collection of a fixed length sequence of data. For each data sequence, a *z* score may then be computed. The mean effect size per trial for the experimental series was very small, but very robust ($es = .0003$, combined $z = 15.58$, $p = 1.8 \times 10^{-35}$) and significantly higher ($z = 4.1$, $p = 0.00004$) than the effect size for the control series ($es = -.00004$). Sixteen study quality measures were investigated; effect size did not significantly co-vary with study quality. The file drawer estimate for this data base is enormous, requiring 54,000 null, unreported studies to reduce the observed effect to chance levels. Given these findings, Radin and Nelson concluded that "it is difficult to avoid the conclusion that under certain circumstances, consciousness interacts with random physical systems" (p. 1512, [24]).

5 Direct mental interactions with living systems (DMILS)

Direct mental interactions with living systems (DMILS) research involves testing procedures where a person (an "agent") is trying to interact with a biological target system, e.g., another person's physiological responses or the behaviour of small animals or fish. In DMILS studies the biological target is located in a sensorially shielded room, providing isolation from any physical contact with the agent. The target's spontaneously fluctuating activity is monitored continuously while the agent, during randomly interspersed influence and noninfluence (control) periods, tries to influence mentally the target's activity in a pre-specified manner. The target system is unaware of timing or goal orientation (i.e., influence or non-influence) of the agent's mental intentions. When human physiological responses are the target system, the target person's only goal during the experimental session is to remain passively alert and to wish mentally that their physiology will unconsciously respond appropriately to the agent's intentions. The mental strategies used by the agent to interact with the remote, shielded target includes wishing and willing the desired changes to manifest in the target, mental imaging of the desired outcome, and in some instances simply paying attention to the target system. The randomised order of the influence or non-influence period is usually conveyed to the agent by a message on a computer monitor; the monitor may also convey to the agent the actual recordings of the target's activity, thereby providing on-going feedback about the effects of their mental intentions upon the remote target system. The experimental design eliminates possible confounding factors such as recording errors, placebo effects, confounding internal rhythms and chance correspondences.

The majority of the recent DMILS research has been conducted by Braud and his colleagues, who published a meta-analytic summary of 37 of their experiments (Braud and Schlitz [26]).

This work involved 13 different experimenters and 655 sessions. These 37 studies examined seven different target systems, including electrodermal activity (EDA) with the agent trying to influence the subject's EDA to increase or decrease (i.e., trying to "calm" or "activate" the subject), blood pressure, fish orientation, mammal locomotion, and the rate of haemolysis of human red blood cells. The overall results from this work have been highly significant (per session overall $es = .33$, Stouffer $z = 7.72$, $p = 2.58 \times 10^{-14}$).

While this work was conducted by 13 different experimenters, it was all performed at the same laboratory. Other laboratories are now attempting to replicate this work, with the initial results generally conforming to those obtained by Braud et al. For example, Delanoy and Sah (1994 [27]) compared EDA responses to conscious responses in a DMILS environment, in which the agent was either remembering and trying to re-experience a very positive, exhilarating emotion ("activate" condition) or was thinking of an emotionally neutral object ("control" condition). The subject's EDA showed significantly greater activity during the activate periods than during the control periods ($es = .31$, $t[31] = 1.77$, $p = 0.04$). However, the subject's conscious responses (i.e., their guesses as to whether the agent was trying to activate or calm them) did not differ from chance expectancy. The finding of a significant physiological effect, with no corresponding effect shown by a conscious response measure, supports similar findings from Tart [28} and Targ and Puthoff [29], and suggests that subtle psi interactions may occur without any conscious recognition on the part of the subject.

6 Relating ESP to personality traits: Two meta-analyses

Parapsychological researchers have long been interested in exploring if there are any factors which might relate to why some people report having more psi experiences in their everyday life than do others. Similarly, while most experimental work is done with volunteer subjects who have not been chosen on the basis of their supposed psi ability, it has been observed that some people appear to do better in experimental psi tests than others. One approach to examining possible reasons for these observed differences has involved exploring the relationship between various personality factors and psi ability.

Two meta-analyses of studies which have looked for correlations between performance on a psi task and different personality traits will be discussed here. One of these involved studies which looked for a relationship between a person's opinion of psi and their own psi abilities with their psi test performance. Research examining what has come to be known as the sheep/goat effect, supported the hypothesis that in experimental psi tests those with positive attitudes ("sheep") tend to score above chance, and those with negative attitudes ("goats") below chance.

Lawrence [30] conducted a meta-analysis of the 73 published studies examining the sheep/goat effect. These studies were conducted by 37 principal investigators, and involved over 4,500 subjects who completed over 685,000 trials. The overall effect size per trial is small ($r = 0.029$), but highly significant over these studies which involved a large number of procedural manipulations and potential modifying variables. The combined Stouffer $z = 8.17$, $p = 1.33 \times 10^{-16}$. Using seven different measure of study quality, Lawrence found that effect size did not covary with study quality. A file-drawer estimate (Rosenthal's "fail-safe N") revealed that 1726 unreported studies with null results (i.e., 23 unreported studies for each of the 73 reported ones) would be required to reduce the significance of the database to chance expectancy.

This database has used a wide range of different sheep/goat scales, ranging from single questions to more lengthy questionnaires. The means of determining belief have also varied, with most focusing upon previous personal psi experiences, self-evaluation of personal psi ability, opinions regarding one's ability to display psi ability in the specific testing situation and/or one's general attitudes towards such phenomena. Lawrence found there was no overall relationship between effect size and the type of measure used, from which he concluded that the sheep/goat effect was quite robust regardless of how it was measured.

Another personality trait that has been studied in relation to psi performance is extraversion/introversion. Honorton, Ferrari and Bem [31] conducted a meta-analysis on the 60 published studies examining this relationship. Prior to this meta-analysis, descriptive reviews of this database had concluded that extraverts performed better than introverts on psi tasks

(Eysenck, [32], Palmer [33], Sargent [34]). However, the ability of meta-analysis to identify flaws and modifying variables led to a different finding in the meta-analysis. While the meta-analysis did find a significant overall effect (r = .09, combined z = 4.63, p = 0.000004), the effect sizes were non-homogeneous. The studies were divided into smaller groups according to various procedural variables in order to discover the source of the non-homogeneity. The authors separated the 45 studies using forced-choice procedures from the 14 studies using free-response methods. Once again, significant but non-homogeneous, effects were found (forced choice: r = .06, combined z = 2.86, p = 0.0042; free-response: r = .20, combined z = 4.82, p = 0.0000015). A further division of these two groups of studies examined whether testing subjects individually or in groups had any impact on the outcomes. This analysis revealed that of the forced-choice studies, 21 studies had tested subjects individually, resulting in a significant, but non-homogeneous effect (r = .15, combined z = 4.54, p = 0.000006). In the 24 forced-choice studies where participants were tested in groups, there was no significant effect (r = .00, z = –0.02), although there was homogeneity .

A flaw analysis showed that the significant effect in the forced-choice database was entirely due to 18 studies in which the extraversion measure had been given after the ESP test, the significance of this correlation being due to 9 of these studies in which the subjects knew how they had performed on their psi task before they completed the extraversion questionnaire. This finding raises the strong possibility that the correlation was due to psychological, as opposed to paranormal, factors. Thus the previous descriptive reviews which had found a significant, positive relationship between extraversion and psi-scoring had failed to uncover the inconsistency in the degree to which this effect was present in these studies, and the flaw which lead Honorton, Ferrari and Bem to conclude that the relationship in forced-choice studies would appear to be artifactual.

In the subset of 14 free-response extraversion studies, a significant (r = .20, combined z = 4.82, p = 0000015) but non-homogeneous effect was obtained. Dividing the studies according to individual or group testing procedures revealed that the 2 studies employing group testing were responsible for the non-homogeneity. The results for the 12 studies which testing subjects individually show homogeneity and a significant correlation (r = .20, combined z = 4.46, p = 0.0000083). Eleven of the studies documented the presentation order of the psi test and extraversion questionnaire. In all of these studies, the extraversion questionnaire was given prior to the ESP test, thereby avoiding the potential problem of subject's knowledge of their ESP results influencing the way that they completed their extraversion questionnaire. These 11 studies show a significant and homogeneous extraversion/ESP correlation (r = .21, combined z = 4.57, p = .000005).

After completing the extraversion/ESP meta-analysis, Honorton et al. examined the autoganzfeld database to see if they could confirm the relationship. For the 221 autoganzfeld trials for which they had extraversion data, they obtained a significant ESP/extraversion correlation (r = .18, t[219 df] = 2.67, p = 0.008). This finding is consistent with those from the free-response extraversion meta-analysis.

7 Discussion

The above seven meta-analyses represent a cross-section of the meta-analyses that have been performed on parapsychological research. They were not chosen to illustrate the greatest effects or to "paint the rosiest picture", but rather to provide a window into the range of effects and variety of methodologies found in psi experimentation. The effect sizes in these studies tend to be very small (RNG–PK) to moderate (i.e., DMILS) in size. However, even the smaller effect sizes appear to be reliably found in the databases. Furthermore, the size of an effect does not provide a good indication of its potential meaningfulness or applicability. For example, a recent medical study investigating whether aspirin could help prevent heart attacks was ended prematurely because the effectiveness of the treatment was so clearly demonstrated after six months of trials that the investigators thought it would be unethical to withhold the treatment further from the control group. Indeed, the findings from the study were heralded as a major medical breakthrough. While the findings from this study were highly significant (χ^2 = 25.01, p

= 0.00001), the effect size is .068, considerably smaller than some of the effect sizes found in the psi literature [12, 13]. It should be noted that small effects have low statistical power [8, 12, 13]. For example, the aspirin study involved over 22,000 subjects. If there had only been 3,000 subjects, the investigators would have had less than a 50 percent chance of finding a conventionally significant effect [13]. Given the small effect sizes which are typical in psi experiments, low replicability is to be expected. Rosenthal [17] notes that "even though controversial research areas are characterised by small effects, that does not mean that the effects are of no practical importance." (p. 324). Indeed, in an article addressing behavioural research in general, Rosenthal [35] warned: "Given the levels of statistical power at which we normally operate, we have no right to expect the proportion of significant results that we typically do expect, even if in nature there is a very real and very important effect" (p. 16).

What can these findings tell us about the functioning of apparent psi abilities? The conceptualisation of ESP as the anomalous input of information into consciousness and PK as the anomalous output of influence are "working" models, which help convey possible interpretations of the obtained phenomena. However, the distinction between ESP and PK is often blurred. If one accepts the precognition database as suggesting that information about an event can be obtained before the occurrence of the event, many of the psi results could be interpreted as representing acts of precognition. For example, while most ganzfeld and DMILS studies are "real-time" and involve an "agent", it is possible that the actual mechanism at work may be the subject obtaining information about the target by "looking" into the future to gain relevant target information and then generating appropriate impressions in the case of ganzfeld studies, or producing the appropriate self-regulatory responses in the case of the DMILS studies. In this context, it should be mentioned that the role of the agent is unclear. No DMILS studies have yet been reported which have not used an agent, but in the case of ganzfeld studies, the recent Edinburgh study [21] found equally significant outcomes in sender and no sender conditions. Similarly, other ESP studies have obtained significant, positive outcomes without using an agent (for a review of this work see Palmer [36]). Findings such as these indicate that a sender appears not to be a necessary component in anomalous information transfer studies, although they may still have a beneficial psychological impact upon the study outcomes [36]. The RNG–PK work has been traditionally conceptualised as representing "influencing" effects, as has the DMILS work which was initially known as "bio-PK". However, alternative interpretations of these apparent effects may involve ESP. May et al. [37] have proposed that apparent RNG–PK effects could be the result of the observer, via precognition, knowing what would be the right moment to initiate a sequence of random event (i.e., when to push the button) to get the desired outcome, thereby making use of the random fluctuations found in RNG systems to create a non-random outcome. Others have questioned the validity of a precognition interpretation of psi data. For example, Morris [38] discusses models based on "real-time" psi effects as possible alternative explanations of precognition. For example, using PK a subject or investigator could influence the random source used to choose the target in precognition studies to obtain a selection consistent with the subject's response.

As the above comments make apparent, the mechanisms which may be involved in the producing the effects found in these databases are still unknown. Process-oriented research is ongoing in parapsychology. In future studies, correlations such as those found in the precognition database may help us better differentiate between the differing theoretical interpretations of these anomalous effects. While this paper has focused upon presenting summaries of experimental data, there are a variety of theoretical models which address these findings. Although it is outside the scope of this paper to review these models, a thorough presentation of theoretical parapsychology is provided by Stokes [39].

In conclusion, the findings from these meta-analyses suggest that consistent trends and patterns are to be found in the database. The consistency of outcomes found in the ganzfeld research, the robust PK effects, the modifying variables revealed by the precognition database, the variety of target systems displaying DMILS effects and the correlations found with personality traits are all indicative of lawful relationships. Given these relationships it is difficult to dismiss the findings as "merely an unexplained departure from a theoretical chance baseline" p. 301 [23]. Whether these effects will prove to represent some combination of currently unrecognised statistical problems, undetected methodological artefacts, or, as seems increasingly likely, a genuinely new, hitherto unrecognised characteristic of mind or consciousness remains to be seen.

Acknowledgements: I would like to thank Robert L. Morris and Jessica Utts for their helpful comments on earlier drafts of this paper.

References

[1] G. V. Glass, *Primary, secondary, and meta-analysis of research*, Educational Researcher **5** (1976), 5–8.
[2] G. W. Snedecor, *Statistical Methods*, Iowa State College Press, 1946.
[3] F. M. Mosteller, and R. R. Bush, *Selected quantitative techniques*, In: G. Lindzey (ed.), Handbook of Social Psychology: Vol. 1. Theory and Method, Addison-Wesley, 1954, 289–334.
[4] G. V. Glass, B. McGaw and M. L. Smith, *Meta-Analysis in Social Research,* Sage Publications Inc., 1981.
[5] L. V. Hedges and I. Olkin, *Statistical Methods for Meta-Analysis,* Academic Press, Ltd., 1984.
[6] F. M. Wolf, *Meta-Analysis: Quantitative Methods for Research Synthesis*, Sage Publications, Inc., 1986.
[7] J. E. Hunter and F. L. Schmidt, *Methods of Meta-Analysis*, Sage Publications, Inc., 1990.
[8] R. Rosenthal, *Meta-Analytic Procedures for Social Research.*, Revised edition, Sage Publications, Inc., 1991.
[9] C. Honorton, *Meta-analysis of psi ganzfeld research: A response to Hyman*, J. of Parapsychology **49**:1 (1985), 51–91.
[10] R. Hyman, *The ganzfeld psi experiment: A critical appraisal*, J. of Parapsychology **49**:1 (1985), 3–50.
[11] J. of Parapsychology **50**:4, 1986.
[12] J. Utts, *Replication and meta-analysis in parapsychology,* Statistical Science **6**:4 (1991), 363–403.
[13] D. J. Bem and C. Honorton, *Does psi exist? Replicable evidence for an anomalous process of information transfer*, Psych. Bulletin **115** (1994),4–18.
[14] R. Hyman, *Anomaly or artifact? Comments on Bem and Honorton*, Psych. Bulletin **115** (1994), 19–24.
[15] C. Honorton, R. E. Berger, M. P. Varvoglis, M. Quant, P. Derr, E. I. Schechter and D. C. Ferrari, *Psi communication in the ganzfeld*, J. of Parapsychology **54**:2 (1990), 99–139.
[16] D. R. Saunders, *On Hyman's factor analysis*, J. of Parapsychology **49**:1 (1985), 86–88.
[17] R. Rosenthal, *Meta-analytic procedures and the nature of replication: The ganzfeld debate*, J. of Parapsychology **50**:4 (1986), 315–336.
[18] S. J. Blackmore, *The extent of selective reporting of ESP ganzfeld studies*, European J. of Parapsychology **3**:3 (1980), 213–220.
[19] R. Hyman and C. Honorton, *A joint communiqué: The psi ganzfeld controversy*, J. of Parapsychology **50**:4 (1986), 351–164.
[20] R. Hyman, *Comment,* Statistical Science **6**:4 (1991), 389–392.
[21] K. S. Dalton, R. L. Morris, D. L. Delanoy, D. Radin, R. Taylor and R. Wiseman, *Security measures in an automated ganzfeld system*, In: Proceedings of the 37th Annual Convention of the Parapsychological Association, Parapsychological Association, 1994, 114–123.
[22] R. Wiseman and R. L. Morris, *Guidelines for testing psychic claimants*, University of Hertfordshire Press, 1995.
[23] C. Honorton and D. C. Ferrari, *"Future telling": A meta-analysis of forced-choice precognition experiments, 1935–1987,* J. of Parapsychology **35** (1989), 281–308.
[24] D. I. Radin and R. D. Nelson, *Evidence for consciousness-related anomalies in random physical systems*, Foundations of Physics **19**:12 (1989), 1499–1514.
[25] D. I. Radin and D. C. Ferrari, *Effects of consciousness on the fall of dice: A meta-analysis*, J. of Scientific Exploration **5**:1 (1991), 61–85.
[26] W. G. Braud and M. J. Schlitz, *Consciousness interactions with remote biological systems: Anomalous intentionality effects*, Subtle Energies 2:1 (1991), 1–46.

[27] D. L. Delanoy and S. Sah, *Cognitive and physiological psi responses to remote positive and neutral emotional states*, In: Proceedings of the 37th Annual Convention of the Parapsychological Association, Parapsychological Association, 1994, 128–138.
[28] C. T. Tart, *Physiological correlates of psi cognition*, Int'l. J. of Parapsychology **5** (1963), 357–386.
[29] R. Targ and H. Puthoff, *Information transmission under conditions of sensory shielding*, Nature **252** (1974), 602–607.
[30] A. R. Lawrence, *Gathering in the sheep and goats... A meta-analysis of forced-choice sheep-goat ESP studies, 1947–1993*, In: Proceedings of the 36th Annual Convention of the Parapsychological Association, Parapsychological Association, 1993, 75–86.
[31] C. Honorton, D. C. Ferrari and D. J. Bem, *Extraversion and ESP performance: A meta-analysis and a new confirmation*, In: L. A Henkel and G. R Schmeidler (eds.) Research in Parapsychology 1990, Scarecrow Press, 1992, 35–38.
[32] H. J. Eysenck, *Personality and extra-sensory perception*, J. of the Society for Psychical Research 44 (1967), 55–70.
[33] J. Palmer, *Attitudes and personality traits in experimental ESP research*. In B B. Wolman (Ed.) Handbook of parapsychology. Van Nostrand Reinhold, 1977, 175–201.
[34] C. L. Sargent, *Extraversion and performance in 'extra-sensory perception' tasks* , J. Personality and Individual Differences 3 (1981), 137–143.
[35] R. Rosenthal, *Replication in behavioral research*, Journal of Social Behavior and Personality **5** (1990), 1–30.
[36] J. Palmer, *Extrasensory perception: Research findings*, In: S. Krippner (ed.), Advances in Parapsychological Research, 2 Extrasensory Perception, Plenum Press, 1978, 59–244.
[37] E. C. May, D. I. Radin, G. S. Hubbard, B. S. Humphrey and J. M. Utts, *Psi experiments with random number generators: An informational model*, In: D. H. Weiner and D. I. Radin (eds.), Research in Parapsychology 1985, Scarecrow Press, 1986, 119–120.
[38] R. L. Morris, *Assessing experimental support for true precognition*, J. of Parapsychology **46** (1982), 321–336.
[39] D. M. Stokes, *Theoretical parapsychology*, In: S. Krippner (ed.), Advances in Parapsychological Research 5, McFarland, 1987, 77–189.
[40] R. Rosenthal, *Replication in behavioral research*, J. of Social Behavior and Personality **5** (1990), 1–30.