DEDICATION

To Mom, Dad, Billy and Ken for always showing me the value and rewards of persistence and hard work.
CONTENTS

TABLES ........................................................................................................................................ iv

ACKNOWLEDGMENTS ................................................................................................................... v

INTRODUCTION .................................................................1

METHOD ..............................................................................................16

RESULTS AND ANALYSES ..............................................................21

DISCUSSION ......................................................................................39

REFERENCES .....................................................................................44

APPENDIX A: PSI GANZFELD STUDY SCRIPT ............................... 50

APPENDIX B: PSI GANZFELD STUDY QUESTIONNAIRE ............... 56

APPENDIX C: SUMMARY OF THREE PUBLISHED PSI GANZFELD META-ANALYSES .......................................................................................... 57

ENDNOTES .......................................................................................... 58
TABLES

TABLE 1: SHORT NON-GANZFELD AND LONG GANZFELD TRIALS HIT RATES ................................................................. 22

TABLE 2: EXACT BINOMIAL TEST OF SHORT NON-GANZFELD TRIALS HIT RATES ................................................................. 23

TABLE 3: EXACT BINOMIAL TEST OF LONG GANZFELD TRIALS HIT RATES ................................................................. 24

TABLE 4: MEAN HIT RATES (IN PERCENTAGE) AND STANDARD DEVIATIONS FOR INITIAL PRIOR BELIEFS AND POSTERIOR PROBABILITIES ....................... 30

TABLE 5: MEAN HIT RATES (IN PERCENTAGE) AND STANDARD DEVIATIONS FOR META-ANALYTIC PRIORS AND POSTERIOR PROBABILITIES .............. 34

TABLE 6: MEANS AND FREQUENCIES FOR PREDICTOR VARIABLES AS A FUNCTION OF PSI HITS AND MISSES ............................................................. 38

TABLE 7: SUMMARY OF LOGISTIC REGRESSION ANALYSIS PREDICTING HITS .................................................................................... 38
ACKNOWLEDGMENTS

I would like to thank the Parapsychology Foundation, Inc. for supporting this study by partially funding it through a research grant.

My friends and cohort have been an important source of support and encouragement. I am grateful for them always being there and sharing the highs and lows of my carrying out this project to fruition. I would also like to thank all the Notre Dame and McNair Program research assistants who have worked diligently at making sure that the study runs smoothly. Finally, I would like to acknowledge my advisor, George S. Howard, for his guidance and patience throughout this entire process.
INTRODUCTION

The empirical study of psi, defined as “anomalous processes of information or energy transfer, processes such as telepathy or other forms of extrasensory perception that are currently unexplained in terms of known physical or biological mechanisms” (Bem & Honorton, 1994, p. 4) has primarily existed on the fringes of mainstream psychological research. The interest in paranormal phenomena by scientists, however, can be traced back to at least the 1850s (Hyman, 1989). J. B. Rhine, as early as the 1920s, was employing and pioneering experimental methods to study the existence of extrasensory perception (Rhine & Rhine, 1929). His classic monograph, Extra-sensory Perception (Rhine, 1935) marked the beginning of the modern era of parapsychology research by initiating the development of a coherent research program for the field. Rhine’s forced-choice card guessing methodology dominated the field until the 1970s (Hyman, 1989). Recently, beginning with a meta-analysis conducted by Daryl Bem and Charles Honorton (1994) and published in Psychological Bulletin, the existence of psi has been thrown in the debate of mainstream psychology (Milton & Wiseman, 1999, 2001; Storm & Ertel, 2001). Central to the debate are the questions, “Does psi exist?” and “To what extent have decades of psi research served to support or refute the existence of it?”
Ganzfeld Procedure and Its Methodological Adequacy

The ganzfeld (meaning “total field”) procedure has, for the last 25 years been the dominant research methodology employed in the study of psi (or more specifically, telepathy; Dalkvist, 2001). The procedure involves sensory deprivation of the participants to enhance the occurrence of psi. Two participants are secluded in separate rooms with one participant (the sender) instructed to telepathically send a target image to the other participant (the receiver). The rationale for the procedure is based on the idea that altered states of consciousness (i.e., dreams or trance-like states) are conducive to psi experiences (Bem & Honorton, 1994; Honorton & Harper, 1974). By creating a homogenous visual and auditory field in the ganzfeld, and thereby reducing sensory input normal to our daily lives, psi signals and transfers are more likely to occur.

Ganzfeld studies first gained prominence in the 1970s (Braud, Wood, & Braud, 1975; Honorton & Harper, 1974; Parker, 1975) and given the extraordinary nature of psi, have unsurprisingly yielded inconsistent and controversial results. Many years of psi ganzfeld studies have found both positive and negative findings for the existence of psi. Critics of psi ganzfeld studies have argued that positive findings can be attributed to methodological flaws with the ganzfeld procedure (Hyman, 1985). Briefly again, the ganzfeld procedure involves the seclusion of two participants in two separate rooms. Typically, the procedure uses a pool of target images grouped in judging sets of 4 images. One image from each set is randomly selected and presented to the sender to send to the receiver. At the end of a trial, the judging set is presented to the receiver for evaluation. This set contains the target image and three decoys presented in a random order. Proper randomization is essential to the selection of the judging set from a larger pool of stimuli.
sets, the selection of the target image for the sender and the presentation of the images from a set for evaluation. Hyman (1985), alarmingly, found studies that used inadequate randomization (i.e. hand-shuffling of images, or coin-tossing) or no randomization at all (these made up 74% of the studies he surveyed).

Other procedural problems can also be construed as contributing to false positive findings in psi ganzfeld studies. For example, 55% of the studies surveyed by Hyman (1985) were found to use single targets. Usage of single targets involves the sender and receiver using the same target image for both sending and judging purposes. Without using duplicate sets of images the procedure may be susceptible to sensory leakage problems. When allowing the receiver to judge from the same set of images used for the receiver, it leaves open the possibility that the receiver may respond correctly by picking up physical clues left by the sender.

After an exchange over the methodological adequacy of past ganzfeld procedures used in psi research, Ray Hyman, a critic, and Charles Honorton, a supporter of psi studies, established what have become recognized guidelines to minimize and prevent criticisms of methodological inadequacy in future studies. The recommendations contained in their famed “joint communiqué” consist of a set of more stringent guidelines that have been since recognized as standard in ganzfeld studies (Hyman & Honorton, 1986). The recommendations put forth in the joint communiqué involve control for sensory leakage, randomization of target images used, procedures used in identifying the target images, problems with multiple analyses and the file-drawer problem, documentation procedures and general statistical analysis issues. In an attempt to establish procedures that adhered to these recommendations more closely, Honorton and
his colleagues (Honorton et al., 1990) conducted a series of studies based on an automated ganzfeld procedure system (“autoganzfeld”) that minimized sensory leakage and other related problems. The autoganzfeld experiments used a fully automated system that minimized the need for monitoring and carrying out of experimental procedures by human experimenters.

Meta-analyses of Ganzfeld Studies

Replicability of research findings has been as much an issue in parapsychology as in other areas of psychology. Most have argued that replication is the key to the successful demonstration of parapsychological phenomena (Krippner et al., 1993). Moreover, researchers have highlighted the idea that replicability should be based on effect size and not on mere agreement of null hypothesis testing conclusions (Krippner et al; Rosenthal, 1986; Utts, 1991). This idea is buttressed by the awareness of difficulties related to having adequate power to detect small effect sizes (Cohen, 1992).

Meta-analytic reviews of existing ganzfeld studies have been attempted by numerous parapsychologists to demonstrate and call attention to the explanation of psi effects. One of the early critical examinations of the database of ganzfeld studies was a meta-analysis conducted by Ray Hyman (1985). Hyman examined 42 ganzfeld psi studies reported between 1974 and 1981. It was previously reported that of these, 23 achieved overall significance (55%) and are suggestive of the existence of psi. Hyman, however, proceeded to evaluate and extricate methodological problems that may have accounted for the positive findings. These included multiple analyses without controlling for Type I error rates and other procedural inadequacies highlighted earlier.
Consequently, he estimated the rate of success to be at most 30%, when 5% is expected under the null. In the same issue of the *Journal of Parapsychology*, Honorton (1985) responded with a competing meta-analysis and rejoinder. In it, he conducted a more focused analysis of the database by applying a uniform test on a common index across a smaller subset of the 42 studies. His analysis found that of the 28 studies that were examined, 43% were found to be significant at the .05 level (combined Stouffer $z$ score was 6.6, $p = 2.1 \times 10^{-11}$). It is important to note that both Hyman and Honorton’s reports are still framed in the context of the vote-counting framework, which in and of itself possesses many problems (Bushman, 1994; Hedges & Olkin, 1980) that are also acknowledged by these researchers.

In a commentary that followed this debate, Rosenthal (1986) tentatively offered his estimate of the psi effect given his review of the database of available studies. Even taking into account the various problems and flaws reported by Hyman (1985) and Honorton (1985), Rosenthal estimated the obtained accuracy rate to be about 33%, when 25% is expected under the null. At this point, the conclusion to be drawn from these meta-analyses appears to be mixed. Recent meta-analyses have arrived at additional conflicting results and heated debates.

Ben and Honorton (1994)\(^1\) arguing that the ganzfeld procedure has produced a database of studies with replicable effect sizes, conducted a meta-analysis yielding a mean effect size ($ES^2$) of 0.162 (Stouffer $Z = 2.52$, $p = 5.90 \times 10^{-3}$, one tailed) suggesting a significant psi effect in the studies they examined. Milton and Wiseman (1999) subsequently attempted to replicate this meta-analysis by only including studies that presumably met the methodological guidelines set forth by Hyman and Honorton (1986).
Milton and Wiseman argued that the exclusion criterion was implemented because studies conducted prior to the joint communiqué are contaminated by experiments with inadequate methodological controls. In their meta-analysis of only studies that began after 1987, Milton and Wiseman found a mean $ES$ of 0.013 (Stouffer $Z = 0.70, p = .24$, one-tailed), refuting the results of Bem and Honorton. Storm and Ertel (2001) subsequently published another meta-analysis, with criticisms of Milton and Wiseman’s meta-analysis methodology, yielding a significant mean $ES$ of 0.138 (Stouffer $Z = 5.66, p = 7.78 \times 10^{-9}$; see Appendix C for more information on these meta-analyses). Subsequent publications by Milton and Wiseman (2001; 2002) and Storm and Ertel (2002) have criticized each other’s methodology and conclusions regarding the true effect size extracted from existing psi ganzfeld research databases.

Although meta-analysis is promising as a quantitative and integrative methodology for research literatures (Rosenthal & DiMatteo, 2001), inconsistencies between meta-analytic findings make the creation of a cumulative scientific enterprise an elusive endeavor (Schmidt, 1992). These inconsistencies in both parapsychology and other areas of psychology are due in part to complex but interrelated factors that include the file drawer problem and the discipline’s reliance on a null hypothesis significance testing as the dominant statistical analysis framework.

File Drawer Problem

Methodological issues concerning the procedure and analysis of psi ganzfeld studies are a source of disagreement between the camp of researchers arguing for the existence and those arguing for the non-existence of psi. Many have addressed numerous
procedural criticisms that may contribute to erroneous interpretations of results (Honorton, 1985; Honorton & Hyman, 1986; Hyman, 1985). Still others differed on the methodological decisions made in meta-analyses of psi ganzfeld studies (e.g., which studies to include/exclude in the meta-analyses and the unidirectionality/bidirectionality of psi effects; Milton & Wiseman, 1999, 2001, 2002; Storm & Ertel 2001, 2002). Underlying these disagreements is a methodological artifact that plagues the studies reviewed by both camps: Rosenthal’s (1979) *file drawer problem*. All things being equal, studies that produce significant results (in the null hypothesis testing sense; e.g., $p < 0.05$) are more likely to be published than studies that do not. Even if the null hypothesis is true in the population of psi phenomena, a proportion of studies equal to alpha will appear to demonstrate a significant effect. The degree to which these studies are being published in the literature is the extent to which the literature reflects an overly optimistic estimate of the psi effect. The file drawer problem can be construed as the degree of uncertainty researchers have of the bias of published psi literature. Hence, the establishment of the existence of psi cannot be clearly judged from published literature without some understanding of the severity of the file drawer problem.

The file drawer problem has been recognized early on by parapsychology associations and researchers. For example, there has been a standing policy against selective reporting of significant results in its meetings and journals by the Parapsychological Association since 1975. Many of the important parapsychology researchers and peer-reviewed journals in the area of parapsychology are affiliated with the Parapsychological Association. In a survey examining the extent of selective reporting, Blackmore (1980) uncovered 19 unpublished studies, although with further
analysis she concluded that this did not constitute a significant bias to the overall proportion of positive findings in the published literature. Using the fail-safe n methodology (Rosenthal, 1979) of estimating the number of studies (averaging a zero effect size) needed to render a significant meta-analytic finding non-significant, Honorton (1985) concluded that the 423 studies needed constituted an unlikely number for studies that did not make it to publication. Storm & Ertel (2001) similarly concluded that the 857 unpublished, non-significant studies needed in the file drawer also constituted evidence against a chance explanation of their meta-analysis findings.

Related to this uncertainty is what Hyman (1985) termed the retrospective study problem. Retrospective studies are studies that were initially exploratory or pilot trials. The tendency for these trials to be counted as successful studies due to findings of significant results (or for these trials to not be counted as successful studies when findings are not significant) serves to inflate the overall success rate of published psi studies. This is particularly a problem given the fact that pilot studies are frequently underpowered and this increases the likelihood of committing Type I errors (Maxwell, 2004). Despite attention to the problem, the extent to which the file drawer problem remains a factor in systematically biasing the published psi effect sizes is unknown.

Limitations of the Null-Hypothesis Significance Testing Framework

The dominance of null-hypothesis significance testing (NHST; also known as the frequentist framework; Gigerenzer, 1993) as a primary statistical analysis framework in the social sciences has resulted in some researchers characterizing it as an “overadoption”
of a methodology by psychologists (Hubbard, Parsa, & Luthy, 1997). Meehl (1978), for example, argued that,

I believe that the almost universal reliance on merely refuting the null hypothesis as the standard method for corroborating substantive theories in the soft areas is a terrible mistake, is basically unsound, poor scientific strategy, and one of the worst things that ever happened in the history of psychology (p. 817).

Meehl is joined by a number of others equally critical of the overreliance of psychologists on NHST (e.g., Cohen, 1994; Falk & Greenbaum, 1995; Gigerenzer, 1993; Oakes, 1986; Schmidt, 1996). In response, supporters have forwarded the argument that problems arise only when NHST is inappropriately used and its results misinterpreted (Chow, 1998; Frick, 1996; Hagan, 1997; Mulaik, Raju & Harshman, 1997; Nickerson, 2000).

The criticisms of the overreliance on NHST is relevant in parapsychology research in that many of the problems associated with its misuse and misinterpretation are possible contributing factors to an inconsistent psi ganzfeld literature. Several major criticisms are briefly reviewed here.

**Binary decisions**

Numerous researchers have objected to the binary decision making process of either accepting or rejecting the null hypothesis (Cohen, 1994; Folger, 1989; Howard et al., 2000; Rosnow & Rosenthal, 1989). This is problematic in several ways. First of all, it has been argued that the null hypothesis is false to begin with (Cohen, 1994; Kupfersmid, 1988). Greenwald (1975) argued that, “If by the null hypothesis one refers to the hypothesis of exactly no difference or exactly no correlation, and so forth, then the initial probability of the null hypothesis being true must be regarded effectively as zero” (p. 6). This results in a situation in which given a large enough sample, any null
hypothesis is theoretically rejectable. Consequently, even small (and theoretically non-significant) effect sizes can be found to be statistically significant. As indicated by many others, as a consequence of this scenario it becomes undesirably likely that problematic conclusions can be erroneously drawn from underpowered studies (Cohen, 1992; Howard et al., 2000; Maxwell, 2004).

Secondly, the binary acceptance or rejection of the null hypothesis has become the basis for editorial decisions on whether an article is accepted for publication (a condition that leads to varying degrees of the file drawer problem; Greenwald, 1975; Kupfersmid, 1988; Rosenthal, 1979). As discussed earlier, this is deleterious to the research enterprise because it contributes to a biased and often confusingly inconsistent literature.

*Probabilistic misinterpretation*

NHST makes it attractive to use *modus tollens* reasoning when obtaining statistical significance in a given analysis. This reasoning is used to deny the antecedent (i.e., null hypothesis) by denying the consequent (i.e., data). NHST, however, cannot be reasoned in this way because of the probabilistic nature of the initial premise (i.e., “If the null hypothesis is correct, then the observed data are highly unlikely”; Cohen, 1994; Pollard, 1993). Falk and Greenbaum (1995) referred to this as the “illusion of probabilistic proof by contradiction” (p. 78). They go on to say, “While a contradiction definitely disproves the premise from which it is drawn, the probabilistic counterpart of that logical deduction does not hold” (p. 78). To illustrate the consequences of this illusion Falk and Greenbaum identified a study conducted by Hardy, Harvie and Koestler (1973) in which the authors presented the outcomes and results of an ESP experiment by
stating, “Taken together, the receivers scored significantly beyond chance . . . with a calculated probability of 3,000 to 1 against it being just chance” (p. 117). Numerous other examples of misinterpretations are widespread and well-documented (Dar, Serlin & Omer, 1994; Kline, 2004; Nickerson, 2000), even those involving respected statisticians such as Jacob Cohen and Sir Ronald Fisher (Cohen, 1994; Gigerenzer, 1993; Oakes, 1986).

Pollard (1993) framed this confusion in a different context, that of the probability of a Type I error in any given analysis. In NHST, Type I error is controlled at a pre-selected level (i.e., \( \alpha = .05 \)). However, this is conditional on the null being true and is what Pollard calls *conditional prior probability*. It is prior because it exists prior to the carrying out of the experiment. What is probably of greater interest to the researcher is the *conditional posterior probability* of Type I error, whereby the conditional refers to the rejection of the null and posterior refers to the fact that it is derived after the examination of empirical data. Pollard emphasized that, “It is crucial to remember that the (posterior) probability of \( H_0 \) being true, given that it has been rejected, is not the same as the (prior) probability of \( H_0 \) will be rejected, given that it is true” (p. 453). It is the former, the probability of a Type I error after the rejection of the null, that is typically of greater interest (which in actuality is undeterminable).

The widespread and common misinterpretation of NHST results and conclusions is in part a function of the practitioner’s desire for results that the approach is unable to provide (Gigerenzer, 1993; Kline, 2004). For example, with NHST nothing can be said of the effect of interest (i.e., \( \psi \)) because the results are framed in such a way that yields the probability of rejecting the data given a null hypothesis (no effect; that is
In any given study, what is of interest is the likelihood of an effect given the findings of the data (\(p[\text{hypothesis/data}]\); Cohen, 1994). Furthermore, in NHST the probability that an experimental hypothesis is true is unanswerable because of the trappings of the binary framework and because the effect of interest is presumed to be constant in the population (Howard et al., 2000).

Bayesian Analysis Alternative

Howard and his colleagues (2000) have argued that meta-analytic and Bayesian approaches are viable alternatives to traditional NHST. They have demonstrated that where NHST is susceptible to misuse and misinterpretation, the two alternatives provide less confusing and more useful information regarding the effects of interest. Although it is possible to arrive at equivalent conclusions with both meta-analytic and Bayesian approaches, there are qualities to the Bayesian framework that makes it a preferable approach. One of the main reasons is that meta-analysis, as conceptualized in the NHST fashion is still subjected to the problems discussed in the NHST section. A brief review of what makes the Bayesian approach preferable is attempted here.

*Probability as belief*

In contrast to NHST, the Bayesian approach conceptualizes probability as a measure of personal belief of the effect of interest. Under the NHST framework, the obtained results yield \(p[\text{data/null hypothesis}]\), which as previously discussed is not what most researchers are interested in. In contrast, by way of Bayes’ theorem, the measure of personal belief is altered as a function of the findings of empirical data. In other words, the Bayesian approach essentially yields the \(p[\text{hypothesis/data}]\), which is what most
interests researchers (Cohen, 1994). Gigerenzer (1993) has aptly termed the frequentist’s desire to obtain this exact probability as Bayesian Id’s wishful thinking. The Bayesian approach moves away from dichotomous acceptance and rejection of hypotheses, allowing for the likelihood of a hypothesis to vary as a function of a probability.

*Cumulative science*

The cumulative function of research findings in the development of theories is an important component of the scientific enterprise (Cooper & Hedges, 1994; Schmidt, 1992). Successful replication (Rosenthal, 1990) and meta-analytic procedures (Rosenthal & DiMatteo, 2001) in psychological research have been promoted in the interest of research synthesis. The Bayesian approach, as suggested earlier, encourages the replication and synthesis of data, as research hypotheses are strengthened with the collection of new data. By moving away from dichotomous acceptance and rejection decisions, the approach also discourages drawing conclusions from a single study. The reliance on observed data only (as opposed to the long-term frequency approach in which non-observed data is accounted for in the calculation of $p$), encourages the collection of more data in the assessment of a hypothesis and avoids statistical analysis problems when experiments are stopped in the frequentist approach (Berger & Berry, 1988).

**Six Psi Ganzfeld Replication Studies**

Six psi ganzfeld studies were conducted with each study consisting of twenty trials. The results were analyzed from NHST, meta-analytic and Bayesian approaches. Although not the main goal of this study, the six replications will also examine the
efficacy of a ganzfeld versus non-ganzfeld procedures and also positive predictors of psi performance.

Ganzfeld Versus Non-Ganzfeld Procedure

Although the ganzfeld procedure has been the dominant methodology used to study psi, there has been a dearth of studies comparing the efficacy of a ganzfeld procedure with a non-ganzfeld procedure. Murre et al. (1988) conducted a ganzfeld psi experiment with a control condition and found neither a significant psi effect nor evidence of an advantage of the ganzfeld condition over a control condition. There are two important components that define the ganzfeld procedure: sensory deprivation and a 30 minute “mentation” period. In the following six studies, a pair of participants was engaged in a psi inducing ganzfeld procedure followed by eight non-ganzfeld procedures. The non-ganzfeld procedure involved the pair of participants switching roles (i.e., the sender in the ganzfeld trial was the receiver in the non-ganzfeld trials, and vice versa) and engaging in eight short (mentation period lasting only 2 minutes) non-ganzfeld (lack of sensory deprivation) trials.

Positive Predictors of Psi Performance

Research has indicated that certain characteristics associated with the participants are positively predictive of psi successes (Dalton, 1997; Honorton, 1997). One of the most consistent predictor is the participants’ personal belief in psi abilities, also known as the “sheep-goat” effect (Lawrence, 1993). Participants who hold a stronger belief in psi abilities show higher success rates than those who relate a lower or no belief in psi.
(Honorton, 1985; Honorton & Schechter, 1987). To assess this predictor and several others, participants completed a questionnaire containing these items.
METHOD

Participants

Participants were recruited from the student population at the University of Notre Dame in South Bend, Indiana. The sample consisted of 240 participants (55.8% female). The mean age of the participants is 19.7 years. A total of 120 long ganzfeld trials and 937 short ganzfeld trials were conducted over six studies with 20 trials in each.

Measure

To assess the predictors of psi performance, a number of questions were included in a questionnaire that was completed by participants prior to their participation in the trials (Appendix B). The questionnaire consisted of self-report items that previous studies have shown to be positive predictors of psi (Dalton, 1997; Honorton, 1997). These positive predictors include various participant characteristics: (a) previous personal psi experiences (Honorton & Schechter, 1987), (b) strong personal belief in the existence of psi (Lawrence, 1993), (c) previous laboratory psi testing experience (Honorton & Schechter, 1987), (d) practice of a mental discipline, such as meditation (Honorton & Schechter, 1987), (e) musically or artistically talented (Schlitz & Honorton, 1992), and (f) when a pair of participants in a ganzfeld study either know each other or are related (Broughton & Alexander, 1995; Honorton et al., 1990). Additionally, demographic information was also included in the questionnaire.
Target stimuli

Target stimuli were images obtained from the *National Geographic* website (http://www.nationalgeographic.org). The images consisted of a variety of subjects including people, nature and animals. Eighty selected images were grouped into twenty judging sets containing four images each. Each set contains images that are chosen by two researchers to be dissimilar in content from each other. For identification purposes, each of these sets is numbered from 1-20. The images within each set are numbered from 1-4. Each image can be identified by a set number and an individual number printed on the back of the image. Both experimenters in the sender and receiver rooms will be using duplicates of the same twenty sets of images.

Procedure

A script for the study was used by the experimenters involved in conducting the trials (Appendix A).

*Introductory period.* Participants in each pair were randomly assigned to be receivers and senders in the long ganzfeld trials. Assignments were switched for the subsequent short non-ganzfeld trials, meaning senders in the long ganzfeld trials became the receivers in the short non-ganzfeld trials, and vice versa. The introductory period consisted of the introduction of the participants to each other and to the experimenters to facilitate a warm and welcoming atmosphere. Participants subsequently completed informed consent forms and the self-report measure on positive predictors of psi. After the introductory period, the participants were secluded in two separate non-adjacent rooms.
Experimental conditions set-up. In both rooms, participants were seated comfortably in a recliner. Prior to the beginning of the trial, both receiver and sender listened to a 10 minute guided relaxation recording through a headphone. At the end of the guided relaxation exercise the receiver was placed in a ganzfeld condition whereby the participant is fitted with translucent hemispheric goggles (made out of halved ping-pong balls) and headphones. A sixty-watt red-filtered floodlight was placed in front of the receiver. For the duration of the entire trial, the receiver listened to a recording of white noise. The white noise and floodlights filtered through the goggles create a homogenous visual and auditory field for the participant.

Presentation of targets. One image from each judging set was randomly selected in each trial to be the target image. The randomized selection of the set and target image from that set was conducted by an experimenter not involved in conducting current trial, ahead of the trial. This procedure prevents problems associated with sensory leakage by one of the experimenters to the participants, as the experimenters involved in running the current trials are unaware of the target selections until the trials begin.

Note-cards identifying randomly selected set numbers (with numbers between 1-20) were used by the experimenters for each trial. Both experimenters in the sender and receiver rooms possessed the same sequence of note-cards. The experimenter in the sender room revealed the set to be used from the note-cards and then referred to a random number table (of numbers 1-4) to determine the image within each set that will serve as the target. The experimenter in the receiver room along with the receiver did not have factual knowledge of the true target image used until the end of the trial. When presented
with the target image, senders were instructed to concentrate on the target image for 30 minutes while the receiver is in the ganzfeld condition.

**Mentation period.** The mentation period is the period when the sender attempts to telepathically send the target image to the receiver and the receiver reports the mental experiences during this period. Specifically, the receiver was instructed to verbally report whatever thoughts, images and feelings that occur during the ganzfeld into a microphone. A one-way radio was used to transmit the receiver’s verbal mentation report to the sender in the other room. During this time, the experimenter in the receiver room took notes of the oral mentation by the receiver. The sender was asked to focus on conveying the image while listening to the receiver’s oral mentation transmitted through the radio.

**Judging period.** The judging period involves the receiver evaluating a set of four images and picking the potential target that was being sent by the sender during the mentation period. At the end of the mentation period, the experimenter in the receiver room revealed the judging set that contained the target image (the same set from which the target image was presented to the receiver earlier) by referring to the randomized note-cards. The experimenter then referred to a table to randomly sequence the images to be presented to the receiver. One of the images in the set is the target and the other three decoys. Prior to the judging of the targets, the experimenter read the written mentation report back to the receiver. The receiver was instructed to process the mentation experience with the help of the mentation notes and to pick the target from the set of four images presented. The experimenter recorded the selected target after the selection is made.
At the end of the judging period, the participants were brought back into the same room and debriefed. The experimenter in the sender room revealed to the receiver the correct target used during the mentation period.

*Short Non-Ganzfeld Trials*

After the long ganzfeld trial, the sender and receiver switched roles and proceeded with eight 2-minute short non-ganzfeld trials. Each participant served as the sender/receiver for all eight trials. These short trials were not conducted with the ganzfeld condition. The two participants were still secluded in separate rooms, but the mentation period did not involve sensory depravation through the use of white noise and red floodlight filtered through goggles. The recording and transmittance of mentation report will also be omitted the short non-ganzfeld trials. All other aspects of the procedure are identical to the long ganzfeld trials.
RESULTS AND ANALYSES

**NHST Approach**

Six studies were conducted and the results showing the number of correct hits for each study is presented in Table 1.

*Short Non-Ganzfeld Trials.* In the first study, an exact binomial test comparing the observed proportion of hits to the null hypothesis proportion of .25 was non-significant, \( p = .96 \). The subsequent analyses of the remaining studies were all non-significant as well. These results are presented in Table 2. When the six studies were combined as one study, the resulting exact binomial test is non-significant, \( p = .58 \). Given these results, it appears that the short non-ganzfeld trials failed to yield significant psi findings. To simplify the comparison and analysis of the study findings from NHST, meta-analytic and Bayesian approaches, the short non-ganzfeld trials are excluded from subsequent analyses.

*Long Ganzfeld Trials.* In the first study, nine of the twenty pairs of participants yielded correct hits. This translates to a 45% hit rate. An exact binomial test yielded statistically significant results, \( p = .04 \) (one-tailed). The subsequent analyses of the remaining studies are presented in Table 3. In summary, the six studies yielded one significant finding and five non-significant findings.
TABLE 1

SHORT NON-GANZFELD AND LONG GANZFELD TRIALS

HIT RATES

<table>
<thead>
<tr>
<th>Study</th>
<th>Total Number of Trials</th>
<th>Number of Correct Hits</th>
<th>Hit Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short Non-Ganzfeld Trials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>146</td>
<td>28</td>
<td>19.2</td>
</tr>
<tr>
<td>2</td>
<td>159</td>
<td>42</td>
<td>26.4</td>
</tr>
<tr>
<td>3</td>
<td>159</td>
<td>47</td>
<td>29.6</td>
</tr>
<tr>
<td>4</td>
<td>153</td>
<td>38</td>
<td>24.8</td>
</tr>
<tr>
<td>5</td>
<td>160</td>
<td>35</td>
<td>21.9</td>
</tr>
<tr>
<td>6</td>
<td>160</td>
<td>42</td>
<td>26.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long Ganzfeld Trials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>9</td>
<td>45.0</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>8</td>
<td>40.0</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>4</td>
<td>20.0</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>5</td>
<td>25.0</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>6</td>
<td>30.0</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>4</td>
<td>20.0</td>
</tr>
</tbody>
</table>
### TABLE 2

**EXACT BINOMIAL TEST OF SHORT NON-GANZFELD TRIALS**

**HIT RATES**

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Hits</th>
<th>Hit Rate (%)</th>
<th>p (1-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>146</td>
<td>28</td>
<td>19.2</td>
<td>.96</td>
</tr>
<tr>
<td>2</td>
<td>159</td>
<td>42</td>
<td>26.4</td>
<td>.37</td>
</tr>
<tr>
<td>3</td>
<td>159</td>
<td>47</td>
<td>29.6</td>
<td>.11</td>
</tr>
<tr>
<td>4</td>
<td>153</td>
<td>38</td>
<td>24.8</td>
<td>.55</td>
</tr>
<tr>
<td>5</td>
<td>160</td>
<td>35</td>
<td>21.9</td>
<td>.84</td>
</tr>
<tr>
<td>6</td>
<td>160</td>
<td>42</td>
<td>26.3</td>
<td>.39</td>
</tr>
<tr>
<td>1-6 (Combined)</td>
<td>937</td>
<td>232</td>
<td>24.8</td>
<td>.58</td>
</tr>
</tbody>
</table>

**NHST Interpretation.** In six separate exact binomial tests, only one of the studies was statistically significant (Study 1). It is plausible that after the significant finding in the first study, researchers examining the data in a strict NHST approach might conclude that the psi phenomenon exists. Although this is uncommon in the psi literature, the larger psychological literature is peppered with examples of single experiment studies whose results are used to definitively support or refute the effect of interest. Moreover, it has been suggested elsewhere that the accumulation of a series of dichotomous accept/reject results do not lend easily to interpretation (Howard, Maxwell & Fleming,
2000) and that there are problems associated with existing methods of accumulating these results (Bushman, 1994; Hedges & Olkin, 1980). With the advent of effect size measures and meta-analytic techniques, focusing on the mere acceptance or rejection of the null hypothesis has become unacceptable by the psychological community.

*Ganzfeld versus non-ganzfeld procedure.* A t-test comparing the hit rates under the long ganzfeld and short non-ganzfeld procedures yielded non-significant results, \( t(1055) = -1.24, p = .21 \) (two-tailed). The overall hit rate under the ganzfeld condition is not significantly higher than the overall hit rate under the non-ganzfeld condition.

### TABLE 3

**EXACT BINOMIAL TEST OF LONG GANZFELD TRIALS**

<table>
<thead>
<tr>
<th>Study</th>
<th>( N )</th>
<th>Hits</th>
<th>Hit Rate (%)</th>
<th>( p ) (one-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>9</td>
<td>45</td>
<td>.04</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>8</td>
<td>40</td>
<td>.10</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>4</td>
<td>20</td>
<td>.77</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>5</td>
<td>25</td>
<td>.59</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>6</td>
<td>30</td>
<td>.38</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>4</td>
<td>20</td>
<td>.77</td>
</tr>
</tbody>
</table>
Meta-analytic Approach

Figure 1 represents the meta-analytic treatment of the long ganzfeld data presented in Table 1. The figure shows the percentage of correct hits as studies are meta-analytically combined. A 95% confidence interval (CI) was constructed around the hit rate.

![Meta-analytic approach interpretation.](image)

Figure 1. Meta-analysis of the six long ganzfeld studies, with the mean hit rate and 95% confidence interval represented. The dotted line represents the hit rate under the null hypothesis.

Meta-analytic approach interpretation. One might immediately notice that the CI in Study 1 contains the null value of 25%. This is the same as saying that Study 1 yielded non-significant results (with $\alpha = .05$). This is inconsistent with the findings from the NHST section in which Study 1 was found to be statistically significant using an exact binomial test. The contradiction is actually a function of the estimation procedure.
employed when constructing CIs for dichotomous data. Whereas a binomial test provides an exact probability under the null hypothesis, constructing CIs for binary variables would generally entail an approximation using a standard normal distribution. For this set of data, the Wald method for constructing CI was used. Agresti and Coull (1998) have demonstrated that coverage probabilities are poor when using such “exact” methods (especially when sample size is small and/or when the observed proportion deviates from .5) and have recommended alternative methods such as the adjusted Wald method or the score method in constructing CIs for binomial proportions. Coverage probabilities using the Wald method tends to be overly conservative, resulting in an overly large coverage. This explains why the CI for Study 1 includes the null value in Figure 1, whereas the binomial test from the NHST section yielded a significant finding. The adjusted Wald and score methods were shown to have coverage probabilities close to the nominal confidence level, even when sample size is small and/or when \( p \) is close to 0 or 1\(^4\). To be consistent with the analyses in the following section, I have opted for the more common Wald method CIs.

After the addition of Study 2 to the analysis, we find that the CI does not contain the null value, therefore suggesting that the hit rate after 2 studies, 42% is statistically significant. With the addition of the third study, we find that the CI once again contains the null value, bringing the hit rate to a non-significant level. With each addition of the subsequent three studies, the analysis remains non-significant. With the addition of the last study, the hit rate is further reduced to 30%. From the perspective of a meta-analyst, it appears that although we may be approximating the effect size of interest, we are at the same time unsure whether this estimate is significantly a chance occurrence\(^5\).
The meta-analytic approach is an improvement over the strict NHST approach in several ways. The meta-analytic focuses the analysis on the effect size of interest rather than dichotomous acceptance/rejection of the null hypothesis. Whereas the interpretation of the results from the previous section might conclude that psi phenomena is unlikely to exist given the dominance of non-significant results, a meta-analysis yields an estimate of that effect. When a CI is computed, the meta-analytic approach also yields information to ascertain the statistical significance at each step of the analysis.

Bayesian Approach

Bayesian priors. The Bayesian approach necessitates the specification of an a priori belief for the effect of interest. These so-called priors are a probabilistic estimation of personal belief. It may be arrived at by empirical means (such as a meta-analysis of existing literature) or by personal declarations of a belief (which in the absence of a directional belief, is called a “noninformative” prior). To demonstrate how discrepant priors are driven by the data, three priors were constructed using a similar methodology employed by Howard, Maxwell and Fleming (2000). Sixteen graduate student members of the psychology department at the University of Notre Dame were surveyed for their personal beliefs in the existence of psi in a ganzfeld experimental context. It was explained to the students that in a psi ganzfeld study a 25% hit rate represented chance occurrence whereas hit rates increasingly higher than the null value represented evidence for psi phenomena. Each student provided (a) a hit rate estimate of the psi phenomena representing their personal beliefs, and (b) the number of studies with results in the opposite direction that it would take to change their mind (an estimate of the confidence in (a)). Participants were given the information that each study consisted of 20 trials and
that 9 trials resulting in hits is statistically significant. The information provided by those
surveyed was used to construct a beta distribution defined by \([a, b]\) (Pruzek, 1997). A
beta distribution’s mean is defined by \(a/(a + b)\) and its variance by \(ab/(a + b)^2(a + b + 1)\).
For example, one of the students surveyed may give a response of 25% hit rate (meaning
she does not believe in the existence of psi) and stating that she needs to observe 10
significant studies for her to change her mind about the existence of psi. This translates
to a beta distribution of \([50, 150]\). Someone who believes strongly in psi might respond
with a 50% hit rate estimate with 50 null studies for him to change his mind. This is
equivalent to a beta distribution of \([500, 500]\). With more studies needed to change a
student’s mind, the variance decreases, and the confidence in the estimate increases. The
estimates given by the individuals surveyed were subsequently rank ordered, with the
median of the top and bottom quartiles representing the priors of “believers” and the
“non-believers” of psi. These hypothetical priors are meant to represent extreme prior
beliefs for the existence of psi. Finally, a third noninformative prior was included in the
analysis as well.

The three priors (believer, non-believer, and noninformative) were used in a
Bayesian analysis of the six studies. In Bayesian analyses, priors are combined with
results from the data (the likelihood) to yield a posterior probability (Howard, Maxwell &
Fleming, 2000). In the case of the following analysis, the posterior probability from
combining the initial priors with the results from Study 1 becomes the prior probability to
be combined with the results from Study 2, and so on. As mentioned earlier, the priors
are defined by a beta distribution \([a, b]\). The likelihood and posterior probabilities are
similarly defined by the same parameters. To distinguish among them, let the prior
probability be defined by \([a', b']\), the likelihood probability by \([a*, b*]\), and the posterior probability by \([a'', b'']\). The posterior beta parameters are simply an additive function of the prior and likelihood parameters, so that \(a'' = a' + a*\), and \(b'' = b' + b*\) (Pruzek, 1997).

Take the non-believer prior as an example. Its beta distribution is defined by \([110, 330]\). The likelihood distribution for Study 1 is defined by \([9, 11]\). The posterior beta parameters are therefore defined by \([(110 + 9), (330 + 11)]\), which yields the posterior distribution \([119, 341]\). The mean of this distribution is 34.5 and the SD is 2. Table 4 represents the initial prior and the posterior mean hit rates and standard deviation after the inclusion of each of the six studies.

**Bayesian approach interpretation.** As shown in Table 4, we can see that the initial priors for the non-believers and believers are quite discrepant. The non-believers hold that there are no psi phenomena and that in ganzfeld studies participants will correctly identify the target image only 25% of the time (at chance level). The hypothetical believers, on the other hand, hold that the hit rate would be at 44.8%, suggesting the existence of psi. In the noninformative prior column, we see that since there is no a priori belief held, it does not factor into the posterior probability after the first study. Hence, the values in that column are equivalent to a meta-analytic treatment of the data.
TABLE 4

MEAN HIT RATES (IN PERCENTAGE) AND STANDARD DEVIATIONS
FOR INITIAL PRIOR BELIEFS AND POSTERIOR PROBABILITIES

<table>
<thead>
<tr>
<th></th>
<th>Noninformative Prior</th>
<th>Non-Believer Prior</th>
<th>Believer Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit Rate</td>
<td>SD</td>
<td>Hit Rate</td>
</tr>
<tr>
<td>Initial prior</td>
<td>-</td>
<td>∞</td>
<td>25.0</td>
</tr>
<tr>
<td>After Study 1</td>
<td>45.0</td>
<td>11</td>
<td>34.5</td>
</tr>
<tr>
<td>After Study 2</td>
<td>42.5</td>
<td>8</td>
<td>36.3</td>
</tr>
<tr>
<td>After Study 3</td>
<td>35.0</td>
<td>6</td>
<td>32.3</td>
</tr>
<tr>
<td>After Study 4</td>
<td>32.5</td>
<td>5</td>
<td>30.9</td>
</tr>
<tr>
<td>After Study 5</td>
<td>32.0</td>
<td>5</td>
<td>30.7</td>
</tr>
<tr>
<td>After Study 6</td>
<td>30.0</td>
<td>4</td>
<td>29.2</td>
</tr>
</tbody>
</table>

The graphical representation of the Bayesian analysis is displayed in Figure 2. Ninety-five percent confidence intervals were constructed around the mean hit rates, as defined by \( \frac{a}{a + b} \pm z_{\alpha/2} \sqrt{\frac{ab}{(a + b)(a + b + 1)}} \) (Pruzek, 1997). One notices that not only are the believer and non-believer priors discrepant, their CIs also do not overlap with each other. This is a function of the high confidence in the two extreme views which result in relatively tight confidence intervals. It is evident from the way that data is subjected to a Bayesian analysis described above and from demonstrations by Howard,
Maxwell and Fleming (2000) that low confidence priors (wide CI) will be more influenced by the data than would be higher confidence priors (narrow CI).

![Graph showing mean hit rates and 95% confidence interval for Bayesian analysis using noninformative (N), non-believer (NB) and believer (B) prior probabilities. The dotted line represents the hit rate under the null hypothesis.](image)

Figure 2. Mean hit rates and 95% confidence interval for Bayesian analysis using noninformative (N), non-believer (NB) and believer (B) prior probabilities. The dotted line represents the hit rate under the null hypothesis.

The Bayesian analysis using a noninformative prior is identical to that of the meta-analysis described in an earlier section. Like the meta-analysis, the hit rate after adding the second study became significant, but remained non-significant with the addition of the remaining four studies.

The non-believer is initially skeptical, but after the addition of the first study the hit rate estimate increases almost ten percentage points, although it still fails to reject the null hypothesis. With the addition of the second study, the mean hit rate again increases.
and the confidence interval continue to miss the rejection level. Adding the remaining three studies, the non-believer’s mean hit rate gradually decreases to 29.2%, but never rejecting the null hypothesis.

The believer began with a 44.8% estimate of the hit rate. With the addition of the first study, it increases slightly and then drops slightly with the addition of the second study. With the addition of the remaining four studies the believer gradually drops to 32.7% with its CI, after the addition of the sixth study, never approaching the statistical non-significance level.

After six studies, the Bayesian analysis using a noninformative prior (also the meta-analysis) resulted in a failure to reject the null hypothesis. The non-believer also fails to reject the null hypothesis, whereas the believer rejects the null hypothesis. As with meta-analyses in general, with the addition of more studies the width of the confidence interval shrinks, representing the growth in confidence in the effect size estimate as more data is considered.

The results so far further demonstrate the improvement of the interpretation of findings from the Bayesian approach over the meta-analytic approach. If interpreted from a strict NHST perspective, the meta-analysis described earlier yields a non-significant finding, failing to provide evidence for psi phenomena. In other words, if the null hypothesis is true (that the magnitude of psi phenomena is zero), then the likelihood of observing the data that we observed (or more extreme) is greater than 5%. One can see the difficulty in making non-convoluted inferences from this logic. Alternatively, the Bayesian interpretation frames this from the perspective of the hit rate estimate, or as the belief held by a researcher. If I held beliefs comparable to the hypothetical believer, I
might, after observing the data from the six studies be still persuaded that psi exists. At the same time, however, due to the findings of the studies I am also much closer to doubting my own beliefs than I did before I began. More importantly, the interpretation of a posterior probability is more straightforward and useful than the conditional probabilities under the NHST framework.

Bias Assessment of the Psi Ganzfeld Literature

As discussed earlier, the extent to which any literature is biased is ultimately unknown. The psi ganzfeld community is arguably more sensitive to problems that contribute to bias than many other areas of psychology, and this may mean that the problem is less severe than in other areas. There are established methods to assess for the likelihood of the publication bias problem, such as using funnel plots (Begg, 1994) or the fail-safe $n$ method (Rosenthal, 1979). At the same time however, these methods may be misleading (Macaskill, Walter & Irwig, 2001; Tang & Liu, 2000) or may yield vague results (i.e., does $n$ studies in the file drawer seem likely for a particular area of research?). Despite the bias problem and difficulties associated with assessing the severity of the problem, current practice in psychology continues to rely on the meta-analysis of published literature to integrate research findings. One might argue that with the accumulation of all new findings, regardless of their results (hence there is no file drawer effect), one would be able to eventually undo the bias.

Let’s assume we can begin to do this by using a set of data whereby we know there is no file drawer effect, that is, the data presented above. Let’s also assume that the psi effect is in reality zero. If we proceed as if a cumulative science is self-correctable, we can examine the impact of adding the six studies to the existing literature. In Table 5,
the Bayesian analysis of three recent meta-analytic results (Bem & Honorton, 1994; Milton & Wiseman, 1999; Storm & Ertel, 2001; see Appendix C for summary of the meta-analyses) with the six current studies is presented. The meta-analytic priors are similar to those presented earlier with the noninformative, believer and non-believer priors, in that there is some discrepancy in their hit rate estimates. The graphical representation of this analysis with 95% confidence intervals is depicted in Figure 3.

TABLE 5

MEAN HIT RATES (IN PERCENTAGE) AND STANDARD DEVIATIONS FOR META-ANALYTIC PRIORS AND POSTERIOR PROBABILITIES

<table>
<thead>
<tr>
<th></th>
<th>BH</th>
<th></th>
<th>MW</th>
<th></th>
<th>SE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit Rate</td>
<td>SD</td>
<td>Hit Rate</td>
<td>SD</td>
<td>Hit Rate</td>
<td>SD</td>
</tr>
<tr>
<td>Meta-analytic</td>
<td>32.2</td>
<td>3</td>
<td>25.9</td>
<td>1</td>
<td>31.0</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>prior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After Study 1</td>
<td>33.0</td>
<td>3</td>
<td>26.2</td>
<td>1</td>
<td>31.1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>After Study 2</td>
<td>33.3</td>
<td>2</td>
<td>26.4</td>
<td>1</td>
<td>31.2</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>After Study 3</td>
<td>32.6</td>
<td>2</td>
<td>26.3</td>
<td>1</td>
<td>31.1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>After Study 4</td>
<td>32.3</td>
<td>2</td>
<td>26.3</td>
<td>1</td>
<td>31.0</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>After Study 5</td>
<td>32.2</td>
<td>2</td>
<td>26.3</td>
<td>1</td>
<td>31.0</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>After Study 6</td>
<td>31.6</td>
<td>2</td>
<td>26.3</td>
<td>1</td>
<td>31.0</td>
<td>&lt; 1</td>
</tr>
</tbody>
</table>

Figure 3. Mean hit rates with 95% confidence intervals for Bayesian analysis using Bem & Honorton (1994) (BH), Milton & Wiseman (1999) (MW) and Storm & Ertel. (2001) (SE) meta-analyses as prior probabilities. The dotted line represents the hit rate under the null hypothesis.

As with the previous analysis, priors with relatively wide CIs will be more impacted by the data. It is not difficult to see that this is a function of the sample size of the meta-analysis. When the sample size (and to be more precise, the number of total trials) of the meta-analysis is large, the variance of the mean estimate is invariably smaller than one with a smaller sample size. Bem and Honorton’s meta-analysis contained the smallest number of trials (see Appendix C) and hence with the addition of the six studies the CIs shrunk more so than the other two meta-analytic priors. The degree to which the point estimate of the prior will be impacted by the data is also in part a function of the variance of the prior.
In proceeding with our goal of “correcting” the literature or “undoing” the bias, we can see from Figure 3, that when a literature is very large, it becomes nearly impossible to accomplish this. In our case with the hypothetical assumption that the psi effect is in reality zero, the effort to drive the two significant findings (BH and SE) to non-significance is even more hopeless. As more null studies are added, the CI is concurrently shrinking such that getting the CIs to eventually contain the null value becomes an impossible task.

Although it may be reasonable to make the assumption that by eliminating the file drawer problem, the six studies conducted will provide unequivocal support for or against a psi effect, the mere elimination of publication bias is not a panacea. This point is strengthened when we examine Bem and Honorton’s (1994) meta-analysis. They reported that their analysis is also not plagued by the file drawer problem as they reported all trials associated with an autoganzfeld research program spanning 6 years. According to that meta-analysis, the effect size is significant with a point estimate of 32.2% (mean $ES = .164$, $p = .005$). Both Bem and Honorton (1994) and the six studies conducted here are free of the publication bias problem, yet their conclusions are not in agreement. The removal of the selective publishing bias does not necessarily yield consistent and conclusive results.

So far I have tried to demonstrate that if a large research literature is flawed and biased, it becomes difficult, if not impossible, to correct or undo the bias by adding more and “better” studies. Furthermore, when conducting studies that are free of publication bias, the results do not unequivocally provide evidence for or against the effect of interest merely because the researchers have ruled out the contaminant of selective reporting.
Positive psi predictors

Only two participants out of the 240 responded affirmatively to Item 3 (previous formal psi laboratory experience; Appendix B) on the questionnaire, so it was excluded from subsequent analyses.

The remaining items (Items 1, 2, 4, 5, 6; Appendix B) were entered into a direct logistic regression, with Items 4, 5 and 6 coded as categorical variables. For Item 5 (“Are you a(n) (please circle all that apply): Artist/Musician/Neither”) a response of “Artist” and/or “Musician” was given a positive value of ‘1’, whereas a “Neither” response was given a value of ‘0’. For Item 6 (“Are you and the other participant: Acquaintances/Close Friends/Related (Family)/None of the Above”), “Close Friends” and “Related (Family)” responses were coded ‘1’, whereas “Acquaintances” and “None of the Above” responses were coded ‘0’. The predictor means and logistic regression results are summarized in Table 6 and Table 7.

The full model containing all five predictors is not reliably different from a constant-only model, $\chi^2 (5, N = 240) = 6.71, p = .24$. None of the predictors in the full model reliably predict psi hits. Although not statistically significant, having a stronger belief in psi and the participants having a kinship or a close friendship are actually negatively predictive of psi.
### TABLE 6

**MEANS AND FREQUENCIES FOR PREDICTOR VARIABLES AS A FUNCTION OF PSI HITS AND MISSES**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Psi Hit</th>
<th>Psi Miss</th>
<th>$\chi^2$ (1) or $t$ (238)</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belief in psi</td>
<td>3.42</td>
<td>3.51</td>
<td>-.44</td>
<td>.66</td>
</tr>
<tr>
<td>Psi experiences</td>
<td>.82</td>
<td>.70</td>
<td>-1.15</td>
<td>.25</td>
</tr>
<tr>
<td>Mental discipline (%)</td>
<td>52.8</td>
<td>45.2</td>
<td>1.15</td>
<td>.28</td>
</tr>
<tr>
<td>Artist/Musician (%)</td>
<td>43.1</td>
<td>32.1</td>
<td>2.62</td>
<td>.11</td>
</tr>
<tr>
<td>Relationship (%)</td>
<td>30.6</td>
<td>33.3</td>
<td>.18</td>
<td>.67</td>
</tr>
</tbody>
</table>

### TABLE 7

**SUMMARY OF LOGISTIC REGRESSION ANALYSIS PREDICTING PSI HITS**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$\beta$</th>
<th>$SE$</th>
<th>Wald Statistic</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belief in psi</td>
<td>-.13</td>
<td>.11</td>
<td>1.36</td>
<td>.24</td>
</tr>
<tr>
<td>Psi experiences</td>
<td>.35</td>
<td>.22</td>
<td>2.52</td>
<td>.11</td>
</tr>
<tr>
<td>Mental discipline a</td>
<td>.32</td>
<td>.29</td>
<td>1.25</td>
<td>.26</td>
</tr>
<tr>
<td>Artist/Musician a</td>
<td>.51</td>
<td>.30</td>
<td>2.94</td>
<td>.09</td>
</tr>
<tr>
<td>Relationship a</td>
<td>-.11</td>
<td>.31</td>
<td>.14</td>
<td>.71</td>
</tr>
</tbody>
</table>

*a Categorical variables
DISCUSSION

The overarching questions that have occupied parapsychological research has been, “Does psi exist?” and “To what extent have decades of psi research served to support or refute the existence of it?” The series of six studies conducted here are an attempt to contribute to this discussion both on a substantial and methodological level.

Is there evidence for the existence of a psi effect? The answer to this question is contingent upon the way that the data is interpreted and is therefore contingent upon the mechanics and perspectives of methodological considerations. Examining only the data collected in this study, one might argue that there is insufficient evidence to reliably conclude that a psi effect exists. The meta-analysis of the six studies yielded a final hit rate estimate of 30%, although this is not statistically significant from the null hypothesis. A power analysis indicate that if Bem and Honorton’s (1994) 32.2% estimate (or mean $ES = .162$) is an accurate assessment of the psi effect, 231 trials would be needed to achieve a power of .80, whereas if Storm and Ertel’s (2001) 31% estimate (or mean $ES = .138$) is accurate, 320 trials would be needed to achieve the same level of power. In other words, this series of studies are underpowered to detect those levels of effect sizes. It is possible that if indeed psi exists at around the level proposed by these two meta-analyses that the six studies conducted here would not be able to reliably detect it. Interestingly though, the final estimate of 30% is relatively close to that of Bem and
Honorton’s (1994) 32.2% and Storm and Ertel’s (2001) 31%. If more trials are conducted, the stability of the estimate can be better assessed as power is increased.

As many others have argued, the synthesis of research findings is important in the assessment of the evidence accumulated across years of psi ganzfeld research (Krippner et al, 1993; Rosenthal, 1986; Utts, 1991). The results from the six studies conducted should be examined in concert with previously analyzed data to assess the psi effect. The meta-analytic treatment of the current data with three previously published meta-analyses is an attempt at synthesizing results across studies. Results indicate that with the addition of the six studies to the meta-analytic databases, the effect size estimates do not change drastically and that conclusions based on the rejection/acceptance of the null hypothesis also do not change as a result of the new findings. Storm and Ertel (2001) with the largest number of studies meta-analyzed saw their 31% estimate fail to change within a tenth of a percentage point with the addition of all six studies. Both Bem and Honorton (1994) and Milton and Wiseman’s (1999) estimates changed within a sixth of a percentage point of the original estimate. The psi meta-analytic literature remains somewhat inconsistent with one negative and two positive estimates.

One of the criticisms of the meta-analytic literature has been the problem associated with biased estimates as a result of the file drawer problem (Hyman, 1985; Kupfersmid, 1988; Rosenthal, 1979). I hope to have demonstrated that the removal of this bias is not in and of itself a panacea. Ruling out the problem of selective reporting does not preclude the need for studies to be conducted in a procedurally and methodologically sound manner. The necessitation for studies to have adequate power (Cohen, 1992) and consistently demonstrate replication of effect sizes (Krippner et al.,
1993; Rosenthal, 1986; Utts, 1991) are two suggestions for developing a coherent and convincing literature.

Howard and his colleagues (Howard et al., 2000, 2002) have argued that the overreliance on NHST in both primary statistical analyses and meta-analyses contribute to some of the inconsistencies and misinterpretation of research findings. I have tried to suggest as they have, that the Bayesian perspective provides an alternative framework upon which to analyze data. The research literature as it currently stands is inherently biased due to the file drawer problem. NHST is one of the reasons that this problem persists. The Bayesian approach is an improvement over NHST framework in that it is both more intuitive and straightforward to interpret and it more effectively facilitates the synthesis of data in the estimation of an effect. One of the more controversial aspects of the approach is the explicit recognition of subjectivity in the analysis of data (Press & Tanur, 2001). Although Howard and his colleagues have pointed out that the subjectivity inherent in priors, after enough data, would ultimately and largely reflect empirical findings, the data presented here demonstrate that the degree to which a hypothesis (expressed as priors) will be driven by progressive collection of new data is strongly dependent on the strength of the priors themselves.

Despite these criticisms, the Bayesian approach is promising as a primary data analytic framework, especially when implemented in the context of research registries. The idea of research registries in medical research has been gaining much exposure in popular media outlets (Vedantam, 2004). Dr. Drummond Rennie, a Journal of the American Medical Association editor was quoted in *Time* magazine saying that one problem with the medical community evaluating the efficacy of new drugs undergoing
clinical trials is that “while journals are very good at evaluating the significance of studies sent to them, what they don’t do well is evaluate what’s not there” (Bjerklie, 2004, p. 42). With the establishment of a research registry, the problem related to biased reporting can be more effectively minimized. Moreover, groups of scientists can establish procedural and methodological guidelines to ensure the consistent quality of studies being conducted and avoid the task of having to fix or undo a flawed literature.

Although existing literature suggests that psi performance is predicted by several participant characteristics, the five variables assessed in this study failed to provide the same evidence. Honorton (1997), for example, forwarded a three factor model involving reported personal psi experiences, scores on the Feeling-Perception scale of the Myers-Briggs Type Indicator, and reported practice of a form of mental discipline as positively predictive of psi performance. When using a sample of Julliard School of the Arts students, others have found that participants produced a 50% hit rate, with musicians alone obtaining a 75% hit rate (Schlitz & Honorton, 1992). It is possible that some items on the survey, such as Item 5, do not adequately capture the same population as those sampled in other studies examining positive psi predictors. Additionally, in an attempt to distinguish between close and distant relationships for Item 6, the wording of the item may not have been effective in making such distinctions. In light of the methodological comments made so far, it would not be wise to assert that the findings here refute the existence of positive predictors of psi conclusively.

The results from the present study indicate that the short non-ganzfeld procedure failed to yield hit rates significantly different from the long ganzfeld procedure. Even if significant differences were observed, it would be impossible, by virtue of the design of
the two conditions to ascertain what aspects of the ganzfeld made a difference in the hit rates. Future research should attend to designing conditions that more precisely allow for the comparison of essential versus non-essential components to a psi-inducing environment.

The answer to the substantive question of whether psi exists and whether years of psi ganzfeld research provide evidence for or against it cannot be adequately provided without addressing the methodological issues raised in this study. The publication practices coupled with the dominance of NHST approaches in psychology largely contribute to an inconsistent and confusing literature. Furthermore, under this scenario, the hope for a self-correcting science may be more of an illusion than reality. Consequently, this study and other researchers (Howard et al., 2000, 2002; Howson & Urbach, 1989; Kline, 2004; Press & Tanur, 2001) have called for radical departures from the status quo of psychological methodology.
REFERENCES


APPENDIX A

Psi Ganzfeld Study Script

<table>
<thead>
<tr>
<th>Night before trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call participants to remind them when they have signed up to participate in the experiment. In your phone call, remind them of the location of the study. If one of the participants cancels then you should notify the other to have them reschedule.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant Sign-Up and Extra-Credit Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>If only one participant signs up for a given time slot, they will not receive credit until they reschedule for another time slot. If this participant signs up for a different time slot and no one else signs up with them, then they can be given the extra credits.</td>
</tr>
<tr>
<td>If two participants sign up for a given time slot, and one of them cancels in time for you to contact the other participant the night before, the same policy (as above) applies. If two participants sign up for a given time slot, and by no-show, only one participant is present at the experiment, then the participant present will get credit for showing.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pre-Trial Set-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn off your cell phones.</td>
</tr>
<tr>
<td>Go to locker and obtain keys to room and also identical set of randomized set number (the note cards that are lettered). Informed consent forms and the questionnaires should be in the experiment rooms.</td>
</tr>
<tr>
<td>Assign yourselves to either sender or receiver rooms (try to balance your experiences out).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Orientation of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wait at least 5-10 minutes for participants to show up. Try calling them if you have a number on file. If they show up late and you don’t have another trial to run immediately after this one, you may go ahead with the present trial. If not, have them reschedule for another time.</td>
</tr>
</tbody>
</table>
If one of the participants does not show up, then you can let the other participant go. He/she is given extra-credit.

Have all participants meet in one room. Introduce everyone (including the experimenters) to the participants. **Attempt to create a friendly and informal social atmosphere** (you can engage each other to make small talk about their majors, etc.). **Attempt to create positive expectations concerning receiver’s ability to identify target.**

Have participants turn off their cell phones.

Have participants complete informed consent form (place form in folder. Folder should be locked in locker) and questionnaire (also in folder and locker). Each participant is given a unique code number (001, 002, 003, etc.). If they are the receiver in the long ganzfeld trial, attach L to their code (e.g. 001L). The other sender in the long trial is given an S. **The code should go on the back of the consent form (upper right hand corner), the top of the questionnaire and the data collection sheet.**

**Background of Study**

Explain the following to participants:
1. Brief history of psi and parapsychology research
   - What is psi?
2. Theoretical background behind ganzfeld method and use in assessing psi
   - What is ganzfeld? Why ganzfeld?
3. Brief overview of the experiment
   - Ganzfeld trial, switch roles, then eight short non-ganzfeld trials etc…

Flip coin and assign participants to either sender or receiver conditions.

<table>
<thead>
<tr>
<th>Long trial (Ganzfeld procedure)</th>
</tr>
</thead>
</table>

Have participants secluded to separate rooms. Explain to them briefly again what will happen next.

For both rooms:
**Before we begin the experiment, you will listen to the instructions to a relaxation exercise for approximately 10 minutes. Please follow the instructions of the recording and try to relax your body and mind. The participant in the other room will also be going through the same exercise. Relaxation before the ESP phase of the experiment is helpful in making ESP phenomena more probable. After the end of the recording you may remove the headphones to signal to me that the recording has ended.**

[Set up CD player so participants can begin listening to the relaxation recording]

For the sender:
[At the end of the relaxation exercise, the sender should remove the headphones to signal to you that the recording has ended. You may put away the CD player and headphones]
and set up the radio next to him/her. **Do not** turn on the radio until you return from the hallway.

*I will now step into the hallway to coordinate with the other team to begin. When I return, I will present to you a target image. Please concentrate on the image for the next 30 minutes and attempt to transmit the image to the receiver in the other room. During this “mentation period”, the receiver will verbally report any thoughts or images that go through his or her mind. You will hear that transmitted through the radio next to you. He/she, however will not hear whatever is said in this room.***

**[Do not reveal to yourself or the sender the target picture until you have returned from the hallway. It is OK to take some time from the 30 minutes to pick out the correct target. When you return from the hallway, you should have coordinated with the other team to begin the 30 minutes countdown. Turn on the radio, go to the note-cards, reveal the set and then go to the random number table to reveal the target to use within the set. Present the target to the sender. Remind him/her that he/she should try to concentrate on the image for the next 30 minutes. Be as quiet as you can during this period. Record the participant number and target image used]**

For the receiver:

**[At the end of the relaxation exercise, remove the CD from the CD player and replace it with the white noise CD. Also help the participant put on the goggles, making sure that he/she is unable to see anything when wearing it. Turn on the red flood light and place it close to the participant so that he/she sees a continuous and uniform red light when wearing the goggles. Put the microphone on the participant by letting it sit on his/her neck with the microphone pointing up in front of the receiver’s mouth. Play the white noise CD and instruct the receiver to turn the volume up as far as it is still comfortable. He/she should not hear anything else when the CD is being played. Turn off the CD and read the following instructions to the receiver:]**

*After I read you these instructions I will go out into the hallway to coordinate with the other team to begin the experiment. During the next phase of the experiment (a 30 minutes period we call the “mentation period”), you will be put in a ganzfeld condition, which is the state in which you will be listening to the white noise CD and wearing the goggles. When I return from the hallway, I will begin playing the white noise CD for you.

*During this phase of the experiment, I want you to think out loud. Report all of the images, thoughts, and feelings that pass through your mind. Do not cling to any of them. Just observe them as they go by. At some point during the session, the sender will send you the target information. Do not try to anticipate or conjure up this information. Just give yourself the suggestion, right now – in the form of making a wish – that the information will appear in consciousness at the appropriate time. Keep your eyes open as much as possible during the session and allow your consciousness to flow through the sound you will hear through the headphones. Now get as comfortable as possible, release all conscious hold of your body, and allow it to relax completely. As soon as you begin observing your mental processes, start thinking out loud. Continue to share your thoughts, images and feelings with us throughout the session.

*[It might be a good idea to let participant know that he/she should not be self-conscious.]*
Mentation period:
During the 30 minutes, try to be as quiet as possible. For the experimenter in the receiver room, maintain a mentation report, by jotting down the general images reported by the receiver. You will read it back to him/her during the judging period. Towards the end of the trial, go to the note-cards and reveal the set and refer to the random number table for the sequence you will present the four pictures to the participant later. When the thirty minutes is up, let the participants know and stop the white noise CD and take off their goggles.

For the sender:
You may stop concentrating on the image now.
[Take the image from the participant and return it to the folders]
At this point, the receiver in the other room will be presented with four images with which to judge. One of the images is the image you attempted to send to him/her. I will step in the hallway to wait for the other team to let me know that that process is completed. You may sit here and relax until I return.
[Step out in the hallway and wait for the other experimenter to let you know that they are ready for you to go to the other room]

For the receiver:
[Stop the CD from playing, remove the goggles and turn off the floodlight]
You may stop concentrating on receiving the image now. At this point I will show you four images in a random order. One of the images is the one the sender has been concentrating on for the last 30 minutes. Please examine them closely and determine for yourself which one is the one that was being sent to you. Prior to making your decision, I will read back to you the mentation notes I kept during the mentation period. These will help remind you what images and thoughts were experienced by you during that period.
[When the receiver picks out the image, record that on the data sheet and go to the hallway to invite the other team back into your room]

Gather all participants in the same room. The experimenter and sender can reveal the correct image to the receiver.

| Short-trials (Ten 2-minute trials) |

When the participants are still in the same room, explain to them that they will now switch roles and rooms. We will now conduct 8 short 2 minute trials identical to the first. The difference will be that the receiver will not be listening to the white noise CD, nor wearing the goggles. At this point seclude the two participants in their respective rooms. Be sure to turn off the two-way radios.

For the receiver:
Please sit comfortably in your chair and when the trial begins, let yourself observe the images, thoughts, and feelings that pass through your mind. At some point during the trial, the sender will send you the target information. Do not try to anticipate or conjure up this information. Just give yourself the suggestion, right now – in the form of making a
wish – that the information will appear in consciousness at the appropriate time. Now get as comfortable as possible, release all conscious hold of your body, and allow it to relax completely. When the two-minutes are up, I will notify you. I will step out in the hallway to coordinate with the other team to begin this trial. When I return the mentation period will begin.

[Head to the hallway to begin the two minutes with the other team; When the two minutes are up, let the receiver know that time is up. During this time, you may cue up the correct set and arrange the pictures in order for presentation].

At this point I will show you four images in a random order. One of the images is the one the sender has been concentrating on for the last 2 minutes. Please examine them closely and determine for yourself which one is the one that was being sent to you.

[When the receiver picks out the image, record that on the data sheet and go to the hallway to invite the other team back into your room. You do not have to read the entire set of instructions to the receiver again. Assume that they understand the process after the first short trial.]

For the sender:
I will now step into the hallway to coordinate with the other team to begin. When I return, I will present to you a target image. Please concentrate on the image for the next 2 minutes and attempt to transmit the image to the receiver in the other room. When this “mentation period” is up, I will let you know.

[Do not reveal to yourself or the sender the target picture until you have returned from the hallway. It is OK to take some time from the 2 minutes to pick out the correct target. When you return from the hallway, you should have coordinated with the other team to begin the 2 minutes countdown. Go to the note-cards, reveal the set and then go to the random number table to reveal the target to use within the set. Present the target to the sender. Be as quiet as you can during this period. Record the participant number and target image used]

You may stop concentrating on the image now.

[Take the image from the participant and return it to the folders]

At this point, the receiver in the other room will be presented with four images with which to judge. One of the images is the image you attempted to send to him/her. I will step in the hallway to wait for the other team to let me know that that process is completed. You may sit here and relax until I return. When I return, we will begin the next two-minute trial.

[Step out in the hallway and wait for the other experimenter to let you know that they are ready for you to go to the other room]

REPEAT THIS PROCESS UNTIL THE 8 SHORT TRIALS ARE UP

Gather all participants in the same room. The experimenter will reveal the number of images that were correctly selected to the receiver.
Post-trial Wrap-up

Cross check the data collection sheet to make sure that both experimenters were using the same set of images for each trial. Make sure that the participant codes are correctly entered.

General things to remember when conducting trials

- Do not conduct the retrieval or arrangement of target images in full view of the participants. In general, you should sit behind the participants so that any activity involving the preparation of the target images occur outside of the participants view.
- When conducting the experiment, remember that you want to prevent criticism that the way that the trial was conducted (i.e. inadvertently letting the experimenter in the other room know what the target image is prior to the mentation period, or your knowledge of what the possible target pictures are tainting with the “mentation” process) prevented psi from happening or somehow leaked the correct target image to the receiver.
  - Do not look at the images (both target and the four containing the target) until you are ready to present them.
  - Do not flip the note-cards until you are at the beginning of the next trial
- The note-cards should be randomized correctly when you get them. When flipping the note-cards to the next number, please make sure that you are only flipping one card at a time. If the two rooms are not coordinated on the same note-card, then the current and subsequent trials are invalid.
APPENDIX B

Age: ______________

Gender:  Female    Male

*Psi* (a scientific term loosely equivalent to ESP) can be described as anomalous processes of information or energy transfer, processes such as telepathy or other forms of extrasensory perception that are currently unexplained in terms of known physical or biological mechanisms. Various forms of psi phenomena include:

*Telepathy* – the ability to read another’s mind through extrasensory perception  
*Clairvoyance* – the ability to see things beyond the power of vision  
*Precognition* – psychic knowledge in advance of its occurrence  
*Psychokinesis* – the production of motion in physical objects by exercising psychic or mental powers

1) On a seven-point scale where “1” indicates strong disbelief and “7” indicates strong belief in psi, circle the degree to which you believe in the existence of psi.

1  2  3  4  5  6  7
Strong Disbelief  Strong Belief

2) If you have had experiences which you thought involved psi, which of the following do you feel you have experienced (please circle all that apply):

Telepathy  Clairvoyance  Precognition  Psychokinesis

3) Have you ever participated in formal laboratory testing of psi phenomena?

Yes  No

4) Have you ever practiced any form of a mental discipline, e.g. meditation, biofeedback, hypnosis, relaxation exercises?

Yes  No

5) Are you a(n)(please circle all that apply):

Artist  Musician  Neither

6) Are you and the other participant:

Acquaintances  Close Friends  Related (Family)  None of the Above

7) Can we contact you for future participation of psi experiments?

Yes  No
APPENDIX C

SUMMARY OF THREE PUBLISHED PSI GANZFELD META-ANALYSES

<table>
<thead>
<tr>
<th>Meta-Analysis</th>
<th>No. of Studies</th>
<th>No. of Trials</th>
<th>Mean ES</th>
<th>Stouffer Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bem &amp; Honorton (1994)</td>
<td>10</td>
<td>329</td>
<td>.164</td>
<td>2.55</td>
<td>.005</td>
</tr>
<tr>
<td>Storm &amp; Ertel (2001)</td>
<td>79</td>
<td>2767</td>
<td>.138</td>
<td>5.66</td>
<td>7.58 x 10^-9</td>
</tr>
</tbody>
</table>

Bem & Honorton (1994)

- Meta-analysis of reduced Honorton et al. (1990) database.

Milton & Wiseman (1999)

- Meta-analysis of studies that “began in 1987 or later and published by February 1997” (p. 388), but curiously excluding Honorton et al.’s (1990) autoganzfeld experiments.

Storm & Ertel (2001)

ENDNOTES

1 The Bem and Honorton (1994) meta-analysis is based on a reduced database of the Honorton et al. (1990) autoganzfeld studies. One study was removed from the original 11 studies because of “response bias”.

2 Effect size ($z/N^{1/2}$, where $N$ is the number of trials in a study), was computed by first obtaining a $z$ score for each study included in the meta-analysis. $Z$ scores were derived from an exact binomial test when the study measured the outcome by comparing probability of the number of “hits” obtained compared to the number expected by chance. A hit is when a participant correctly identifies the target image sent to him/her by the sender. When more than one outcome was reported in a study, the $z$ score associated with the main outcome was used. The mean effect size ($ES$) was obtained by cumulating the effect size scores and dividing that by the number of studies meta-analyzed (Milton & Wiseman, 1999; Storm & Ertel, 2001). The mean $ES$ ($\_ [z/n^{1/2}] / k$, where $k$ is the number of studies meta-analyzed) measure is an estimate of $r$ (Rosenthal, 1994; Storm & Ertel, 2001).

3 Although a total of 960 short non-ganzfeld trials would have been conducted (120 pairs of participants and 8 short trials for each pair), 23 trials were dropped due to the senders and receivers not using the same sets of images for sending and judging during those particular trials.

4 When CIs were constructed using one of the recommended methods (Agresti & Coull, 1998), the adjusted Wald method, the CIs for Study 1 and for Study 1 + 2 did not contain the null value and is considered statistically significant. With the inclusion of the each of the remaining four studies, the CIs all contain the null value.

5 As mentioned earlier, a failure to reject the null hypothesis is often mistakenly interpreted as evidence for the null hypothesis, which in this case, that there are no psi phenomena. Such NHST pitfalls embodied in many meta-analyses contribute to a troubling and confusing literature.

6 As indicated earlier, the participants who provided estimates for the priors were instructed to respond in units of studies with each study containing 20 trials. The formula provided by Pruzek (1997) conceptualizes the beta distribution in units of trials. Therefore, the responses made by each participant were multiplied by 20 to yield the parameters to calculate the beta distribution. In this example, the response of 10 significant studies translates to 200 trials, in which 25% are hits, yielding the prior beta distribution of [50, 150].
Although a majority of the meta-analyses examined used mean $ES$ as the effect size measure for psi, I have chosen to use the hit rate instead. Not all psi ganzfeld studies employ a four-choice judging set such that a 25% hit rate is equivalent to chance occurrence. Nevertheless, by expressing the meta-analytic findings in such a way, the reader can more intuitively understand the effect size than when using mean $ES$. The effect sizes expressed as hit rates were reported directly by Bem and Honorton (1994) and Storm and Ertel (2001). Together with the reported number of trials meta-analyzed, the information was used to derive a beta distribution prior using the method described in an earlier section. Milton and Wiseman (1999) did not directly report a hit rate for their meta-analysis. To create a consistent and intuitive effect size measure, a hit rate was derived by combining the reported $z$ scores (unweighted Stouffer $z$ method) for each of the studies analyzed by Milton and Wiseman and calculating the number of hits out of the total trials that would result in an exact binomial significance equivalent to that $z$ score.

The resulting inconsistency of the psi meta-analytic literature is not simply a function of the synthesis with the current six studies but also because of the different methodological criteria used to include and exclude studies in the three respective meta-analyses summarized in Appendix C.