# Advances in Parapsychological Research 9

Edited by Stanley Krippner, Adam J. Rock, Julie Beischel, Harris L. Friedman and Cheryl L. Fracasso

# How Much Evidence Is Necessary to Define the Reality of a Phenomenon? Frequentist, Bayesian, and Quantum Modeling of Ganzfeld ESP

**Patrizio E. Tressoldi**

In this essay, I summarize the strength of the evidence for extrasensory perception (ESP) with the receiver in a Ganzfeld state, namely one in which there is by an undifferentiated auditory and visual field. (Note: The term ganzfeld, derived from German ganz ["whole, entire"] and feld ["field, area"] was coined as a generic term for the unstructured visual field.) I focus on how much evidence is necessary to state that this phenomenon is real or, at least, very probable, and whether that amount is present in the research. I also describe the main statistical results obtained from analyzing the ESP Ganzfeld database using a frequentist, Bayesian, and new quantum modeling approach.

## How Much Evidence Is Necessary to State That a Phenomenon Is Real?

"Extraordinary claims require extraordinary evidence" was a phrase made popular by Carl Sagan who reworded Laplace's principle which stated that "the weight of evidence for an extraordinary claim must be proportioned to its strangeness" (Laplace, n.d.). This statement is at the heart of the scientific method, and provides a model for critical thinking, rational thought, and appropriate skepticism. However, no standards, quantitative or otherwise, have been agreed upon in order to define whether or not sufficient extraordinary evidence has been obtained. Consequently, the measures of extraordinary evidence are completely reliant on subjective evaluation and the acceptance of extraordinary claims never unequivocal. In science, the definition of extraordinary evidence is more a social agreement than an objective evaluation, even if most scientists might state the contrary.

However, if we shift from this problem to the definition of how much evidence is necessary to recommend medical or psychological interventions for human health, we realize that a consensus has been obtained. Thanks to an evidence-based medicine (EBM) or evidence-based practice (EBP) approach, which aims to apply the best available evidence gained from the scientific method to clinical decision making, it is possible to assess the strength of evidence on the risks and benefits of treatments (including lack of treatment). These approaches use diagnostic tests to help clinicians learn whether or not a treatment will do more good than harm. Evidence quality can be assessed based on the source type (from meta-analyses and systematic reviews of double-blind, placebo-controlled clinical trials at the top end to conventional wisdom at the bottom end) as well as based on other factors including statistical validity, clinical relevance, currency, and peer-review acceptance.

The strongest evidence for therapeutic interventions is provided by the systematic review of randomized, triple-blind (clinicians, patients, and evaluators of the treatment outcomes), placebo-controlled trials with allocation concealment and complete follow-up involving a homogeneous patient population and medical condition. In contrast, patient testimonials, case reports, and even expert opinions are seen as having little value as proof because of the placebo effect, the biases inherent in the observation and reporting of cases, difficulties in ascertaining who is an expert, and more. Consequently, it turns out that if one meta-analysis or at least two well-designed controlled clinical trials give clear evidence about the efficacy of a clinical intervention, it is possible to recommend the adoption of this intervention to practitioners and physicians. Why this standard is not sufficient to support the reality of a parapsychological phenomenon, such as ESP with the receiver in a Ganzfeld state, remains a puzzle. Is more evidence required than the amount needed to recommend clinical practices for human health where lives, and quality of life, are at stake?

### *Replication*

Among the requirements to state the reality of a phenomenon, there is a substantial consensus that replication is one of the more fundamental (Schmidt, 2009). In other words, a phenomenon may be considered real or very probable when it has been observed many times and preferably by different people or research groups. Whereas a failure to replicate is quite expected in cases of conceptual replication or when the experimental procedure or materials entail relevant modifications, a failure in case of an exact or quasi-exact replication poses serious concerns about the reality of the phenomenon under investigation. Within mainstream psychology, the lack of replications has once more become the focus of significant controversy (see Roediger, 2012), although the cry for more replication studies has been raised throughout the history of psychology.

Replications are particularly needed in the more controversial areas of psychology, such as parapsychology (see Krippner and Friedman, 2010). As Krippner and Friedman pointed out, the main critique of parapsychology by many mainstream skeptics involves the supposed lack of replication of findings in this area, when in fact replication rates in parapsychology are not that much different from many mainstream areas of psychological research. This is not to say that parapsychology is on solid empirical ground but, rather, many areas of psychological science are on similarly shaky foundations compared to those of parapsychology.

However, what represents a replication in psychology? Typically, studies using human participants investigate a particular phenomenon within a sample of participants and infer the probability of observing the same results in the population by testing a null hypothesis. The Null Hypothesis Significant Testing approach, even if hotly criticized (e.g., Nickerson, 2000), is still the most used statistical approach. Within this approach, a replication is observed when the null (or nil) hypothesis is refuted, usually with a probability set by convention at $p$ = 0.05 to avoid committing a Type I error; that is, to state erroneously that the hypothesis is rejected. Among the many limitations of this approach, there is the often neglected problem of statistical power.

Statistical power depends on three classes of parameters: (a) the significance level (i.e., the Type I error probability) of the test, (b) the size(s) of the sample(s) used for the test, and (c) an effect size (*ES*) parameter defining the active hypothesis (H1) and, thus, indexing the degree of deviation from the null hypothesis (H0) in the underlying population. Power analysis can be used prospectively to calculate the minimum sample size required, so that one can be reasonably likely to detect an effect of a given size. Power analysis can also be used to calculate the minimum *ES* that is likely to be detected in a study using a given sample size.

In most experimental designs, the accepted probability of making a Type I error, as stated above, is $\alpha$ = 0.05 and the desired power is not less than 0.80. However, in order to define how to obtain such a level of power, it is necessary to know the *ES* of the phenomena being identified. It is intuitive that the smaller the effect, the greater should be the effort needed to detect it. This analogy is similar to the signal/noise relationship. The smaller the signal, the stronger must be the means to detect it in the noise. In psychological experiments, these means are often the number of participants taking part in the study and the number of trials they are requested to perform. Given that power = 1- $\beta$ (the percentage of negative events) = $ES * \sqrt{N}$ (number of participants) $/SD$ (standard deviation) * $\alpha$, if the estimated *ES* of a phenomenon is known, after the definition of the desired power and the $\alpha$ level, the only free parameter is $N$; that is, the number of participants or trials (Keppel and Wickens, 2004).

From the evidence of ESP with the Ganzfeld database, the estimated effect size is very small, ranging from 0.11 to 0.14 in standard units. Some simple calculations to estimate the minimum number of participants required to detect this effect size by setting $\alpha$ at 0.05 and power at 0.80 or higher suggest that they should be at least 100 (see Tressoldi, 2012). How many studies which failed to replicate a parapsychological phenomenon may be due to simply a lack of power? This is one of the reasons why the assessment of available evidence related to ESP would benefit from meta-analytic studies, which bolster the power by combining many studies.

## *Three or More Sigmas?*

Another approach to define the strength of evidence is the measure of sigmas or standard deviations from the mean (Chassin, 1998; Schroeder, Linderman, Liedtke, and Choo, 2008). For example, Six Sigma is a business management strategy, originally developed by Motorola, USA, in 1986, that is widely used in many sectors of industry, associated with statistical modeling of manufacturing processes. The maturity of a manufacturing process can be described by a *sigma* rating indicating its yield, or the percentage of defect-free products it creates. A six sigma process is one in which 99.99966 percent of the products manufactured are statistically expected to be free of defects (3.4 defects per million). Motorola set a goal of six sigma for all of its manufacturing operations, and this goal became a byword for the management and engineering practices used to achieve it. This approach to measure the strength of evidence is also used in physics when a statistical significance of six-sigma is often the physicists' way of saying that the measurement of a phenomenon is certainly correct (i.e., Opera collaboration, 2011). Do we have similar evidence supporting the reality of ESP?

## *The Evidence for ESP from the Ganzfeld Database*

This line of research started almost 40 years ago conceptualizing ESP as a weak signal that is normally masked by internal somatic and external sensory "noise." By reducing ordinary sensory input, the signal-to-noise ratio is raised, thereby enhancing a person's ability to detect the information (Honorton, 1977). To test the hypothesis that a reduction of sensory input itself facilitates ESP performance, investigators turned to the Ganzfeld procedure (Braud, Wood, and Braud, 1975; Honorton and Harper, 1974), a procedure originally introduced into experimental psychology during the 1930s to test propositions derived from gestalt theory (Avant, 1965).

The meta-analyses of the accumulated evidence started in 1985 by Hyman (1985) and Honorton (1985) and continued with increasing frequency after a publication by Bem and Honorton (1994): Milton and Wiseman (1999, 2002); Storm and Ertel (2001, 2002); Bem, Palmer, and Broughton (2001); Storm (2006); Storm, Tressoldi, and Di Risio (2010a, b); Utts, Norris, Suess, and Johnson (2010); Tressoldi (2011); Williams (2011); Kruschke (2011c); Rouder and Morey (~~in press~~); Storm and Tressoldi (~~in press, a~~); and Tressoldi and Khrennikov (2012). To facilitate the interpretation of the state of the art of available evidence, results obtained with the more recent meta-analyses are presented separately for the frequentist, the Bayesian, and a new quantum modeling statistical approach.

**Frequentist statistical approach.** The classical frequentist statistical approach to meta-analysis was introduced by Glass and colleagues in the early

1980s (Glass, McGaw, and Smith, 1981; Hedges and Olkin, 1985). It consists of a weighted inverse variance average of standardized measures (ESs) observed in all of the available studies relating to a specific topic (i.e., medical, educational, psychological, etc.). The strength of the evidence is demonstrated by the number of studies retrieved and the measure of the average effect sizes with their confidence intervals and the associated probability of the null hypothesis being rejected. According to the fixed-effect model, it is assumed that there is one true *ES* (hence the term "fixed effect") which underlies all of the studies in the analyses and that any differences between this value and the observed effects are due to sampling errors. In contrast, under the random-effects model, it is assumed that the true effect could vary from study to study as a consequence of the influence of so-called moderator variables (e.g., participants or stimuli characteristics). The *ES*s in the studies that were actually performed are assumed to represent a random sample of the *ES*s, leading to the term "random effects" (Borenstein, Hedges, Higgins, and Rothstein, 2010).

Three Ganzfeld ESP studies using the classical frequentist statistical approach to meta-analysis are discussed. The database analyzed by Storm, Tressoldi, and Di Risio (2010a) and Tressoldi (2011) comprised all available studies up to 2009 for a total of 108 effect sizes. The null hypothesis was rejected with the following results: Stouffer $z = 8.31$; $p = 9.5x\ 10^{-17}$; Fixed effect: $z = 19.36$; $p = 1.67x^{-83}$; Random effect: $z = 6.39$; $p = 1.65x^{-10}$. Williams' (2011) review is a basic assessment of 59 Ganzfeld ESP studies reported in the period following the publication of a stringent set of methodological guidelines and recommendations by Hyman and Honorton (1986).

The assessment indicates that these 59 studies have a combined hit rate of approximately 30 percent, which is significantly above the chance expected hit rate of 25 percent. A comparison of the hit rates across four Ganzfeld meta-analyses, as well as across 15 laboratories, seems to further indicate replication of the Ganzfeld ESP effect by a broad group of independent researchers. Overall, $z = 7.37$, $p = 8.6 \times 10^{-14}$. Table 1 summarizes the basic statistics obtained by the three previously described meta-analyses. From the frequentist statistics of three meta-analyses, it is clearly evident that the six-sigmas criterion has been met by all three as can be seen from the z values in the last column of Table 1.

**Bayesian statistical approaches.** Many statistical experts consider frequentist or Null Hypothesis Significant Testing (NHST) inadequate for analyzing data (i.e., Wagenmakers, Wetzels, Borsboom, and van der Maas, 2011). One alternative to NHST is the Bayesian statistical approach. In the following, I summarize the main differences between NHST and the Bayesian statistical approach to help with comprehension of three such meta-analyses carried out on ESP with Ganzfeld.

Unlike NHST, Bayesian formulations of data-analytic questions provide rational and richly informative answers (Kruschke, 2011a). When the question is about null values, there are two Bayesian formulations that ask the question at different levels and provide correspondingly different types of information.

| *Reference* | *n. studies* | *MCE* | *Mean Observed Hits* | *ES (0.95CI)* | *z* |
|---|---|---|---|---|---|
| Storm, Tressoldi and Di Risio, 2010 | 108 | 0.25 | 0.33 | 0.14 (-0.04-0.33) | 8.31* |
| Tressoldi, 2011 | 108 | 0.25 | 0.33 | 0.13° (0.09-0.17) | 6.39 |
| Williams, 2011 | 59 | 0.25 | 0.31 | 0.11 (0.09-0.129) | 7.37 |

°= Random effect; *=Stouffer's *z*

**Table 1. Summary of Frequentist Descriptive and Inferential Statistics**

Bayesian inference is the reallocation of credibility across a space of possibilities. In one Bayesian approach to assessing null values, the analyst sets up two competing models of the possible values. One model posits that only the null value is possible. The alternative model posits that a broad range of other values is also possible. Bayesian inference is used to compute which model is more credible, given the data. This method is called *Bayesian model comparison* and it is based on the following equation as indicated in Figure 1:

This equation shows that the Bayes Factor (BF) converts the prior odds of the models, $p(M1)=p(M2)$, to the posterior odds of the models, $p(M1\times D)=p(M2\times D)$. As the BF increases more than 1.0, the value indicating an identical probability of the two models, the evidence increases in favor of model M1 over model M2. The convention for interpreting the magnitude of the BF is that there is "substantial" evidence for model M1 when the BF exceeds 3.0 and "extreme evidence" when the BF is greater than 100, and, equivalently, "substantial" evidence for model M2 when the BF is less than 0.3 and "extreme evidence" when the BF is less than 0.01 (Wetzels, Matzke, Lee, Rouder, Iverson, and Wagenmakers, 2011).

$$\frac{p(M_1|D)}{p(M_2|D)} = \underbrace{\frac{p(D|M_1)}{p(D|M_2)}}_{\text{BF}} \frac{p(M_1)}{p(M_2)}$$

**Legend for Figure 1: p = probability; M1, M2 = model 1 and model 2; D= data.**

Rouder and Morey (2011) developed a meta-analytic version of the Bayes Factor *t*-test and used it to assess the evidence across multiple experiments. Tressoldi (2011) applied this approach to the database of all studies available until 2009 for a total of 108 ESs. The obtained Bayes Factor, related to the comparison between H1 (evidence favoring ESP) and H0 (evidence not favoring ESP), yielded the following result: $BF_{(H1/H0)}$= 18,861,051. A similar analysis carried out using a different algorithm, proposed by Rouder and Morey (~~in press~~), yielded the following result: $BF_{(H1/H0)} = 9.91 \times 10^{24}$.

In a second Bayesian approach to assessing null values, the analyst sets up a range of candidate values, including the null value, and uses Bayesian inference to compute the relative credibility of all the candidate values. This method is called *Bayesian parameter estimation*. The model-comparison approach emphasizes the Bayes Factor, whereas the parameter-estimation approach emphasizes the explicit posterior distribution on the parameter values, such as mean, standard deviation, etc.

Bayesian estimation, with its explicit parameter distribution, is not only more informative than Bayesian model comparison, but also more robust (Kruschke, 2011b). In Bayesian parameter estimation, the analyst establishes the credibility for each value of the parameter before observing new data. These parameter-value credibilities are called the prior distribution or "prior" for short. A major advantage of the estimation approach is that there is an explicit distribution on the parameter values. The analysis explicitly reveals uncertainty about the underlying accuracy in each experiment and across experiments. The hierarchical structure also lets the estimate of accuracy in each experiment be informed by data from other experiments. On the other hand, the Bayesian model comparison often provides only a Bayes Factor, which informs about the relative credibility of the point null hypothesis and another specific non-null prior hypothesis, without demonstrating what the parameter values could be.

Utts et al. (2010) adopted this approach when analyzing a database of studies that met certain criteria for methodological rigor and adherence to standard Ganzfeld procedures. These included 16 of the studies mentioned by Dawson (1991), with 8 eliminated because of unresolved allegations of methodological flaws, as well as all 11 studies drawn from Bem and Honorton (1994) and the 29 studies with a "standardness score" of more than four as analyzed by Bem, Palmer, and Broughton (2001), for a total of 56 studies. Utts et al. (2010) used a Bayesian hierarchical model that assumes a constant probability of a hit within a study but the possibility of different probabilities across studies, as in letting $p_i$ be the probability of a hit for Study $i$, $i = 1, 2, \ldots, 56$. $n_i$ and $Xi$ are the number of sessions and number of hits, respectively, for study $I$, assuming $Xi$ has a binomial $(n_i, p_i)$ distribution. They tested different prior beliefs by using four different possible prior distributions for the median of the distribution of $p_i$'s, the possible hit probabilities across all studies. The four different priors were:

1. Non-informative prior for μ puts equal probability on all real numbers (improper). This is probably not realistic; the true probability of a hit is unlikely to be as low as 0 or as high as 1.
2. Open-minded prior, which uses median $(p) = 0.25$ as the best guess, and for which we are 90 percent sure the median$(p)$ is between 0.12 and 0.41.
3. Psi believer's prior, which uses median $(p) = 0.33$ and for which we are 90 percent sure the median$(p)$ is between 0.30 and 0.36.

4. Psi skeptics prior, which uses median ($p$) = 0.25 and for which we are 90 percent sure the median($p$) is between 0.245 and 0.255.

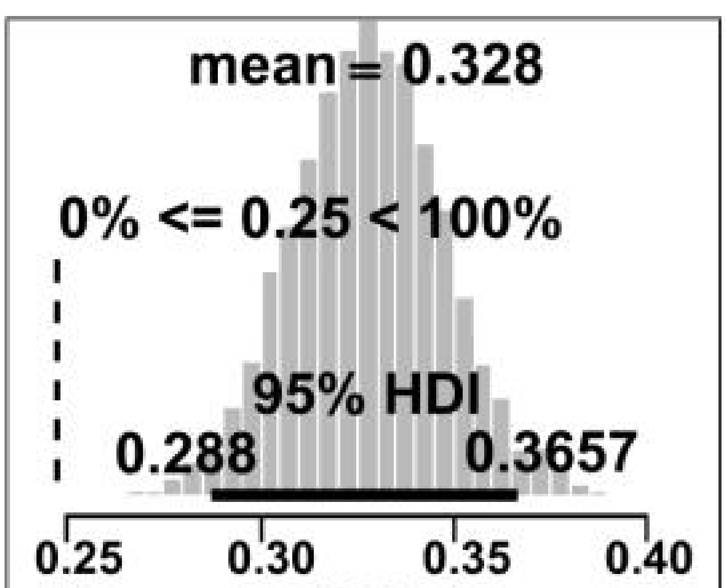
The summary of Utts et al. (2010) statistical analysis is reported in Table 2.

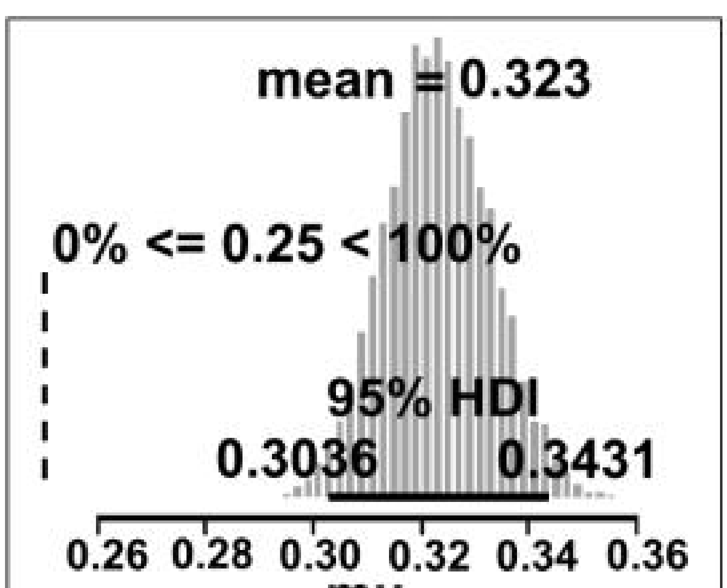
| *Type of Prior* | *Prior median (p); 90% sure* | *2.5%* | *50%* | *97.5%* | *95% range* |
|---|---|---|---|---|---|
| Non informative | N/A | 0.30 | 0.33 | 0.36 | 0.19-0.49 |
| Open-minded | 0.25; 0.12-0.41 | 0.29 | 0.33 | 0.36 | 0.17-0.51 |
| Psi beliver | 0.33; 0.30-0.36 | 0.308 | 0.326 | 0.345 | 0.28-0.37 |
| Psi skeptic | 0.25; 0.245-0.255 | 0.251 | 0.257 | 0.262 | 0.254-0.26 |

**Table 2. Median and 95 percent intervals for the posterior distribution of median (*p*); 95 percent range for *p*.**

The second analysis using the parameters modeling Bayesian approach was carried out by Kruschke (2011c). This author analyzed the 29 studies related to ESP with Ganzfeld included in the Storm et al. (2010a) meta-analysis in which expected responding chance was 0.25. The across-experiments $\mu$ (see Figure 1) is clearly above chance, with 100 percent of the posterior sample falling above 0.25. The 95 percent Highest Density Interval[1] (HDI) goes from 0.288 to 0.366, as shown in Figure 2. A replication with all 114 available studies up to 2011 was carried out by Tressoldi (unpublished) yielded the results presented in Figure 3.

The across-experiments in Figure 3 is clearly above chance, with 100 percent of the posterior sample falling above 0.25. The 95 percent Highest Density Interval (HDI) goes from 0.30 to 0.34. Both Bayesian statistical approaches, models comparison and parameters estimation, confirm that ESP in Ganzfeld is possible, and the results are well above the mean chance expected.




***Left:*** **Figure 2. Highest Density Interval (HDI) of mean (mu) estimation of the studies related to ESP with Ganzfeld included in the Storm et al. (2010) meta-analysis.** ***Right:*** **Figure 3. Highest Density Interval (HDI) of mean (mu) estimation of the studies related to ESP with Ganzfeld database from 1974 to 2011.**

**Quantum modeling statistical approach.** Tressoldi and Khrennikov (2012) analyzed the ESP with Ganzfeld research, using a database updated until 2011, employing a statistical approach used in quantum physics to analyze the Remote State Preparation (RSP) protocol. Quantum Mechanics (QM) RSP is a variant of teleportation where Alice[1] has full knowledge of the state she intends to prepare at Bob's location. The goal of RSP is to prepare a quantum state at a distant location, without sending the actual state. Alice, the sending party, knows exactly the target state that she wants Bob, the receiving party, to have. Several features are usually desired in an RSP protocol: Bob should have been limited to no knowledge of the state Alice is trying to prepare and the required communication resources (classical and/or quantum) should be limited. Perhaps most importantly, the protocol should yield output states at Bob's location which closely match the target states which Alice intended to prepare.

If we change Alice with sender and Bob with receiver, it is similar to the typical ESP protocol used to investigate telepathy. Tressoldi and Khrennikov (2012) named this protocol Remote State Preparation of Mental Information (RSPMI). However, it is important to take into account not only similarity but also the differences between a typical RSP protocol used to connect physical information between two electronic devices and the typical ESP protocol used to connect mental information between two humans. Table 3 summarizes this comparison. From 87 experiments and a total of 3,338 events (trials), an average of 1,120 hits, corresponding to 33.82 percent of coincidences, were observed. Weight-

| | *RSP* | *RSPMI* |
|---|---|---|
| Alice identity | Electronic device | Electronic device or Human being |
| Bob identity | Electronic device | Human being |
| Alice initial knowledge | Target complete knowledge | Target complete knowledge |
| Bob initial knowledge | Target zero knowledge | Target zero knowledge |
| Information type | *q*bits | Classical (i.e. images, video clips) |
| Entanglement mode | i.e. Parametric downconversion | Mental connection |
| Transmitted cbits | 2(3) | 0 |
| Locality loophole | Partially closed | Closed for sensory information |
| Fair-sampling loophole | open | closed |
| Events per experiment | thousands | Usually less than one hundred |
| Time per event | Fraction of seconds | 15 to 30 minutes |
| Coincidence counts | Electronic device | Electronic device or independent judge |

**Table 3. Similarities and differences between the standard RSP and RSPMI protocol.**

ing the result of each study for √number of trials, the average fidelity is $F$ = 0.808969964 ± 0.001463, violating the benchmark $F$ = 0.7482029 expected with a $p$ = 0.25 by 41.5 standard units.

## Conclusion

All statistical analyses of the Ganzfeld ESP databases examined, using frequentist, Bayesian, and quantum modeling statistical approaches, converge in showing clear evidence of ESP, satisfying the highest statistical standards required to support an extraordinary claim with extraordinary evidence. Is this evidence sufficient to state that this phenomenon is real? From a statistical point of view, it seems hard to challenge this assertion. How many other mental phenomena have obtained this level of evidence? Three different statistical approaches carried out from different authors converge in supporting that hits obtained in a Ganzfeld condition are superior to those expected by chance, satisfying all usual criteria to define that a phenomenon may be real or at least very probable.

Is this statistical evidence sufficient to convince lay people and scientists alike? Probably not; all those people who consider ESP too discordant from their conception of the human mind [e.g., those who cannot accept that the human mind may manifest non-local properties or who believe that the human mind is constrained by its biological (neural) correlates] will likely remain skeptical. Of course, the history of science is full of strong denials of phenomena that today are considered "normal." Even though the conception of human mind as non-local or extended may be found in various philosophies (e.g., panpsychism, mental monism, etc.) and religions (Buddhism, Hinduism, etc.), the dominant theory within the scientific community is strictly local. The possibility that the human mind may have non-local properties, even if its biological correlates (brain activity) seem to be constrained by local interactions, poses serious challenges to the mind-brain identity or material (biological) monism assumed by most. For those who see a violation of physical laws, it is important to remind them that physics, as in all scientific disciplines, still has many unsolved problems, and it is suggested that they look carefully at the non-local properties of physical objects empirically investigated in quantum physics. If human information may express non-local properties like physical ones have been shown to express, it is an open theoretical and empirical question (i.e., Tressoldi and Khrennikov, 2012).

The evidence coming from the ESP studies with the Ganzfeld databases underline the importance of mental noise reduction as a fundamental moderator. If we compare the evidence of ESP using different protocols (e.g., forced-choice with participants in normal states of consciousness; see the recent meta-analysis of Storm, Tressoldi, and Di Risio~~, in press~~), the average $ES$ of 0.014 is almost one tenth of those observed with participants in a Ganzfeld state. ESs similar or better than those observed with ESP with Ganzfeld are observed only in those studies investigating implicit (not conscious) responses; for example, in the so-called "anticipatory physiological responses" to random events as demonstrated by Mossbridge, Tressoldi, and Utts (2012) or the "retrocausal effects" using Bem's protocol (Bem, 2011).

To summarize the content of this essay, I can say that studying ESP using a Ganzfeld protocol is one of the more successful stories about the scientific investigation of the human mind, and we should be grateful to the scientific prowess of Charles Honorton, the father of this line of investigation. The evidence available to date appears to satisfy the statistical restrictive criteria required to support this as an "exceptional claim." However, its widespread acceptance now hinges on cultural changes that would allow for respecting the data.

## NOTES

1. The HDI indicates which points of a distribution are believed in most strongly, and which cover most of the distribution. Thus, the HDI summarizes the distribution by specifying an interval that spans most of the distribution, say 95 percent of it, such that every point inside the interval has higher believability than any point outside the interval.

2. The names Alice and Bob are commonly used placeholder names for archetypal characters in fields such as cryptography and physics. The names are used for convenience; for example, "Alice sends a message to Bob encrypted with his public key" is easier to follow than "Party A sends a message to Party B encrypted by Party B's public key." Following the alphabet, the specific names have evolved into common parlance within these fields—helping technical topics to be explained in a more understandable fashion. In typical implementations of these protocols, it is understood that the actions attributed to characters such as Alice or Bob need not always be carried out by human parties directly but also by a trusted automated agent (such as a computer program) on their behalf.

## REFERENCES

Avant, L. L. (1965). Vision in the ganzfeld. *Psychological Bulletin, 64*, 246–258.

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

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

Bem, D. J., Palmer, J., and Broughton, R. S. (2001). Updating the Ganzfeld database: A Victim of Its Own Success? *Journal of Parapsychology, 65*, 207–218.

Borestein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*, 97–111.

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

Chassin, M. R. (1998). Is health care ready for Six Sigma quality? *The Milbank Quarterly, 76*(4), 565–591.
Dawson, R. (1991). Comment on "Replication and Meta-Analysis in Parapsychology" by J. Utts, *Statistical Science, 6*(4), 382–385.
Glass, G. V, McGaw, B., and Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage.
Hedges, L. V., and Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
Honorton, C. (1977). Psi and internai attention states. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 435–472). New York: Van Nostrand Reinhold.
Honorton, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology, 49*, 51–91.
Honorton, C., and Harper, S. (1974). Psi-mediated imagery and ideation in an experimental procedure for regulating perceptual input. *Journal of the American Society for Psychical Research, 68*, 156–168.
Hyman, R. (1985). The ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology, 49*, 3–49.
Hyman, R., and Honorton, C. (1986). A joint communiqué: The psi ganzfeld controversy. *Journal of Parapsychology, 50*, 351–364.
Keppel, G., and Wickens, T. D. (2004). *Design and analysis. A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson Education International.
Krippner, S., and Friedman, H. (Eds.). (2010). *Debating psychic experiences: Human potential or human illusion?* Santa Barbara: Praeger.
Kruschke, J. K. (2011a). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science, 6*(3), 299–312.
Kruschke, J. K. (2011b). *Doing Bayesian data analysis: A tutorial with R and BUGS*. Burlington, MA: Academic Press/Elsevier.
Kruschke, J. K. (2011c). Extrasensory Perception (ESP): Bayesian estimation approach to meta-analysis. Available online at: http://doingbayesiandataanalysis.blogspot.com.
Laplace, P. (n.d.). Quote. *Today in science history* (n.p.). Available online at: http://www.todayinsci.com/L/Laplace_PierreSimon/LaPlacePierreSimon-Quotations.htm.
Milton, J., and Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin, 125*, 387–391.
Milton, J., and Wiseman, R. (2002). A response to Storm and Ertel (2002). *Journal of Parapsychology, 66*, 183–186.
Mossbridge, J., Tressoldi, P., and Utts, J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: a meta-analysis. *Frontiers in Psychology, 3*, 390. doi:10.3389/fpsyg.2012.00390.
Nickerson, R. S. (2000). Null Hypothesis Significance Testing: A review of an old and continuing controversy. *Psychological Methods, 5*(2), 241–301.
Opera collaboration (2011). Measurement of the neutrino velocity with the OPERA detector in the CNGS beam. Available online at: http://arxiv.org/abs/1109.4897.
Roediger, H. (2012). Psychology's woes and a partial cure: The value of replication. *American Psychological Society Observer, 25*(2), 27–29.
Rouder, J. N., and Morey, R. D. (in press). A Bayes-factor meta-analysis of recent ESP experiments: A comment on Storm, Tressoldi, and Di Risio (2010). *Psychological Bulletin*.
Rouder, J. N., and Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin Review*. doi:10.3758/s13423-011-0088-7.
Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is

neglected in the social sciences. *Review of General Psychology, 13*, 90–100. doi:10.1037/a0015108

Schroeder, R. G., Linderman, K., Liedtke, C., and Choo, A.S. (2008). Six Sigma: Definition and underlying theory. *Journal of Operations Management, 26*, 536–554.

Storm, L. (2006). Technical paper no. 11. Meta-analysis in parapsychology: I. The Ganzfeld domain. *Australian Journal of Parapsychology, 6*, 35–53.

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

Storm, L., and Ertel, S. (2002). The ganzfeld debate continued: A response to Milton and Wiseman (2001). *Journal of Parapsychology, 66*, 73–82.

Storm, L., Tressoldi, P. E. (~~in press, a~~). Testing the Storm et al. (2010) meta-analysis using Bayesian and frequentist approaches: A Response to Rouder et al. (2012). *Psychological Bulletin.*

Storm, L., Tressoldi, P. E., and Di Risio, L. (~~in press, b~~). Meta-analysis of ESP studies, 1987–2010: Assessing the success of the forced-choice design in parapsychology. *Journal of Parapsychology.*

Storm, L., Tressoldi, P. E., and Di Risio, L. (2010a). Meta-analysis of free-response studies, 1992–2008: Assessing the noise reduction model in parapsychology. *Psychological Bulletin, 136*, 471–485.

Storm, L., Tressoldi, P. E., and Di Risio, L. (2010b). A meta-analysis with nothing to hide: Reply to Hyman (2010). *Psychological Bulletin, 136*, 491–494.

Tressoldi, P. E. (2011). Extraordinary claims require extraordinary evidence: The case of non-local perception, a classical and Bayesian review of evidences. *Frontiers in Psychology* I*2*(117), 1–5. doi:10.3389/fpsyg.2011.00117

Tressoldi, P. E. (2012). Replication unreliability in psychology: Elusive phenomena or "elusive" statistical power? *Frontiers in Psychology 3*, 218. doi:10.3389/fpsyg.2012.00218

Tressoldi, P. E., and Khrennikov, A. (2012). Remote state preparation of mental information: A theoretical model and a summary of experimental evidence. *NeuroQuantology, 10*(3), 394–402.

Utts, J., Norris, M., Suess, E., and Johnson, W. (2010). The strength of evidence versus the power of belief: Are we all Bayesians? In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society*. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July 2010), Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute. Available online at: https://www.stat.auckland.ac.nz/~iase/publications.php.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., and van der Maas, H. (2011). Why psychologists must change the way they analyze their data: The case of psi. *Journal of Personality and Social Psychology, 100*, 426–432.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J., Iverson, G., and Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6*(3), 291–298.

Williams, B. J. (2011). Revisiting the Ganzfeld ESP debate: A basic review and assessment. *Journal of Scientific Exploration, 25*(4), 639–661.