Psychological Bulletin
2013, Vol. 139, No. 1, 241–247

0033-2909/13/$12.00 DOI: 10.1037/a0029008

COMMENT

A Bayes Factor Meta-Analysis of Recent Extrasensory Perception Experiments: Comment on Storm, Tressoldi, and Di Risio (2010)

Jeffrey N. Rouder
University of Missouri

Richard D. Morey
University of Groningen

Jordan M. Province
University of Missouri

Psi phenomena, such as mental telepathy, precognition, and clairvoyance, have garnered much recent attention. We reassess the evidence for psi effects from Storm, Tressoldi, and Di Risio's (2010) meta-analysis. Our analysis differs from Storm et al.'s in that we rely on Bayes factors, a Bayesian approach for stating the evidence from data for competing theoretical positions. In contrast to more conventional analyses, inference by Bayes factors allows the analyst to state evidence for the no-psi-effect null as well as for a psi-effect alternative. We find that the evidence from Storm et al.'s presented data set favors the existence of psi by a factor of about 6 billion to 1, which is noteworthy even for a skeptical reader. Much of this effect, however, may reflect difficulties in randomization: Studies with computerized randomization have smaller psi effects than those with manual randomization. When the manually randomized studies are excluded and omitted studies included, the Bayes factor evidence is at most 330 to 1, a greatly attenuated value. We argue that this value is unpersuasive in the context of psi because there is no plausible mechanism and because there are almost certainly omitted replication failures.

Keywords: psi phenomena, ESP, Bayes factor, Bayesian meta-analysis

The term *psi* refers to a class of phenomena more colloquially known as extrasensory perception, and includes telepathy, clairvoyance, and precognition. Although psi has a long history at the fringes of psychology, it has recently become more prominent with Bem's (2011) claim that people may literally feel the future and Storm, Tressoldi, and Di Risio's (2010) meta-analytic conclusion that there is broad-based evidence for psi in a variety of domains. In previous work, we critiqued Bem's demonstration on statistical grounds and showed that the provided evidence was not convincing (Rouder & Morey, 2011; see also Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). In this article, we assess the evidence in Storm et al.'s meta-analysis.

Jeffrey N. Rouder, Department of Psychological Sciences, University of Missouri; Richard D. Morey, Faculty of Behavioral and Social Sciences, University of Groningen, Groningen, the Netherlands; Jordan M. Province, Department of Psychological Sciences, University of Missouri.

This research is supported by National Science Foundation Grant SES-1024080. We thank Patrizio Tressoldi for graciously sharing the data and computations in Storm et al. (2010). This research would not have been possible without his openness and professionalism.

Correspondence concerning this article should be addressed to Jeffrey N. Rouder, Department of Psychological Sciences, University of Missouri, 212D McAlester Hall, Columbia, MO 65211. E-mail: rouderj@missouri.edu

Our main concern is that Bem (2011) and Storm et al. (2010) do not provide principled measures of the evidence from their data. Bem, for example, relies on conventional null hypothesis significance testing (NHST). NHST has a well-known and important asymmetry: The researcher can only accumulate evidence for the alternative, and the null serves as a straw-man hypothesis that may only be rejected. In assessments of psi, the null hypothesis corresponds to the plausible and reasonable position that there is no psi. It is problematic that such a reasonable position may only be rejected and never accepted in NHST. Storm et al. performed a conventional meta-analysis where the goal was to estimate the central tendency and dispersion of effect sizes across a sequence of studies, as well as to provide a summary statement about these effect sizes. They found a summary z score of about 6, which corresponds to an exceedingly low p value. Yet, the interpretation of this p value was conditional on never accepting the null, effectively ruling out the skeptical hypothesis a priori (see Hyman, 2010).

Problems with the interpretation of NHST are well known in the statistical community, and there are many authors who advocate Bayes factor as a principled approach for assessing evidence from data (Berger & Berry, 1988; Jeffreys, 1961; Kass, 1992). The Bayes factor, first proposed by Laplace (1986), is the probability of the data under one hypothesis relative to the probability of the data under another. These hypotheses may be null or alternatives, and in this manner, there is no asymmetry in the treatment of the null. The Bayes factor describes the degree to which researchers and readers should update their beliefs about the relative plausibility of the two hypotheses in light of the data. Many authors, including Bem, Utts, and Johnson (2011); Edwards, Lindman, and Savage (1963); Gallistel (2009); Rouder, Speckman, Sun, Morey, and Iverson (2009); and Wagenmakers (2007), advocate inference by Bayes factors in psychological settings.

In our assessment of Bem's (2011) data, we found Bayes factor values ranging from 1.5 to 1 to 40 to 1 in favor of a psi effect, with the value dependent on the type of stimulus. Consider the largest value, 40 to 1, which is the evidence for a psi effect with emotionally evocative, nonerotic stimuli. Researchers who held beliefs that a psi effect was as likely to exist before observing the data, should hold beliefs that favor a psi effect by a factor of 40 after observing them. We, however, remain skeptical. Given the lack of mechanism for the feeling-the-future hypothesis, and its discordance with well-established principles in physics, we agree with Bem that it is prudent to hold a priori beliefs that favor the nonexistence of psi, perhaps by several orders of magnitude. Against this appropriate skepticism, the factor of 40 from the data is unimpressive. We emphasize here that a Bayes factor informs the community about how beliefs should change. Different researchers with different a priori beliefs may hold different a posterior beliefs while agreeing on the evidence from data. The goal in this article is to provide a Bayes factor assessment of the evidence for psi provided by Storm et al.'s (2010) large meta-analysis. A similar endeavor is undertaken by Tressoldi (2011), though our conclusions differ substantially from his.

A Reassessment of Storm et al. (2010)

Storm et al. (2010) provided a meta-analysis of 67 psi experiments conducted from 1992 to 2008. These experiments typically involve three people: a *sender,* a *receiver,* and a *judge.* The sender telepathically broadcasts an item to the receiver, who is isolated from the sender. The receiver then describes his or her thoughts about the item in a free-report format. The judge, who is also isolated from the sender, hears the free report from the receiver and decides which of several possible targets this free report best matches. One of these targets is the sent item, and the judge is said to be correct if he or she chooses this target as the best match.

Table 1 shows a Bayes factor analysis for a number of data sets and models. The rows of the table indicate the data set, and the columns indicate which models are compared. For now, we focus on the first row, for full set, and the first column, for one effect and informed prior. The full set includes all 67 studies analyzed by Storm et al. (2010), and the details of the one-effect informed prior model are discussed subsequently. The Bayes factor is about 6 billion to 1, which is a large degree of statistical support. These values indicate that readers should update their priors by at least nine orders of magnitude, which is highly noteworthy. The value we obtain is larger than the 19-million-to-1 Bayes factor reported by Tressoldi (2011) on an expanded set of 108 studies.[1] In summary, there is ample evidence in the data set as constituted to sway a skeptical but open-minded reader. As discussed next, however, there is reason to suspect that perhaps the data set is not well constituted.

Issues With Storm et al.'s (2010) Data Set

We carefully examined the nine studies that provide the highest degree of support for psi.[2] Some of these studies are documented thoroughly and appear to use standard and accepted experimental controls (e.g., Del Prete & Tressoldi, 2005; Smith & Savva, 2008; Tressoldi & Del Prete, 2007; Wezelman, Gerding, & Verhoeven, 1997). Nonetheless, the following key problems were evident either in the studies themselves or in their treatment in the Storm et al. meta-analysis.

Lack of Internal Validity

May (2007) provided seemingly strong evidence for psi; he reported 64% accuracy across 50 three-choice trials ($z = 4.57$, $p < .001$). May's statistical procedures, however, are opaque. He constructed an idiosyncratic and difficult-to-interpret statistic that he called "the figure of merit." Unfortunately, May presented no theoretical sampling distribution of the figure-of-merit statistic under the null. Instead, he constructed this null sampling distribution from the performance of three participants contributing 15 trials each. Hence, the distribution under the null has unaccounted-for variability, and cannot be used to standardize performance in psi conditions. We exclude this experiment because it lacks sufficient internal validity.

Shaping the Randomization Process

One of the key methodological components in exploring psi is proper randomization of trials (Hyman & Honorton, 1986). Storm et al. (2010) stated that they included only studies in which randomization was proper and was performed only by computer algorithm or with reference to random-number tables. Yet, we found examples of included studies that either did not mention how randomization was achieved (e.g., Dalton, 1997) or added an extra step of discarding "atypical" sequences. Consider, for example, Targ and Katra (2000), who stated: "These pictures were selected randomly, and then filtered to provide a representative mixture of possible targets to avoid any accidental stacking that could occur if, for example, we had an overrepresentation . . . of [a particular picture]" (p. 110). Clearly, such shaping can only have negative consequences, as it disrupts the randomization that lies at the heart of the experimental method (Hyman & Honorton, 1986).

Fortunately, Storm et al. (2010) indicated in their spreadsheet whether each study was *computer randomized* or *manually randomized.* Manual randomization is a heterogeneous class of studies including those where randomization is not mentioned (e.g., Dalton, 1997) or was shaped (e.g., Targ & Katra, 2000). If manual randomization is innocuous, then there should be no difference in

[1] Tressoldi (2011) used our meta-analytic Bayes factor (Rouder & Morey, 2011) in which it is assumed that the data are normally rather than binomial distributed. The normal model may be less efficient because it contains two base parameters (mean, variance) rather than one.

[2] We originally set out to survey the 12 studies referenced in Storm et al. (2010) that yielded z scores over 2.0. Unfortunately, it is difficult to obtain these studies as they are neither carried by many academic institutions nor available through interlibrary loan.

Table 1
Bayes Factor Assessment of Storm et al.'s (2010) Data Sets

	One effect		Multiple effects		Three effects	
Data set	Informed	Uniform	Informed	Uniform	Informed	Uniform
Full set	5.59×10^{9}	1.69×10^{9}	3.08×10^{11}	1.05×10^{-16}	2.40×10^{14}	7.30×10^{12}
Revised Set 1	63.3	17.7	1.25×10^{-6}	5.58×10^{-28}	2,973	76.3
Revised Set 2	31.7	8.77	5.45×10^{-8}	1.95×10^{-30}	328	7.85

performance across computerized and manual randomization procedures.

Before we assess whether performance varied across randomization strategies, the status of Lau (2004) needs consideration. In one of his experiments, Lau ran an unusually large number of number of trials, 937, which is more than 20% of the total number of trials in the data set and more than 7 times larger than the next largest experiment (128 trials). Storm et al. (2010) classified Lau's studies as manually randomized, and the study with 937 trials accounts for 49% of the total number of manually randomized trials. Yet, in the introduction to his studies, Lau discussed the importance of proper randomization. In the method section, however, he provided no further detail. We contacted Lau and learned through personal communication that he generated random number sequences via the Research Randomizer website (http://www.randomizer.org), which uses the Math.random JavaScript function. Hence, we have reclassified his studies as computer randomized.

Figure 1 shows the distribution of accuracy across the 63 studies where the judge had four choices. As can be seen, manual randomization leads to better psi performance than computerized randomization. We performed a Bayes factor analysis of all studies except May (2007) and found that the evidence for a difference in performance is about 6,350 to 1. We discuss the construction of this Bayes factor subsequently. A reasonable explanation for this difference is that there is a flaw in at least some of the manual randomization studies, leading to predictable dependencies between experimental trials. No psi is needed to explain higher-than-chance performance under these conditions.

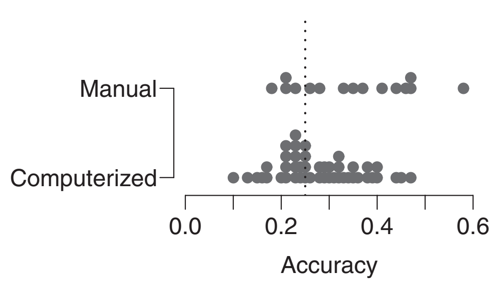

Figure 1. Distribution of accuracy across psi experiments as a function of the implementation of randomization. In computerized randomization, computers drew random numbers without any human filtering. In manual randomization, either there was filtering for atypical sequences or the method of randomization was not mentioned. The figure shows those studies with four choices, and chance performance corresponds to .25.

Selection of Studies

We noticed in our brief survey that not all the data in the reports were included in the Storm et al. (2010) meta-analysis. Consider the work of Del Prete and Tressoldi (2005), who ran two extrasensory perception conditions: one standard and one under hypnosis. In the hypnosis condition, Del Prete and Tressoldi observed 45 successes out of 120 trials (37.5%) in four-choice trials (chance baseline performance of 25%). In the condition with no hypnosis, there were 29 successes out of 120 trials (24.2%). Storm et al. included the first condition but not the second. This exclusion is surprising in the context of their meta-analysis because the no-hypnosis condition is similar to other included studies. Another example of selectivity comes from the treatment of Tressoldi and Del Prete (2007), who also ran psi experiments under hypnosis. These researchers used two sets of instructions, one to imagine an out-of-body experience and a second with more standard remote-viewing instructions. Instructions were manipulated within subjects in an AB design; half the participants had the out-of-body instructions first and the remote-viewing instructions second. The other half had the reverse. There was no effect of the instructions, but there was an unexpected effect of order. There was a psi effect for the first block of trials (a combined 40 successes out of 120 four-choice trials) but not for the second (a combined 29 successes out of 120 four-choice trials). Storm et al. included only the first block of trials but not the second. We see no basis for such an ad hoc exclusion given the criteria set out by Storm et al. These two omissions are examples of a selection artifact.

Analysis of Revised Data Sets

A prudent course is to analyze the set with the manual randomization studies excluded.[3] Of the original set of 67 studies, we excluded May (2007; insufficient internal validity) and 19 others that had manual randomization (see Appendix). We include two sets from Lau (2004), as these used computer randomization without any human filtering. We call this reduced set of 47 studies Revised Set 1. We also constructed a second revised set, Revised Set 2, by including the omitted conditions from Del Prete and Tressoldi (2005) and Tressoldi and Del Prete (2007). The additional rows in Table 1 provide Bayes factors for these two revised

[3] We do not wish to imply that Storm et al. (2010) are imprudent in their inclusion of the manual randomization studies. Claims of psi are sufficiently theoretically important and controversial that the community benefits from multiple analyses with these studies included and excluded, as we have done here.

sets. As can be seen, the Bayes factor in the first column is no longer a towering value of several orders of magnitude. Instead, it is around 63 to 1 and 32 to 1 for the two sets, respectively. Context for this value, as well as others in the table, is provided subsequently.

Bayes Factor Analysis

In this section, we describe the computation of the Bayes factor and the development of psi alternative hypotheses. The Bayes factor is the ratio of the probability of data under competing hypotheses H_1 and H_0:

$$B = \frac{\Pr\ (\mathrm{Data}|H_1)}{\Pr\ (\mathrm{Data}|H_0)}.$$

Let Y_i, N_i, and K_i denote the number of correct responses, the number of trials, and the number of choices per trial for the ith study, $i = 1, \ldots, I$. In this case, the binomial is a natural model of the data. One property of the Storm et al. (2010) data set is that the studies span a range of number of choices. Y_i is modeled as

$$Y_i \sim \mathrm{Binomial}(N_i, p_i),$$

where

$$p_i = \frac{1}{K_i} + \left(1 - \frac{1}{K_i}\right)\eta_i.$$

The free parameter η_i denotes the performance on the ith study, with higher values of η_i corresponding to better true performance. Parameter η_i ranges from 0 to 1, and these anchors denote floor and ceiling levels of performance, respectively.

One key property of Bayes factors is that they are sensitive to prior assumptions about parameters. Although some critics consider this dependency may be problematic (e.g., Gelman, Carlin, Stern, & Rubin, 2004; Liu & Aitkin, 2008), we consider it an opportunity to explore several different types of prior assumptions about psi effects. This strategy of exploring a range of psi alternatives is also used by Bem et al. (2011) in their Bayes factor analysis.

Under the no-psi null hypothesis, the prior on η_i has all the mass at the point $\eta_i = 0$ for all studies. With this prior,

$$\Pr(\mathrm{Data}|H_0) = \prod_{i=1}^{I} f(Y_i, N_i, K_i^{-1}),$$

where f is the probability mass function of the binomial distribution.[4]

Specifying priors that include psi effects is more complicated than specifying priors for the no-psi null. One could specify an alternative hypothesis by committing a priori to a specific known performance level, say, $\eta_i = .10$ for all studies. This commitment, however, is too constraining to be persuasive. Fortunately, in Bayesian statistics, one can specify an alternative that encompasses a range of prior values for η_i. We first develop priors for the case there is a single unknown performance parameter η for all studies, that is, $\eta_1 = \cdots = \eta_I = \eta$. Let $\pi(\eta)$ denote a prior density for η. Two examples of $\pi(\eta)$ are given in Figure 2A. The solid line, which is a uniform distribution, shows the case where η takes on values with equal density. The dashed line is a different prior that favors smaller values of η over larger ones. This is an informative prior that captures the belief that psi effects should be small. Both priors in Figure 2A are beta distributions, which is a flexible and convenient form when data are binomially distributed.[5] The corresponding priors on p, the probability of success, is shown for the four-choice studies ($k = 4$) in Figure 2B.

With these specifications:

$$\Pr(\mathrm{Data}|H_1) = \int_0^1 \left[\prod_i f\left(Y_i, N_i, \frac{1}{K_i} + \left(1 - \frac{1}{K_i}\right)\eta\right)\right]\pi(\eta)d\eta,$$

where π is the probability density function of the uniform or informed beta distribution. The one-dimensional integral may be performed accurately and quickly by numeric methods such as Gaussian quadrature (Press, Teukolsky, Vetterling, & Flannery, 1992). The resulting Bayes factor for both priors is shown in Table 1 in the columns labeled "One effect." There is no penalty or correction needed for considering multiple alternative models with Bayes factor; one may consider as many priors as one desires without any loss. The resulting Bayes factor is always qualified by the reasonable or appropriateness of the prior. We believe in this case that the one-effect informed prior is perhaps the most appropriate of those we explore here.

In these one-effect priors, there is a single true-performance parameter for all studies. This degree of homogeneity, however, may be unwarranted. We constructed multiple-effect priors that allowed a separate parameter η_i for each study. The prior on each performance parameter η_i is an independent and identical beta distribution. We considered a uniform ($\alpha = \beta = 1$) and informed prior ($\alpha = 1$, $\beta = 4$) for each η_i. The resulting marginal probability is shown at the bottom of the page.

$$\Pr(\mathrm{Data}|H_1) = \int_0^1 \cdots \int_0^1 \prod_i \left[f\left(Y_i, N_i, \frac{1}{K_i} + \left(1 - \frac{1}{K_i}\right)\eta_i\right)\pi(\eta_i)\right] d\eta_1 \cdots d\eta_I$$

$$= \prod_i \left[\int_0^1 f\left(Y_i, N_i, \frac{1}{K_i} + \left(1 - \frac{1}{K_i}\right)\eta_i\right)\pi(\eta_i)d\eta_i\right].$$

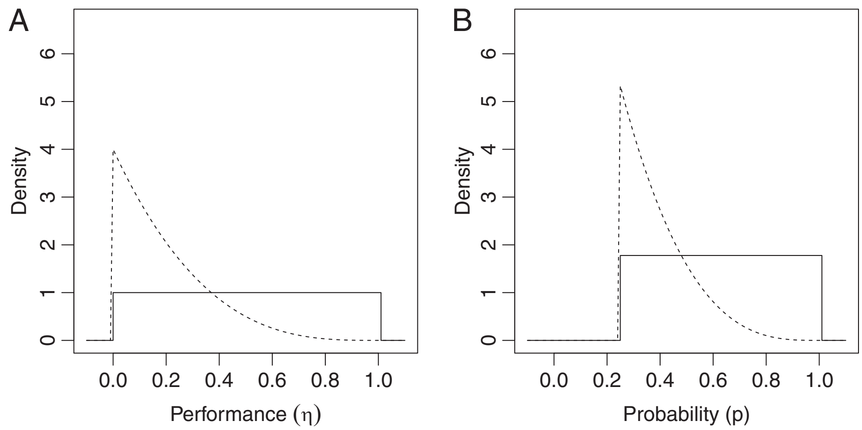

Figure 2. The informed prior (dashed lines) and uniform prior (solid lines) used in analysis: priors on performance parameter η (A) and priors on probability parameter p for a four-choice experiment (B).

The last expression is the product of one-dimensional integrals and may be conveniently evaluated with standard numerical techniques. The resulting Bayes factors are shown in Table 1 under the columns labeled "Multiple effects." Multiple-effect priors with multiple performance parameters fare relatively poorly. They are too richly parameterized and too flexible for the simple structure and relatively small sample sizes of the studies in the data set. For this set, it is more appropriate to consider one-effect models than multiple-effect models.

We also considered priors in which there are three effects rather than many. The motivation for this choice comes from Storm et al. (2010), who divided the experiments in the meta-analysis into three categories based on the conscious state of the receiver in the experiment. In one category, the receivers were in their normal waking state of consciousness. In the other two categories, receivers were in altered state of consciousness. In the second category, consciousness was altered by the ganzfeld procedure; in the final category consciousness was altered by some other technique such as hypnosis or advanced relaxation. To model this difference in conscious state, we allowed all experiments within a category common performance parameter, but there were separate performance parameters across the three categories. As before, informed and uniform prior settings were used on performance parameters, and the results are shown in the last two columns labeled "Three effects." These three-effect priors yielded the strongest support for psi, about 330 to 1 for Revised Set 2. Interpretation and qualifications are provided in the Conclusion.

As discussed previously, we also performed a Bayes factor analysis to assess the difference in performance between the 47 studies with computer randomization and the 19 studies with manual randomization. This analysis was performed assuming one common performance parameter for computer-randomized studies and a different common performance parameter for manually randomized studies. The prior on each of these performance parameters was the informed prior in Figure 2A (dashed line). The resulting value of 6,350 to 1 provides evidence for the proposition that studies with manual randomization had higher performance than those with computerized randomization.

Conclusion

We agree with Storm et al. (2010) and Tressoldi (2011) that uncritical consideration of full set of recent psi experiment provides strong statistical evidence for a psi effect. The Bayes factor, the ratio of the probability of the data under competing hypotheses, is on the order of billions to one or higher in favor of an effect, and the magnitude of this factor implies that even skeptics would need to substantially revise their beliefs. Nonetheless, closer examination of the data set reveals that the method of randomization affects performance. Experiments with manual randomization resulted in higher performance than those with computerized randomization (Bayes factor of 6,350 to 1). When these manually randomized experiments are excluded, the evidence for psi is attenuated by at least eight orders of magnitude (hundred million). Moreover, this attenuation does not take into account the possibility of file-drawer selectivity artifacts. In our brief review of just eight notable psi experiments, we found two data sets from Del Prete and Tressoldi (2005) and Tressoldi and Del Prete (2007), that should have been included. When these two sets are included, the largest Bayes factor for psi is 330 to 1, and this value is conditional on psi differences across altered states of consciousness. Although this degree of support is greater than that provided in many routine studies in cognition (Wetzels et al., 2011), we nonetheless remain skeptical of the existence of psi for the following two reasons:

[4] The probability mass function of a binomial distribution for y successes in N trials with probability parameter p is

$$f(y, n; p) = \binom{n}{y} p^y (1-p)^{n-y} \quad 0 \le p \le 1.$$

[5] The probability density function of a beta distribution for probability p with parameters α and β is

$$f(p; \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 \le p \le 1,\ \alpha, \beta > 0,$$

where B is the beta function (Press et al., 1992). For the uniform prior, $\alpha = \beta = 1$; for the informed prior, $\alpha = 1$ and $\beta = 4$.

1. The Bayes factor describes how researchers should update their prior beliefs. Bem (2011) and Tressoldi (2011) provided the appropriate context for setting these prior beliefs about psi. They recommended that researchers apply Laplace's maxim that extraordinary claims require extraordinary evidence. Psi is the quintessential extraordinary claim because there is a pronounced lack of any plausible mechanism. Accordingly, it is appropriate to hold very low prior odds of a psi effect, and appropriate odds may be as extreme as millions, billions, or even higher against psi. Against such odds, a Bayes factor of even 330 to 1 seems small and inconsequential in practical terms. Of course for the unskeptical reader who may believe a priori that psi is as likely to exist as not to exist, a Bayes factor of 330 to 1 is considerable.

2. Perhaps more importantly, the Bayes factors in Table 1 should be viewed as upper bounds on the evidence from Storm et al. (2010). We are struck in that reviewing only eight studies, we found a host of infelicities including missing data sets from Del Prete and Tressoldi (2005) and Tressoldi and Del Prete (2007). Including these two studies reduced the three-effect model Bayes factor by a factor of 9. In all likelihood, these are not the only two missing sets, and it is reasonable to worry about the existence of others. Our concern differs from Storm et al., who concluded there would have to be at least 86 null studies missing from the meta-analysis to account for their significant findings. This computation, however, rests on the full set, which is seemingly contaminated by studies without proper randomization. As an aside, we are not convinced that either the philosophical or distributional assumptions in Storm et al. are the most satisfying (see, e.g., Givens, Smith, & Tweedie, 1997, for a Bayesian approach to estimating the number of missing studies in a meta-analysis). We simply note here that the obtained Bayes factors are upper bounds and the true value may be less favorable for psi.

In summary, although Storm et al.'s (2010) meta-analysis seems to provide a large degree of support for psi, more critical evaluation reveals that it does not. In our view, the evidence from Storm et al. for psi is relatively equivocal and certainly not sufficient to sway an appropriately skeptical reader.

References

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., Utts, J., & Johnson, W. O. (2011). Must psychologists change the way they analyze their data? *Journal of Personality and Social Psychology, 101,* 716–719. doi:10.1037/a0024777

Berger, J. O., & Berry, D. A. (1988). Statistical analysis and the illusion of objectivity. *American Scientist, 76,* 159–165.

Dalton, K. (1997). Exploring the links: Creativity and psi in the ganzfeld. In *Proceedings of the 40th Annual Convention of the Parapsychological Association* (pp. 119–134). Durham, NC: Parapsychological Association.

Dalton, K., Steinkamp, F., & Sherwood, S. J. (1999). A dream GESP experiment using dynamic targets and consensus vote. *Journal of the American Society for Psychical Research, 96,* 145–166.

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

da Silva, F. E., Pilato, S., & Hiraoka, R. (2003). Ganzfeld vs. no ganzfeld: An exploratory study of the effects of ganzfeld conditions on ESP. In *Proceedings of the 46th Annual Convention of the Parapsychological Association* (pp. 31–49). Durham, NC: Parapsychological Association.

Del Prete, G., & Tressoldi, P. E. (2005). Anomalous cognition in hypnagogic state with OBE induction: An experimental study. *Journal of Parapsychology, 69,* 329–339.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242. doi:10.1037/h0044139

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review, 116,* 439–453. doi:10.1037/a0015251

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London, England: Chapman & Hall.

Givens, G. H., Smith, D. D., & Tweedie, R. L. (1997). Publication bias in meta-analysis: A Bayesian data-augmentation approach to account for issues exemplified in the passive smoking debate. *Statistical Science, 12,* 221–250. doi:10.1214/ss/1030037958

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

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

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). New York, NY: Oxford University Press.

Kass, R. E. (1992). Bayes factors in practice. *The Statistician, 42,* 551–560.

Laplace, P. S. (1986). Memoir on the probability of the causes of events. *Statistical Science, 1,* 364–378. doi:10.1214/ss/1177013621

Lau, M. (2004). *The psi phenomena: A Bayesian approach to the ganzfeld procedure.* (Unpublished master's thesis). University of Notre Dame, South Bend, IN.

Liu, C. C., & Aitkin, M. (2008). Bayes factors: Prior sensitivity and model generalizability. *Journal of Mathematical Psychology, 52,* 362–375. doi:10.1016/j.jmp.2008.03.002

May, E. C. (2007). Advances in anomalous cognition analysis: A judge-free and accurate confidence-calling technique. In *Proceedings of the 50th Annual Convention of the Parapsychological Association* (pp. 57–63). Petaluma, CA: Parapsychological Association.

Parker, A., & Westerlund, J. (1998). Current research in giving the ganzfeld an old and a new twist. In *Proceedings of the 41st Annual Convention of the Parapsychological Association* (pp. 135–142). Durham, NC: Parapsychological Association.

Parra, A., & Villanueva, J. (2004). Are musical themes better than visual images as ESP-targets? An experimental study using the ganzfeld technique. *Australian Journal of Parapsychology, 4,* 114–127.

Parra, A., & Villanueva, J. (2006). ESP under the ganzfeld, in contrast with the induction of relaxation as a psi-conducive state. *Australian Journal of Parapsychology, 6,* 167–185.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, F. P. (1992). *Numerical recipes in C: The art of scientific computing* (2nd ed.). Cambridge, England: Cambridge University Press.

Roe, C. A., & Flint, S. (2007). A remote viewing pilot study using a ganzfeld induction procedure. *Journal of the Society for Psychical Research, 71,* 230–234.

Roe, C. A., McKenzie, E. A., & Ali, A. N. (2001). Sender and receiver creativity scores as predictors of performance at a ganzfeld ESP task. *Journal of the Society for Psychical Research, 65,* 107–121.

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review, 18,* 682–689. doi:10.3758/s13423-011-0088-7

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. doi:10.3758/PBR.16.2.225

Simmonds-Moore, C., & Holt, N. J. (2007). Trait, state, and psi: A comparison of psi performance between clusters of scorers on schizotypy in a ganzfeld and waking control condition. *Journal of the Society for Psychical Research, 71,* 197–215.

Smith, M. D., & Savva, L. (2008). Experimenter effects in the ganzfeld. In *Proceedings of the 51st Annual Convention of the Parapsychological Association* (pp. 238–249). Columbus, OH: Parapsychological Association.

Storm, L. (2003). Remote viewing by committee: RV using a multiple agent/multiple percipient design. *Journal of Parapsychology, 67,* 325–342.

Storm, L., & Barrett-Woodbridge, M. (2007). Psi as compensation for modality impairment—A replication study using sighted and blind participants. *European Journal of Parapsychology, 22,* 73–89.

Storm, L., & Thalbourne, M. A. (2001). Paranormal effects using sighted and vision-impaired participants in a quasi-ganzfeld task. *Australian Journal of Parapsychology, 1,* 133–170.

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

Targ, R., & Katra, J. E. (2000). Remote viewing in a group setting. *Journal of Scientific Exploration, 14,* 107–114.

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

Tressoldi, P. E., & Del Prete, G. (2007). ESP under hypnosis: The role of induction instructions and personality characteristics. *Journal of Parapsychology, 71,* 125–137.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14,* 779–804. doi:10.3758/BF03194105

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100,* 426–432. doi:10.1037/a0022790

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspectives on Psychological Science, 6,* 291–298. doi:10.1177/1745691611406923

Wezelman, R., Gerding, J. L. F., & Verhoeven, I. (1997). Eigensender ganzfeld psi: An experiment in practical philosophy. *European Journal of Parapsychology, 13,* 28–39.

Appendix

List of Studies Excluded From the Full Set to Form Revised Set 1

Study	No. of trials	No. correct	No. of choices
Dalton (1997)	128	60	4
Dalton et al. (1999)	32	15	4
Dalton et al. (2000)	16	7	4
da Silva et al. (2003), ganzfeld condition	54	18	4
da Silva et al. (2003), nonganzfeld condition	54	10	4
May (2007)	50	32	3
Parker & Westerlund (1998), serial study	30	7	4
Parker & Westerlund (1998), Study 4	30	14	4
Parker & Westerlund (1998), Study 5	30	11	4
Parra & Villanueva (2004), picture	54	25	4
Parra & Villanueva (2004), music clips	54	19	4
Parra & Villanueva (2006), ganzfeld condition	138	57	4
Parra & Villanueva (2006), nonganzfeld condition	138	57	4
Roe & Flint (2007)	14	4	8
Roe et al. (2001)	24	5	4
Simmonds & Holt (2007)	26	8	4
Storm (2003)	10	5	5
Storm & Barrett-Woodbridge (2007)	76	16	4
Storm & Thalbourne (2001)	84	22	4
Targ & Katra (2000)	24	14	4

Received June 8, 2011
Revision received April 5, 2012
Accepted April 19, 2012 ■