



STUDY PROTOCOL

**REVISED** Stage 1 Registered Report: Anomalous perception in a Ganzfeld condition - A meta-analysis of more than 40 years investigation [version 3; peer review: 1 approved, 2 approved with reservations]

Patrizio E. Tressoldi <sup>1</sup>, Lance Storm<sup>2</sup>

<sup>1</sup>Science of Consciousness Research Group, Università degli studi di Padova, Padova, ITALY, 35131, Italy

<sup>2</sup>School of Psychology, University of Adelaide, Adelaide, Australia, 5005, Australia

**V3** First published: 30 Jul 2020, 9:826  
<https://doi.org/10.12688/f1000research.24868.1>  
 Second version: 16 Oct 2020, 9:826  
<https://doi.org/10.12688/f1000research.24868.2>  
 Latest published: 26 Feb 2021, 9:826  
<https://doi.org/10.12688/f1000research.24868.3>

**Abstract**

This meta-analysis is an investigation into anomalous perception (i.e., conscious identification of information without any conventional sensorial means). The technique used for eliciting an effect is the ganzfeld condition (a form of sensory homogenization that eliminates distracting peripheral noise). The database consists of peer-reviewed studies published between January 1974 and June 2020 inclusive. The overall effect size will be estimated using a frequentist and a Bayesian random-effect model. Moderators analyses will be used to examine the influence of level of experience of participants, the type of task and the peer-review level. Publication bias will be estimated by using four different tests. Trend analysis will be conducted with a cumulative meta-analysis and a meta-regression model with Year of publication as covariate.

**Keywords**

meta-analysis, ganzfeld; anomalous cognition, publication bias; consciousness

**Open Peer Review**

**Reviewer Status** ✓ ? ?

	Invited Reviewers		
	1	2	3
<b>version 3</b> (revision) 26 Feb 2021			? report
<b>version 2</b> (revision) 16 Oct 2020			? report
<b>version 1</b> 30 Jul 2020	✓ report	? report	

- Jessica Utts**, University of California, Irvine, Irvine, USA
- Julia Haaf**, University of Amsterdam, Amsterdam, The Netherlands
- Stefan Schmidt** , University of Freiburg, Freiburg, Germany  
 Institute of Frontier Areas of Psychology and Mental Health, Freiburg, Germany

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Patrizio E. Tressoldi ([patrizio.tressoldi@unipd.it](mailto:patrizio.tressoldi@unipd.it))

**Author roles:** **Tressoldi PE:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Writing – Original Draft Preparation; **Storm L:** Conceptualization, Data Curation, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** The author(s) declared that no grants were involved in supporting this work.

**Copyright:** © 2021 Tressoldi PE and Storm L. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Tressoldi PE and Storm L. **Stage 1 Registered Report: Anomalous perception in a Ganzfeld condition - A meta-analysis of more than 40 years investigation [version 3; peer review: 1 approved, 2 approved with reservations]**

F1000Research 2021, 9:826 <https://doi.org/10.12688/f1000research.24868.3>

**First published:** 30 Jul 2020, 9:826 <https://doi.org/10.12688/f1000research.24868.1>

**REVISED Amendments from Version 2**

- Clarified and revised V2 taking in account the comments of the reviewer
- Expanded the description of effect size estimation

**Any further responses from the reviewers can be found at the end of the article**

## Introduction

The possibility of identifying pictures or video clips without conventional (sensorial) means, in a ganzfeld environment, is a decades old controversy, dating back to the pioneering investigation of Charles Honorton, William Braud and Adrian Parker between 1974 and 1975 (Parker, 2017).

In the ganzfeld, a German term meaning ‘whole field’, participants are immersed in an homogeneous sensorial field where peripheral visual information is masked out by red light diffused by translucent hemispheres (often split halves of ping-pong balls or special glasses) placed over the eyes, while a relaxing rhythmic sound, or white or pink noise, is fed through headphones to shield out peripheral auditory information. Once participants are sensorially isolated from external visual and auditory stimulation, they are in a favourable condition for producing inner mental contents about a randomly selected target hidden amongst decoys. The mentation they produce can either be used by the participant to guide his/her target selection, or it can be used to assist in an independent judging process.

In the prototypical procedure, participants are tested in a room isolated from external sounds and visual information. After they made themselves comfortable in a reclining armchair, they receive the instructions related their task during the ganzfeld condition. Even if there are different verbatim versions, the instructions describe what they should do mentally in order to detect the information related to the target and how to filter out the mental contents not related to it. This information will be described aloud and recorded for playback before or during the target identification phase. After the relaxation phase, they are exposed to the ganzfeld condition for a period ranging from 15 to 30 minutes. During this phase, participants describe verbally all images, feelings, emotions, they deem related to the target usually a picture or a short videoclip of real objects or events.

Once the ganzfeld phase is completed, participants are presented with different choices (e.g. the target plus three decoys) of the same format, e.g. picture or videoclip, and they must choose which one is the target (binary decision). Alternatively, they may be asked to rate all four (e.g., from 0 to 100), to indicate the strength of relationship between the information detected during the ganzfeld phase and the images or video clips contents.

A variant of the judgment phase is to send the recording of the information retrieved during the ganzfeld phase to an

external judge for independent ratings of the target. In order to prevent voluntary or involuntary leakage of information about the target by the experimenters, the research assistant who interact with the participants must be blind to the target identity until the participants’ rating task is over. The choice of the target and the decoys is usually made using automatic random procedures, and scores are automatically fed onto a scoring sheet.

There are three different ganzfeld conditions:

- Type 1: the target is chosen after the judgment phase;
- Type 2: the target is chosen before the ganzfeld phase;
- Type 3: the target is chosen before the ganzfeld phase and presented to a partner of the participant isolated in a separate and distant room. From an historical perspective, this last type is considered the typical condition.

These differences are related to some theoretical and perceptual concepts we will discuss later. It is important to note that type of task makes no difference to the participant who only engages in target identification *after* the ganzfeld phase.

## Review of the Ganzfeld Meta-Analyses

It is interesting to note that most of the cumulative findings (meta-analyses) of this line of investigation were periodically published in the mainstream journal *Psychological Bulletin*.

Honorton (1985) undertook one of the first meta-analyses of the many ganzfeld studies completed by the mid-1980s. In total, 28 studies yielded a collective hit rate (correct identification) of 38%, where mean chance expectation (MCE) was 25%. Various flaws in his approach were pointed out by Hyman (1985), but in their joint-communiqué they agree that “there is an overall significant effect in this database that cannot reasonably be explained by selective reporting or multiple analysis” (Hyman & Honorton, 1986, p. 351).

A second major meta-analysis on a set of ‘autoganzfeld’ studies was performed by Bem & Honorton (1994). These studies followed the guidelines laid down by Hyman & Honorton (1986). Moreover the autoganzfeld procedure avoids methodological flaws by using a computer-controlled target randomization, selection, and judging technique. They overall reported hit rate of 32.2% exceeded again the mean chance expectation.

Milton & Wiseman (1999) meta-analysed further 30 studies collected for the period 1987 to 1997; reporting an overall nonsignificant standardized effect size of 0.013. However, Jessica Utts (personal communication, December 11, 2009) using the exact binomial test on trial counts only ( $N = 1198$ ; Hits = 327), found a significant hit rate of 27% ( $p = 0.036$ ).

Storm & Ertel (2001) comparing Milton & Wiseman’s (1999) database with Bem & Honorton’s (1994) one, found the two did not differ significantly. Furthermore Storm and Ertel went

on to compile a 79-study database, which had a statistically significant average standardized effect size of 0.138.

*Storm et al. (2010)*, meta-analysed a database of 29 ganzfeld studies published during the period 1997 to 2008, yielding an average standardized effect size of 0.14. *Rouder et al. (2013)* reassessing *Storm et al.'s (2010)* meta-analysis, with a Bayesian meta-analysis found evidence for the existence of an anomalous perception in the original dataset observing a Bayes Factor of 330 in support to the alternative hypothesis (p. 241). However, they contended the effect could be due to “difficulties in randomization” (p. 241), arguing that ganzfeld studies with computerized randomization had smaller effects than those with manual randomization. The reanalysis by *Storm et al.'s (2013)* showed that this conclusion was unconvincing as it was based on *Rouder et al.'s* faulty inclusion of different categories of study.

In the last meta-analysis by *Storm & Tressoldi (2020)*, related to the studies published from 2008 to 2018, the average standardized effect size was 0.133; 95%CI: 0.06 - 0.18.

### This study

The main aim of this study is to meta-analyse all available ganzfeld studies dating from 1974 up to June 2020 in order to assess the average effect size of the database with the more advanced statistical procedures that should overcome the limitations of the previous meta-analyses. Furthermore, we aim to identify whether there are moderator variables that affect task performance. In particular, we hypothesize that participant type and type of task are two major moderators of effect size (see Methods section).

## Methods

### Reporting guidelines

This study will follow the guidelines of the APA Meta-Analysis Reporting Standard (*Appelbaum et al., 2018*) and the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P, *Moher et al., 2015*).

### Studies retrieval

Retrieval of studies related to anomalous perception in a Ganzfeld environment is simplified, firstly by the fact that most of these studies have already been retrieved for previous meta-analyses, as cited in the introduction. Secondly, this line of investigation is carried out by a small community of researchers. Thirdly, most of the studies of interest to us are published in specialized journals that adopted the editorial policy of accepting paper with results that are statistically non-significant (according to the frequentist approach). This last condition is particularly relevant because it reduces the publication bias due to non publication (file drawer effect) of studies with statistically non significant results often as a consequence of a reduced statistical power.

Furthermore, in order to integrate the previous retrieval method, we will carry-out an online search with [Google Scholar](#),

[PubMed](#) and [Scopus](#) databases of all papers from 1974 to 2020 including in the title and/or the abstract the word “ganzfeld” (e.g. for PubMed: Search: ganzfeld[Title/Abstract] Filters: from 1974 – 2020).

### Studies inclusion criteria

The following inclusion criteria will be adopted:

- Studies related to anomalous perception in a ganzfeld environment;
- Studies must use human participants only (not animals);
- Number of participants must be in excess of two to avoid the inherent problems that are typical in case studies;
- Target selection must be randomized by using a Random Number Generator (RNG) in a computer or similar electronic device, or a table of random numbers. Randomization procedures must not be manipulated by the experimenter or participant;
- Studies must provide sufficient information (e.g., number of trials and outcomes) for the authors to calculate the direct hit-rates and effect size values, so that appropriate statistical tests can be conducted.
- Peer reviewed and not-peer reviewed studies, e.g. published in proceedings or doctoral dissertations.

### Variables coding

For each included study, one of the authors, expert in meta-analyses, will code the following variables:

- Authors;
- Year of publication;
- Number of trials;
- Number of hits;
- Number of choices of each trial;
- Task type (Type 1,2 or 3);
- Participants type (selected vs. unselected). The authors of the study will score as selected all participants that were screened for one or more particular characteristic deemed favourable for the performance in this type of task.
- Peer-Review level: level = 0 for studies published in conference proceedings; level = 1, for the studies published in scientific journals with full peer-review

The second author will independently check all studies, and the data will be compared with those extracted by the other author. Discrepancies will be corrected by inspecting the original papers.

The complete database will be made available through open access posting within the dedicated project in the Open Science Framework (<https://osf.io/t7sya/>) platform.

### Effect size measures

As standardized measure of effect size, we will apply that used in Storm *et al.* (2010) and Storm & Tressoldi (2020) : Binomial Z score/ $\sqrt{\text{number of trials}}$  using the number of trials, the hits score and the chance probability as raw scores. The exact binomial Z score will be obtained applying the formula implemented online at <http://vassarstats.net/binomialX.html>. When this algorithm will not compute z when either number of trials or number of hits is low, we will use the one-tailed exact binomial p-value, to find the inverse normal z by using the online app at <https://www.wolframalpha.com/widgets/gallery/view.jsp?id=540d8e149b5e7de92553fdd7b1093f6d>

As standard error we will use the formula:  $\sqrt{(\text{hit rate} * (1-\text{hit rate})/\text{trials} * \text{chance percentage} *(1-\text{chance percentage}))}$ .

In order to take in account the effect size overestimation bias in small samples, the effect sizes and their standard errors, will be transformed in the Hedge's g effect sizes, with the corresponding standard errors by applying the formula presented in Borenstein *et al.* (2009, pp. 27–28:  $g=(1-(3/(4df-1))) * d$ ).

### Overall effect size estimation

In order to take in account the between-studies heterogeneity, the overall effect size estimation of the whole database will be calculated by applying both a frequentist and a Bayesian random effect model for testing its robustness.

### Frequentist random-effect model

Following the recommendations of Langan *et al.* (2019), we will use the restricted maximum likelihood (REML) approach to estimate the heterogeneity variance with the Knapp and Hartung method for adjustment to the standard errors of the estimated coefficients (Rubio-Aparicio *et al.*, 2018).

Furthermore, in order to control for possible influence of outliers, we will calculate the median and mode of the overall effect size applying the method suggested by Hartwig *et al.* (2020).

These calculations will be implemented in the R statistical environment with the `metafor` package v. 2.4 (Viechtbauer, 2017). See syntax provided as extended data (Tressoldi & Storm, 2020).

### Bayesian random-effect model

As priors for the overall effect size we will use a normal distribution with Mean = 0.01; SD =0.03, constrained positive, lower bound = 0 (Haaf & Rouder, 2020), given our expectation of a positive value. As prior for the tau parameter we will use an inverse gamma distribution with shape = 1, scale = 0.15.

This Bayesian meta-analysis will be implemented with the `MetaBMA` package v. 0.6.3 (Heck *et al.*, 2017).

### Publication bias tests

Following the suggestions of Carter *et al.* (2019), we will apply four tests to assess publication bias:

- the 3-parameter selection model (3PSM), as implemented by Coburn & Vevea (2019) with the package 'weightr' v.2.0.2;
- the p-uniform\* (star) v. 0.2.2 test as described by van Aert & van Assen (2019),
- the sensitivity analysis using the Mathur & VanderWeele (2020) package `PublicationBias` v.2.2.0.
- The Robust Bayesian meta-analysis test implemented with the `RoBMA` package v.1.0.5 (Bartoš & Maier, 2020).

The three parameters model represent the average true underlying effect,  $\delta$ ; the heterogeneity of the random effect sizes,  $\tau^2$  and the probability that there is a nonsignificant effect in the pool of effect sizes. The probability parameter is modeled by a step function with a single cut point at  $p = 0.025$  (one-tailed), which corresponds to a two-tailed  $p$  value of 0.05. This cut point divides the range of possible  $p$  values into two bins: significant and nonsignificant. The three parameters are estimated using maximum likelihood (Carter *et al.*, 2019, p. 124).

The  $p$ -uniform\* test, is an extension and improvement of the  $p$ -uniform method. P-uniform\* improves upon  $p$ -uniform giving a more efficient estimator avoiding the overestimation of effect size in case of between-study variance in true effect sizes, thus enabling estimation and testing for the presence of between-study variance in true effect sizes.

Sensitivity analysis as implemented by Mathur & VanderWeele (2020), assumes a publication process such that “statistically significant” results are more likely to be published than negative or “nonsignificant” results by an unknown ratio,  $\eta$  (eta). Using inverse-probability weighting and robust estimation that accommodates non-normal true effects, small meta-analyses, and clustering, it enables statements such as: “For publication bias to shift the observed point estimate to the null, ‘significant’ results would need to be at least 30-fold more likely to be published than negative or ‘nonsignificant’ results” (p. 1). Comparable statements can be made regarding shifting to a chosen non-null value or shifting the confidence interval.

The Robust Bayesian meta-analysis test is an extension of Bayesian meta-analysis obtained by adding selection models to account for publication bias. This allows model-averaging across a larger set of models, ones that assume publication bias and ones that do not. This test allows to quantify evidence for the absence of publication bias estimated with a Bayes Factor. In our case we will compare only two models, a random-effect model assuming no publication bias and a random-model assuming publication bias.

See Syntax Details in the Supporting Information

### Cumulative meta-analysis

In order to study the overall trend of the cumulative evidence we will perform a cumulative effect size estimation. Furthermore, we will estimate the overall effect size taking the variable “year of publication” as covariate using a meta-regression model.

### Moderators effects

We will compare the influence of the following three moderators: (i) Type of participant, (ii) Type of task and (iii) Level of peer-review.

As described in the Variable Coding paragraph, the variable Type of participant will be coded in a binary way: selected vs unselected. Type of task will be coded as Type 1, Type 2, and Type 3, as described in the Introduction and Level of Peer-review as 0 for studies published without a full peer-review or 1, for the studies published after a full peer-review.

### Statistical power

Once the overall effect size and its precision are estimated, we will calculate the number of trials necessary to achieve a statistical power of at least .80 with an  $\alpha = .05$ . With this estimation we can examine how many studies in the database reached this threshold. The overall statistical power will be estimated with the R package *metameta* v.0.1.1. (Quintana, 2020).

### Reporting

The search and selection of the studies will be presented by using a PRISMA flowchart.

### Descriptive statistics

Descriptive statistics will be produced related to the variables, trials, hits, participant type, and peer-review level task types.

### Overall effect size

We will present the estimated average effect size along with the corresponding 95% Confidence Intervals or Credible Intervals of both the frequentist and Bayesian random-model effect as described in the Methods section. We will calculate the values of  $\tau^2$  and  $I^2$  (Higgins & Thompson, 2002), and their confidence intervals, as measures of between-study variance.

### Publication bias tests

We will present the results of the four publication bias tests described in the Methods section.

### Cumulative effect size

The results of the cumulative meta-analysis will be represented with a cumulative forest plot.

### Moderator effects estimation and comparison

We will estimate and compare the average effect size along with the corresponding 95% Confidence Intervals of the two types the participant, the three task types and the two peer-review level, both with a parameter comparison of the overlap of their 95% CIs and with a focused hypothesis testing statistic e.g. ANOVA.

### Dissemination of information

Apart the Registered Report, all information related to this study will be made available open access at Open Science Framework <https://osf.io/t7sya>.

### Study status

The study has not started yet.

### Discussion

We will discuss the robustness of the overall results in order to determine a degree of confidence in the evidence for anomalous perception. In case of an insufficient degree of confidence in the evidence, we will consider whether it is worthwhile pursuing such a line of investigation and offer solutions to improve the evidence.

However, even if the overall results show a sufficient degree of evidence, we will discuss how this line of investigation can instil greater confidence by using a preregistration registry as proposed by Watt & Kennedy (2016) in order to reduce so-called questionable research practices (John *et al.*, 2012), and provide more transparent procedures during data collection and analysis (see for example, the Transparent Psi Project; Kekecs *et al.*, 2019).

### Data availability

#### Underlying data

No data are associated with this article

### Extended data

Figshare: Anomalous perception in a ganzfeld condition: a meta-analysis of more than 40 years of investigation. <https://doi.org/10.6084/m9.figshare.12674618.v2> (Tressoldi & Storm, 2020)

- Syntax Details.docx (Syntax related to all statistical analyses)

### Reporting guidelines

Figshare: PRISMA-P checklist for ‘Stage 1 Registered Report: Anomalous perception in a Ganzfeld condition - A meta-analysis of more than 40 years investigation’ <https://doi.org/10.6084/m9.figshare.12674618.v2> (Tressoldi & Storm, 2020)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).



## References

- Appelbaum M, Cooper H, Kline RB, *et al.*: **Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report.** *Am Psychol.* 2018; **73**(1): 3–25.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bartoš F, Maier M: **RoBMA: An R Package for Robust Bayesian Meta-Analyses.** R package version 1.0.4. 2020.  
[Reference Source](#)
- Bem DJ, Honorton C: **Does psi exist? Replicable evidence for an anomalous process of information transfer.** *Psychol Bull.* 1994; **115**(1): 4–18.  
[Publisher Full Text](#)
- Borenstein M, Hedges LV, Higgins JPT, *et al.*: **Introduction to Meta-Analysis.** Chichester, UK: John Wiley & Sons, Ltd. 2009.  
[Publisher Full Text](#)
- Carter E, Schönbrodt F, Gervais W, *et al.*: **Correcting-bias-in-psychology.** *Adv Methods Pract Psychol Sci.* 2019; **2**(2): 115–144.  
[Publisher Full Text](#)
- Coburn KM, Vevea JL: **Package 'weightr'. Estimating Weight-Function Models for Publication Bias.** 2019.  
[Reference Source](#)
- Haaf JM, Rouder JN: **Does Every Study? Implementing Ordinal Constraint in Meta-Analysis.** *PsyArXiv.* 2020.  
[Publisher Full Text](#)
- Hartwig FP, Smith GD, Schmidt AF, *et al.*: **The median and the mode as robust meta-analysis estimators in the presence of small-study effects and outliers.** *Res Synth Methods.* 2020; **11**(3): 397–412.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Heck DW, Gronau QF, Wagenmakers E: **metaBMA: Bayesian model averaging for random and fixed effects meta-analysis.** 2017.  
[Publisher Full Text](#)
- Higgins JPT, Thompson SG: **Quantifying heterogeneity in a meta-analysis.** *Stat Med.* 2002; **21**(11): 1539–1558.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hyman R: **The ganzfeld psi experiment: A critical appraisal.** *J Parapsychol.* 1985; **49**(1): 3–49.  
[Reference Source](#)
- Hyman R, Honorton C: **Joint communiqué: The psi ganzfeld controversy.** *J Parapsychol.* 1986; **50**(4): 351–364.  
[Reference Source](#)
- Honorton C: **Meta-analysis of psi ganzfeld research: A response to Hyman.** *J Parapsychol.* 1985; **49**(1): 51–91.  
[Reference Source](#)
- John LK, Loewenstein G, Prelec D: **Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling.** *Psychol Sci.* 2012; **23**(5): 524–532.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kekacs Z, Aczel B, Palfi B, *et al.*: **Raising the value of research studies in psychological science by increasing the credibility of research reports: The Transparent Psi Project - Preprint.** 2019.  
[Publisher Full Text](#)
- Langan D, Higgins JPT, Jackson D, *et al.*: **A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses.** *Res Synth Methods.* 2019; **10**(1): 83–98.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mathur MB, VanderWeele TJ: **Sensitivity analysis for publication bias in meta-analyses.** *J R Stat Soc Ser C Appl Stat.* 2020; 1–29.  
[Publisher Full Text](#)
- Milton J, Wiseman R: **Does psi exist? Lack of replication of an anomalous process of information transfer.** *Psychol Bull.* 1999; **125**(4): 387–391.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Moher D, Stewart L, Shekelle P, *et al.*: **Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement.** *Syst Rev.* 2015; **4**(1): 1.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Parker A: **'Ganzfeld'.** Psi Encyclopedia. London: The Society for Psychical Research. 2017.  
[Reference Source](#)
- Quintana D: **dsquintana/metameta: 0.1.1 (beta) (Version 0.1.1).** *Zenodo.* 2020.  
[Publisher Full Text](#)
- Rouder JN, Morey RD, Province JM: **A Bayes factor meta-analysis of recent extrasensory perception experiments: Comment on Storm, Tressoldi, and Di Risio (2010).** *Psychol Bull.* 2013; **139**(1): 241–247.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Rubio-Aparicio M, López-López JA, Sánchez-Meca J, *et al.*: **Estimation of an overall standardized mean difference in random-effects meta-analysis if the distribution of random effects departs from normal.** *Res Synth Methods.* 2018; **9**(3): 489–503.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Storm L, Ertel S: **Does psi exist? Comments on Milton and Wiseman's (1999) meta-analysis of ganzfeld research.** *Psychol Bull.* 2001; **127**(3): 424–433, discussion 434–8.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Storm L, Tressoldi PE, Di Risio L: **Meta-analyses of free-response studies, 1992–2008: Assessing the noise reduction model in parapsychology.** *Psychol Bull.* 2010; **136**(4): 471–485.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Storm L, Tressoldi PE, Utts J: **Testing the Storm *et al.* (2010) meta-analysis using Bayesian and frequentist approaches: Reply to Rouder *et al.* (2013).** *Psychol Bull.* 2013; **139**(1): 248–254.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Storm L, Tressoldi P: **Meta-Analysis of Free-Response Studies 2009–2018: Assessing the Noise-Reduction Model Ten Years On.** *J Soc Psych Res.* 2020; **84**(4): 193–219.  
[Publisher Full Text](#)
- Tressoldi P, Storm L: **Anomalous perception in a Ganzfeld condition: A meta-analysis of more than 40 years investigation.** *figshare.* Online resource. 2020.  
<http://www.doi.org/10.6084/m9.figshare.12674618.v2>
- van Aert RCM, van Assen MALM: **Correcting for publication bias in a Meta-Analysis with the P-Uniform\* method.** 2019.  
[Publisher Full Text](#)
- Viechtbauer W: **The metafor Package.** 2017.  
[Reference Source](#)
- Watt CA, Kennedy JE: **Options for Prospective Meta-Analysis and Introduction of Registration-Based Prospective Meta-Analysis.** *Front Psychol.* 2016; **7**: 2030.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

# Open Peer Review

Current Peer Review Status:   

---

Version 3

Reviewer Report 09 March 2021

<https://doi.org/10.5256/f1000research.54788.r80300>

© 2021 Schmidt S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

 **Stefan Schmidt** 

<sup>1</sup> Department of Psychosomatic Medicine and Psychotherapy, University of Freiburg, Freiburg, Germany

<sup>2</sup> Institute of Frontier Areas of Psychology and Mental Health, Freiburg, Germany

I am not satisfied with the revised version 3 in reply to my review. The authors have reacted to some of the minor issues that were have raised, but the reply regarding the two major issues and some of the other minor issues have not fully addressed my concerns.

## Major Issue 1:

- The issue with the garbage in - garbage out problem was not addressed. The reply of the authors does not deal with the new problem of studies with likely poorer methodology before 1986.

## Major Issue 2:

- The issue of using different effect sizes that belong to different classes of effect sizes is still pending. My plea for a short clarifying paragraph in the introduction was not taken up. Also, the results of prior meta-analyses are still not described by the same variables. I understand that these meta-analyses have used different approaches but I think it will be helpful to the reader if this is made explicit.

Regarding some of the procedures the authors apply, I do not see that they meet the criteria of specification. This refers to applying Google scholar as a research database. This database is not suitable since it is not transparent regarding content and also not regarding updates. The algorithms may even be influenced by cookies, IP addresses, etc. So if two people do the same research in this database we cannot guarantee to have the same results. The same situation is true regarding the webpage [www.wolframalpha.com](http://www.wolframalpha.com). Since you do not know how exactly this is operating or when it will change its mode, you cannot guarantee that you have transparently specified your procedures.

I do not understand the answer of the authors regarding the issue of peer-review of proceedings.



Does this mean proceedings are regarded as peer-reviewed or not?

With respect to moderator comparison, the authors write: "...with a focused hypothesis testing statistic e.g. ANOVA." I would be happy if this could be prespecified in an unambiguous matter.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** clinical and experimental research on mindfulness, meditation, consciousness and parapsychology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

---

## Version 2

Reviewer Report 12 February 2021

<https://doi.org/10.5256/f1000research.30185.r76033>

© 2021 Schmidt S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Stefan Schmidt** 

<sup>1</sup> Department of Psychosomatic Medicine and Psychotherapy, University of Freiburg, Freiburg, Germany

<sup>2</sup> Institute of Frontier Areas of Psychology and Mental Health, Freiburg, Germany

The present report describes a study protocol for conducting a meta-analysis on all Ganzfeld studies published so far. While there are many meta-analyses of Ganzfeld studies so far, this is the first one since 1985 that also includes the early studies from the beginning in 1974 to 1985. On the other end, the authors will include new studies from 2018 to 2020. The objective is to have for the first time the full ganzfeld database available in order to study moderators. This is a very sound aim and the resulting database will be of large value for future research.

Many issues have been already raised by the two other reviewers and the authors have revised and improved the protocol accordingly.

I have two major issues and some minor comments.

### 1. Methodological Quality:

Reviewer J. Utts has already suggested including a rating for methodological quality. The author's reply stated that have already used a quality rating in earlier meta-analyses and have not found any correlation.

Now regarding this study, the crucial difference with respect to earlier meta-analyses is that the authors also include the pre-communiqué studies before 1986. These studies have been already

criticized for methodological quality and that is also the reason why they have not been included in earlier meta-analyses. Therefore, the planned meta-analysis may likely include studies with lower quality into the existing database. If these studies are given the same weight as studies with assumed higher quality, then the estimation of the aggregated effect size might be worse than before. This problem is known in the literature as “garbage in-garbage out”.

Thus, I suggest to rate the study quality on a rating system that also codes for issues mentioned in the joint-communiqué from 1986.

Just as an example: I have performed in 2002, a meta-analysis including all DMILS studies (direct mental interaction in living systems; a different experimental protocol in parapsychology)[ref1]. DMILS started at approx. the same time in the mid-1970s. I made a detailed rating of study quality and found that study quality was inversely related with effect-size. This resulted in excluding four weak studies and also in weighting the remaining studies in the meta-analysis according to their quality.

This means that the authors will also need to pre-specify a procedure on how to deal with a likely correlation of study quality and effect size in the whole database and/or with a significant difference in study quality before 1986. The protocol needs to take care that the overall effect-size is not affected by studies with low quality or questionable procedures.

## 2. Type of effect size:

There is some confusion with the type of effect size that is applied. Authors speak about “standardized effect size” or “mean standardized effect size” (e.g. page 4). Usually, all effect sizes are standardized, so this expression does not make much sense. There is the expression ‘standardized mean difference’ if one compares two means (not the case here) which refers to the fact that the difference is standardized to the standard deviation of the means. In principle, there are many different types of effect sizes depending on the kind of data that they are needed for. The ganzfeld case is not a standard case since here statistics is based on comparison to chance probabilities which is rather rare compared to other fields of science. Some researchers (e.g. Rosenthal) have grouped effect-sizes into families (d-type family, r-type family, etc.). This helps the reader to interpret the effect size (r ranges from -1 to +1, d-types can get larger than -1/+1, etc.). Also, rules of thumb are usually given for the interpretation of the effect sizes of different families. Thus, it is suggested that the authors use consistent terminology throughout their protocol. This refers to their own effect-size computation (here it looks like they apply a d-type effect size since it can be transformed to Hedges’ g, which belongs to the d-type family), as well as to the description of earlier meta-analyses. In addition, I would be happy about a small paragraph in the introduction that explains on what different types of effect sizes have been used in the history of ganzfeld meta-analyses (e.g. Cohen’s h) how they relate to each other and why the effect size issue here is not a trivial one.

## Minor comments:

Page 3:

Three types of Ganzfeld. These three types look like equivalent ones while they are in fact not from a historical perspective. From such a view type 3 would be the standard condition and the other ones special cases (no sender, target selected later). While this is of no importance for computation of the meta-analysis I suggest providing this information in order to make the publication more accessible for readers not familiar with parapsychology.

In displaying other ganzfeld meta-analyses the description is inconsistent, sometimes hit rates and sometimes effect sizes are provided, sometimes p-values and sometimes confidence intervals.

This should be streamlined, so the reader can compare the results. Maybe also a table would be of use?

Page 4:

I do not understand the sentence "...because it reduces the publication bias due to the non-rejection of the statistical null hypothesis often consequent to reduced statistical power." I have a slight idea what you want to express. Please clarify this, e.g. by making two sentences.

Regarding databases for literature research please also include PsychInfo, and more important, Lexscien.

With respect to study inclusion as well as variable coding a good standard is that this is done by two independent researchers. This should be also mentioned in the protocol.

I am not entirely satisfied with the variable peer-review level. E.g. proceedings of the Parapsychological Associations are peer-reviewed. In the period before 2006 or 2008 there were full papers submitted and peer reviewed. This would be a different procedure than in the earlier or later times when only short proceedings were published.

Page 5:

Effect size calculation: the binominal distribution is approximated by the normal distribution. However, for small numbers, the exact binomial probability will be used. Please specify the cut-off for this procedure. Just referring to a web-site for this decision does not guarantee that others could replicate this procedure later on.

Please provide the formula for the transformation into Heges'g instead of giving a reference.

You are applying two aggregation models. Please specify how you would interpret your findings in case they diverge.

Same for the four methods on publication bias estimation. What will be the interpretation if they result in different findings?

"The Robust Bayesian meta-analysis test" on the lower part of the page should not be a headline.

Page 6:

Please specify your methodological approach on how to test for incline or decline effect. Or is the following sentence starting with "Furthermore,..." this description?

Moderator effects (left column): Study quality needs to be assessed as moderator (see above).

Moderator effects (right column): please specify the tests for the three (four) moderator effects.

Extended data:

There is a spelling error in the word 'ganzfeld'

## References

1. Schmidt S, Schneider R, Utts J, Walach H: Distant intentionality and the feeling of being stared at:

two meta-analyses. *Br J Psychol.* 2004; **95** (Pt 2): 235-47 [PubMed Abstract](#) | [Publisher Full Text](#)

**Have the authors pre-specified sufficient outcome-neutral tests for ensuring that the results obtained can test the stated hypotheses, including positive controls and quality checks?**

Yes

**Is the rationale for, and objectives of, the study clearly described?**

Yes

**Is the study design appropriate for the research question?**

Partly

**Are sufficient details of the methods provided to allow replication by others?**

Partly

**Are the datasets clearly presented in a useable and accessible format?**

Not applicable

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** clinical and experimental reseach on mindfulness, meditation, consciousness and parapsychology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 17 Feb 2021

**Patrizio Tressoldi**, Università degli studi di Padova, Padova, Italy

### **1. Methodological Quality:**

.....Now regarding this study, the crucial difference with respect to earlier meta-analyses is that the authors also include the pre-communiqué studies before 1986. These studies have been already criticized for methodological quality and that is also the reason why they have not been included in earlier meta-analyses. Therefore, the planned meta-analysis may likely include studies with lower quality into the existing database.

.....This means that the authors will also need to pre-specify a procedure on how to deal with a likely correlation of study quality and effect size in the whole database and/or with a significant difference in study quality before 1986. The protocol needs to take care that the overall effect-size is not affected by studies with low quality or questionable procedures.

**Reply: As replied to Jessica Utts', comments, in two previous meta-analyses (Storm, Tressoldi & Di Risio, 2010; Storm & Tressoldi, 2020) we didn't find a statistical significant correlation between study quality assessed with an ad hoc system and effect size. We then plan to use the classical peer-review level as a conventional**

**measure of studies quality. Furthermore, with our planned cumulative meta-analysis and regression meta-analysis with the variable Year of publication as covariate, it will be possible to examine the influence of the old studies on the overall effect size.**

## **2. Type of effect size:**

.....There is some confusion with the type of effect size that is applied. Authors speak about "standardized effect size" or "mean standardized effect size" .

.....Thus, it is suggested that the authors use consistent terminology throughout their protocol. This refers to their own effect-size computation (here it looks like they apply a d-type effect size since it can be transformed to Hedges' g, which belongs to the d-type family), as well as to the description of earlier meta-analyses.

.....I would be happy about a small paragraph in the introduction that explains on what different types of effect sizes have been used in the history of ganzfeld meta-analyses (e.g. Cohen's h) how they relate to each other and why the effect size issue here is not a trivial one.

**Reply: The type of effect size that we will use is described in the "Effect size measures" paragraph. In the introduction, we now used the term "average effect size", when related to the overall results of the different meta-analyses. We also have clarified what raw data are used to compute the effect size.**

## **Minor comments:**

Page 3:

Three types of Ganzfeld. These three types look like equivalent ones while they are in fact not from a historical perspective. From such a view type 3 would be the standard condition and the other ones special cases (no sender, target selected later). While this is of no importance for computation of the meta-analysis I suggest providing this information in order to make the publication more accessible for readers not familiar with parapsychology.

**Reply: On pag. 3, in the description of the three different ganzfeld conditions, we added "From an historical perspective, this last type is considered the typical condition.**

In displaying other ganzfeld meta-analyses the description is inconsistent, sometimes hit rates and sometimes effect sizes are provided, sometimes p-values and sometimes confidence intervals. This should be streamlined, so the reader can compare the results. Maybe also a table would be of use?

**Reply: Unfortunately, the average effect size were estimated with different methods and are not comparable.**

Page 4:

I do not understand the sentence "...because it reduces the publication bias due to the non-rejection of the statistical null hypothesis often consequent to reduced statistical power. " I have a slight idea what you want to express. Please clarify this, e.g. by making two sentences.

**Reply: Now we changed the sentence as: "This last condition is particularly relevant**

**because it reduces the publication bias due to non publication (file drawer effect) of studies with statistically non significant results often as a consequence of a reduced statistical power.”**

Regarding databases for literature research please also include PsychInfo, and more important, Lexscien.

**Reply: Google Scholar includes all PsychInfo items. Lexscien is not open access and it does not allow a search with keywords.**

With respect to study inclusion as well as variable coding a good standard is that this is done by two independent researchers. This should be also mentioned in the protocol.

**Reply: now we have specified “The second author will independently check all studies, and the data will be compared with those extracted by the other author”**

I am not entirely satisfied with the variable peer-review level. E.g. proceedings of the Parapsychological Associations are peer-reviewed. In the period before 2006 or 2008 there were full papers submitted and peer reviewed. This would be a different procedure than in the earlier or later times when only short proceedings were published.

**Reply: As if only short proceeding were published, authors had to submit a full paper.**

Page 5:

Effect size calculation: the binominal distribution is approximated by the normal distribution. However, for small numbers, the exact binomial probability will be used. Please specify the cut-off for this procedure. Just referring to a web-site for this decision does not guarantee that others could replicate this procedure later on.

**Reply: Our binomial z score was always an exact binomial probability. Our description of how we obtained these values, makes our procedure replicable.**

Please provide the formula for the transformation into Hedges'g instead of giving a reference.

**Reply: now added.**

You are applying two aggregation models. Please specify how you would interpret your findings in case they diverge.

**Reply: we suppose you are referring to the cumulative effect size and the regression model with Year of publication as covariate. Given that both methods are based on different algorithms, the divergence of their results will be commented according to their difference.**

Same for the four methods on publication bias estimation. What will be the interpretation if they result in different findings?

**Reply: the four planned publication bias tests, are based on different algorithms, hence they different findings will be commented comparing their differences.**

“The Robust Bayesian meta-analysis test” on the lower part of the page should not be a headline.

**Reply: Fixed**



Page 6:

Please specify your methodological approach on how to test for incline or decline effect. Or is the following sentence starting with "Furthermore,..." this description?

**Reply: we have rewritten this paragraph as "In order to study the overall trend of the cumulative evidence we will perform a cumulative effect size estimation.**

**Furthermore, we will estimate the overall effect size taking the variable "year of publication" as covariate using a meta-regression model."**

Moderator effects (right column): please specify the tests for the three (four) moderator effects.

**Reply: the comparisons among the different moderators categories, are described in the paragraph "Moderator effects estimation and comparison"**

Extended data:

There is a spelling error in the word 'ganzfeld'

**Reply: fixed, thank you.**

**Competing Interests:** I'm the corresponding author

---

## Version 1

Reviewer Report 05 October 2020

<https://doi.org/10.5256/f1000research.27439.r70878>

© 2020 Haaf J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### Julia Haaf

Department of Psychology, Psychological Methods, University of Amsterdam, Amsterdam, The Netherlands

In this stage 1 registered report the authors describe a planned meta-analysis to target the Ganzfeld effect as is found in the parapsychological literature. Summarizing all conducted studies on the topic seemingly is a relevant research objective, though I might add that I do wonder about the quality of the conducted studies and their reporting. While the authors plan to conduct publication bias correction, to this point it is virtually impossible to fully account and correct for all the biases baked into the literature, let alone the parapsychological literature. I have a few comments on the statistical analysis, but some of these require some additional work by the authors.

1. **Effect size measure of interest.** The authors plan to report effect size measures based on

the Binomial test. The binomial z-score seems like an appropriate choice given the models they plan to use. I wonder, however, why the authors decided to divide the binomial z-score by the square root of n, the sample size, given that the binomial z is calculated from the binomial mean and standard deviation, both dependent on n. In addition, if the binomial z corresponds to Fisher's z then we know the standard error is  $1/\sqrt{n-3}$ . What is the standard error for  $z/\sqrt{n}$ , and how do we transform it to the standard error of Hedge's g, as planned by the authors? Both frequentist and Bayesian meta-analysis requires the calculation of standard errors to weigh the study effects which is how meta-analysis accounts for sample size/precision. *This point must be addressed to make the article scientifically sound.*

2. **Random model or random-effects model.** The authors plan to use a model to account for between-study heterogeneity. In the meta-analytic literature, these models are called random-effects models (not random models). The wording of fixed-effect model vs. random-effects model from this literature is a bit unfortunate because it does not correspond to what is typically considered fixed vs. random effects in the statistical literature (Gelman, 2004, p. 20)<sup>1</sup>. It might be better to describe the so-called random-effects model simply as a model accounting for between-study heterogeneity.
3. **Bayesian model with ordinal constraints.** The authors reference Rouder *et al.* (2019). I think this reference is perhaps misplaced. It does not really correspond to the sentence. Rouder *et al.* propose instead of interpreting the mean effect size across studies to focus on the distribution of true effect sizes. Therefore, the ordinal constraint is placed on each study's true effect simultaneously. The way the authors describe it they plan to (only) apply an ordinal constraint on the overall effect. If the authors are interested in the question of whether all studies show an effect in the same direction, and I think this would be an interesting question for this application, I might shamelessly refer them to some of my recent work in this area (Haaf & Rouder, preprint)<sup>2</sup>.
4. **Priors.** If the authors want to use a model that accounts for between-study heterogeneity (aka a random-effects model) they need to specify an additional prior distribution on that heterogeneity parameter. I would suggest adding this prior to this stage of the registered report. In the Haaf & Rouder preprint I mentioned above there are suggestions for priors on Fisher's z that might be useful here. *This point must be addressed to make the article scientifically sound.*
5. **Which publication bias correction method is the best?** The authors plan to implement three ways of correcting for publication bias. If the three methods diverge in results, how will they interpret the results? Is there an ordering of method quality, or a way of combining them? Additionally, there have been newer development on publication bias corrections for Bayesian meta-analysis (Maier *et al.*, preprint)<sup>3</sup>. Maybe this is also an option.
6. I really like the idea of a cumulative meta-analysis for this application! In Jasp (JASP Team, 2020) there is also an option to apply a cumulative Bayesian meta-analysis, maybe as a nice addition.

---

## References

1. Gelman A: Analysis of variance—why it is more important than ever. *The annals of statistics*. 2005; **33** (1): 1-53
2. Haaf J, Rouder J: Does Every Study? Implementing Ordinal Constraint in Meta-Analysis. 2020. [Publisher Full Text](#)
3. Maier M, Bartoš F, Wagenmakers E: Robust Bayesian Meta-Analysis: Addressing Publication Bias with Model-Averaging. 2020. [Publisher Full Text](#)

**Have the authors pre-specified sufficient outcome-neutral tests for ensuring that the results obtained can test the stated hypotheses, including positive controls and quality checks?**

Partly

**Is the rationale for, and objectives of, the study clearly described?**

Yes

**Is the study design appropriate for the research question?**

Yes

**Are sufficient details of the methods provided to allow replication by others?**

Partly

**Are the datasets clearly presented in a useable and accessible format?**

Not applicable

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Quantitative and mathematical psychology, expertise in Bayesian statistics, multilevel modeling and meta-analysis.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 10 Oct 2020

**Patrizio Tressoldi**, Università degli studi di Padova, Padova, Italy

Thank you for your comments and suggestions. Here follows our replies;

1. **Effect size measure of interest.** The authors plan to report effect size measures based on the Binomial test. The binomial z-score seems like an appropriate choice given the models they plan to use. I wonder, however, why the authors decided to divide the binomial z-score by the square root of n, the sample size, given that the binomial z is calculated from the binomial mean and standard deviation, both dependent on n. In addition, if the binomial z corresponds to Fisher's z then we know the standard error is  $1/\sqrt{n-3}$ . What is the standard error for  $z/\sqrt{n}$ , and how do we transform it to the standard error of Hedge's g, as planned by the authors? Both frequentist and Bayesian meta-analysis requires the calculation of standard errors to weigh the study effects which is how meta-analysis accounts for sample

size/precision. *This point must be addressed to make the article scientifically sound.*

**Reply: In the paragraph “Effect size measures” we have added how the effect sizes standard errors will be computed and how both will be transformed with the Hedges’s formulas.**

1. **Random model or random-effects model.** The authors plan to use a model to account for between-study heterogeneity. In the meta-analytic literature, these models are called random-effects models (not random models). The wording of fixed-effect model vs. random-effects model from this literature is a bit unfortunate because it does not correspond to what is typically considered fixed vs. random effects in the statistical literature (Gelman, 2004, p. 20)<sup>1</sup>. It might be better to describe the so-called random-effects model simply as a model accounting for between-study heterogeneity.

**Reply: In the “Overall effect size estimation”, we have added this clarification**

1. **Bayesian model with ordinal constraints.** The authors reference Rouder *et al.* (2019). I think this reference is perhaps misplaced. It does not really correspond to the sentence. Rouder *et al.* propose instead of interpreting the mean effect size across studies to focus on the distribution of true effect sizes. Therefore, the ordinal constraint is placed on each study’s true effect simultaneously. The way the authors describe it they plan to (only) apply an ordinal constraint on the overall effect. If the authors are interested in the question of whether all studies show an effect in the same direction, and I think this would be an interesting question for this application, I might shamelessly refer them to some of my recent work in this area (Haaf & Rouder, preprint)<sup>2</sup>.

**Reply: In the “Bayesian random-effect model” paragraph we have corrected this reference.**

1. **Priors.** If the authors want to use a model that accounts for between-study heterogeneity (aka a random-effects model) they need to specify an additional prior distribution on that heterogeneity parameter. I would suggest adding this prior to this stage of the registered report. In the Haaf & Rouder preprint I mentioned above there are suggestions for priors on Fisher’s  $z$  that might be useful here. *This point must be addressed to make the article scientifically sound.*

**Reply: In the “Bayesian random-effect model” paragraph we have added the tau parameter prior distribution (already available in the Syntax details file).**

1. **Which publication bias correction method is the best?** The authors plan to implement three ways of correcting for publication bias. If the three methods diverge in results, how will they interpret the results? Is there an ordering of method quality, or a way of combining them? Additionally, there have been newer development on publication bias corrections for Bayesian meta-analysis (Maier *et al.*, preprint)<sup>3</sup>. Maybe this is also an option.

**Reply: The available literature hasn’t found yet the “best” publication bias for all conditions. We will analyse the robustness of our findings comparing the results of all the publication bias tests. As a fourth test, we will add the RoBMA test as suggested.**

1. I really like the idea of a cumulative meta-analysis for this application! In Jasp (JASP Team, 2020) there is also an option to apply a cumulative Bayesian meta-analysis, maybe as a nice addition.

**Reply: as a further test of the decline effect we will perform a meta-regression analysis using "Year of publication" as covariate (see " Cumulative meta-analysis" paragraph).**

**Competing Interests:** No competing interests were disclosed.

Reviewer Report 09 September 2020

<https://doi.org/10.5256/f1000research.27439.r68427>

© 2020 Utts J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### Jessica Utts

Department of Statistics, University of California, Irvine, Irvine, CA, USA

This article outlines a meta-analysis the authors plan to conduct of all studies meeting certain criteria for the experimental realm in parapsychology called "ganzfeld." The first ganzfeld experiments were conducted in the early 1970s. There have been multiple meta-analyses of ganzfeld studies over the years, but none have covered the entire research period, as the authors plan to do with this one. In addition to estimating an overall effect size, the proposed meta-analysis will examine two additional questions. One is whether the timing and/or participation of a "sender" makes a difference. This question will be examined by classifying the sessions into one of 3 types – target selected after the session, target selected before the session but with no sender, target selected before the session and a sender used. The other question of interest is whether there is a difference in effect size when the participant in the session was selected for characteristics thought to enhance performance.

The authors are to be commended for addressing so many different issues that arise in meta-analysis, and for planning to use both frequentist and Bayesian methods. However, there are some details missing from the report that led to my answer of "partially provided" for the question about sufficient details to allow replication by others. Here are some details that are not provided as completely as needed in the paper if someone were to try to replicate their analyses. (It's possible that they are provided in one of the references on protocols and reporting of meta-analysis, but not in the paper.)

- Will effect sizes be weighted by the size of the study? Obviously they should be. It makes no sense to give equal weight to a study with  $n = 20$  and  $n = 100$ . But it isn't clear in the methodology part of the report how the effect sizes will be combined.
- Will any measure of study quality be used? Some of the earlier studies were criticized for possible methodological problems. Or will studies that don't meet certain quality criteria be

omitted? Or is the plan to omit the studies that didn't use proper randomization methods sufficient?

- Will studies that did not use standard targets (photographs, videos, locations) be excluded? For instance, at least one study used music instead of photographs or videos. Those probably should be excluded, because they represent possibly testing a different ability.
- The reference to using Hedges  $g$  to reduce bias for small studies is not clear. Hedge's  $g$  is usually used for comparing means.
- It is not clear exactly what effect size measure will be used, but if I understand it correctly, it will be  $z/\sqrt{n}$  where  $z$  is found using the normal approximation to the binomial with continuity correction. Although that method gives results very close to using an exact binomial probability for sample sizes of perhaps 20 or more, it may not work well for small sample sizes. In fact the computation website mentioned in the report (<http://vassarstats.net/binomialX.html>) won't even compute  $z$  if either  $np$  or  $nq$  is less than 5. In such cases, an effect size could be found by using the exact binomial  $p$ -value, then finding the inverse normal  $z$  that gives that area in the upper tail. There is an effect size measure specially intended for proportions (Cohen's  $h$ ) but it may not be applicable if a study uses ratings instead of direct hits.
- It isn't clear how the three types of studies will be compared. Will analysis of variance be used? Or, as mentioned, only looking at 95% confidence intervals for each type?

**Have the authors pre-specified sufficient outcome-neutral tests for ensuring that the results obtained can test the stated hypotheses, including positive controls and quality checks?**

Yes

**Is the rationale for, and objectives of, the study clearly described?**

Yes

**Is the study design appropriate for the research question?**

Yes

**Are sufficient details of the methods provided to allow replication by others?**

Partly

**Are the datasets clearly presented in a useable and accessible format?**

Not applicable

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Statistical analysis and methods, with applications to various disciplines including parapsychology; statistics education.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**



Author Response 10 Oct 2020

**Patrizio Tressoldi**, Università degli studi di Padova, Padova, Italy

Thank you for your comments and suggestions. Here follows our replies:

- Will effect sizes be weighted by the size of the study? Obviously they should be. It makes no sense to give equal weight to a study with  $n = 20$  and  $n = 100$ . But it isn't clear in the methodology part of the report how the effect sizes will be combined.

**Reply: the random-effect model explained in the "Frequentist random-effect model" paragraph. weights the studies by using the inverse of their variance plus an estimate of the heterogeneity of the studies  $\tau^2$ ;  $w_i = 1/(\tau^2 + v_i)$**

- Will any measure of study quality be used? Some of the earlier studies were criticized for possible methodological problems. Or will studies that don't meet certain quality criteria be omitted? Or is the plan to omit the studies that didn't use proper randomization methods sufficient?

**Reply: In our 2010 and 2020 meta-analyses we assessed study quality using two judges whose ratings were highly correlated. We did not find a statistically significant correlation between study quality and ES. As to the proper randomization methods, our oldest studies applied the proper randomisation according to the Honorton-Hyman's joint communiqué.**

**As a new way to test the correlation between study quality and ES we added a comparison between the studies published in journals with a full peer-review and the studies published in conference proceedings that usually have a less complete peer-review.**

- Will studies that did not use standard targets (photographs, videos, locations) be excluded? For instance, at least one study used music instead of photographs or videos. Those probably should be excluded, because they represent possibly testing a different ability.

**Reply: There are no theoretical reasons why targets different from images, pictures or video clips, cannot be used. We could assess whether their use will generate ES outliers.**

**We have assessed dynamic vs. static vs. objects/music in our 2020 study (no statistical differences in ES). Objects/music category is a heterogeneous group, but ES for the single musical target study is not statistical different to the ES for objects.**

- The reference to using Hedges  $g$  to reduce bias for small studies is not clear. Hedge's  $g$  is usually used for comparing means.

**Reply: Hedges'  $g$  can be applied to all continuous effect size like Cohen's  $d$  independently from the experimental design (e.g. one and two-groups) see Borenstein et al (2009) pag. 30**

- It is not clear exactly what effect size measure will be used, but if I understand it correctly, it will be  $z/\sqrt{n}$  where  $z$  is found using the normal approximation to the binomial with continuity correction. Although that method gives results very close to using an exact binomial probability for sample sizes of perhaps 20 or more, it may not work well for small sample sizes. In fact the computation website mentioned in

the report (<http://vassarstats.net/binomialX.html>) won't even compute  $z$  if either  $np$  or  $nq$  is less than 5. In such cases, an effect size could be found by using the exact binomial  $p$ -value, then finding the inverse normal  $z$  that gives that area in the upper tail. There is an effect size measure specially intended for proportions (Cohen's  $h$ ) but it may not be applicable if a study uses ratings instead of direct hits.

**Reply: In the effect size measures paragraph, we added where that is the case, we will use wolframalpha calculator available online:**

<https://www.wolframalpha.com/widgets/gallery/view.jsp?id=540d8e149b5e7de92553fdd7b1093f6d>

- It isn't clear how the three types of studies will be compared. Will analysis of variance be used? Or, as mentioned, only looking at 95% confidence intervals for each type?

**Reply: I the "Moderator effects" paragraph we added we will compare the moderators effect by comparing the overlap of their 95% CIs and with a focused hypothesis testing statistic e.g. ANOVA.**

**Competing Interests:** No competing interests were disclosed.

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**