Check for updates

STUDY PROTOCOL

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

Patrizio E. Tressoldi ¹, Lance Storm²

¹Science of Consciousness Research Group, Università degli studi di Padova, Padova, ITALY, 35131, Italy

²School of Psychology, University of Adelaide, Adelaide, Australia, 5005, Australia

V1 **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 model and a Bayesian random model. Moderator analysis will be used to examine the influence of level of experience of participants and the type of task. Publication bias will be estimated by using three different tests. Trend analysis will be conducted on the cumulative database.

Keywords

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

Open Peer Review

Approval Status ✓ ? ?

	1	2	3
version 3			
(revision)			?
26 Feb 2021			view
version 2			
(revision)			?
16 Oct 2020			view
version 1	✓	?	
30 Jul 2020	view	view	

1. **Jessica Utts**, University of California, Irvine, Irvine, USA
2. **Julia Haaf**, University of Amsterdam, Amsterdam, The Netherlands
3. **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)

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: © 2020 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 1; peer review: 1 approved, 1 approved with reservations]**

F1000Research 2020, 9:826 <https://doi.org/10.12688/f1000research.24868.1>

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

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 our choices (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.

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 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 mean 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 a 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 overall 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 overall effect size of the database and determine 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 the non-rejection of the statistical null hypothesis often consequent to reduced statistical power. However, 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 studies even if published in proceedings;

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.

The second author will randomly check 10% of 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 use the following formula: Binomial Z score/ $\sqrt{\text{number of trials}}$. The Binomial Z score will be obtained applying the formula implemented online at <http://vassarstats.net/binomialX.html>

In order to take in account the effect size overestimation bias in small samples, the effect size will be transformed in the Hedge's *g* effect size, obtaining the corresponding variance by applying the formula presented in Borenstein *et al.* (2009, pp. 27–28).

Overall effect size estimation

The overall effect size estimation of the whole database will be calculated by applying both a frequentist and a Bayesian random model for testing its robustness.

Frequentist random 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 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 (Rouder *et al.*, 2019), given our expectation of a positive value. This Bayesian meta-analysis will be implemented with the *MetaBMA* package v. 0.6.3 (Heck *et al.*, 2017). See Syntax in the Supporting Information.

Publication bias tests

Following the suggestions of Carter *et al.* (2019), we will apply three 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), and
- the sensitivity analysis using the Mathur & VanderWeele (2020) package *PublicationBias* v.2.2.0.

The three parameters model represent the average true underlying effect, δ ; the heterogeneity of the random effect sizes, τ^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 & Vander-Weele (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, η (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. See Syntax in the Supporting Information

Cumulative meta-analysis

In order to study the overall trend of the cumulative evidence and in particular for testing the presence of a decline or incline effect, we will perform a cumulative effect size estimation (see Syntax in the Supporting Information)

Moderators effects

We will compare the influence of the following two moderators: (i) Type of participant, and (ii) Type of task.

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. See Syntax in the Supporting Information.

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.

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 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 Random Models as described in the Methods section. We will calculate the values of τ^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 three 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

We will present and compare the average effect size along with the corresponding 95% Confidence Intervals of the two types the participant, and of the three task types.

Dissemination of information

Apart the Registered Report, all information related to this study will be made available open access at Open Science Framework.

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.v1> (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.v1> (Tressoldi & Storm, 2020)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (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](#)
- 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](#)
- 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](#)
- Kekecs 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 [preprint].** 2020.
[Reference Source](#)
- 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](#)
- 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](#)
- Rouder JN, Haal JM, Davis-Stober CP, *et al.*: **Beyond overall effects: A Bayesian approach to finding constraints in meta-analysis.** *Psychol Methods.* 2019; **24**(5): 606–621.
[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. *PsyArxiv.* 2020.

[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.v1>

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 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)¹. 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)².
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)³. 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' 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)¹. 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)².

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)³. 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 τ^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 statistically different to the ES for objects.

- The reference to using Hedges' g to reduce bias for small studies is not clear. Hedges' 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: In 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

F1000Research