

SUBMISSION FORM RESEARCH PROPOSAL**1. General guidelines of use**

- ❖ **This submission form may be used for an individual research proposal and for experimental research lines. Only principal investigators of a project can submit a proposal for evaluation. These principal investigators should be employed by the Tilburg School of Social and Behavioral Sciences (TSB) of Tilburg University.**

NB. When you are planning a series of experiments that are similar in content in similar populations and with similar procedures, it is sometimes to be preferred to submit a research line, as this saves time and effort. This is not obligatory though. The submission of a research line does costs more work, as the committee expects a more extensive description of the series of experiments.

- ❖ Ethical approval of a research project is valid for the indicated duration of the project or until a change occurs in study population, data collection or other procedures. The ethical approval of a research line is valid for a maximum of 5 years (or until a change occurs in study population, data collection or other procedures) after which a new proposal needs to be submitted.
- ❖ Researchers are obliged to inform the Psychology Ethics Committee on the progression of the study and the activity status. When a project has been completed, the researcher needs to inform the Psychology Ethics Committee upon completion. The committee will send reminders to those researchers that have neglected to send updates on their project once a year.
- ❖ Below mentioned researchers and other involved personnel commit themselves to treat all study participants according to the most recent version of the Helsinki declaration (<http://www.wma.net/en/30publications/10policies/b3/>).
- ❖ Researchers and other involved personnel also guarantee that the study participant may discontinue their participation at all times without any consequences. The researchers and other involved personnel commit themselves to maximize the quality of the research, statistical analysis and the reports and to respect specific rules and regulations concerning specific methodologies (e.g., fMRI) .
- ❖ With this electronic signature the undersigned declares to have described the research project truthfully, with special attention to the ethical aspects of the project.

For agreement:

Name: CHJ Hartgerink

ANR (employee number):

5	9	6	6	3	2
---	---	---	---	---	---

Date: 24-08-2015

- PSYCHOLOGY ETHICS COMMITTEE TSB –

GENERAL INFORMATION

Title	Validation of statistical methods to detect (potential) data fabrication
Principal investigator(s)	CHJ Hartgerink
Other investigators (Phd student, research assistant, research nurse, etc.)	MALM van Assen JM Wicherts + potential research assistant
Contact information of PI (address, email, telephone)	Tilburg University P1.774B, Warandelaan 2 PO BOX 90153 LE Tilburg
Grant supplying funding organization & grant number	-
Where does the study take place?	<input type="checkbox"/> Tilburg University <input type="checkbox"/> Hospital, i.e. <input type="checkbox"/> Other medical institution, i.e. <input type="checkbox"/> At the homes of the study participants <input checked="" type="checkbox"/> Elsewhere, i.e. participant's personal computer and participant's preferred location.
Nature of the study	<input checked="" type="checkbox"/> Experimental <input type="checkbox"/> Observational <input type="checkbox"/> Intervention <input type="checkbox"/> Retrospective
This proposal concerns a	<input type="checkbox"/> Research project <input checked="" type="checkbox"/> Research line
Is the research project bound by the Dutch WMO law? ¹	<input type="checkbox"/> yes ⇒ the project will need to be judged by a certified medical ethics committee <input checked="" type="checkbox"/> no, it concerns non-medical research <input type="checkbox"/> no, it concerns medical research that is not bound by the Dutch WMO law (according to a medical ethics committee) <input type="checkbox"/> I am not sure

¹ www.ccmo.nl

- PSYCHOLOGY ETHICS COMMITTEE TSB -

Planned starting date <i>(please consider that the ethics approval will take 4-6 weeks)</i>	01/09/2015 (dd/mm/yyyy) <i>Note: includes ethics application period</i>
Planned end date	31/08/2016 (dd/mm/yyyy)

Medical research: scientific research within the medical setting, with patients as study participants and aiming to answer a research question that is relevant within the health sciences. Study procedures that make a research project bound by WMO law are for example: randomization to treatment arms, drawing of (additional) blood, collection of urine for longer periods of time, questionnaires on trauma, taboo subjects or sexuality). The patient needs to be involved in the study, retrospective studies and data collection from medical records only are bound by other regulations and general privacy laws. This latter research is evaluated by the Psychology Ethics Committee.

NB. Executing a WMO-bound study without the appropriate approval is punishable by law.

SUMMARY

Give a summary of the proposed study/research line (max 1500 words) Make sure you give sufficient information on the data collection procedures (manipulations, stimuli, questionnaires, certainly when they may be ethically sensitive).

Background:

Research misconduct (fabrication, falsification, and plagiarism) undermines the epistemological pursuit of science and negatively affects the validity of published findings, fairness in the scientific reward system, and trust in science.

Consequently, many journals have begun using tools to detect plagiarism and image manipulation in submitted papers (Bosch, Hernández, Pericas, Doti, & Marušić, 2012). Although no systematic data on the results of these screenings are currently available, numerous case studies (e.g., in the Journal of Cell Biology) and their widespread use suggest that these screening tools are effective and useful. In the social and behavioral sciences image detection tools are less useful because data are primarily quantitative and based on observations of behavior, questionnaires, (cognitive) tests, computer tasks, etc.

Moreover, current estimates of the prevalence of research misconduct are based on self-report surveys (Collins & Bornmann, 2013). These estimates range from .3% through 4.9%, with a weighted mean estimate of 1.97% (Fanelli, 2009). These estimates are based on survey questions that are subject to under-reporting, but also ask the wrong question for estimating true prevalence. These surveys ask whether a researcher has, for example, fabricated data *at least* once. This gives an estimate of the prevalence of fabricating researchers, but not the prevalence of research that is affected by fabrication.

Hence, there is a clear need to develop and validate statistical tools to detect potentially problematic data. These tools can operate on the basis of either summary data reported in articles in the social and behavioral sciences (e.g., significance results, means and standard deviations) or raw data (e.g., from computerized tasks or questionnaires) and have proven useful in ad hoc misconduct investigations. We note that in the proposed studies we will *neither* apply these statistical tools to actively seek out new misconduct cases nor to encourage a state of distrust within the social sciences; the development of these tools is for scientific purposes in facilitating further research on data fabrication and providing reviewers of papers with validated tools to detect problematic results and/or data in accordance with the relevant ethical regulations (which include confidentiality and proper

- PSYCHOLOGY ETHICS COMMITTEE TSB -

procedures). In other words, the proposed studies seek to investigate data patterns in fabricated data rather than be related to any researcher who fabricates these data. To study fabricated data, our study participants (researchers) will be asked to fabricate statistics and data anonymously.

Research question(s):

- 1) What is the diagnostic performance of different statistical tools to detect fabricated behavioral data?
- 2) What are the modi operandi of fabricating data?

Study design:

One condition design for both study 1 and study 2

Procedure & materials:

Study 1

We will invite participants for an online experiment via email (see Appendix for proposed invitation format). Potential respondents are psychological researchers who have published in high-impact peer-reviewed psychology journals in 2015. We aim to collect data from at least 36 respondents. Before participating in the study, the respondent is presented with a consent form that clearly describes that the study requires the respondent to openly fabricate data fabrication purely for research purposes. The form also indicates that these openly fabricated data are subsequently subjected to statistical tools to detect data fabrication. The online survey ensures anonymity by removing all information that enables respondent identification. Upon giving informed consent, participants will be prompted to keep notes during the study with respect to how they fabricate data. It will also be explicated that respondents will not (and cannot due to masked responses) be accused of any wrongdoing or anything unethical as a result of this study.

In the proposed study, respondents are asked to generate summary statistics for four studies on the well-known anchoring effect (Jacowitz & Kahneman, 1995). The anchoring effect refers to the tendency of people to use (irrelevant) quantitative information when making decisions or estimations. For example, when asking “The Nile is more than 1,500 miles long. How long do you think it is?” people tend to underestimate the length of the Nile, but tend to overestimate when the anchor is 10,000 miles. Each study includes one of the four following questions, as in Klein et al. (2014): (i) distance from San Francisco to New York, (ii) population of Chicago, (iii) height of Mt. Everest, and (iv) number of babies born per day in the United States. For each question, there were two different conditions: a high and a low anchor compared to the true answer. Respondents are provided with the true value.

Thirty-six independent respondents are asked to each generate an effect size, two group sizes, two means and two standard deviations (sd) for each of the four studies in a within-subjects fashion. That is, $2 \text{ (conditions)} \times 2 \text{ (sex)} \times 4 \text{ (question)} \times 2 \text{ (mean and sd)} + 2 \text{ (sex)} \times 2 \text{ (conditions)} \times 4 \text{ (group sizes)} = 48$ statistics are generated by each respondent (within-subjects design). We instruct respondents to generate results for the hypotheses (i) no effect of sex, (ii) main effect of condition, and (iii) randomly assigned group sizes, under the restriction that total sample size equals 100. We provide an Excel file that computes the effect size and test statistic based on sample size, means, and standard deviations, but respondents are free to generate the statistics in any way they like.

Subsequently, respondents are asked how they generated data. This inquiry is twofold. First, respondents are asked questions with respect to their statistical knowledge via an online self-report rating of their statistical knowledge (1, poor, through 10, excellent) and which statistical packages they frequently use (e.g., SPSS, R, SAS). Second, a semi-structured, online, and anonymous inquiry is conducted. First, the respondent is asked to freely describe how the results were generated. Second, structured questions are posed with respect to the application of data simulation (yes/no), fabricating an entire dataset (yes/no), and using trial-and-error to come to

- PSYCHOLOGY ETHICS COMMITTEE TSB -

the wanted result (yes/no). In order to allow the respondents to add to their description previously given, a final open question is given for further remarks on their data generation process, where they are encouraged to provide any details that come to them.

Upon completion, respondents will be provided with a debrief screen. First, they are asked whether they ever ran an anchoring experiment. Second, the definition of scientific misconduct and an excerpt from the Singapore Statement on Research Integrity are presented. We also stress that the generation of fictitious data within this study was with the purpose of developing methods to detect data fabrication. Therefore, code of conduct reminders have been indicated to decrease dishonest behaviors (Mazar, Amir, & Ariely, 2008) and we implement.

Study 2

Twenty Dutch and Flemish researchers are approached individually with a request to participate in our study (approach is adapted from invitation letter for Study 1; see Appendix). These researchers are randomly selected from peer-reviewed papers that incorporate the Stroop task. They are provided with a formal invitation to schedule a session where an experimenter (applicant or research assistant) visits the researcher. They are also provided with an information leaflet after scheduling the session, which informs them of the general procedure. This includes the informed consent form (similar to Project 1) and explains the goal of the study. This leaflet also explains that the three fabricated datasets that are the hardest to detect will get a bonus reward.

During the session, the general instruction will explicate the hypotheses of the effect for which the dataset is to be fabricated and the timeframe (i.e., 45 minutes). For this project we use the Stroop effect (Stroop, 1935), which is a classic and widely used research paradigm in psychology that focuses on participants' response times. In this paradigm, participants are asked to determine the color a word is presented in (i.e., word colors), but the word also reads a color (i.e., color words). The presented word color (i.e., "red", "blue", or "green") can be either presented in a congruent color (e.g., "red" presented in red) or incongruent color (i.e., "red" presented in green). The dependent variable in the Stroop task is the response latency, where latency is on average higher for incongruent than for congruent words. Respondents are asked to fabricate four statistics (mean and standard deviation of latency of both congruent and incongruent conditions) of 25 individuals up to two decimals, yielding a total of 100 data points per respondent. We provide example means and standard deviations for both congruent and incongruent trials and provide a link to the Stroop task.

Respondents are also requested to keep notes of the data generation process, similar to Study 1. Upon having fabricated data, we conduct a semi-structured inquiry of approximately twenty minutes as to how they fabricated the data (similar as in Project 1; audio-recorded and then transcribed for anonymity). Subsequently, respondents are posed the following questions:

- (i) How did you check whether the fabricated results corresponded with the provided expectation for which results were to be fabricated?
- (ii) What strategy did you apply to fabricate the data? (e.g., copy paste a few responses, first fabricate the means, etc.)
- (iii) How did you fabricate response latencies for the current dataset?
- (iv) Did you use any statistical packages in the fabrication process? If so, please specify which and how you used them.
- (v) Are you familiar with any statistical tools to detect data fabrication in data?
- (vi) Is there anything you would like to note about how you fabricated the results?

After answering these open-ended questions, respondents are debriefed as in study 1.

Calculation and argumentation of sample size (per experiment in case of a research line): (please try to use appropriate software (e.g. G*Power) for the calculation of sample sizes).

The unit of analysis in this research line are the generated data by each participant (i.e., within-subjects). Power to detect fictitious data is thus dependent on how much data one respondent generates, and not on how many respondents are included in the study. The methods that we apply are unknown with respect to their power

- PSYCHOLOGY ETHICS COMMITTEE TSB –

levels. More precisely, examining the statistical power of the tools is the primary focus of the two studies. The number of results are generated by respondents (48 in Study 1, 100 in Study 2) were determined by a considerate weighing of how taxing the task would be to the respondent and the information we think we need.

In Study 1, 36 respondents is aimed for because there are 36 labs of which we have genuine data from the Many Labs project. For Study 2, we target 20 respondents because there are 20 labs with genuine data on the Stroop task from the Many Labs 3 project. In this way, we have an equal amount of genuine and fabricated data sets in both studies. Power could not be calculated because there is no straightforward way to stipulate true effect sizes; note that the statistical tools do not make use of standard statistical tests for which power analyses have been developed.

Proposed statistical analyses:

- Digit analysis (tests whether the rightmost digits follow a uniform distribution)
- Variance analysis (tests whether the variation of the variances across conditions adheres to principles from sampling theory)
- Reverse Fisher method (tests for excess of p-values just below 1)
- ROC-curve analysis (calibrates to optimal decision level alpha for the sample)
- Multivariate correlations (tests whether higher order relations between variables are adhered to as observed in genuine data)

Scientific and societal relevance:

Our two studies improve the understanding of mechanisms used to generate data, and facilitates further research on the prevalence of data fabrication throughout the social sciences. The use of statistical tools examined in the proposed studies also would be of interest to different stakeholders, such as the ORI in the US or the Dutch National Board for Research Integrity (LOWI), journal editors, academic publishers, peer-reviewers, or (potential) whistleblowers. Currently, editors and peer-reviewers do not actively look for scientific misconduct whilst reviewing (Bornmann, Nast, & Daniel, 2008). These statistical tools also help whistleblowers evaluate their suspicions in an easy and standardized way. The use of tools to detect potential scientific misconduct is already implemented by the ORI by providing forensic tools for detecting plagiarism and image manipulation. It is highly important that we know their performance in diagnosing potential data fabrication and that these tools are used sensibly.

SUBJECTS

1. Please check the relevant study population:

- ☐ Students
- ☐ General population without complaints
- ☐ General population with specific “complaints”, i.e. stress, dementia, pregnancy, medically unexplained complaints
- ☐ Patients, i.e.

.....

X Other, i.e. researchers

- PSYCHOLOGY ETHICS COMMITTEE TSB –

2. Do you use patients or persons from the general population with specific “complaints”? Please indicate below why it is necessary to execute your study in this study population.

X N/A

3. Age category of the study population:

☐ <12 yrs

☐ 12-17 yrs

X \geq 18 yrs

4. Are the proposed participants able to give informed consent?

X Yes

☐ No *

* Ability to give informed consent: According to Dutch law, persons younger than 12 are not able to give informed consent, and both parents or caretakers need to sign for participation. In the age category of 12-17 yrs one of the parents as well as the adolescent need to sign the informed consent to be able to participate. In case an adult is unable to give informed consent, the legal guardian needs to sign for participation.

5. Is your study population younger than 12 years of age? Please give a reason for the inclusion of this young research population in your study.

X N/A

6. Organization where the recruitment of participants will take place:

☐ Tilburg University

- PSYCHOLOGY ETHICS COMMITTEE TSB –

X Other, i.e. US research institutes, Flemish and Dutch research institutes

☐ N.V.T. , want

7. Reward for participation (per experiment)

☐ None

☐ Reimbursement of travel expenses

☐ Course credit

X Financial reward, i.e., fixed-pay for time spent, plus bonus depending on quality of generated data (latter only in study 2)

8. Describe in detail the expected burden of the experiment/assessment occasions for the study participants with respect to time, mental and physical burden:

Study 1 is estimated to last approximately 30 minutes and the mental/ethical burden to the participant is that they are requested to fabricate research results. [this will be informally piloted]
Study 2 is estimated to last approximately 90 minutes and the mental/ethical burden to the participant is that they have to fabricate research results. [this will be informally piloted]

9. Describe potential negative consequences of participation for the study participants:

☐ N/A because

.....
.....

Participants could experience dissonance when requested to generate fictitious results due to academic codes of conduct.

10. Describe measures that have been taken to protect the study participant (e.g. insurance, debriefing, etc.):

☐ N/A because

.....
...

Participants will be thoroughly informed before providing consent explaining why we are asking them to generate fictitious research results. Additionally, debriefing will remind them of the academic code of conduct and address any concerns participants might have.

- PSYCHOLOGY ETHICS COMMITTEE TSB –

11. In case of a research line, is it allowed for subjects to participate in more than 1 experiment within the same line?

☐ N/A this is a research proposal for 1 study only

☐ Yes, because

.....
.....

x No, because independent samples are made (US in study 1 and Flemish/Dutch in study 2).

12. Are participants subjected to acts? Indicate which ones, and with what purpose.

Examples: intervention, denials (subjects are asked not to smoke, drink alcohol or eat within a certain time frame preceding the experiment), dietary requests, invasive procedures (venipuncture to draw blood), medical (e.g. exercise test, fMRI or PET scans) or neuropsychological tests, admissions into hospital/institution, intelligence tests.

The act of generating fictitious research results for the purpose of evaluating the diagnostic value of statistical tools to detect fictitious data.

INFORMATION, DATA, ARCHIVING AND PRIVACY

1. When applicable, are there terms/conditions set by the funding organization with respect to information, privacy and reporting?

☐ Yes, i.e.

.....

☐ No

x N/A

2. Method of recruitment (multiple options may be checked):

☐ Advertisement

☐ Conversation with medical doctor/psychiatrist/psychologist/social worker

☐ Voluntary application, i.e.

x Other, i.e. purposive sampling of researchers.

- PSYCHOLOGY ETHICS COMMITTEE TSB –

3. How much time is given to eligible participants/parents/caretakers/guardians to decide about participation after the participant has received the participant information letter?

1 week

4a. Does this study deviate from standard rules and regulation concerning information giving and privacy?

Standard rules regarding information: participants or their legal representative are

1. informed study in writing and in advance completely about the nature of the study
2. asked to give written informed consent by means of a consent form
3. debriefed afterwards (in writing and orally) about the goals of the study and reason for potential misleading elements during the experiment

Standard rules regarding research data:

4. data are processed in a coded fashion (and anonymous if possible) and stored confidentially
5. a participant may always look into their own data (except when a study is completely anonymous, then there is no link between personal information and study data)
6. all data must be available for inspection for all investigators involved in the project

☒ No

☐ yes, this study deviates with respect to rule(s):

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 (multiple options possible)

4b. If the study deviates from the above stated rules on one or more points, please describe how the study deviates from the standard rules per deviation:

4c. The study deviates from the above stated rules on one or more points, please describe the reasons for deviation from the standard rules

- PSYCHOLOGY ETHICS COMMITTEE TSB –

ADDITIONAL INFORMATION

Please use this space to add information that is important to your project but was not asked about in the form.

The proposed statistical analyses are new or relatively unknown.

EMAILING THE PROPOSAL TO THE ETHICS COMMITTEE

The committee would like to receive the following documents in addition to this research proposal (if applicable):

☐ Advertisement

☒ Participant information letter (precedes participation)

☐ Informed consent form

☐ Written debriefing

☐ Written consent of external (outside university) institution to recruit participants

- PSYCHOLOGY ETHICS COMMITTEE TSB –

Appendix

Dear fellow scientist,

We are conducting a study on methods to distinguish genuine psychological data from faked data. To this end, we are looking for published authors who are willing to generate statistics and data that we can subsequently use to study these methods. Because you published a peer-reviewed article in Psychology in 2015, we would like to kindly invite you to help us with our study, which is part of Chris Hartgerink's doctoral research. Your participation in the study will last approximately XX [this needs to be piloted] minutes and is entirely anonymous. This study has been approved by the Psychological Ethics Committee of TSB of Tilburg University [pending approval].

The goal of this study is to validate statistical methods to discern between true research results and fictitious research results, in order to improve our knowledge of how to detect such fictitious research results in the scientific literature.

All responses are fully anonymized, which includes no logging of IP-addresses.

If you are willing to participate, you will receive an Amazon gift card (\$TBD) when the study has ended (INSERT DATE OF COMPLETION). Your e-mail address will be separated from the dataset, for privacy reasons.

If you have any questions, please feel free to reply to this e-mail.

Kind regards from Tilburg University,
Chris Hartgerink

Marcel van Assen

Jelte Wicherts