

Running Head: Data from a pre-publication independent replication initiative examining ten moral judgement effects

Title

Data from a pre-publication independent replication initiative examining ten moral judgement effects

Authors

Warren Tierney¹, Martin Schweinsberg¹, Jennifer Jordan², Deanna M. Kennedy³, Israr Qureshi⁴, S. Amy Sommer⁵, Nico Thornley¹, Nikhil Madan¹, Michelangelo Vianello⁶, Eli Awtrey³, Luke Lei Zhu⁷, Daniel Diermeier⁸, Justin E. Heinze⁹, Malavika Srinivasan¹⁰, David Tannenbaum⁸, Eliza Bivolaru¹, Jason Dana¹¹, Clinton P. Davis-Stober¹², Christilene du Plessis¹³, Quentin F. Gronau¹⁴, Andrew C. Hafenbrack¹⁵, Eko Yi Liao¹⁶, Alexander Ly¹⁴, Maarten Marsman¹⁴, Toshio Murase¹⁷, Michael Schaerer¹, Christina M. Tworek¹⁸, Eric-Jan Wagenmakers¹⁴, Lynn Wong¹, Tabitha Anderson¹⁹, Christopher W. Bauman²⁰, Wendy L. Bedwell²¹, Victoria Brescoll¹¹, Andrew Canavan¹⁹, Jesse J. Chandler²², Erik Cheries²³, Sapna Cheryan³, Felix Cheung^{24,25}, Andrei Cimpian^{18,35}, Mark A. Clark²⁶, Diana Cordon¹⁹, Fiery Cushman¹⁰, Peter H. Ditto²⁰, Alice Amell¹⁹, Sarah E. Frick²¹, Monica Gamez-Djokic²⁷, Rebecca Hofstein Grady²⁰, Jesse Graham²⁸, Jun Gu³⁰, Adam Hahn³¹, Brittany E. Hanson³², Nicole J. Hartwich³¹, Kristie Hein¹⁹, Yoel Inbar³³, Lily Jiang³, Tehlyr Kellogg¹⁹, Nicole Legate¹⁹, Timo P. Luoma³¹, Heidi Maibeucher¹⁹, Peter Meindl²⁹, Jennifer Miles²⁰, Alexandra Mislin²⁶, Daniel C. Molden²⁷, Matt Motyl³², George Newman¹¹, Hoai Huong Ngo³⁴, Harvey Packham²⁵, P. Scott Ramsay²¹, Jennifer L. Ray³⁵, Aaron M. Sackett³⁶, Anne-Laure Sellier⁵, Tatiana Sokolova^{5,9}, Walter Sowden⁹, Daniel Storage¹⁸, Xiaomin Sun³⁷, Jay J. Van Bavel³⁵, Anthony N. Washburn³², Cong Wei³⁷, Erik Wetter³⁸, Carlos T. Wilson¹⁹, Sophie-Charlotte Darroux¹, Eric Luis Uhlmann¹

[in press at *Nature: Scientific Data*]

Affiliations

1 INSEAD, France and Singapore

2 IMD, Lausanne, Switzerland

3 University of Washington Bothell, United States

4 IE Business School, IE University, Spain

5 HEC Paris, France

6 University of Padova, Italy

3 University of Washington Bothell, United States

7 University of Manitoba, Canada

8 University of Chicago, United States

9 University of Michigan, United States

10 Harvard University, United States

11 Yale University, United States

12 University of Missouri, United States

13 Rotterdam School of Management, Erasmus University, The Netherlands

- 14 University of Amsterdam, The Netherlands
- 15 UCP - Católica Lisbon School of Business & Economics, Portugal
- 16 Hang Seng Management College, Hong Kong
- 17 Roosevelt University, United States
- 18 University of Illinois at Urbana-Champaign, United States
- 19 Illinois Institute of Technology, United States
- 20 University of California, Irvine, United States
- 21 University of South Florida, United States
- 22 Mathematica Policy Research, Institute for Social Research, University of Michigan,
United States
- 23 University of Massachusetts Amherst, United States
- 24 Michigan State University, United States
- 25 University of Hong Kong, Hong Kong
- 26 American University, United States
- 27 Northwestern University, United States
- 28 University of Southern California, United States
- 29 University of Pennsylvania, United States
- 30 Monash University, Australia
- 31 Social Cognition Center Cologne, University of Cologne, Germany
- 32 University of Illinois at Chicago, United States
- 33 University of Toronto, Canada
- 34 Université Paris Ouest Nanterre la Défense, France
- 35 New York University, United States
- 36 University of St. Thomas, United States
- 37 Beijing Normal University, China
- 38 Stockholm School of Economics, Sweden

Corresponding authors:

Correspondence concerning this paper should be addressed to Warren Tierney or Martin Schweinsberg, INSEAD Organizational Behavior Area, Boulevard de Constance, 77305 Fontainebleau, France, warren.tierney@insead.edu, warren.tierney@hotmail.com, martin.schweinsberg@insead.edu, martin.schweinsberg@gmail.com, or to Eric Luis Uhlmann, INSEAD Organizational Behavior Area, 1 Ayer Rajah Avenue, 138676 Singapore, eric.uhlmann@insead.edu, eric.luis.uhlmann@gmail.com.

Abstract

We present the data from a crowdsourced project seeking to replicate findings in independent laboratories before (rather than after) they are published. In this Pre-Publication Independent Replication (PPIR) initiative, 25 research groups attempted to replicate 10 moral judgment effects from a single laboratory's research pipeline of unpublished findings. The 10 effects were investigated using online/lab surveys containing psychological manipulations (vignettes) followed by questionnaires. Results revealed a mix of reliable, unreliable, and culturally moderated findings. Unlike any previous replication project, this dataset includes the data from not only the replications but also from the original studies, creating a unique corpus that researchers can use to better understand reproducibility and irreproducibility in science.

Background & Summary

The replicability of findings from scientific research has garnered enormous popular and academic attention in recent years¹⁻³. Results of replication initiatives attempting to reproduce previously published findings reveal that the majority of independent studies do not produce the same significant effects as the original investigation¹⁻⁵.

There are many reasons why a scientific study may fail to replicate besides the original finding representing a false positive due to publication bias, questionable research practices, or error. These include meaningful population differences between the original and replication samples (e.g., cultural, subcultural, and demographic variability), overly optimistic estimates of study power based on initially published results, study materials that were carefully pre-tested in the original population but are not as well suited to the replication sample, a lack of replicator expertise, and errors in how the replication was carried out. Nonetheless, the low reproducibility rate has contributed to a crisis of confidence in science, in which the truth value of even many well-established findings has suddenly been called into question⁶.

The present line of research introduces a collaborative approach to increasing the robustness and reliability of scientific research, in which findings are replicated in independent laboratories *before*, rather than after, they are published^{7,8}. In the Pre-Publication Independent Replication (PPIR) approach, original authors volunteer their own findings and select expert replication labs with subject populations they expect to show the effect. PPIR arguably increases the informational value of replications, since common alternative explanations for failures to replicate such as a lack of replicator expertise and theoretically anticipated population differences are addressed. Sample sizes are also much larger than is common in the field, and the analysis plan is pre-registered⁹, allowing for more accurate effect size estimates and identification of unexpected population differences. An effect has been overestimated and is quite possibly a false positive if it consistently fails to replicate in PPIRs. Pre-publication independent replication also has the benefit of ensuring published findings are reliable before they are widely disseminated, rather than only checking after-the-fact.

In this first crowdsourced Pre-Publication Independent Replication (PPIR) initiative, 25 laboratories attempted to replicate 10 unpublished moral judgment findings in the research “pipeline” of the last author and his collaborators (see Table 1). The original authors selected replication laboratories with directly relevant expertise (e.g., moral judgment researchers), and access to subject populations theoretically expected to show the effect. A pre-set list of replication criteria were applied¹⁰: whether the original and replication effect were in the same direction, whether the replication effect was statistically significant, whether the effect size was significant meta-analyzing the original and replication studies, whether the original effect size was within the confidence interval of the replication effect size, and finally the small telescopes criterion (a replication effect size large enough to be reliably captured by the original study¹¹). Of the 10 original findings, 6 replicated according to all criteria, two studies failed to replicate entirely, one study replicated but with a smaller effect size than the original study, and one study replicated in United States samples but not outside the United States (see ref. 7 for a full empirical report).

Unique among the replication initiatives thus far, the pipeline project corpus includes the data from not only the replications but also all of the original studies targeted for replication. This creates a unique opportunity for future analysts to better understand reproducibility and irreproducibility in science, since the data from the original studies can be reanalyzed to better understand why a particular effect did or did not prove reliable. The dataset is complemented by both socioeconomic and demographic information on the research participants, and contains data from 6 countries (the United States, Canada, the Netherlands, France, Germany, and China) and replications in 4 languages (English, French, German, and Chinese). The Pre-Publication Independent Replication Project dataset is publicly available on the Open Science Framework (Data Citation 1) and is accompanied by SPSS syntax which can be used to reproduce the analyses. This array of data will serve as a resource for researchers interested in research reproducibility, statistics, population differences, culturally-moderated phenomena, meta-science, moral judgments, and the complexities of replicating studies. For example, the data can be re-analyzed using meta-regression techniques in order to better understand if certain study characteristics or demographics moderate effect sizes. A re-analyst could also try out different analytic techniques and see how robust certain effects are to different specifications.

Methods

Insert Table 1 about here

Participants

The Pre-Publication Independent Replication Project corpus includes three datasets. The first dataset (PPIR 1.sav: Data Citation 1) contains data from 3 original studies and their replications, a second dataset (PPIR 2.sav: Data Citation 1) contains data from 3 original studies and their replications, and a third dataset (PPIR 3.sav: Data Citation 1) contains data from 4 original studies and their replications. In total data were collected from 11,805 participants. The first SPSS file (PPIR 1.sav: Data Citation 1) contains data from 3,944 participants (including 514 from the original studies), while the second SPSS file (PPIR 2.sav: Data Citation 1) contains data from 3,919 participants (including 351 from original studies) and the final SPSS file (PPIR 3.sav: Data Citation 1) contains data from 3,829 participants (including 582 from original studies). An additional replication dataset collected in France contained 113 participants. No participants were removed from either the original or replication studies. All participants agreed to the informed consent form and the studies were in accordance with ethics regulations of the respective universities.

Testing procedure

The data were collected using both online and paper-pencil surveys from the respective laboratories and participants. The replications used the same materials and measurements as in the original studies (with the exception that the materials were translated into multiple languages). In the online version of the replications, Qualtrics was used to collect the data. This online platform allowed us to randomize the order by which the studies were presented. In order to prevent participant fatigue, studies were administered in one of three batches, each batch contained three to four studies, and study order was counterbalanced between subjects. Once subjects agreed to participate in the study, they read vignettes (see below for an example of the vignette from the Cold-Hearted Prosociality study) and completed survey questions assessing their reactions. Thereafter, the participants were thanked for their participation and debriefed.

Karen works as an assistant in a medical center that does cancer research. The laboratory develops drugs that improve survival rates for people stricken with breast cancer. As part of Karen's job, she places mice in a special cage, and then exposes them to radiation in order to

give them tumors. Once the mice develop tumors, it is Karen's job to give them injections of experimental cancer drugs.

Lisa works as an assistant at a store for expensive pets. The store sells pet gerbils to wealthy individuals and families. As part of Lisa's job, she places gerbils in a special bathtub, and then exposes them to a grooming shampoo in order to make sure they look nice for the customers. Once the gerbils are groomed, it is Lisa's job to tie a bow on them.

Although the majority of the data were collected as described above, there were some exceptions. Specifically, as opposed to counterbalancing the order in which the study was presented, participants at Northwestern University were randomly allocated to a survey which either contained one longer study or three shorter studies that were presented in a fixed order. Participants at Yale University did not complete one study as the researchers felt that the participants may be offended by it. Also, there was a translation error in one study run at the INSEAD Paris laboratory which required that study to be re-run separately. Finally, study order for participants at HEC Paris was not counterbalanced but rather fixed. Table 2 includes an outline of the number of replications and conditions, a brief synopsis of study and instructions for creating variables. Detailed reports of each original study and the complete replication material are available on the OSF in the Supplementary File 1 (00.Supplemental_Materials_Pipeline_Project_Final_10_24_2015.pdf: Data Citation 1). Ideally, any person seeking to replicate the findings would use this document in combination with the Supplementary File 2, which outlines all the names and measurement details used in the study (PPIR_Codebook.xlsx: Data Citation 1).

Insert Table 2 about here

Data Records

All data records listed in this section are available from the Open Science Framework (Data Citation 1) and can be downloaded without an OSF account. The datasets were anonymized to remove any information that could identify the participant responses, such as identification numbers from Amazon's Mechanical Turk. The analysis was conducted with SPSS version 20 and detailed SPSS syntax (including comments) are provided to help with data analysis. In total there are 3 datasets and 11 syntax files available. These datasets are also accompanied

by a codebook which describes the variables, the coding transformations necessary to replicate the analyses, and a synopsis of the respective studies.

First dataset

Location: (PPIR 1.sav: Data Citation 1)

File format: SPSS Statistic Data Document file (.sav)

This file contains basic demographic information and responses to the items measured in the Moral Inversion study (SPSS Syntax files/PPIR 1 - 2 moral inversion.sps: Data Citation 1), Intuitive Economics study (SPSS Syntax files/PPIR 1 - 4 intuitive economics.sps: Data Citation 1), and Burn in Hell study (SPSS Syntax files/PPIR 1 - 7 burn in hell.sps: Data Citation 1).

Second dataset

Location: (PPIR 2.sav: Data Citation 1)

File format: SPSS Statistic Data Document file (.sav)

This file contains basic demographic information and responses to the items measured in the Presumption of Guilt study (SPSS Syntax files/PPIR 2 - 1 presumption of guilt.sps: Data Citation 1), The Moral Cliff study (SPSS Syntax files/PPIR 2 - 3 moral cliff.sps: Data Citation 1), and Bad Tipper study (SPSS Syntax files/PPIR 2 - 9 bad tipper.sps: Data Citation 1).

Third dataset

Location: (PPIR 3.sav: Data Citation 1)

File format: SPSS Statistic Data Document file (.sav)

This file contains basic demographic information and responses to the items measured in the Higher Standard Effect studies (SPSS Syntax files/PPIR 3 - 5 higher standard - Charity.sps: Data Citation 1; SPSS Syntax files/PPIR 3 - 5 higher standard - Company.sps: Data Citation 1), Cold Hearted Prosociality (SPSS Syntax files/PPIR 3 - 6 cold-hearted.sps: Data Citation 1), Bigot-Misanthrope study (SPSS Syntax files/PPIR 3 - 8 bigot misanthrope.sps: Data Citation 1), and Belief-Act Inconsistency study (SPSS Syntax files/PPIR 3 - 10 belief act consistency.sps: Data Citation 1).

Codebook

Location: (Data descriptor – Codebook/PPIR _Codebook.xlsx: Data Citation 1)

File format: Microsoft Excel Worksheet (.xlsx)

Introduction to PPIR project, outline of transformations, description and labels for variables from three datasets.

Technical Validation

The studies include an array of original measurements which must be calculated to test the concepts of interest. These measures range from a single item to aggregated measures with multiple items, some of which must be reverse coded. (See Figure 1, for an example of the items measuring candidate evaluations in the Higher Standards study; note that item 5 must be reverse coded prior to averaging the items into a composite). Instructions for how to create the study variables, the relevant conditions, and a synopsis of what concepts the variables measure, can be found in Table 2.

Insert Figure 1 about here

Figure 1 outlines a typical questionnaire that was administered to the subjects to assess their attitudes and beliefs toward the characters depicted in the vignettes. The subjects were required to write next to the statement the number that best indicated how much they believed the statement was representative of Lisa's or Karen's characteristics.

Acknowledgements

The authors gratefully acknowledge financial support from an R&D grant from INSEAD.

Author Contributions

Prepared the dataset for publication: Warren Tierney, Martin Schweinsberg

Wrote the data descriptor: Warren Tierney, Martin Schweinsberg, Eric Luis Uhlmann

Analysis co-pilots for data publication: Warren Tierney, Jennifer Jordan, Deanna M. Kennedy, Israr Qureshi, Martin Schweinsberg, Amy Sommer, Nico Thornley

Designed the pre-publication independent replication project and wrote the original project proposal: Eric Luis Uhlmann

Coordinators for the pre-publication independent replication project: Martin Schweinsberg, Nikhil Madan, Michelangelo Vianello, Amy Sommer, Jennifer Jordan, Warren Tierney, Eli Awtrey, Luke (Lei) Zhu, Eric Luis Uhlmann

Contributed original studies for replication: Daniel Diermeier, Justin E. Heinze, Malavika Srinivasan, David Tannenbaum, Eric Luis Uhlmann, Luke Zhu

Translated study materials: Adam Hahn, Nicole Hartwich, Timo Luoma, Hoai Huong Ngo, Sophie-Charlotte Darroux

Analyzed data from the replications: Michelangelo Vianello, Jennifer Jordan, Amy Sommer, Eli Awtrey, Eliza Bivolaru, Jason Dana, Clinton P. Davis-Stober, Christilene du Plessis, Quentin F. Gronau, Andrew C. Hafenbrack, Eko Yi Liao, Alexander Ly, Maarten Marsman, Toshio Murase, Israr Qureshi, Michael Schaerer, Warren Tierney, Nico Thornley, Christina M. Tworek, Eric-Jan Wagenmakers, Lynn Wong

Carried out the replications: Eli Awtrey, Jennifer Jordan, Amy Sommer, Tabitha Anderson, Christopher W. Bauman, Wendy L. Bedwell, Victoria Brescoll, Andrew Canavan, Jesse J. Chandler, Erik Cheries, Sapna Cheryan, Felix Cheung, Andrei Cimpian, Mark A. Clark, Diana Cordon, Fiery Cushman, Peter Ditto, Thomas Donahue, Sarah E. Frick, Monica Gamez-Djokic, Rebecca Hofstein Grady, Jesse Graham, Jun Gu, Adam Hahn, Brittany E. Hanson, Nicole J. Hartwich, Kristie Hein, Yoel Inbar, Lily Jiang, Tehlyr Kellogg, Deanna M. Kennedy, Nicole Legate, Timo P. Luoma, Heidi Maibuecher, Peter Meindl, Jennifer Miles, Alexandra Mislin, Daniel Molden, Matt Motyl, George Newman, Hoai Huong Ngo, Harvey Packham, Philip S. Ramsay, Jennifer Lauren Ray, Aaron M. Sackett, Anne-Laure Sellier, Tatiana Sokolova, Walter Sowden, Daniel Storage, Xiaomin Sun, Christina M. Tworek, Jay J. Van Bavel, Anthony N. Washburn, Cong Wei, Erik Wetter, Carlos T. Wilson

Competing Interests

The authors declare no competing financial interests. Warren Tierney had full access to all of the data and takes responsibility for the integrity and accuracy of the analysis.

References

1. Begley, C.G., & Ellis, L.M. Drug development: Raise standards for preclinical cancer research. *Nat.* **483**, 531–533 (2012).
2. Open Science Collaboration. Estimating the reproducibility of psychological science. *Sci.* **349**, aac4716. (2015).
3. Ebersole, C. R. *et al.* Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *J. Exp. Soc. Psychol.*, (2016).
4. Prinz, F., Schlange, T., & Asadullah, K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug. Discov.* **10**, 712–713 (2011).
5. Pashler, H., & Wagenmakers, E. J. Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? *Perspect. Psychol. Sci.* **7**, 528-530 (2012).
6. Klein, R. A. *et al.* Investigating variation in replicability: A "many labs" replication project. *Soc. Psychol.* **45**, 142–152 (2014).
7. Schweinsberg, M. *et al.* The pipeline project: Pre-publication independent replications of a single laboratory's research pipeline. *J. Exp. Soc. Psychol.*, (2016).
8. Schooler, J. W. Metascience could rescue the 'replication crisis'. *Nat.*, **515**, 9 (2014).
9. Wagenmakers, E.-J. *et al.* An agenda for purely confirmatory research. *Perspect. Psychol. Sci.*, **7**, 627-633 (2012).
10. Brandt, M. J. *et al.* The replication recipe: What makes for a convincing replication? *J. Exp. Soc. Psychol.*, **50**, 217-224 (2014).
11. Simonsohn, U. Small telescopes: Detectability and the evaluation of replication results. *Psychol. Sci.*, **26**, 559-569 (2015).

Data Citations

Tierney, W., Schweinsberg, M. & Uhlmann, E. L. *Open Science Framework*
<http://dx.doi.org/10.17605/OSF.IO/G7CU2> (2016).