

README — Replication Package

Does Unfairness Hurt Women? The Effects of Losing Unfair Competitions

Stefano Piasenti, Marica Valente, Roel van Veldhuizen, Gregor Pfeifer

Economic Journal | April 2026

1. Overview

This package contains all data and code necessary to replicate the empirical results reported in the paper and its appendices. The package includes anonymised experimental data from three studies (one main experiment and two survey experiments) and four R analysis scripts.

Data sources used:

- Main experiment: online experiment on Prolific (Spring 2021, N = 2,086)
- Survey Experiment 1: Prolific / Qualtrics survey (May 2024, N = 670)
- Survey Experiment 2: Prolific / Qualtrics survey (October 2025, N = 170)

Software: R (version ≥ 4.1 for scripts 01-03; R 4.2.1 required for script 04). All required packages are installed automatically by 00_install_packages_windows.R or 00_install_packages_unix.R.

Approximate total runtime: under 5 minutes on a standard desktop PC (script 04 requires approximately 3 minutes on an 8-core machine; all other scripts complete in under 1 minute each).

2. Data Availability and Provenance Statements

2.1 Statement about Rights

The authors of the manuscript have legitimate access to and permission to use all data used in this manuscript. All data were collected by the authors via the Prolific platform and via Qualtrics under the standard terms of those platforms. The authors have documented permission to redistribute the anonymised data contained within this replication package. No third-party data requiring special redistribution agreements are used. The data are made available under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence. Appropriate permissions are documented in the LICENSE.txt file.

2.2 Summary of Availability

- All data used in this paper are publicly available in this replication package.

Note: Raw (un anonymised) Prolific participant files are not redistributed. The anonymised versions included in the package contain all variables needed to reproduce every result in the paper.

2.3 Details on Each Data Source

2.3.1 Main Experiment Data (Prolific, Spring 2021)

Online experiment conducted on Prolific between May and June 2021. Participants completed a real-effort task (letter counting), a tournament in Stage 2 with meritocratic, unfair, or feedback-only treatment assignment, and a tournament entry decision in Stage 3. Data collected in 11 separate sessions (men-only and women-only cohorts, plus one mixed session).

N = 2,086 participants (1,089 female, 997 male). Median completion time: 13 min 35 sec. Participation fee: GBP 1.50; median total payment: GBP 2.10.

Pre-registration (AEA Social Science Registry, trial 7651):

<https://www.socialscienceregistry.org/trials/7651>

Included in package: Yes -- anonymised. Raw session files are in replication_data/raw_data_anonymous/ (11 session subfolders). The merged analysis dataset dat_numeric.csv (2,086 rows x 375 columns) is provided pre-built.

Anonymisation applied: PROLIFIC_PID transformed by a deterministic rule; IPAddress, ResponseId, and all name/contact fields set to NA. All experimental outcome variables, performance scores, treatment assignments, and Prolific demographic variables are untouched. The anonymisation script (anonymize.R) is included; the reverse-mapping key is not distributed.

2.3.2 Survey Experiment 1 Data (Prolific / Qualtrics, May 2024)

Prolific survey experiment programmed in Qualtrics. Participants rated two tournament procedures (Meritocratic and Unfair) on three dimensions: fairness, merit, and random chance. N = 670 participants. Median completion time: 3 min 40 sec. Participation fee: GBP 1.50.

Pre-registration (OSF): <https://osf.io/4f97v>

Included in package: Yes -- anonymised. Files in replication_data/survey_data_anonymous/: survey1_data_anonymous.xlsx and demographics_survey1_anonymous.csv.

Anonymisation: StartDate, EndDate, IPAddress, ResponseId, all name/contact fields, free-text responses, and Prolific submission metadata removed. All rating and demographic variables are untouched. Participant IDs transformed using the same deterministic rule as the main experiment.

2.3.3 Survey Experiment 2 Data (Prolific, October 2025)

Prolific survey experiment. Participants rated two procedures (Meritocratic and Risk-only) on the same three dimensions. N = 170 participants. Median completion time: 4 min 28 sec. Participation fee: GBP 1.50.

Pre-registration (OSF): <https://osf.io/hbv7q>

Included in package: Yes -- anonymised. File: replication_data/survey_data_anonymous/survey2_data_anonymous.xlsx. Same anonymisation procedure as Survey Experiment 1.

3. Experimental Documentation

This paper uses data generated via online experiments. The following documentation is provided in accordance with the Economic Journal replication requirements.

- Experiment design: described in Section 2 and Appendix A of the paper.
- Participant instructions: a copy of the instructions shown to participants during the experiment, in the original language (English), is attached. See also Appendix H of the paper.
- Participant selection and eligibility: participants were recruited on Prolific. Eligibility criteria: fluent English speakers, UK residents, no prior participation in related studies. See Section 2 of the paper for full details.
- IRB / Ethics approval: the institutional ethics approval obtained for this study is attached.

4. Dataset List

File (relative path)	Source	Notes	Provided
replication_data/dat_numeric.csv	Authors (merged from raw sessions)	Main analysis dataset; 2,086 rows x 375 cols	Yes
replication_data/raw_data_anonymous/ (11 session subfolders)	Authors / Prolific (Spring 2021)	Anonymised session XLSX + Prolific CSV/TXT files	Yes
replication_data/survey_data_anonymous/survey1_data_anonymous.xlsx	Authors / Prolific + Qualtrics (May 2024)	Survey 1; N = 670; Meritocratic vs Unfair	Yes
replication_data/survey_data_anonymous/demographics_survey1_anonymous.csv	Authors / Prolific (May 2024)	Prolific demographics for Survey 1 participants	Yes
replication_data/survey_data_anonymous/survey2_data_anonymous.xlsx	Authors / Prolific (October 2025)	Survey 2; N = 170; Meritocratic vs Risk-only	Yes

5. Computational Requirements

5.1 Software Requirements

R version $\geq 4.1.0$ is required for scripts 01, 02, and 03. R version 4.2.1 (2022-06-23) is required for exact replication of the causal forest estimates in script 04. Results are sensitive to the number of CPU cores; the paper uses R 4.2.1 on an 8-core Windows 10/11 PC.

R packages required (all installed automatically by 00_install_packages_windows.R for Windows users and 00_install_packages_unix.R for Linux/macOS users):

Package	Version	Used in script(s)	Source
readxl	CRAN latest	01, 03	CRAN
zoo	CRAN latest	01	CRAN
tidyr	CRAN latest	01	CRAN
dplyr	CRAN latest	01, 03	CRAN
reshape2	CRAN latest	01	CRAN
spatstat	CRAN latest	01	CRAN
tibble	CRAN latest	01	CRAN
ggplot2	CRAN latest	02, 03, 04	CRAN
stargazer	CRAN latest	02	CRAN
lmtest	CRAN latest	02	CRAN
sandwich	CRAN latest	02	CRAN
car	CRAN latest	02	CRAN
vtable	CRAN latest	02	CRAN
xtable	CRAN latest	02, 04	CRAN
plyr	CRAN latest	02	CRAN
moments	CRAN latest	02	CRAN
hdm	0.3.2 (pinned)	02, 04	CRAN archive
grf	1.2.0 (pinned)	04	CRAN archive

Note: hdm 0.3.2 and grf 1.2.0 are fetched from the CRAN archive by 00_install_packages_windows.R or 00_install_packages_unix.R to ensure exact replication. Do not update these two packages after installation.

5.2 Controlled Randomness

Scripts 01-03 do not use random number generation. The causal forest estimation in script 04 uses internally seeded random number generation within the grf package. Exact replication of the causal forest estimates requires R 4.2.1, grf 1.2.0, and an 8-core machine. Estimates may differ on machines with a different number of CPU cores, as grf is multi-threaded.

5.3 Memory, Runtime, and Storage Requirements

Estimated runtime per script:

- 00_install_packages_windows.R -- 2-10 minutes on first run; under 5 seconds thereafter
- 00_install_packages_unix.R -- 2-10 minutes on first run; under 5 seconds thereafter
- 01_code_merge.R -- under 1 minute
- 02_code_main_analysis.R -- under 30 seconds
- 03_code_survey.R -- under 30 seconds
- 04_code_machine_learning_part.R -- approximately 3 minutes (8-core PC)

Storage: under 100 MB for the full replication package.

RAM: 4 GB is sufficient for all scripts.

Hardware used by authors: Windows 10/11, 64-bit, 8-core CPU, 16 GB RAM.

6. Description of Programs/Code

6.1 Folder Structure

replication_code/

- README.docx -- this file
- 00_install_packages_windows.R -- installs all R packages for Windows users (run once)
- 00_install_packages_unix.R -- installs all R packages for Linux/macOS users (run once)
- 01_code_merge.R -- merges raw session data into dat_numeric.csv
- 02_code_main_analysis.R -- main OLS results, tables, and figures
- 03_code_survey.R -- survey experiment results and figures
- 04_code_machine_learning_part.R -- causal forest / ML results and figures

replication_data/

- dat_numeric.csv -- pre-built main analysis dataset (2,086 x 375)
- raw_data_anonymous/ -- 11 anonymised session subfolders
- survey_data_anonymous/ -- anonymised survey XLSX and demographics CSV
- figures/ -- output: figures from script 02
- figures_ML/ -- output: figures from script 04
- figures_survey/ -- output: figures from script 03
- tables/ -- output: tables from script 02
- tables_ML/ -- output: tables from script 04

6.2 Script Descriptions

00_install_packages_windows.R and 00_install_packages_unix.R

Installs all R packages required by scripts 01-04. Prints a status line for each package: [OK - already installed], [NEWLY INSTALLED], [VERSION MISMATCH], or [FAILED]. Forces installation of hdm version 0.3.2 and grf version 1.2.0 from the CRAN archive. Issues a warning if the R version is not 4.2.1. Run once before any analysis script.

01_code_merge.R

Input: replication_data/raw_data_anonymous/ (11 session subfolders)

Output: replication_data/dat_numeric.csv

Merges anonymised Prolific CSV and TXT files (some sessions were exported in different formats and more than one file) and experimental XLSX files from all 6 experimental sessions into a single analysis-ready dataset (2,086 rows x 375 columns). Filters to approved participants, harmonises column names across sessions, and constructs numeric dummy variables for all treatments and outcomes. Runtime: under 1 minute.

02_code_main_analysis.R

Input: replication_data/dat_numeric.csv

Output: Figures saved to replication_data/figures/; tables saved to replication_data/tables/

Figures saved to replication_data/figures/:

- Figure 2 (top panel) -- tournament entry rates, Meritocratic + Unfair [Figure2_top.pdf]
- Figure 2 (bottom panel) -- tournament entry rates, Feedback treatment [Figure2_bottom.pdf]
- Figure A5 -- theoretical Bayesian posteriors (Appendix C) [FigureA5.pdf]
- Figure A6 (2 files) -- performance score distributions by gender [FigureA6_perf_women.pdf, FigureA6_perf_men.pdf]
- Figure A7 (2 files) -- tournament entry by performance quartile [FigureA7_entry_by_perf_merit.pdf, FigureA7_entry_by_perf_unfair.pdf]

Tables saved to replication_data/tables/:

- Table 1 -- summary statistics, full sample by gender
- Table A2 -- summary statistics, Meritocratic treatment by gender
- Table A3 -- summary statistics, Unfair treatment by gender
- Table A4 -- summary statistics, Feedback treatment by gender
- Table 2 -- main OLS regressions (tournament entry)
- Table A5 -- Romano-Wolf MHT-adjusted p-values
- Table A6 -- Benjamini-Hochberg MHT-adjusted p-values
- Table A9 -- belief updating by gender and treatment

Table A7 -- optimal behaviour in the experiment (Appendix E)

Table A8 -- conditional optimal behaviour (Appendix E)

Also prints all in-text statistics for Sections 4.2-4.4. Runtime: under 30 seconds.

03_code_survey.R

Input: replication_data/survey_data_anonymous/ (XLSX and CSV files)

Output: Figures saved to replication_data/figures_survey/; statistics printed to R console

Figures saved to replication_data/figures_survey/:

- Figure A1 -- within-person rating differences, Survey 1 [FigureA1_survey1.pdf]
- Figure A2 -- Meritocratic treatment ratings by gender, Survey 1 [FigureA2_merit_gender.pdf]
- Figure A3 -- Unfair treatment ratings by gender, Survey 1 [FigureA3_unfair_gender.pdf]
- Figure A4 -- fairness and merit ratings, Survey 2 [FigureA4_survey2.pdf]

Also prints all in-text statistics for Section 4.1 (fairness, merit, and random-chance ratings by treatment and gender). Runtime: under 30 seconds.

04_code_machine_learning_part.R

Input: replication_data/dat_numeric.csv

Output: Figures saved to replication_data/figures_ML/; tables saved to replication_data/tables_ML/

Figures saved to replication_data/figures_ML/:

- Figure 3 (3 files) -- CATE distributions by gender, one per treatment [Figure3_merit.pdf, Figure3_unfair.pdf, Figure3_feed.pdf]
- Figure A8 (3 files) -- CATE distributions, broader covariate set [FigureA8_merit.pdf, FigureA8_unfair.pdf, FigureA8_feedback.pdf]

Tables saved to replication_data/tables_ML/:

- Table A10 -- main predictors of the effect of losing via LASSO (main covariate set, Appendix F.1)
- Table A11 -- main predictors of the effect of losing via LASSO (broader covariate set, Appendix F.2)

Also prints in-text p-values for gender differences in CATEs by treatment (Section 4.5 and Appendix F). Runtime: approximately 3 minutes on an 8-core PC. Requires R 4.2.1 and grf 1.2.0 for exact replication.

6.3 Configuration

Each analysis script (01-04) contains a single path variable that must be updated to point to the replication_data/ folder on the replicator's machine. Specifically:

- 01_code_merge.R: edit the line `BASE <- "..."` near the top (line 40)
- 02_code_main_analysis.R: edit the `setwd(...)` call near the top (line 60)
- 03_code_survey.R: edit the line `BASE <- "..."` near the top (line 49)
- 04_code_machine_learning_part.R: edit the `setwd(...)` call near the top (line 72)

No other configuration is required. The path string must use forward slashes or double backslashes on Windows (e.g., "C:/Users/YourName/replication_data").

7. Instructions to Replicators

Step 1 -- Install R

Download and install R from <https://cran.r-project.org/>. Any version ≥ 4.1 works for scripts 01-03. For script 04 (exact causal forest replication), install R 4.2.1: <https://cran.r-project.org/bin/windows/base/old/4.2.1/>. Both versions can coexist on the same machine.

Step 2 -- Install RStudio (recommended)

Download from <https://posit.co/download/rstudio-desktop/>. Windows users should also install Rtools (<https://cran.r-project.org/bin/windows/Rtools/>) to allow R to compile packages from source. Choose the Rtools version matching your R version. Mac and Linux users do not need Rtools.

Step 3 -- Install all packages (run once)

Open either 00_install_packages_windows.R (Windows user) or 00_install_packages_unix.R (Linux/macOS user) in RStudio and click Source (top-right of the editor pane), or press Ctrl+Shift+S (Windows/Linux) / Cmd+Shift+S (Mac). Alternatively, open and source the platform script directly:

Windows: 00_install_packages_windows.R

Linux / macOS: 00_install_packages_unix.R

The script will:

Check your R version and warn if it is not 4.2.1 (required for exact replication of Script 04 only; Scripts 01–03 run on any recent R)

Print a status line for every package: [OK - already installed], [NEWLY INSTALLED], [VERSION MISMATCH], or [FAILED]

Automatically fetch hdm 0.3.2 and grf 1.2.0 from the CRAN archive (these two packages are version-pinned for exact replication — do not update them after installation)

Finish with a summary table showing the installed version of every package

If you see [FAILED] for hdm or grf:

Windows: Install the Rtools version that matches your R installation — R 4.2.x requires Rtools42, R 4.3.x requires Rtools43, R 4.4.x requires Rtools44. Download from: <https://cran.r-project.org/bin/windows/Rtools/>. Restart R and re-run the script.

macOS: Install Xcode Command Line Tools (xcode-select --install in Terminal) and gfortran (via Homebrew: brew install gcc, or from <https://mac.r-project.org/tools/>). Restart R and re-run the script.

Linux: Install compiler tools (sudo apt-get install r-base-dev on Debian/Ubuntu, or sudo dnf install R-devel gcc-gfortran on Fedora/RHEL). Restart R and re-run the script.

Run the installer once per R installation; it is safe to re-run at any time.

Step 4 -- Set the data path

Open each analysis script (01 through 04) in RStudio. At the top of each file, locate and update the single path line to the correct absolute path of the replication_data/ folder on your machine. The specific lines to edit are: `BASE <- "..."` in scripts 01 and 03 (lines 40 and 46 respectively), and `setwd(...)` in scripts 02 and 04 (lines 58 and 72). Use forward slashes in path strings on all platforms (e.g. "C:/Users/YourName/replication_data"). This is the only edit required in any script.

Step 5 -- Run scripts in order

Run scripts in this order: 01 -> 02 -> 03 -> 04. Scripts 02-04 can also be run directly using the pre-built `dat_numeric.csv` (already included), without first running script 01.

Note: For exact replication of the causal forest estimates in script 04, use R 4.2.1 and grf 1.2.0 on an 8-core machine. Estimates may differ on machines with a different number of CPU cores.

8. List of Tables and Figures

Output	Script	File name / location
Table 1 (summary statistics, full sample by gender)	02_code_main_analysis.R	tables/Table1.tex
Table A2 (summary statistics, Meritocratic treatment)	02_code_main_analysis.R	tables/TableA2.tex
Table A3 (summary statistics, Unfair treatment)	02_code_main_analysis.R	tables/TableA3.tex
Table A4 (summary statistics, Feedback treatment)	02_code_main_analysis.R	tables/TableA4.tex
Table 2 (main OLS regressions: tournament entry)	02_code_main_analysis.R	tables/Table2.tex
Table A5 (Romano-Wolf MHT-adjusted p-values)	02_code_main_analysis.R	tables/TableA5.tex
Table A6 (Benjamini-Hochberg MHT-adjusted p-values)	02_code_main_analysis.R	tables/TableA6.tex
Table A9 (belief updating by gender and treatment)	02_code_main_analysis.R	tables/TableA9.tex
Table A7 (optimal behaviour in the experiment, Appendix E)	02_code_main_analysis.R	tables/TableA7.tex
Table A8 (conditional optimal behaviour, Appendix E)	02_code_main_analysis.R	tables/TableA8.tex
Figure 2 top panel (entry rates, Meritocratic + Unfair)	02_code_main_analysis.R	figures/Figure2_top.pdf
Figure 2 bottom panel (entry rates, Feedback treatment)	02_code_main_analysis.R	figures/Figure2_bottom.pdf
Figure A5 (theoretical Bayesian posteriors, Appendix C)	02_code_main_analysis.R	figures/FigureA5.pdf
Figure A6 -- women (performance score distributions)	02_code_main_analysis.R	figures/FigureA6_perf_women.pdf
Figure A6 -- men (performance score distributions)	02_code_main_analysis.R	figures/FigureA6_perf_men.pdf
Figure A7 -- Meritocratic	02_code_main_analysis.R	figures/FigureA7_entry_by_perf_merit.pdf

(entry by performance quartile)		
Figure A7 -- Unfair (entry by performance quartile)	02_code_main_analysis.R	figures/FigureA7_entry_by_perf_unfair.pdf
Figure A1 (within-person rating differences, Survey 1)	03_code_survey.R	figures_survey/FigureA1_survey1.pdf
Figure A2 (Meritocratic treatment ratings by gender, Survey 1)	03_code_survey.R	figures_survey/FigureA2_merit_gender.pdf
Figure A3 (Unfair treatment ratings by gender, Survey 1)	03_code_survey.R	figures_survey/FigureA3_unfair_gender.pdf
Figure A4 (fairness and merit ratings, Survey 2)	03_code_survey.R	figures_survey/FigureA4_survey2.pdf
Figure 3 -- Meritocratic (CATE distribution by gender)	04_code_machine_learning_part.R	figures_ML/Figure3_merit.pdf
Figure 3 -- Unfair (CATE distribution by gender)	04_code_machine_learning_part.R	figures_ML/Figure3_unfair.pdf
Figure 3 -- Feedback (CATE distribution by gender)	04_code_machine_learning_part.R	figures_ML/Figure3_feed.pdf
Figure A8 -- Meritocratic (broader covariate set)	04_code_machine_learning_part.R	figures_ML/FigureA8_merit.pdf
Figure A8 -- Unfair (broader covariate set)	04_code_machine_learning_part.R	figures_ML/FigureA8_unfair.pdf
Figure A8 -- Feedback (broader covariate set)	04_code_machine_learning_part.R	figures_ML/FigureA8_feedback.pdf
Table A10 (LASSO predictors, main covariate set, Appendix F.1)	04_code_machine_learning_part.R	tables_ML/TableA10.tex
Table A11 (LASSO predictors, broader covariate set, Appendix F.2)	04_code_machine_learning_part.R	tables_ML/TableA11.tex

9. References

Piasenti, S., Valente, M., van Veldhuizen, R., and Pfeifer, G. (2026). Does Unfairness Hurt Women? The Effects of Losing Unfair Competitions. Economic Journal (forthcoming).

Pre-registration -- Main experiment (AEA Social Science Registry, trial 7651):

<https://www.socialscienceregistry.org/trials/7651>

Pre-registration -- Survey Experiment 1 (OSF): <https://osf.io/4f97v>

Pre-registration -- Survey Experiment 2 (OSF): <https://osf.io/hbv7q>

Athey, S. and Wager, S. (2019). Estimating Treatment Effects with Causal Forests: An Application. Observational Studies, 5(2), 37-51.

Chernozhukov, V., Hansen, C., and Spindler, M. (2016). hdm: High-Dimensional Metrics. R Journal, 8(2), 185-199.

*Social Science Data Editors (2024). Template README for Social Science Replication Packages.
https://social-science-data-editors.github.io/template_README/*