ECONSTOR Make Your Publications Visible.

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Frese, Joris

Working Paper Report for the Nature Human Behaviour Mass Reproduction Initiative: Seeing Racial Avoidance on New York City Streets

I4R Discussion Paper Series, No. 202

Provided in Cooperation with: The Institute for Replication (I4R)

Suggested Citation: Frese, Joris (2025) : Report for the Nature Human Behaviour Mass Reproduction Initiative: Seeing Racial Avoidance on New York City Streets, I4R Discussion Paper Series, No. 202, Institute for Replication (I4R), s.l.

This Version is available at: https://hdl.handle.net/10419/311307

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

INSTITUTE for **REPLICATION**

No. 202 I4R DISCUSSION PAPER SERIES

Report for the Nature Human Behaviour Mass Reproduction Initiative: Seeing Racial Avoidance on New York City Streets

Joris Frese

February 2025



I4R DISCUSSION PAPER SERIES

I4R DP No. 202

Report for the Nature Human Behaviour Mass Reproduction Initiative: Seeing Racial Avoidance on New York City Streets

Joris Frese¹

¹European University Institute, Fisole/Italy

FEBRUARY 2025

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and metascientific work in the social sciences. Provided in cooperation with EconStor, a service of the <u>ZBW – Leibniz Information Centre for Economics</u>, and <u>RWI – Leibniz Institute for Economic Research</u>, I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur University of Ottawa Anna Dreber Stockholm School of Economics Jörg Ankel-Peters RWI – Leibniz Institute for Economic Research

E-Mail: joerg.peters@rwi-essen.de RWI – Leibniz Institute for Economic Research Hohenzollernstraße 1-3 45128 Essen/Germany www.i4replication.org

Report for the Nature Human Behaviour Mass Reproduction Initiative: Seeing Racial Avoidance on New York City Streets*

Joris Frese

June 24, 2024

Abstract

This report investigates the computational reproducibility and robustness of the paper "Seeing Racial Avoidance on New York City Streets" (Dietrich and Sands 2023). These reproduction efforts are part of the mass reproduction initiative for articles published in Nature Human Behaviour (NHB), which is jointly organized by the Institute for Replication and NHB. In the original NHB paper, Dietrich and Sands analyze a field experiment in New York City, finding that "pedestrians deviate by, on average, 3.43% of the sidewalk width [...] or around 4 inches, in the presence of black confederates" (compared to white confederates), signalling a statistically significant racial avoidance of black people. For this report, I first conduct a step-by-step reproduction of the original replication materials, followed by robustness checks including 1) an analysis without outliers, 2) analyses with alternative seeds for the bootstrapped standard errors, and 3) an analyis with non-bootstrapped standard errors. I find that the original results are fully reproducible and that they are robust to many, but not all, alternative specifications.

KEYWORDS: Computational Reproducibility; Robustness; I4R; Racial Avoidance

^{*}Author: Joris Frese, Department of Political and Social Sciences, European University Institute. E-mail: joris.frese@eui.eu. The author declares no financial support for this report and no conflict of interest. The replication files for the robustness reproduction documented in this report can be found at https://github.com/fresej/I4R-NHB-Reproduction.

1 Introduction

In January 2024, the Institute for Replication (I4R) announced a large-scale collaboration with the academic journal Nature Human Behaviour (NHB), one of the most widely cited journals across the social and behavioural sciences. As part of this initiative, collaborators are invited to reproduce or replicate papers published in NHB from 2023 onwards. In this reproduction report, I document the results of my reproduction and robustness checks for the paper "Seeing Racial Avoidance on New York City Streets" by Bryce Dietrich and Melissa Sands, which was published in NHB in 2023.

For this paper, Dietrich and Sands analyzed data including more than 3500 pedestrians in New York City who were experimentally exposed to either two young white men conversing on the boardwalk, or two young black men conversing on the boardwalk, or to a baseline object as a control condition. The outcome of interest in their study is the "standardized pedestrian deviation", meaning "the deviation of a pedestrian from the confederate location as a proportion of the sidewalk width" (Dietrich and Sands 2023). Higher distance to one pair of confederates (the two young men in conversation) compared to the other then signals racial avoidance, as all other features besides the confederates' race are held constant.

Precisely, five hypothesis tests are reported in the paper and its Appendix: (1) The obstruction avoidance hypothesis, which states that pedestrians keep a wider distance to either confederate pair than to the control object, (2) the racial avoidance hypothesis, which states that pedestrians keep a wider distance to the black compared to the white confederates, (3) the neighborhood outgroup salience hypothesis, which states that this width gap will be larger in white neighborhoods, (4) the pedestrian outgroup salience hypothesis, which states that this width gap will be larger for white compared to black pedestrians, and (5) the gender heterogeneity hypothesis, which states that female pedestrians will keep a wider distance to either confederate pair compared to male pedestrians. The authors find support for hypotheses 1, 2, and 5, but not for hypotheses 3 and 4.

In the following sections, I first document my computational reproduction of all main and supplementary analyses for this paper, focusing both on the equivalence of the produced outputs and the user-friendliness of the provided code. Next, Institute for Replication

I4R DP No. 202

I conduct three additional analyses to assess the robustness of the paper's main finding (the increased deviation of pedestrians from black compared to white confederates) to alternative specifications: 1) an analysis without outliers, 2) analyses with alternative seeds for the bootstrapped standard errors, and 3) an analysis with non-bootstrapped standard errors. Finally, I conclude with an overall assessment of the paper's reproducibility and robustness.

2 Computational Reproducibility

The replication files contain separate R scripts for each table and figure in the main paper and the supplementary materials, as well as all the necessary datasets in CSV format. To assess the computational reproducibility of all results from the paper and the supplementary materials, I ran all R scripts and compared the reproduced results to the original ones. I managed to get all scripts to run and the results I obtained were identical to those originally reported by the authors, indicating a fully reproducible workflow and transparently and truthfully reported results.

Apart from the equivalence of outputs, which is the most important factor of this initial reproduction effort, several aspects of the replication files warrant further discussion: the structure of the provided datasets, the working directories in the provided scripts, and the object storage in the provided scripts. I will quickly go through each of these three aspects and provide suggestions for improvements. None of these discussions affect or question the results of the paper in any way. They are merely concerned with the user-friendliness of the replication files. I argue that this aspect is important because user-friendly replication files encourage replication attempts and increase transparency.

Most of the datasets provided in the replication files (except the "nyc_data.csv" dataset for table 1) do not contain original data, but merely pre-processed summary statistics. As an example, Figure 1 below shows the full dataset needed to replicate figure 2 from the original paper. The dataset contains only 12 observations: the summary statistics (including coefficients, t-values, p-values, confidence intervals, etc) behind the 12 coefficients plotted in that figure. Thus, an independent replication of most analyses is not possible. Obtaining identical results to the original

I4R DP No. 202

analyses is the only possible outcome of any computational reproduction attempt because the only course of action is to plot and tabularize the results obtained by the authors themselves. While the dataset "nyc_data.csv" contains the full data, it should not be the replicator's duty to try and independently write up the code behind the analyses of all figures and tables (except for table 1). In some cases, the script for table 1 contains references to lines of code that prepare the data for figure 2 and have been commented out according to the authors, but these lines are not aligned with the published script anymore (e.g., they refer to lines 46-47, but line 47 is empty). To increase transparency, the analyses behind all outputs should be provided for replication, not just the outputs of the analyses themselves.

A second, less important critique that nevertheless concerns the user-friendliness of the replication files, is related to the working directories in the scripts. Each of the more than 15 separate scripts refers to working directories on the original authors' computers. These directories need to be manually changed in each script. To increase user-friendliness, one of the following two options would suffice: either a master script should be provided to manage the directories and dependencies across all scripts, or an R project should be included in the replication files to make the provision of working directories redundant.

One final, minor critique is related to the object storage in the script for table 1. The results of each analysis are stored in objects of the same name (e.g., "mod3", or "boot_se"). Thus, after each analysis, the object containing the results of the previous analysis is overwritten. While the results are still saved for the final HTML output, I would recommend storing all results in R in objects with unique names to allow for easier accessibility (e.g., in case a replicator wants to run additional analyses or consult the more detailed model outputs in these objects).

3 Robustness Checks

As part of the instructions for the NHB mass reproduction intitiative, every replicator was told to submit a pre-re-analysis plan documenting a total of three robustness checks to be conducted. As shown in the pre-re-analysis plan in Appendix 7.1, I registered three analyses for a robustness replication of the papers main finding,

I4R DP No. 202

which states that, compared to white confederates, "pedestrians deviate by, on average, 3.43% of the side-walk width [...] or around 4 inches, in the presence of black confederates" (Dietrich and Sands 2023). This statistically significant finding indicates racial avoidance of black people. To assess the robustness of this finding, I conducted 1) an analysis without outliers, 2) analyses with alternative seeds for the bootstrapped standard errors, and 3) an analysis with non-bootstrapped standard errors. The results of all these robustness replications are documented in Table 1, alongside the original result, which I successfully reproduced. I will now briefly go through the setup and the results of each robustness replication in the following three subsections.

3.1 Excluding Outliers

The first robustness check that I pre-registered is based on the observation that there is an unusual spike of outliers in the density distribution of the main outcome variable: the difference in the distance from the pedestrians to the confederates and the mean distance from the pedestrians to a baseline object. As shown in Figure 2 below, the outcome variable is almost normally distributed. However, at around 0.4 (or 40% of the boardwalk width), there is a substantial spike in observations. I assume that this spike is due to people walking on the opposite side of the boardwalk, close to the wall/building. To see whether the results might be driven by these people who are simply walking on the opposite side of the boardwalk, I decided to conduct a robustness check where I re-analyze the main model after excluding outliers with outcomes above 0.4 and below -0.4. If the results do not hold without these outliers, that weakens the conclusions of the paper, as it would mean that results are mainly driven by pedestrians walking far away from the confederates anyway. Whether someone walks at 1m distance or 1.1m distance barely matters, whereas the difference between 5cm and 15cm distance might be quite substantial.

As shown in rows 3 and 4 of Table 1 below, the coefficient of this robustness analysis still has the same sign as the original one in both specifications (with and without control variables), but it is substantially smaller and not statistically significant any longer. This suggests that the results are indeed mainly driven by pedestrians walking at the opposite side of the boardwalk, far away from the pedestrians.

3.2 Setting Alternative Seeds for Bootstrapping

The main analysis relies on bootstrapping to estimate standard errors that are clustered at the 15-minute-location-level. As this bootstrapping involves a random process, the authors rightfully included a command to set a random seed in their code. This seed seems to have been arbitrarily chosen (it is: 70984). To ensure that the results are not driven by this arbitrarily chosen random seed, I re-analyzed the main model with three alternative, pre-registered seeds (111; 123; 2023). As shown in Table 1, rows 5-10, the results are unaffected by these alternative seeds. The standard errors are identical at the third decimal place in all specifications.

3.3 Estimating Non-Bootstrapped Standard Errors

Lastly, I pre-registered an analysis without bootstrapped standard errors, instead using the lm_robust canned routine in R to estimate clustered standard errors. In this specification, the standard errors are again clustered at the 15-minute-locationlevel as in all the other specifications, but lm_robust estimates non-bootstrapped, heteroskedasticity-robust standard errors.¹ With this approach, the results still hold in the model with control variables, but they are not statistically significant any longer in the model without control variables, as shown in rows 11 and 12 of Table 1.

4 Conclusion

The results reported in the NHB article "Seeing racial avoidance on New York City streets" (Dietrich and Sands 2023) show that pedestrians in New York City tend to keep a larger distance to black people than to white people, suggesting racial avoidance of black people. In this pre-registered robustness reproduction report, which is part of the large-scale collaboration between the I4R and NHB to assess the replicability of all NHB articles from 2023 onwards, I re-analyzed the original code provided by the authors and conducted a series of robustness checks to see if the original results hold across alternative model specifications.

¹Using the argument se_type = "stata".

I4R DP No. 202

All original results are fully computationally reproducible, but I also document room for improvement with regard to the structure of the provided datasets, the working directories in the provided scripts, and the object storage in the provided scripts.

When re-analyzing the main model with alternative specifications, the results hold across most, but not all specifications. Alternative seeds for bootstrapping do not affect the results in any way. The results also hold in one specification estimating non-boostrapped standard errors, but only when control variables are included. When excluding extreme outlier pedestrians who are walking at the far edge of the board walk, the estimated coefficients still have the same sign as before, but they are much smaller and not statistically significant, suggesting that the main results are mostly driven by these pedestrians walking on the opposite side of the boardwalk.

References

Dietrich, B. and Sands, M.: 2023, Seeing racial avoidance on New York City streets, *Nature Human Behaviour* **7**(8), 1275–1281.

5 Figures

Figure 1: Excerpt from the replication files: full dataset behind figure 2 in the original paper

1	location	pedestrians $\hat{}$	controls 🔅	df [‡]	fit [‡]	t_value	p_value	lower90 $^{\diamond}$
1	Both Locations	All	Y	3322	0.0342876578469357	5.16656435216738	2.52492728171492e-07	0.0233686202906213
2	Both Locations	All	N	3417	0.0322435295591867	2.40303034917575	0.0163128274983258	0.0101671254522813
3	Both Locations	Non-Black	Y	3149	0.0329705906520977	3.88985656896277	0.00010237720873827	0.0190246382725037
4	Both Locations	Non-Black	N	3241	0.0295296808651084	1.97153333489005	0.048747833397057	0.00488597370319235
5	Upper East Side	All	Y	424	0.0392648818398897	1.22343709258894	0.221844133103437	-0.0136404978759991
6	Upper East Side	All	N	514	0.0319048823911172	0.750501745277272	0.453295988617679	-0.0381464087796556
7	Upper East Side	Non-Black	Y	396	0.0393544972546106	1.39980732007736	0.162353931569866	-0.00699772272390366
8	Upper East Side	Non-Black	N	483	0.0216781941013288	0.572871340732434	0.566998394832271	-0.0406848117150845
9	Midtown	All	Y	2895	0.0346587279817908	5.12848140665955	3.11341934433362e-07	0.0235391044067798
10	Midtown	All	N	2901	0.0324353046101683	2.61980240536313	0.00884392146385982	0.0120641611033153
11	Midtown	Non-Black	Y	2751	0.0322839167305173	3.76231078495169	0.00017186410409173	0.0181648796361398
12	Midtown	Non-Black	N	2756	0.0311848395792776	2.01470289880199	0.0440327393710998	0.0057161989248932

Figure 2: Prevalence of outliers



Histogram of nyc_data\$distance_diff

6 Tables

Model	Controls	Beta	T-Statistic	Standard Error	DF	P-Value	CI
Original	Y	0.034	5.167	0.007	3322	< 0.001	[0.021, 0.047]
Original	Ν	0.032	2.403	0.013	3417	0.016	[0.006, 0.059]
Excluding Outliers	Υ	0.006	1.231	0.005	2801	0.218	[-0.004, 0.016]
Excluding Outliers	Ν	0.005	0.517	0.009	2896	0.605	[-0.014, 0.023]
Alternative Seed 1	Υ	0.034	4.948	0.007	3322	< 0.001	[0.021, 0.048]
Alternative Seed 1	Ν	0.032	2.551	0.013	3417	0.011	[0.007, 0.057]
Alternative Seed 2	Υ	0.034	5.169	0.007	3322	< 0.001	[0.021, 0.047]
Alternative Seed 2	Ν	0.032	2.446	0.013	3417	0.014	[0.006, 0.058]
Alternative Seed 3	Υ	0.034	5.171	0.007	3322	< 0.001	[0.021, 0.047]
Alternative Seed 3	Ν	0.032	2.485	0.013	3417	0.013	[0.007, 0.058]
Non-Bootstrapped SE	Υ	0.034	4.046	0.008	3322	< 0.001	[0.017, 0.052]
Non-Bootstrapped SE	Ν	0.032	1.309	0.025	3417	0.206	[-0.019, 0.084]

Table 1: Main results and robustness replications

Note: All values are rounded to 3 decimal places.

7 APPENDIX

7.1 Pre-Re-Analysis Plan (08.03.2024)

PRAP written by Joris Frese (European University Institute) **Pre-Re-Analysis Plan** For a robustness reproduction of "Seeing Racial Avoidance on New York City Streets" (Dietrich and Sands 2023) In the framework of the large-scale collaboration between the Institute for Replication (I4R) and Nature Human Behaviour (NHB), I will conduct a robustness reproduction of the analyses for the article "Seeing Racial Avoidance on New York City Streets" by Bryce Dietrich and Melissa Sands, which was published in NHB in 2023. This effort will be structured along the following lines: 1. Testing the computational reproducibility of all original code and data provided in the replication files. 2. Testing the robustness of the main results (Table 1, Rows 1+2 (Both Locations → All Pedestrians → With Controls & Without Controls)) to alternative specifications: - Robustness Check 1: Re-analyzing the main model excluding extreme outliers (i.e., excluding distance differences below -0.4 and above 0.4). - Robustness Check 2: Re-analyzing the main model with different seeds for the bootstrap (i.e., changing the arbitrarily chosen seed (70984) to three different arbitrarily chosen, but more common seeds (111; 123; 2023)). Robustness Check 3: Re-analyzing the main model with non-bootstrapped _ clustered standard errors (i.e., using lm_robust instead of bootstrapping for clustering). Any deviations from this PRAP will be made transparent in the final reproduction report.