When Is Discrimination Unfair

Replication Package

2_Main Analyses

Description of Files

This folder (2_Main Analyses) contains 8 data files:

d1_when_is_discrimination_unfair_data_set.dta (subjects' survey responses)
d2_google_trends.csv (google trends data for Figure A2.1)
d3_acs2019.dta[1]
d4_gss2020.dta[2]
d5_acs2019_analysis.dta (*optional*, output from do file 2_**prep_acs**.do)
d6_gss2020_analysis.dta (*optional*, output from do file 2_**prep_gss**.do)
d7_state_pop.dta
d8_open_response_handlabels.csv (This file provides hand coded labels/categories for the self-reported, open text responses giving subjects' reasons for their fairness assessments. It is used to generate Tables A6.1 and A6.2.)

and 7 Stata do files:

0_**homedir**.do
1_**master**.do
2_**prep_acs**.do
2_**prep_gss**.do
3_**figures**.do
4_**tables**.do
5_**pre_analysis_plan**.do (It generates the results from our pre-analysis plan in Appendix P.)

"**homedir**" must be run first, which sets the home directory and install packages used by all the remaining do files. Notice that package "grc1leg" has to be manually installed using "search".

"**master**" generates almost all the results in the paper and appendices by invoking "**figures**", "**tables**" and "**pre_analysis_plan**" do files. Note that "**pre_analysis_plan**" generates the results from our pre-analysis plan (in Appendix P).

If you choose not to run "**master**", instead, you run "**figures**", "**tables**" and "**pre_analysis_plan**" separately, you should notice that they can be executed in any order.

[1] Source: CEPR Data – ACS, "cepr_acs_2019.zip," https://ceprdata.org/acs-uniform-data-extracts/acs-data/. Same as "acs2019.dta" in Folder 1_Dataset Construction/jj_Materials for Weights.
[2] Source: General Social Survey, "GSS 2016-2020 Panel," https://gss.norc.org/get-the-data/stata. Same as "gss2020.dta" in Folder 1_Dataset Construction/jj_Materials for Weights.

The only other results in the paper (confidence intervals for α in Appendix 7) are generated by standalone R files in the folder 3_Additional Analyses.

Note: Tables are in .rft format, which should be opened using Microsoft Word.

Executing the Files

1. Copy the preceding 15 files into a folder. This folder will be Stata's home directory – homedir—for this replication exercise.

2. Open 0_homedir.do in Stata's do-file editor. As indicated there, modify the path to the home directory from the one on my computer to the one on yours. Save the modified do file, execute it, and check to see that all the listed items are in present your home directory.

3. To hold the results, create three folders called Figures, Tables, and PAP in this home directory. If you run "**homedir**", it will automatically create those folders.

4. If necessary, install the following packages to your Stata: coefplot, outreg2, estout, and grc1leg. If you run "**homedir**", it will automatically install coefplot, outreg2, and estout packages. But you have to manually install "grc1leg" using command "search grc1leg", then click "grc1leg from http://www.stata.com/users/vwiggins", and click "(click here to install)".

5. (Optional) Run "**master**" do file, which automatically run the remaining do files ("**prep_acs**", "**prep_gss**", "**figures**", "**tables**" and "**pre_analysis_plan**")

6. (Optional) If you want to re-generate d5_acs2019_analysis.dta and d6_gss2020_analysis.dta, you can run "**prep_acs**" and "**prep_gss**" do files.

7. Run any of the remaining do-files ("**figures**", "**tables**" and "**pre_analysis_plan**") in any order.

Explanation on variables in the main data set

(d1_when_is_discrimination_unfair_data_set.dta)

- **randomid**: the respondent's randomly generated 5-digit identifier.
- **durationinseconds**: number of seconds it takes for the respondent to complete the survey.
- **gender**: a string that indicates the gender of the respondent**.**
- **age**: a string that indicates the age category of the respondent (e.g., aged 18-24).
- **education**: a string that indicates the education level of the respondent (e.g., 4-year college or university).
- **race**: a string that indicates the race of the respondent (e.g., White, Black, Asian, etc.).
- **party**: a string that indicates the U.S. political party preference of the respondent (e.g., democrat).
- **spectrum**: a string that indicates the political leaning of the respondent (e.g., slightly liberal).
- **black**: a binary variable that takes on a unit value if the encountered scenario belongs to the Black race treatment.

- **white**: a binary variable that takes on a unit value if the encountered scenario belongs to the White race treatment.
- **stat**: a binary variable that takes on a unit value if the encountered scenario belongs to the statistical treatment.
- **taste**: a binary variable that takes on a unit value if the encountered scenario belongs to the taste-based treatment.
- **low_just**: a binary variable that takes on a unit value if the encountered scenario illustrates the "low-justifiable" version of taste-based or statistical discrimination.
- **high_just**: a binary variable that takes on a unit value if the encountered scenario illustrates the "high-justifiable" versions of taste-based or statistical discrimination.
- **open_response**: the respondent's justification (in words) for providing their fairness assessment to the last scenario they encountered. The entry for this variable is only populated among scenarios corresponding to the fourth (i.e., last) on a respondent encountered.
- **fair**: the respondent's fairness assessment on a numerical scale of -3 (very unfair) to 3 (very fair).
- **bro**: the respondent's assessment on the relative economic opportunities between Black and White people in the U.S. – it is also numerical on a scale of -3 (Black people have much less opportunity…) to 3 (Black people have much more opportunity…).
- **non_white**: a binary variable that takes on a unit value if the respondent's race is not White.
- **prefer_no_answer**: a binary variable that takes on a unit value if the respondent preferred not to answer the follow-up open response question in Stage 3.
- **male**: a binary variable that takes on a unit value if the respondent is a male.
- **female**: a binary variable that takes on a unit value if the respondent is a female.
- **race_xxx**: a binary variable that takes on a unit value if the respondent is from some racial group. E.g., xxx could be "white," black," "indigenous" (for Indigenous American), "other" (other racial category), etc.
- **other**: a binary variable that takes on a unit value if the respondent's race is non-White but not Black.
- **agexx_yy:** a binary variable that takes on a unit value if the respondent comes from some age category. E.g., xx_yy could be 18_24, 25_34, etc.
- **education_zzz:** a binary variable that takes on a unit value if the respondent has some education level. E.g., zzz could be "hs," (for high school), "some_college" (2-year or some college), "four_college" (4-year college or university), or "higher" (graduate school).
- **party_ppp**: a binary variable that takes on a unit value if the respondent has some U.S. political party preference. E.g., ppp could be "democrat," or "indep_other" (independent or some other party).
- **conserv**: a binary variable that takes on a unit value if the respondent is "very conservative" or "conservative."
- **mod**: a binary variable that takes on a unit value if the respondent is "slightly liberal," "moderate," or "slightly conservative." I.e., we lump these categories into a "moderate" group for our analyses.
- **lib**: a binary variable that takes on a unit value if the respondent is "very liberal" or "liberal."

- **census_ccc**: a binary variable that takes on a unit value if the respondent is from some Census region. E.g., ccc can be “west,” “midwest,” “south,” and “northeast.”
- **survey_weight:** survey weights based on the 2019 ACS.
- **survey_weight2:** survey weights based on the 2020 GSS.
- **black1**: a binary variable that takes on a unit value if the respondent encountered the Black treatment in Stage 1.
- **black2**: a binary variable that takes on a unit value if the respondent encountered the Black treatment in Stage 2.
- **white1**: a binary variable that takes on a unit value if the respondent encountered the White treatment in Stage 1.
- **white2**: a binary variable that takes on a unit value if the respondent encountered the White treatment in Stage 2.
- **taste1**: a binary variable that takes on a unit value if the respondent encountered the taste-based treatment in Stage 1.
- **taste2**: a binary variable that takes on a unit value if the respondent encountered the taste-based treatment in Stage 2.
- **ordering**: takes on a numerical value of 1-4 and indicates the order in which the scenario was encountered by the respondent.
- **low_just1**: takes on a binary variable if the respondent encountered the low-justifiable scenario first within Stage 1.
- **low_just3**: takes on a binary variable if the respondent encountered the low-justifiable scenario first within Stage 2.
- **stat1**: a binary variable that takes on a unit value if the respondent encountered the statistical treatment in Stage 1.
- **stat2**: a binary variable that takes on a unit value if the respondent encountered the statistical treatment in Stage 2.
- **black_only**: a binary variable that takes on a unit value if the respondent only encounters the Black treatment throughout the survey.
- **white_only**: a binary variable that takes on a unit value if the respondent only encounters the White treatment throughout the survey.
- **switcher**: a binary variable that takes on a unit value if the respondent switches race treatments between stages.

The data set also contains binary variables corresponding to certain U.S. states (e.g., **ca** for California or **pa** for Pennsylvania). **ind** corresponds to Indiana.