Appendix D - Weighting Methodology and Results

A plausible concern about the results of our study is that MTurk workers are not representative of the US population and the main results could be biased to the degree that responses to our experimental conditions are correlated with systematic differences between MTurk workers and the population at-large. Although experimental conditions were randomized and therefore uncorrelated with MTurk worker attributes, *being* an MTurk worker is subject to potential selection bias. To assess the degree to which selection bias affects our estimates, we reweight survey responses to better approximate US population statistics across multiple important domains including age, race, sex, income, and education.

The impact of our reweighting technique on our focal outcomes is minimal, suggesting that the observed differences between our sample of MTurk workers and the US population are unlikely to overturn our results. However, we may only condition estimated sampling weights on individual population distributions, not joint distributions (i.e., just age and race singularly, not the joint distribution of age *and* race). This is predominantly an issue with available population characteristics and not statistical considerations, as the entropy balancing algorithm used to calculate sampling weights could easily calculate weights conditioned on joint distributions if such information were available.

Calculating Sampling Weights

The degree to which MTurk workers differ from the US population is well-documented (Levay et al., 2016) and our sample characteristics are in accord with prior research: survey respondents are more often White, more likely to have a college degree, and generally have lower incomes (see Table 1 in the main text for details). US population estimates were taken

from the US Census Bureau American Community Survey (2018) and are reported in Table D1

alongside descriptive statistics for our sample of MTurk workers.

[Insert Table D1 about here]

We used the "ebalance" function (Hainmueller and Xu, 2013) in Stata 15.1 to estimate

sampling weights for our MTurk worker sample. The "ebalance" function implements an entropy

balancing algorithm which was conceived as a substitute for propensity score matching (PSM)

procedures. In addition to its utility as an alternative to PSM, it can also be used to reweight

observations toward a series of population benchmarks, thus producing a sample which, when

weighted using the "svy" suite of functions in Stata, can approximate average US population

characteristics. Specifically, entropy balancing estimates the following equation:

$$H(\omega) = \sum_{[i|D=0]} \omega_i log(\omega_i/q_i)$$

Where ω_i is the weight assigned to each observation which satisfies the conditions imposed upon

the balancing algorithm - in this example, that sample distributions match those in the

population. The exact Stata syntax we use to implement this algorithm follows below:

```
** Define macro for census matching vars

    local census race_ethn1 race_ethn2 race_ethn3 race_ethn4 race_ethn5 ///

    age_cat1 age_cat2 age_cat3 ///

    region1 region2 region3 ///

    sex1 ///

    educ1 educ2 educ3 educ4 educ5 ///

    income1 income2 income3 income4 income5 income6


** Define macro for desired sample means

    local means 0.6156 0.1157 0.1231 0.0036 0.0867 ///

    0.1258 0.3424 0.3389 ///
```

```
      0.1794 0.2110 0.3752 ///

      0.5133 ///

      0.1260 0.2730 0.2080 0.0830 0.1910 ///

      0.0520 0.2115 0.2128 0.1680 0.1157 0.1000


** Estimate sample weights using ebalance command

      ebalance `census', manualtargets(`means') generate(samp_wght)


** Re-estimate regression using sample weights

      svyset samp_wght

      svy: regress callback i.race_crec_cred
```

The initial two steps simply define local macros for later use in the ebalance function - the first

macro lists categories of the variables for which we obtain population estimates. The second

macro lists population parameters for each category - e.g., race_ethn1 corresponds to the White,

Non-Hispanic race category, which represents 61.56% of the US population according to the

most recent ACS (U.S. Census Bureau; 2018) estimates. The following line of syntax

implements the entropy balancing algorithm, specifying population parameters within the

"manualtargets" option and returning a weight stored in the "samp_wght" variable for each

observation using the "generate" option. After estimating this syntax, the "svy" suite of

commands is then used to re-estimate our analysis incorporating sampling weights. The results of

this sensitivity analysis are discussed below.

Results After Weighting

Table D2 reports weighted and unweighted average callback willingness across all

experimental conditions and separated into respondents who evaluated White and African-

American applicants. As is evident in these results, weighted averages which account for

sampling weights are very similar to unweighted averages from the original sample. In fact, the maximum difference we observe is only .02 points on the callback willingness scale, suggesting that we should not expect the results reported in the main text to differ when sampling weights are incorporated into our t-tests. A regression analysis using sampling weights (not shown, available by request) provides estimates very similar to those reported in Tables 2 and 3, so we conclude here that our primary results are robust to sample selection issues related to the observed differences between our sample and the population, though that robustness does not necessarily extend to joint distributions of the weighting variables.

[Insert Table D2 about here]

REFERENCES

Hainmueller, Jens and Yiqing Xu. 2013. ebalance: A Stata package for entropy balancing. *Journal of Statistical Software*, 57(7). https://www.jstatsoft.org/article/view/v054i07

Levay, Kevin E., Jeremy Freese, and James N. Druckman. 2016. The demographic and political composition of Mechanical Turk samples. *Sage Open* https://doi.org/10.1177/2158244016636433

U.S. Census Bureau. 2018. *2013-2017 American Community Survey 5-Year Estimates*. Available online: https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml

Table D1. U.S. Population, Sample, and Weighted Sample Descriptive Statistics

Characteristic	US Population Mean	Unweighted MTurk Sample (n=5822) Mean	Weighted MTurk Sample (n=5784) Mean	Difference (Unweighted – Weighted)
Race & Ethnicity				
White, NH	61.56	73.94	61.57	12.37
African-American, NH	12.31	7.87	12.31	-4.44
Other, NH	8.67	9.82	8.67	1.15
White, Hisp	11.57	5.60	11.57	-5.97
African-American, Hisp	0.36	0.67	0.36	0.31
Other, Hisp	5.53	2.10	5.52	-3.42
Age Category				
18 to 24 years	12.58	8.57	12.58	-4.01
25 to 44 years	34.24	70.16	34.27	35.89
45 to 64 years	33.89	18.26	33.87	-15.61
65 years and above	19.29	3.01	19.28	-16.27
Sex				
Male	48.67	45.81	48.66	-2.85
Female	51.33	54.19	51.34	2.85
Education				
Less than HS	12.60	0.52	12.60	-12.08
HS Degree or GED	27.30	8.41	27.29	-18.88
Some college	20.80	24.17	20.80	3.37
Associate's Degree	8.30	11.77	8.30	3.47
Bachelor's Degree	19.10	39.89	19.11	20.78
Graduate Degree	11.80	15.24	11.90	3.34
Income				
$14,999 or less	5.20	15.89	5.24	10.65
$15,000 to $29,999	21.15	18.67	21.14	-2.47
$30,000 to $44,999	21.28	20.84	21.27	-0.43
$45,000 to $59,999	16.80	16.44	16.80	-0.36
$60,000 to $74,999	11.57	11.53	11.56	-0.03
$75,000 to $99,999	10.00	9.22	9.99	-0.77
$100,000 or more	14.10	7.41	14.00	-6.59
Region				
Northeast	17.94	19.11	17.94	1.17
Midwest	21.10	20.88	21.10	-0.22
South	37.52	38.55	37.52	1.03
West	23.44	21.45	23.44	-1.99

Notes. U.S. Population statistics taken from the US Census Bureau American Community Survey (2018).

Table D2: Original and Weighted Callback Averages for All Experimental Conditions

Criminal Record Condition	Credential Conditions	White Applicants		Black Applicants	
		Unweighted (n=2900)	Weighted (n=2879)	Unweighted (n=2922)	Weighted (n=2905)
		Mean	Mean	Mean	Mean
No Record	No Credential	4.54	4.54	4.56	4.56
Drug Felony	No Credential	2.73	2.74	3.04	3.03
	Involuntary Job Training	3.01	2.99	3.42	3.43
	Voluntary Job Training	3.37	3.36	3.46	3.46
	Occupational License	3.41	3.42	3.40	3.42
	Reference Letter	4.07	4.08	4.06	4.06
Violent Felony	No Credential	2.73	2.72	2.78	2.78
	Involuntary Job Training	3.02	3.03	3.04	3.04
	Voluntary Job Training	3.33	3.33	3.43	3.43
	Occupational License	3.32	3.33	3.28	3.29
	Reference Letter	4.11	4.13	3.92	3.92