Supplementary Information for Labels vs. Pictures: Treatment-Mode Effects in Experiments About Discrimination

Marisa A. Abrajano* Christopher S. Elmendorf† Kevin M. Quinn‡

September 26, 2017

*UC San Diego Department of Political Science
†UC Davis School of Law
‡University of Michigan Department of Political Science

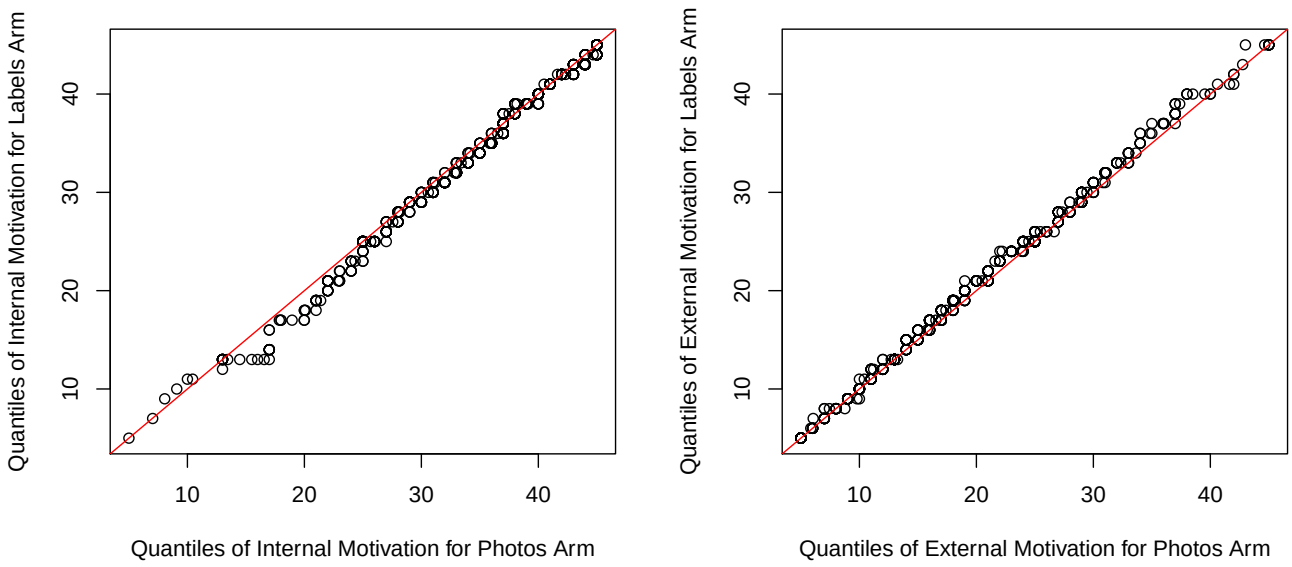

Figure 1: *QQ-plots of Internal and External Motivation to Control Stereotyping Compared by Treatment Mode.* A bootstrap Kolmogorov-Smirnov test of the equality of distributions across treatment modes for the internal motivation measure yields a test statistic of $D = 0.29$ and a corresponding p-value of 0.90. A similar test of the equality of distributions across treatment modes for the external motivation measure yields a test statistic of $D = 0.30$ and a corresponding p-value of 0.88.

1 Motivation to Control Stereotyping by Treatment Mode

Because we measure the key moderator (motivation to control stereotyping) after the subjects are given their six choice tasks, one might be concerned that the treatment mode (photos vs. labels) might exert a causal effect on the moderator. Table 1 presents summary statistics for the internal motivation and external motivation to control stereotyping variables broken down by treatment mode. Here we see that the medians are identical and that the means are not significantly different across treatment modes. This gives us some reassurance that treatment mode is not exerting an effect our measures of motivations to control stereotyping.

	Photo Median	Labels Median	Photos Mean	Labels Mean	t-stat.	p-value
Internal Motiv.	37	37	35.29	34.85	1.01	0.31
External Motiv.	18	18	18.29	18.71	-0.87	0.38

Table 1: *Motivation to Control Stereotyping by Treatment Mode.* The t-stat. column gives the t-statistic for a test of the null hypothesis of the equality of the means of the relevant measure of motivation to control stereotyping in the labels arm and the photos arm of the experiment. The p-value column gives the corresponding p-value for that two-sample t-test.

We can also compare the full distributions of the motivation to control stereotyping variables across treatment modes. Figure 1 presents QQ-plots that make these comparisons. Using a bootstrap Kolmogorov-Smirnov test we again do not see statistically significant differences between the

relevant distributions.

2 Non-Response by Treatment Mode

We may also be concerned that the treatment mode assigned to subject has an impact on that subject's willingness to answer subsequent questions—either the six choice questions given to each subject or the motivation to control stereotyping questions. Tables 2, 3, 4, and 5 all present information on the fraction of missing responses by treatment mode. In no case are there statistically significant differences in the fractions of missingness across treatment modes.

	Choice Not Missing	Choice Missing
Photo Arm	4723	185
Labels Arm	4611	183

Table 2: *Non-Response to the Choice Questions by Treatment Mode (Aggregated Across All Six Matchups).* The χ^2 test for the equality of proportions (with continuity correction) yields a test statistic of $\chi^2 = 0.005$ and an associated p-value of 0.94.

	First	Second	Third	Fourth	Fifth	Sixth
Photo Arm Choice Observed	791	787	788	791	784	782
Photo Arm Choice Missing	27	31	30	27	34	36
Labels Arm Choice Observed	767	766	771	765	772	770
Labels Arm Choice Missing	32	33	28	34	27	29

Table 3: *Non-Response to the Choice Questions by Treatment Mode and Matchup Position (First Through Sixth).* The χ^2 test for the equality of the 12 missingness proportions (without continuity correction) yields a test statistic of $\chi^2 = 3.68$ and an associated p-value of 0.98.

	Int. Motiv. Not Missing	Int. Motiv. Missing
Photo Arm	802	16
Labels Arm	779	20

Table 4: *Non-Response to the Internal Motivation to Control Stereotyping Questions by Treatment Mode.* The χ^2 test for the equality of proportions (with continuity correction) yields a test statistic of $\chi^2 = 0.33$ and an associated p-value of 0.56.

	Ext. Motiv. Not Missing	Ext. Motiv. Missing
Photo Arm	801	17
Labels Arm	778	21

Table 5: *Non-Response to the External Motivation to Control Stereotyping Questions by Treatment Mode.* The χ^2 test for the equality of proportions (with continuity correction) yields a test statistic of $\chi^2 = 0.32$ and an associated p-value of 0.57.

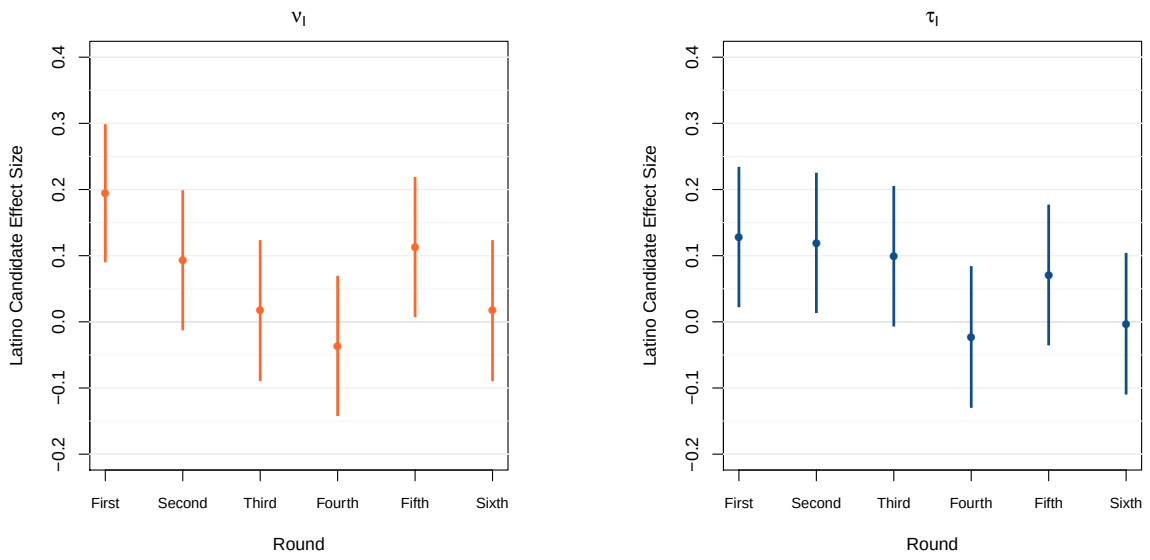

Figure 2: *Estimates of ν_I and τ_I by Round.* Recall that ν_I is the Latino ethnicity effect within the photos arm of the experiment for the subset of respondents that we label as internally motivated to control stereotyping and that τ_I is the Latino ethnicity effect within the labels arm of the experiment for the subset of respondents that we label as internally motivated to control stereotyping. We can reject the null hypothesis of no difference in ν_I across the six matchup at the 0.02 level. We are not able to reject the null of no difference in τ_I across the six matchups at conventional levels ($p = 0.17$).

3 Effect Estimates by Round

As noted in the main body of the paper, there was one situation in which we found statistically significant differences in the Latino ethnicity effect across the six matchups. This was within the group of respondents who are internally motivated to control stereotyping and who were exposed to the photos arm of the experiment. The corresponding estimand is what we labeled ν_I. The p-value from a Wald test of the equality of the six effects was 0.02.

In Figure 2 we plot the estimates of ν_I broken down by the chronological order of the matchups. What we see is an initial, very strong, positive effect of Latino ethnicity in the first matchup seen by respondents and then diminishing effects in most later matchups.

To provide a point of comparison, Figure 2 also plots the effects by round for the respondents who were internally motivated to control stereotyping but who were exposed to the labels arm of the experiment. The estimand here is what we labeled τ_I. The p-value from a Wald test of the equality of the six effects was 0.17.

4 Estimating Population Quantities

In response to a request from a referee, in this section we report estimates of the population counterparts to the in-sample estimands that we define in the pre-analysis plan and the paper. Since we have the survey weights from GfK, we can use these weights to construct estimates of the population versions of the full sample estimands τ, ν, and δ. We do not have the appropriate

weights to estimate population versions of subgroup-specific estimands such as τ_I, etc. Further, we cannot calculate these weights ourselves because we do not know the distribution of the motivation to control stereotyping variables in the U.S. population.

Estimand	Estimate	SE	p-value
τ (Labels Treatment)	-0.056	0.018	0.001
ν (Photos Treatment)	0.002	0.016	0.900
$\tau - \nu$	-0.058	0.024	0.015

Table 6: Population-weighted results based on all 6 matchups. The p-value column gives the p-values from z-tests of the null hypothesis that the estimand in question is equal to 0.

Estimand	Estimate	SE	p-value
τ (Labels Treatment)	-0.058	0.038	0.120
ν (Photos Treatment)	0.100	0.037	0.007
$\tau - \nu$	-0.158	0.053	0.003

Table 7: Population-weighted results based on just the first matchup. The p-value column gives the p-values from z-tests of the null hypothesis that the estimand in question is equal to 0.

Estimand	Estimate	SE	p-value
τ (Labels Treatment)	-0.056	0.019	0.003
ν (Photos Treatment)	-0.018	0.018	0.326
$\tau - \nu$	-0.038	0.026	0.142

Table 8: Population-weighted results using data from matchups 2 through 6. The p-value column gives the p-values from z-tests of the null hypothesis that the estimand in question is equal to 0.