

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

Marisa A. Abrajano¹, Christopher S. Elmendorf² and Kevin M. Quinn³

¹ Department of Political Science, University of California, San Diego, La Jolla, CA 92093-0521, USA

² School of Law, University of California, Davis, Davis, CA 95616, USA

³ Department of Political Science, University of Michigan, Ann Arbor, MI 48109, USA. Email: kmq@umich.edu

Abstract

Does treatment mode matter in studies of the effects of candidate race or ethnicity on voting decisions? The assumption implicit in most such work is that such treatment mode differences are either small and/or theoretically well understood, so that the choice of how to signal the race of a candidate is largely one of convenience. But this assumption remains untested. Using a nationally representative sample of white voting-age citizens and a modified conjoint design, we evaluate whether signaling candidate ethnicity with ethnic labels and names results in different effects than signaling candidate ethnicity with ethnically identifiable photos and names. Our results provide strong evidence that treatment-mode effects are substantively large and statistically significant. Further, these treatment-mode effects are not consistent with extant theoretical accounts. These results highlight the need for additional theoretical and empirical work on race/ethnicity treatment-mode effects.

Keywords: experimental studies, survey experiments, treatment-mode effects, conjoint experiments

1 Introduction

Researchers conducting experiments about racial discrimination typically manipulate the apparent race or ethnicity of persons whom a decision-maker is asked to serve or choose among. *How* the researcher signals race/ethnicity varies considerably from study to study. Sometimes the treatments consist of real people, or photographs of people (Terkildsen 1993; Yinger 1995; Ondrich, Ross, and Yinger 2003; Pager 2007; Iyengar *et al.* 2010; Weaver 2012; Doleac and Stein 2013; Stephens 2013; Moehler and Conroy-Krutz 2016; Visalvanich 2016); sometimes the researcher employs racially or ethnically identifiable names (e.g., Bertrand and Mullainathan 2004; McConaughy *et al.* 2010; Butler and Broockman 2011; DeSante 2013; Agan and Starr 2016); and sometimes researchers use explicit labels, such as the words “African-American” and “White” (e.g., Hainmueller, Hopkins, and Yamamoto 2014; Hainmueller and Hopkins 2015; Sen 2015; see also Sigelman *et al.* 1995; Philpot and Walton 2007).

The tacit assumption underlying much of this work is that the form of the race/ethnicity treatment is not important. Names, photographs, and labels are largely interchangeable (but see Sen and Wasow 2016). The strong grip of this assumption is reflected in the lack of research on treatment-mode differences. To the best of our knowledge, not a single choice-object experiment in any social scientific discipline has investigated whether these treatment modes

Political Analysis (2018)

vol. 26:20–33

DOI: 10.1017/pan.2017.36

Corresponding author

Kevin M. Quinn

Edited by

Jonathan N. Katz

© The Author(s) 2018. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

Authors' note: We gratefully acknowledge support from Time-sharing Experiments for the Social Sciences (TESS) and the National Science Foundation (grant SES 16-59922). In addition, earlier versions of this paper benefited from comments from Ryan Enos, Don Green, Taeku Lee, Efrén Pérez, Nazita Lajevardi, Tyler Reny, and workshop participants at the USC Gould School of Law and the University of Michigan Political Science Department. Replication data and code can be found at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DFEH8S>.

make a difference.¹ There are reasons to expect the effects to diverge. Racial and ethnic labels seem likely to elicit cognitive reactions (“Do I, or should I, value the labeled characteristic positively or negatively, and to what extent?”), whereas pictures may elicit more automatic or emotional reactions (“I feel more comfortable with her”; or, “That guy looks untrustworthy.”). A substantial body of work in psychology and neuroscience suggests that whites who are committed in principle to antidiscrimination norms may nonetheless have aversive, gut-level reactions to racial and ethnic minorities (e.g., Blair and Banaji 1996; Kubota, Banaji, and Phelps 2012; Perez 2015). The same intuition undergirds political science research on racial campaign appeals, which posits and has often found that relatively subtle, “implicit” appeals to racial sentiments are more effective than “explicit” appeals (e.g., Mendelberg 2001; Huber and Lapinski 2006; Hutchings and Jardina 2009; McIlwain and Caliendo 2011).² One might therefore surmise that the treatment effect of race/ethnicity conveyed by images will be larger than the corresponding effect of written labels.

This conjecture—if borne out—would have important implications for the burgeoning body of research using conjoint experimental designs (e.g., Bechtel, Hainmueller, and Margalit 2015; Carlson 2015; Crowder-Meyer *et al.* 2015; Franchino and Zucchini 2015; Hainmueller, Hangartner, and Yamamoto 2015; Hainmueller and Hopkins 2015; Sen 2015; Bansak, Hainmueller, and Hangartner 2016; Bechtel, Genovese, and Scheve 2016; Brown *et al.* 2016; Carnes 2016; Gallego and Marx 2016; Horiuchi, Smith, and Yamamoto 2016; Ono and Yamada 2016; Wright, Levy, and Citrin 2016)—nearly all of which represent choice objects (typically persons) with a table of written attributes.³

We implement a vote-choice experiment informed by the standard conjoint design. Our respondents, a nationally representative sample of white voting-age citizens, evaluate six pairs of hypothetical candidates. One candidate in each pair is depicted as Latino and the other as white. The candidates in our experiment are designed to be roughly equally appealing in other respects. Candidate attributes apart from ethnicity are presented in the tabular form used in conjoint studies. We randomize which candidate in each pair is Latino, and, at the level of the respondent, whether candidate race/ethnicity is conveyed using names and a Race/Ethnicity row in the table of the candidate attributes, or names and photographs.

We hypothesized that the candidates depicted as Latino would receive less voter support when race/ethnicity was communicated using photographs as opposed to written labels, and we expected the differential treatment to be most pronounced among respondents who, by standard social-psychology metrics, are “internally motivated to control stereotyping” (Plant and Devine 1998).⁴ Such people try to avoid treating minorities badly because they think it is wrong. By contrast, people who are “externally motivated to control stereotyping” check themselves because they do not want others to think poorly of them.

Contrary to our main hypothesis, our findings reveal that white respondents who are internally motivated to control stereotyping gave almost exactly the same vote share to Latino candidates in the “labels” and “pictures” branches of our study. Yet respondents who are *not* so motivated chose the white candidate by landslide margins in the labels condition, while giving only slightly

- 1 There is a related body of work on racial campaign appeals in which researchers manipulate whether a campaign advertisement shows photographs of African Americans or refers in words to African Americans, but in these studies the dependent variable is usually the respondent’s policy preference, or candidate preference in an election between white candidates. The respondent is not asked to make a decision about the person whose apparent race or ethnicity is manipulated.
- 2 But see Valentino, Neuner, and Vandenbroek (2016) which finds that the differential effectiveness of implicit and explicit racial appeals has diminished.
- 3 This format is embodied in the software for conjoint experiments produced by Hainmueller, Hopkins and Yamamoto. Apart from our own work, Crowder-Meyer *et al.* (2015) is to our knowledge the only conjoint study that has represented race/ethnicity using pictures instead of labels.
- 4 Pretests on a convenience sample of Amazon Mechanical Turk subjects corroborated this hypothesis when the minority candidate was presented as Latino, but our pretests revealed no differential treatment effect when the minority candidate was presented as Black.

more support to white candidates than Latino candidates in the pictures condition. Overall, Latino candidates fared significantly worse when their ethnic identity was conveyed via labels rather than pictures. These findings confirm that treatment-mode differences do exist where the focal attribute is race/ethnicity, and reinforce the need for additional research on the subject matter.

The hypotheses and principal data analysis presented in this paper follow a preanalysis plan registered with Evidence in Governance and Politics prior to fielding the experiment.⁵ See Abrajano, Elmendorf, and Quinn (2017) for replication data.

2 Study Design

Our study builds on Hainmueller, Hopkins, and Yamamoto (2014)'s conjoint design for stated-preference experiments. Respondents are presented with six candidate matchups. Each candidate is defined by five discrete-valued attributes: Race/Ethnicity, Education, Military Service, Political Party, and Other Information. The choice task is, "If you were voting in this election, which candidate do you think you would prefer?"

Respondents are randomly assigned to the "labels" or "pictures" branch of the study. In the labels condition, Race/Ethnicity is very explicitly signaled. It is presented as a row in the table of candidate attributes, with labels "Latino/Hispanic" and "White." In the pictures branch, the Race/Ethnicity row is omitted, and a photograph of the candidate is included at the top of his column in the table. In both branches, the candidates are referred to by name, and the names are ethnically identifiable.

We include ethnically identifiable names in both branches of the study for realism and to minimize the risk of differential treatment effects due to variation in the clarity of the signal of ethnicity. Although the photographs we employ are meant to be ethnically unambiguous (more on this below), we thought some respondents might fail to perceive ethnicity from photographs alone. We use the same ethnically identifiable names in both branches of the study, so that any differential effect of the pictures treatment relative to labels can be interpreted as the marginal effect of pictures, rather than the compound effect of pictures-plus-names relative to labels-without-names. Pretests with Amazon Mechanical Turk ("MTurk") workers yielded similar, very high rates of "correct" classification of ethnicity for both the white-photo-plus-name and the Latino-photo-plus-name profiles in our study.⁶

The screen shots in Figures 1 and 2 show how we presented candidates to respondents in, respectively, the labels and pictures conditions.

In pretests with inexpensive MTurk samples, we followed Hainmueller, Hopkins, and Yamamoto (2014) and randomized all candidate attributes. The fully randomized design is, however, a low-power design for detecting differences between pictures and labels Race/Ethnicity treatments. It is often the case that one candidate in a pair ends up being much more desirable than the other candidate on non-ethnic grounds. These differences dampen the average treatment effect of race/ethnicity and introduce additional estimation uncertainty.

Before fielding our study on a nationally representative sample of voting-age citizens, we modified the standard conjoint design to obtain more power. Using data from MTurk pretests, we preselected six pairs of candidate profiles, fixing all attribute levels other than Race/Ethnicity for each profile. The profiles were chosen so that each candidate in a pair would be roughly equally attractive in terms of his valence traits. To ensure that party effects would not swamp

⁵ <http://egap.org/registration/2027>

⁶ We asked MTurk respondents to classify the candidates in the final matchup by ethnicity. 93.8% of the Latino-candidate observations were correctly classified, and 95.6% of the white-candidate observations were correct. (Respondents in the pretest were shown an even mix of black, Latino, and white candidates. The response options were "Black," "White," "Latino/Hispanic," and "unsure." About 2% of the white-candidate observations and 3% of the Latino-candidate observations were "unsures.") We are therefore fairly confident that any difference between the treatment effects of labels and pictures is not due to imprecision in the signal of ethnicity conveyed by the pictures.

The following candidates are running for **City Council**:

Candidate	Carlos Gonzalez	Matthew Olson
Military Service	No military service	Served in military; honored for valor in combat
Education	Graduated from an Ivy League university	Did not graduate from college
Political Party	Democratic Party	Democratic Party
Race / Ethnicity	Latino/Hispanic	White
Other Information	Was the starting quarterback for college football team	Volunteers with parent-teacher association

If you were voting in this election for city council, which candidate do you think you'd vote for?

Select one answer only

- ☐ Carlos Gonzalez
- ☐ Matthew Olson

Figure 1. Presentation of candidate profiles to respondents: “labels” condition.

The following candidates are running for **City Council**:

Candidate	Ramon Cervantes	Andy Schaefer
		
Military Service	Served in military; honored for valor in combat	Served in military
Education	Did not graduate from college	Graduated from a state university
Political Party	Republican Party	Republican Party
Other Information	Honored as whistle-blower for exposing corruption at a public agency	Volunteers with parent-teacher association

If you were voting in this election for city council, which candidate do you think you'd vote for?

Select one answer only

- ☐ Ramon Cervantes
- ☐ Andy Schaefer

Figure 2. Presentation of candidate profiles to respondents: “pictures” condition.

race effects, we also made all matchups intraparty—Democrat vs. Democrat or Republican vs. Republican.⁷ This modified design retains what is for our purpose the most important feature

7 We told respondents that the candidates were running in a city council election, and that in city council elections it is sometimes occurs that the leading candidates are both affiliated with the same political party.

of the conjoint setup: the candidates in a pair differ from one another in several respects other than Race/Ethnicity, such that the respondent's choice of (say) the white candidate over the Latino candidate does not *necessarily* mean that the respondent preferred the chosen candidate on the basis of ethnicity. This should tend to limit social desirability effects.⁸

For each respondent, we randomized (1) the order of presentation of the six candidate pairs, (2) which candidate in each pair is shown in the left column of the table of candidate profiles, (3) which candidate in each pair is Latino (the other candidate is always white), and (4) conditional on candidate ethnicity, the assignment of names to candidates. In the pictures branch of the survey, we also randomized the assignment of photographs to candidates, again conditional on candidate ethnicity. Finally, in the labels branch of the survey, we randomized at the level of the respondent the position of the Race/Ethnicity row in the table of candidate attributes.

After voting on the six candidate matchups, respondents answered a few demographic questions, and a battery of ten questions developed by social psychologists to measure internal and external motivations to control stereotyping of racial and ethnic minorities (Plant and Devine 1998).⁹ We considered several more general measures of social sensitivity, but because there appears to be no consensus in the political science or psychology literature about the best such measure, we opted for a measure that targets racial/ethnic stereotyping. The distributions of the motivation-to-control-stereotyping variables, summarized in the Supplementary Information available at <https://doi.org/10.1017/pan.2017.36>, are similar across the photos-and-labels treatment arms, alleviating possible concerns about posttreatment bias.¹⁰

2.1 Choice of photos and names

We designed a photo-selection protocol to obtain several typical yet clearly ethnically identifiable photos of serious candidates, while minimizing researcher discretion. Choosing photos presents a dilemma. On the one hand, there is a risk that the white and Latino photos will differ from one another in some vote-relevant dimension apart from ethnicity, such as perceived attractiveness or seriousness. On the other hand, if the minority is negatively stereotyped, precoding photographs for similarity on such dimensions may result in a minority treatment condition that includes only visually extraordinary candidates.

The best solution would be to use photographs drawn from the pool of “potential candidates,” defined as people who might run for office in a world without race discrimination, but that population is unobservable. We use an imperfect but observable proxy: currently serving state legislators.

We hired MTurk workers to precode male state legislator photographs for age, race/ethnicity, and quality as a campaign photo.¹¹ We asked for campaign-photo-quality ratings as a way to minimize the risk of imbalance with respect to non-racial signals about professionalism, seriousness, or competence communicated by the photographs.¹² Each image was rated by an

8 On the other hand, because all matchups feature a Latino candidate running against a white candidate of the same political party, some respondents may think that choosing one candidate over the other is tantamount to disclosing their ethnic preference.

9 The questions were originally developed to address stereotyping of and discrimination against blacks. We used the original question wording but replaced “Blacks” with “Latinos/Hispanics.” We also randomized the order of the ten motivation-to-control-stereotyping questions.

10 In a perfect world, we would have asked the motivation-to-control-stereotyping questions in a follow-up survey several weeks or months after the experiment, so as to minimize concerns about the treatments possibly affecting responses to the motivation questions. But a two-wave design would have been more costly.

11 Photos were downloaded from Project VoteSmart's website. We limit our study to male legislators because we do not want gender to confound our results, and because the greater number of male state legislators makes it easier to obtain matches across racial groups. More on this below.

12 Like judgments of likability or attractiveness, judgments of photo quality are made by coders who are aware of the apparent race of the person in the photo, so there is some risk of bias from coders' racial stereotypes. But we think the risk of bias is less severe when the coder is asked to evaluate an attribute of the picture itself, rather than an attribute of the person in the picture.

average of 37 MTurk workers. On the basis of the codings we excluded very young and very old legislators, as well as legislators who were not clearly identified as Latino or white. We then used the GenMatch package in R to find white matches for the remaining Latinos, matching on perceived age and photo quality and winnowing the pool down to eight well-matched whites and Latinos (16 photos in all).

To select the names, we matched the surnames of current state legislators to the 1,000 most common surnames in the United States, excluded non-matches, and selected from the remaining surnames the eight most ethnically unique Anglo and Latino names.¹³ We then created two pools of presumptively ethnic first names, selecting for each pool the first names of all state legislators whose surnames matched one of the eight we had chosen for the ethnic group in question.¹⁴ We randomly assigned first names to surnames (within ethnic groups), subject to the constraint that no first-name/surname pair match an actual legislator's name. This protocol was designed to ensure that the names were ethnically identifiable without also being highly improbable for a serious candidate for elective office.

We then administered a conjoint survey to MTurk workers, using photographs to represent the candidates, and after the final matchup, we asked respondents which candidate in that matchup had the higher quality campaign photo. This choice task departs from our photo prescreening surveys in a couple of respects: respondents make relative rather than absolute judgments of photo quality, and each picture comes with a candidate profile including an ethnically identifiable name. We discovered that Latino candidates' photos were picked as higher quality somewhat less often than white candidates' photos. Though this might be a race-discrimination effect, it could also reflect a problem with our prescreening survey.¹⁵ To assuage concerns that any race-treatment effect documented in our study might really be a photo-quality treatment effect, we dropped two relatively unfavorable Latino photos and two relatively favorable white photos. This restores balance in average photo quality across racial groups, per the "matchup" measure of photo quality.

Two potential sources of bias should be acknowledged. First, because we used only clearly ethnically identifiable photos, the Latino candidates shown in the photos condition may be darker-skinned and more stereotypically Latino in appearance than the Latinos in the unobserved potential-candidate population. To the extent that discrimination varies with phenotypic prototypicality, our estimated "pictures" treatment effects may overstate the average level of discrimination against a more phenotypically representative sample. Cutting the other way, if Latinos seeking elective office do face discrimination, the Latinos elected to a state legislature may be visually extraordinary relative to the unobserved population of potential Latino candidates. If so, our estimated "pictures" treatment effect of Latino ethnicity is probably upwardly biased. However, these potential sources of bias are unlikely to affect our investigation of the *difference* in the pictures vs. labels treatment effects across respondents who are/are not internally motivated to control stereotyping. We have no *a priori* reason to think that voters in one of these groups would value extraordinary (versus ordinary) Latino potential candidates differently than voters in the other group.

2.2 The experimental population, and relevant subgroups

Our experimental population consists of a probability sample of 1,617 white voting-age citizens, provided by GfK. Constructed using address-based sampling of U.S. Postal Service files, GfK's

¹³ We used ethnic uniqueness ratings derived from the 2000 census by Word *et al.* (2008).

¹⁴ As the first name, we used the "preferred name" field in the Project VoteSmart database.

¹⁵ For example, measurement error may be greater when respondents are asked to rate campaign photos in isolation from one another, rather than indicating which of two pictures is the higher quality campaign photo.

panel is generally regarded as among the highest quality for online surveys.¹⁶ Our survey was fielded from August 16, 2016 to August 22, 2016. A random sample of 2,666 panel members was drawn from GfK's KnowledgePanel®. A total of 1,617 (excluding breakoffs) members responded to the invitation and 1,617 qualified for the survey, yielding a final stage completion rate of 60.7% and a qualification rate of 100.0%.¹⁷

Following Plant and Devine (1998), we create additive indices for internal and external motivation-to-control stereotyping, summing responses to the five Likert-type questions for each motivation. We use median splits to classify respondents as “high” or “low” for each motivation.

The results reported below are unweighted, as the estimands of interest are sample average treatment effects (Franco *et al.* 2016).

3 Notation and Estimands

Let $i \in \mathcal{I}$ index survey respondents. Three subsets of respondents are of interest to us. First is the set of respondents who are internally motivated to control stereotyping of Latinos/Hispanics. We use $\mathcal{I}_I$ to denote this subset of respondents. Second we are interested in respondents who are not internally motivated to control stereotyping of Latinos/Hispanics. These are simply all respondents in $\mathcal{I}$ who are not in $\mathcal{I}_I$. We use $\mathcal{I}_{\setminus I}$ to denote this set of respondents. Third, we are interested in those respondents who are externally, but not internally, motivated to control stereotyping of Latinos/Hispanics.¹⁸ We use $\mathcal{I}_{E \setminus I}$ to denote this subset of respondents.

Let $p \in \mathcal{P}$ index pairs of candidate profiles. Each candidate profile is defined by these attributes (apart from race/ethnicity): military service, education, political party, and other information. As noted above, attribute levels for the profiles in each pair were selected to make the profiles roughly equally appealing to respondents. We make use of six pairs of candidates in this experiment, i.e. $|\mathcal{P}| = 6$. We adopt the convention of referring to one of the candidates in a pair as “candidate A” and the other as “candidate B.”

We consider two manipulations, which are randomly assigned to respondents and candidate pairs. The first intervention is the depiction of candidate ethnicity via textual labels (“Latino/Hispanic” or “White”) and first and last names. We use T to denote this first textual manipulation. Let T_{ip} denote the manipulation applied to respondent i when viewing candidate pair p . T_{ip} can take two values: 0 and 1:

$$T_{ip} = \begin{cases} 1 & \text{if } i \text{ receives textual information that candidate A in } p \text{ is Latino/Hispanic and} \\ & i \text{ receives textual information that candidate B in } p \text{ is white} \\ 0 & \text{if } i \text{ receives textual information that candidate A in } p \text{ is white and} \\ & i \text{ receives textual information that candidate B in } p \text{ is Latino/Hispanic.} \end{cases}$$

The second intervention is the depiction of candidate ethnicity via photographs and first and last names. We use V to denote this second visual manipulation and let V_{ip} denote the manipulation applied to respondent i when viewing candidate pair p . V_{ip} can take two values: 0 and 1:

¹⁶ For details on the sampling methodology, see

https://www.gfk.com/fileadmin/user_upload/dyna_content/US/documents/KnowledgePanel_Methodology.pdf.

¹⁷ The recruitment rate for this study, reported by GfK, was 14.5% and the profile rate was 63.6%, for a cumulative response rate of 5.6%.

¹⁸ We operationalize this as being above the sample median on the external motivation-to-control-stereotyping index of Plant and Devine (1998) but below the sample median on the internal motivation-to-control-stereotyping index.

$$V_{ip} = \begin{cases} 1 & \text{if } i \text{ receives visual information that candidate } A \text{ in } p \text{ is Latino/Hispanic and} \\ & i \text{ receives visual information that candidate } B \text{ in } p \text{ is white} \\ 0 & \text{if } i \text{ receives visual information that candidate } A \text{ in } p \text{ is white and} \\ & i \text{ receives visual information that candidate } B \text{ in } p \text{ is Latino/Hispanic.} \end{cases}$$

Our outcome variable is the revealed preference for either candidate A or B in each candidate pair by each respondent. Using potential outcomes notation, we let $Y_{ip}(T_{ip} = t)$ denote the choice of respondent i when viewing candidate pair p and T_{ip} is set equal to t by our intervention. We adopt the convention that $Y_{ip}(T_{ip} = t) = 1$ denotes a choice of candidate A and $Y_{ip}(T_{ip} = t) = 0$ is a choice of candidate B . We define $Y_{ip}(V_{ip} = v)$ analogously. We let Y_{ip} denote the observed choice behavior of respondent i when evaluating candidate pair p and let $Y_{ip} = 1$ denote a choice of A and $Y_{ip} = 0$ a choice of B .

We are interested in the following sample average treatment effects.

$$\tau = \frac{1}{|I|} \frac{1}{|\mathcal{P}|} \sum_{i \in I} \sum_{p \in \mathcal{P}} [Y_{ip}(T_{ip} = 1) - Y_{ip}(T_{ip} = 0)]$$

and

$$\nu = \frac{1}{|I|} \frac{1}{|\mathcal{P}|} \sum_{i \in I} \sum_{p \in \mathcal{P}} [Y_{ip}(V_{ip} = 1) - Y_{ip}(V_{ip} = 0)].$$

The sample average treatment effect τ is the effect of candidate ethnicity conveyed by textual labels on choice behavior. It can be interpreted as the expected change in the vote margin of a white candidate matched up against a copartisan Latino opponent (of similar quality) that would occur if the candidates' ethnicities were transposed while holding constant the other candidate attributes, that is, if the white candidate were Latino and the Latino candidate were white. A negative value of τ implies that voters, on average, disfavor Latino/Hispanic candidates relative to white candidates when the ethnicity of candidates is conveyed via textual labels. The interpretation of ν is similar except for the fact that the intervention here is the provision of photographic information regarding candidate ethnicity.

We are also interested in the analogous effects defined for subsets of I corresponding to those who are internally motivated to control stereotyping of Latinos/Hispanics:

$$\tau_I = \frac{1}{|I_I|} \frac{1}{|\mathcal{P}|} \sum_{i \in I_I} \sum_{p \in \mathcal{P}} [Y_{ip}(T_{ip} = 1) - Y_{ip}(T_{ip} = 0)]$$

and

$$\nu_I = \frac{1}{|I_I|} \frac{1}{|\mathcal{P}|} \sum_{i \in I_I} \sum_{p \in \mathcal{P}} [Y_{ip}(V_{ip} = 1) - Y_{ip}(V_{ip} = 0)]$$

those not internally motivated to control stereotyping of Latinos/Hispanics:

$$\tau_{\setminus I} = \frac{1}{|I_{\setminus I}|} \frac{1}{|\mathcal{P}|} \sum_{i \in I_{\setminus I}} \sum_{p \in \mathcal{P}} [Y_{ip}(T_{ip} = 1) - Y_{ip}(T_{ip} = 0)]$$

and

$$\nu_{\setminus I} = \frac{1}{|I_{\setminus I}|} \frac{1}{|\mathcal{P}|} \sum_{i \in I_{\setminus I}} \sum_{p \in \mathcal{P}} [Y_{ip}(V_{ip} = 1) - Y_{ip}(V_{ip} = 0)]$$

and those who are externally motivated to control stereotyping of Latinos/Hispanics but not internally motivated:

$$\tau_{E \setminus I} = \frac{1}{|I_{E \setminus I}|} \frac{1}{|\mathcal{P}|} \sum_{i \in I_{E \setminus I}} \sum_{p \in \mathcal{P}} [Y_{ip}(T_{ip} = 1) - Y_{ip}(T_{ip} = 0)]$$

and

$$\nu_{E \setminus I} = \frac{1}{|I_{E \setminus I}|} \frac{1}{|\mathcal{P}|} \sum_{i \in I_{E \setminus I}} \sum_{p \in \mathcal{P}} [Y_{ip}(V_{ip} = 1) - Y_{ip}(V_{ip} = 0)].$$

4 Hypotheses

Our first primary hypothesis is that $(\tau - \nu) > 0$. In words, candidate profiles assigned Latino ethnicity are expected to be chosen more often when ethnicity is represented textually rather than visually. This motivating assumption is that some “ethnically ambivalent” respondents will choose the Latino candidate when his ethnicity is conveyed with labels (triggering cognitive processing), but not when his ethnicity is conveyed via photographs, which are expected to induce more of a gut-level reaction.

Our second primary hypothesis is that $(\tau_I - \nu_I) > (\tau_{\setminus I} - \nu_{\setminus I})$. In words, the difference between the label and photo variants of the Latino-ethnicity treatment effect is expected to be larger for respondents who are more-internally motivated to control stereotyping of Latinos.

Our third primary hypothesis is that $(\tau_{E \setminus I} - \nu_{E \setminus I}) = 0$. That is, among respondents who are motivated to control stereotyping to maintain their reputation, but not for internal reasons, we expect the label and picture versions of the Latino-ethnicity treatment to have the same effect. These respondents are not thought to experience dissonance between their gut-level and cognitive reactions to Latino ethnicity, dissonance that in other respondents may cause the treatment effects of labels and pictures to diverge. To be sure, externally motivated respondents may be reluctant to say publicly they prefer candidates labeled white to roughly similar candidates labeled Latino, but in the context of an anonymous online survey—especially where the candidate profiles list several attributes other than Race/Ethnicity—we expect no dissembling from these respondents, and thus similar treatment effects from labels and pictures.

A fourth, exploratory hypothesis is that $\tau < 0$. That is, Latino ethnicity is dispreferred in biethnic matchups between copartisan candidates of similar quality when ethnicity is labeled.

A fifth, exploratory hypothesis is that $\nu < 0$. That is, Latino ethnicity is dispreferred in biethnic matchups between copartisan candidates of similar quality when ethnicity is represented visually.

5 Analysis and Results

Given the random assignment of the manipulations to respondents and candidate pairs, unbiased estimators of all estimands can be constructed from sample averages. Consequently, we test the above hypotheses using linear regression with clustered standard errors in a manner consistent with Hainmueller, Hopkins, and Yamamoto (2014).

To maximize the power of our tests, we would like to pool across the six choice tasks given to each respondent, on the assumption that respondents' preference as between any two candidate profiles is independent of where that matchup occurs in the sequence of six matchups presented to the respondent. It turns out, however, that the behavior of one class of respondents (those who are internally motivated to control stereotyping) assigned to one treatment arm (the photos condition) varies significantly across the six matchups. See the Supplemental Information for details. For now, we report the tests of our primary hypotheses in three ways: first, pooling across all six matchups (see Table 1); second, using data only from the first matchup (see Table 2); and

Table 1. 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)	-0.059	0.017	<0.001
ν (Photos Treatment)	-0.002	0.016	0.882
$\tau - \nu$	-0.056	0.023	0.014
τ_I	0.065	0.022	0.003
ν_I	0.066	0.022	0.003
$\tau_{\setminus I}$	-0.167	0.023	<0.001
$\nu_{\setminus I}$	-0.060	0.022	0.006
$(\tau_I - \nu_I) - (\tau_{\setminus I} - \nu_{\setminus I})$	0.106	0.045	0.018
$\tau_{E \setminus I}$	-0.227	0.029	<0.001
$\nu_{E \setminus I}$	-0.092	0.031	0.003
$\tau_{E \setminus I} - \nu_{E \setminus I}$	-0.135	0.042	0.001

Table 2. 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)	-0.048	0.037	0.189
ν (Photos Treatment)	0.082	0.036	0.021
$\tau - \nu$	-0.130	0.051	0.011
τ_I	0.128	0.053	0.016
ν_I	0.194	0.052	<0.001
$\tau_{\setminus I}$	-0.203	0.049	<0.001
$\nu_{\setminus I}$	-0.012	0.049	0.808
$(\tau_I - \nu_I) - (\tau_{\setminus I} - \nu_{\setminus I})$	0.125	0.102	0.218
$\tau_{E \setminus I}$	-0.272	0.064	<0.001
$\nu_{E \setminus I}$	-0.062	0.067	0.350
$\tau_{E \setminus I} - \nu_{E \setminus I}$	-0.209	0.093	0.024

finally, for completeness, using data from matchups two through six (see Table 3). By construction, choices made in the first matchup are free of spillover effects from the later matchups.

Our results contradict our first hypothesis, namely, that Latino candidates are more likely to be chosen when ethnicity is conveyed by labels and names rather than pictures and names ($\tau - \nu > 0$). We find instead that being Latino is a significant disadvantage when ethnicity is represented by labels, but irrelevant when ethnicity is represented through photos. The difference between the labels and pictures treatments is statistically significant at the 0.014 level. If we look only at data from the first matchup, we see that $(\tau - \nu)$ is again estimated to be negative. This difference is significantly different from 0 at the 0.011 level. In the first matchup τ is not significantly different from 0 at conventional levels while ν is significantly different from 0.

Our second primary hypothesis concerns the relative size of the labels/pictures treatment gap, for respondents who are “high” and “low” in internal motivation-to-control stereotyping: $(\tau_I - \nu_I) > (\tau_{\setminus I} - \nu_{\setminus I})$. When we pool the data for statistical power this hypothesis is confirmed ($p = 0.018$), but not in the way we expected.¹⁹ We hypothesized that $(\tau_I - \nu_I)$ would be positive, in line with our rationale for our first hypothesis. But as Table 1 shows, this expression is

¹⁹ If we use only data from the first matchup, $(\tau_I - \nu_I)$ is estimated to be larger than $(\tau_{\setminus I} - \nu_{\setminus I})$, but the difference is not statistically significant. See Table 2.

Table 3. 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.

Estimand	Estimate	SE	p -value
τ (Labels Treatment)	-0.061	0.018	0.001
ν (Photos Treatment)	-0.019	0.017	0.268
$\tau - \nu$	-0.041	0.025	0.099
τ_I	0.053	0.025	0.038
ν_I	0.041	0.025	0.103
$\tau_{\setminus I}$	-0.160	0.024	<0.001
$\nu_{\setminus I}$	-0.070	0.024	0.004
$(\tau_I - \nu_I) - (\tau_{\setminus I} - \nu_{\setminus I})$	0.103	0.049	0.038
$\tau_{E \setminus I}$	-0.218	0.031	<0.001
$\nu_{E \setminus I}$	-0.098	0.034	0.004
$\tau_{E \setminus I} - \nu_{E \setminus I}$	-0.120	0.046	0.009

estimated to be almost exactly equal to zero, while $(\tau_{\setminus I} - \nu_{\setminus I})$ is strongly negative. Transposing the candidates' ethnicities (candidate A from white to Latino, and candidate B from Latino to white) *increases* candidate A 's vote share among the more-internally motivated half of the sample by 6.5 percentage points regardless of whether ethnicity is conveyed by pictures or labels. The same transposition of racial/ethnic cues *reduces* candidate A 's vote share among the less-internally motivated half of the sample by about the same margin when ethnicity is signaled via photos. Yet among these less-internally motivated respondents the reduction in support for candidate A under the counterfactual scenario that he is Latino is *2.5 times as large* (a 16.7 percentage-point vs. a 6.6 percentage-point shock to his vote margin) when ethnicity is conveyed with labels rather than photos. This result confirms our expectation that variations in the way the racial/ethnic cue is signaled can lead to substantively different outcomes, though not for the specific group we hypothesized.

Our third primary hypothesis concerns the behavior of respondents whose gut-level and cognitive reactions to Latino ethnicity were expected to be similar, yet who acknowledge misrepresenting their ethnic preferences in public settings—those who are externally but not internally motivated to control stereotyping. Given the anonymous, online nature of our survey, and the fact that the candidate profiles featured several attributes other than ethnicity, we expected no dissembling, i.e., $(\tau_{E \setminus I} - \nu_{E \setminus I}) = 0$.

Looking at the results in Table 1 we see that both $\tau_{E \setminus I}$ and $\nu_{E \setminus I}$ are estimated to be negative and significantly different from 0 at conventional levels. However, the effect of conveying racial/ethnic cues via labels is about 2.5 times larger than the photos effect (a 22.7 percentage-point change in vote share vs. a 9.2 percentage-point change in vote share). The difference between these two effects is significantly different from 0 at the 0.001 level. Many of our more socially sensitive respondents were clearly willing to express their preference for white candidates in the labels condition. And rather surprisingly, these same respondents were not equally hostile to Latinos when ethnicity was conveyed visually rather than textually.

6 Discussion

Our results defy conventional wisdom about white voters' responses to racial/ethnic images and written labels. The absence of race/ethnicity labels in the pictures condition makes this treatment more implicit than the other, so we expected a more negative Latino-ethnicity treatment effect in this condition, especially among respondents who are internally motivated to control stereotyping. What we found suggests that internally motivated respondents have

the same reaction to photographic and labeled signals of ethnicity, but that respondents who are not so motivated react much more strongly—and more negatively—to the Latino labels-and-names treatment than to the corresponding photos-and-names treatment. Pooling across all respondents and all six matchups, Latino ethnicity was not a statistically significant disadvantage unless it was labeled ($\nu = -0.002$ with $p = 0.882$; $\tau = -0.059$ with $p < 0.001$).

What could explain the attenuation of discrimination against Latinos in the photos condition among these voters whom we categorize as not internally motivated to control stereotyping (the respondents in $I_{\setminus I}$ and $I_{E \setminus I}$)? One possibility is that the Latino candidates depicted in the treatment photos are visually extraordinary relative to the white candidates depicted in the photos. As we explained above, selection effects in a world in which most voters discriminate against Latino candidates could result in Latino state legislators being visually exceptional relative to white state legislators. (Recall that our treatment photos are drawn from the population of state legislators.) As a further check, we had MTurk workers rate the treatment photos for attractiveness. It turns out that our Latino and white photos are well-balanced on this dimension, except for one Latino who appears significantly *less* attractive than he “should” be (for balance). So any bias from the correlation between attractiveness and ethnicity in the treatment photos should make more negative the estimated effect of Latino ethnicity in the photos condition.²⁰

Another possibility is that the photos-and-names treatment conveyed a less precise signal of ethnicity than the textual labels. But in our pretests, MTurk workers had no trouble correctly inferring candidate ethnicity in the pictures condition. The fuzzy-signal hypothesis is also hard to square with the behavior of internally motivated respondents, who gave the same bonus to Latino ethnicity in the labels and pictures conditions. If the signal were fuzzy in the pictures condition, the treatment effect of Latino identity represented visually should be biased toward zero among internally motivated respondents too.

Third, there may be something humanizing about the candidates’ faces. Perhaps discriminatory sentiments have less effect on vote choice when candidates are seen as real people rather than as abstractions.²¹

A final interpretation is that many white voters, even those predisposed to vote against Latino candidates on the basis of the candidate’s ethnicity, do not dwell on ethnicity unless it is explicitly primed. The 2016 presidential election engendered much discussion of “white identity politics,” and it may be the case that white voters at the low end of the motivation-to-control-stereotyping scale simply respond more negatively to minority race/ethnicity the more they are encouraged to consider it.²²

Whatever may prove to be best explanation for our results, they clearly do not support the idea that the choice of treatment mode (labels & names vs. photos & names) is of no consequence in studies of racial and ethnic discrimination. In this paper, we have made the case that researchers need to be cognizant of how race/ethnicity is conveyed in conjoint studies conducted in the United States. Scholars employing conjoint studies in contexts outside of the United States to study racial/ethnic politics would be wise to do the same.

While we have demonstrated that the way in which ethnicity is signaled matters, more work needs to be done to understand the mechanisms behind these treatment-mode effects. We would not be surprised if multiple mechanisms are at work—some of which may be contextually

²⁰ For reasons explained above, we do not think it is appropriate to adjust for attractiveness, because it is rated by respondents who may be reacting to Latino ethnicity.

²¹ Voters who are predisposed not to support a Latino candidate may also conjure negative mental images of Latinos when they see the word “Latino.” The actual image of a professional-looking Latino candidate may “outperform” these voters’ mental prototype of a Latino candidate, whereas the picture of a professional-looking white candidate does not. Thanks to Dan Klerman for suggesting this idea.

²² Note in this regard that White (2007) finds that black voters respond more strongly to explicit than implicit racial appeals.

dependent. We strongly encourage future researchers to explore the mechanisms underlying treatment-mode differences, such as those described in this study.

Supplementary material

For supplementary material accompanying this paper, please visit

<https://doi.org/10.1017/pan.2017.36>.

References

- Abrajano, Marisa, Elmendorf Christopher, and Quinn Kevin. 2017. Replication Data for: Labels vs. pictures: Treatment-mode effects in experiments about discrimination. doi:10.7910/DVN/DFEH8S, Harvard Dataverse, V1, UNF:6:abrrhxR2xkGTYmtBg5vQGw==.
- Agan, Amanda Y., and Sonja B. Starr. 2016. Ban the box, criminal records, and statistical discrimination: A field experiment. *U of Michigan Law & Econ Research Paper* (16-012). URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2795795.
- Bansak, Kirk, Jens Hainmueller, and Dominik Hangartner. 2016. How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers. *Science* aag2147.
- Bechtel, Michael M., Federica Genovese, and Kenneth Scheve. 2016. Interests, norms, and mass support for international climate policy. Available at SSRN 2528466.
- Bechtel, Michael M., Jens Hainmueller, and Yotam M. Margalit. 2015. Policy design and domestic support for international bailouts. Available at SSRN 2163594.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review* 94(4):991–1013.
- Blair, Irene V., and Mahzarin R. Banaji. 1996. Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology* 70(6):1142.
- Brown, Madeline M., John Carey, Yusaku Horiuchi, and Lauren K. Martin. 2016. Are University Communities Deeply Divided Over the Value of Diversity on Campus? An Application of Conjoint Analysis. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2775464.
- Butler, Daniel M., and David E. Broockman. 2011. Do politicians racially discriminate against constituents? A field experiment on state legislators. *American Journal of Political Science* 55(3):463–477.
- Carlson, Elizabeth. 2015. Ethnic voting and accountability in Africa: A choice experiment in Uganda. *World Politics* 67(02):353–385.
- Carnes, Nicholas. 2016. Keeping workers off the ballot. Working Paper.
- Crowder-Meyer, Melody, Shana Kushner Gadarian, Jessica Trounstein, and Kau Vue. 2015. Complex interactions: Candidate race, sex, electoral institutions, and voter choice. Working Paper.
- DeSante, Christopher D. 2013. Working twice as hard to get half as far: Race, work ethic, and America's deserving poor. *American Journal of Political Science* 57(2):342–356.
- Doleac, Jennifer L., and Luke C. D. Stein. 2013. The visible hand: Race and online market outcomes. *The Economic Journal* 123(572):F469–F492.
- Franchino, Fabio, and Francesco Zucchini. 2015. Voting in a multi-dimensional space: A conjoint analysis employing valence and ideology attributes of candidates. *Political Science Research and Methods* 3(02):221–241.
- Franco, Annie, Neil Malhotra, Gabor Simonovits, and L. J. Zigerell. 2016. Developing standards for post-stratification weighting in population-based survey experiments. Working Paper.
- Gallego, Aina, and Paul Marx. 2016. Multi-dimensional preferences for labour market reforms: A conjoint experiment. *Journal of European Public Policy* 24:1–21.
- Hainmueller, Jens, and Daniel J. Hopkins. 2015. The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants. *American Journal of Political Science* 59(3):529–548.
- Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2014. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis* 22(1):1–30.
- Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto. 2015. Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences* 112(8):2395–2400.
- Horiuchi, Yusaku, Daniel M. Smith, and Teppei Yamamoto. 2016. Identifying voter preferences for politicians' personal attributes: A conjoint experiment in Japan. Available at SSRN 2827969.
- Huber, Gregory A., and John S. Lapinski. 2006. The race card revisited: Assessing racial priming in policy contests. *American Journal of Political Science* 50(2):421–440.
- Hutchings, Vincent L., and Ashley E. Jardina. 2009. Experiments on racial priming in political campaigns. *Annual Review of Political Science* 12:397–402.

- Iyengar, Shanto, Solomon Messing, Jeremy Bailenson, and Kyu S. Hahn. 2010. Do explicit racial cues influence candidate preference? The case of skin complexion in the 2008 campaign. Paper presented at the 2010 Annual Meeting of the American Political Science Association.
- Kubota, Jennifer T., Mahzarin R. Banaji, and Elizabeth A. Phelps. 2012. The neuroscience of race. *Nature Neuroscience* 15:940–948.
- McConaughy, Corrine, Ismail White, David Leal, and Jason Casellas. 2010. A latino on the ballot: Explaining coethnic voting among latinos and the response of white Americans. *Journal of Politics* 34:571–584.
- McIlwain, Charlton, and Stephen M. Caliendo. 2011. *Race appeal. How candidates invoke race in U.S. political campaigns*. Philadelphia, PA: Temple University Press.
- Mendelberg, Tali. 2001. *The race card: Campaign strategy, implicit messages, and the norm of equality*. Princeton, NJ: Princeton University Press.
- Moehler, Devra, and Jeffrey Conroy-Krutz. 2016. Eyes on the ballot: Priming effects and ethnic voting in the developing world. *Electoral Studies* 42:99–113.
- Ondrich, Jan, Stephen Ross, and John Yinger. 2003. Now you see it, now you don't: Why do real estate agents withhold available houses from black customers? *Review of Economics and Statistics* 85(4):854–873.
- Ono, Yoshikuni, and Masahiro Yamada. 2016. Do Voters Prefer Gender Stereotypic Candidates?: Evidence from a Conjoint Survey Experiment in Japan. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2772966.
- Pager, Devah. 2007. The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future. *Annals of the American Academy of Political and Social Sciences* 1(609):104–133.
- Perez, Efrén O. 2015. *Unspoken politics: Implicit attitudes and political thinking*. Cambridge: Cambridge University Press.
- Philpot, Tasha S., and Hanes Walton. 2007. One of our own: Black female candidates and the voters who support them. *American Journal of Political Science* 51(1):49–62.
- Plant, E. Ashby, and Patricia G. Devine. 1998. Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology* 3(75):811–832.
- Sen, Maya. 2015. How political signals affect public support for judicial nominations: Evidence from a conjoint experiment. Working Paper.
- Sen, Maya, and Omar Wasow. 2016. Race as a bundle of sticks: Designs that estimate effects of seemingly immutable characteristics. *Annual Review of Political Science* 19:499–522.
- Sigelman, Carol K., Lee Sigelman, Barbara J. Walkosz, and Michael Nitz. 1995. Black candidates, white voters: Understanding racial bias in political perceptions. *American Journal of Political Science* 1(39):243–265.
- Stephens, LaFleur Nadiyah. 2013. The effectiveness of implicit and explicit racial appeals in a post-racial America. PhD thesis, University of Michigan.
- Terkildsen, Nayda. 1993. When white voters evaluate black candidates: The processing implications of candidate skin color, prejudice, and self-monitoring. *American Journal of Political Science* 37:1032–1053.
- Valentino, Nicholas A., Fabian G. Neuner, and L. Matthew Vandenbroek. 2016. The changing norms of racial political rhetoric and the end of racial priming. *The Journal of Politics*. doi:10.1086/694845.
- Visalvanich, Neil. 2016. Asian candidates in America: The surprising effects of positive racial stereotyping. *Political Research Quarterly* 70:1–24.
- Weaver, Vesla M. 2012. The electoral consequences of skin color: The hidden side of race in politics. *Political Behavior* 34(1):159–192.
- White, Ismail K. 2007. When race matters and when it doesn't: Racial group differences in response to racial cues. *American Political Science Review* 101(02):339–354.
- Word, David L., Charles D. Coleman, Robert Nunziata, and Robert Kominski. 2008. Demographic aspects of surnames from census 2000. Unpublished manuscript. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download>.
- Wright, Matthew, Morris Levy, and Jack Citrin. 2016. Public attitudes toward immigration policy across the legal/illegal divide: The role of categorical and attribute-based decision-making. *Political Behavior* 38(1):229–253.
- Yinger, John. 1995. *Closed doors, opportunities lost: The continuing costs of housing discrimination*. New York: Russell Sage Foundation.