# Conjoint Experiments and Social Desirability in a Sensitive Topic: The Case of Race and Welfare Stereotypes in the United States

Kirill Zhirkov[1], Kristin Lunz Trujillo[2,3], and C. Daniel Myers[4]

[1] University of Virginia

[2] Northeastern University

[3] Harvard University

[4] University of Minnesota

[Draft. Please don't cite or circulate]

**Abstract**

Conjoint experimental design infers preferences from stated choices and thus limits concerns about social desirability bias. The latter conjecture has been put forward in the foundational conjoint studies in the discipline—but there have been few systematic attempts to test it. We address this question by evaluating the effect of using of one of the most common ways that social scientists attempt to reduce social desirability bias: signaling race implicitly using racially distinctive names. As part of a conjoint experiment measuring stereotypes of welfare recipients, we use an experiment-in-experiment design that randomly assigns respondents to explicit or implicit conditions. In the explicit condition, race is signaled to participants openly: profiles are described as white, black, or Hispanic. In the implicit condition, race is signaled through racially distinct names. We compare the average marginal component effects across the two conditions and find no differences in the effects of the race attribute. Our results support the current practice of including potentially sensitive attributes, such as race, in conjoint experiments using explicit labels.

*Keywords:* conjoint experiments, race, social desirability, stereotypes, survey methods, welfare

## 1. Introduction

Introduced to political science less than a decade ago, conjoint experiments quickly gained popularity in the discipline. One reason is their potential to reduce social desirability bias (SDB): In conjoint experiments, respondents are not asked to reveal their preferences directly. Instead, such preferences are inferred by researchers from observed choices. Further, the inclusion of a large number of attributes in each profile is thought to reduce social desirability as this "provide[s] respondents with multiple reasons to justify any particular choice or rating" (Hainmueller, Hopkins, and Yamamoto 2014, 3), a logic similar to the one underlying list experiments (Blair and Imai 2012). Recently, suggestive evidence of SDB reduction in conjoint experiments has been provided (Horiuchi, Markovich, and Yamamoto 2021).

In this letter, we examine the same problem from a different perspective, by evaluating one of the most common ways that social scientists seek to avoid SDB in experiments: implicitly signally race using racially distinctive names. Race is a particularly socially sensitive topic for Americans. Social scientists conducting vignette or audit experiments that manipulate the race of an individual have long feared that labeling it explicitly can depress the effect of interest by priming the social norm of race-neutral decision making. As a result, the convention across disciplines has been to signal race implicitly, frequently by using racially distinctive names (Butler and Homola 2017). Curiously, most conjoint experiments that use race as an attribute do not follow this convention, instead labeling the race of a profile explicitly (Carnes and Lupu 2016; Hainmueller, Hopkins, and Yamamoto 2014; Ono and Burden 2019; Zhirkov 2021; but see Doherty, Dowling, and Miller 2019). This choice may be based on the assumption that conjoint experiments are not subject to social desirability bias, and thus that explicitly listing a person's race in a conjoint profile will have the same effect as signaling this attribute implicitly.

While plausible, if this assumption is not true it threatens the validity of a number of prominent studies.

We test the validity of this assumption using an experiment that varies how race is signaled in a conjoint profile. As part of an experiment measuring the content of Americans' stereotypes of welfare recipients,[1] we implement an experiment-in-experiment design, in which respondents are randomly assigned to one of two variants of the conjoint task. In one condition, race is listed directly; in the other, race is signaled via racially distinctive names. We find no differences in average marginal component effects across the two conditions, providing additional evidence regarding the resilience of conjoint designs to the direct inclusion of sensitive attributes. Practically, our results suggest that researchers have nothing to gain by abandoning direct signaling of race in conjoint experiments—which is currently the dominant practice.

## 2. Experimental Design

We recruited 1,280 non-Hispanic white U.S. adults in January 2021 using Lucid online panel.[2] Our design constituted an experiment within an experiment. In the conjoint task, respondents were presented with profiles of hypothetical persons and asked to assess their typicality as welfare recipients (see Supplementary Material for the exact instructions). Each respondent was asked to rate 30 profiles, the highest number suggested in the literature (Bansak et al. 2018), to maximize statistical power for the comparison of interest.[3] Profiles were described in terms of seven attributes: race, gender, marital status, number of children, immigration status,

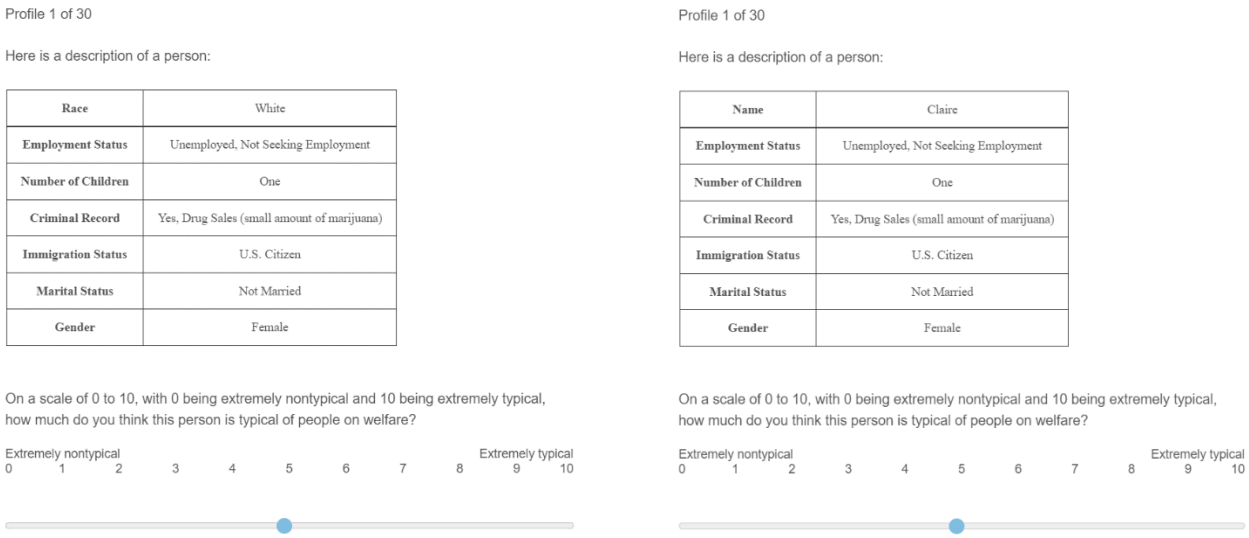
---

1 Conjoint experiments have been originally used to study preferences but are increasingly used to measure the content of individuals' stereotypes of groups (Goggin, Henderson, and Theodoridis 2020).

2 This number excludes 37 respondents who gave similar ratings to all conjoint profiles or failed attention checks. Lucid is known to well approximate national probability samples on the key demographic characteristics (see Coppock and McClellan 2019). Demographics for our sample can be found in Supplementary Material.

3 Some respondents rated fewer profile than 30 profiles (but no fewer than 27). They were kept in the analysis.

employment status, and criminal record (order randomized). One half of respondents was randomly assigned to rating profiles where race was signaled directly (explicit condition). The other half rated profiles with race signaled by racially distinctive names (implicit condition). Sample profiles in the two conditions as presented to respondents are shown in Figure 1.




**(a)** Explicit condition **(b)** Implicit condition

**Figure 1.** Conjoint design: sample profiles

The names used in the implicit condition are presented in Table 1. To create this list, we selected five male and five female names commonly associated with each racial category, chosen from the list of names found in previous research to be rated as stereotypically white, black, or Hispanic by at least 90% of respondents (Gaddis 2017a, 2017b).[4] The race and gender of each profile were independently randomized, according to standard conjoint procedure. For profiles in the implicit condition one of the five names that matched the race--gender combination was

[4] Names in the original studies by Gaddis have been taken from the State of New York's birth record data for all births from 1994–2012.

randomly selected and the "Race" attribute was replaced by the "Name" attribute that displayed this name.

**Table 1.** Randomized values for the race/ethnicity attribute for profiles in conjoint experiment

| | Values |
|---|---|
| Explicit | White<br>Black<br>Hispanic |
| Implicit (names) | *White names:* Brad, Brett, Dustin, Graham, Luke, Claire, Erin, Joan, Katie, Megan<br>*Black names:* DeShawn, Jamal, Keyshawn, Lamar, Terrell, Ebony, Kenya, Lakisha, Latonya, Shanice<br>*Hispanic names:* Alejandro, Carlos, Felipe, Juan, Pablo, Alejandra, Florencia, Guadalupe, Juanita, Valentina |

*Note.* Other randomized attributes: gender, marital status, number of children, immigration status, employment status, and criminal record. See Table S1 in Supplementary Material for the full list of attributes and values. Names were programmed to only appear along with the appropriate gender attribute

## 3. Results

We estimate average marginal component effects (AMCEs) for the two conditions using the standard procedure: an OLS regression with standard errors clustered by respondent (Hainmueller, Hopkins, and Yamamoto 2014). Since all attribute values in our conjoint design have been randomized independently, we estimate the model including only the race attribute categories with "White" as the reference. Results presented in Table 2 show no meaningful differences in AMCEs for the race attribute values across the two conditions. If anything, the AMCE for "Hispanic" is somewhat stronger in the explicit condition, although not reliably so. An additional test also does not allow rejecting the null hypothesis that the two differences are jointly zero ($p$ = .568).

**Table 2.** AMCEs of the race/ethnicity attribute values on stereotype ratings by condition

| | Explicit | Implicit | Absolute difference |
|---|---|---|---|
| AMCE: Black | 0.11 | 0.11 | <0.01 |
| | [−0.01, 0.22] | [0.01, 0.22] | [−0.16, 0.15] |
| AMCE: Hispanic | 0.11 | 0.05 | 0.06 |
| | [0.00, 0.23] | [−0.06, 0.15] | [−0.09, 0.22] |
| Observations (rated profiles) | 18,960 | 19,382 | 38,342 |
| Clusters (respondents) | 633 | 647 | 1,280 |

*Note.* 95% confidence intervals in brackets. Standard errors clustered by respondent. "White" is the reference category for AMCE calculation

## 4. Conclusion

In this note, we have further investigated how conjoint designs perform when applied to sensitive topics. Our study focused on race and welfare stereotypes in the United States. Using an experiment within an experiment, we have demonstrated that AMCE estimates do not differ depending on whether race is signaled explicitly (racial group labels) or implicitly (racially distinctive names).

There can be different reasons for this finding. One interpretation is that the conjoint design reduces social desirability bias that otherwise would depress the effects in the explicit race condition. But it is also possible that signaling race with names (rather than directly) increases cognitive load on respondents thus depressing the effects in the implicit condition. Our design does not allow discriminating between these two explanations. However, this alternative explanation points to some of the potential downsides to signaling race implicitly. Implicitly signaling race may artificially reduce effect sizes because of this increased cognitive load, and because it relies on respondents to correctly connect names to races. This is a particular problem for studies that may try to compare the effect of race to the effect of other attributes that are signaled explicitly. Further, racially distinctive names may also signal other attributes, such as socioeconomic status (Fryer and Levitt 2004; but see Butler and Homola 2017).

At the same time, results reported in this note have an important practical implication for applied conjoint-experimental research. Specifically, our findings suggest that scholars have nothing to gain by abandoning the currently dominant practice of signaling race in conjoint experiments directly via racial group labels.

## References


Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. 2018. "The Number of Choice Tasks and Survey Satisficing in Conjoint Experiments." *Political Analysis* 26 (1): 112–19.

Blair, Graeme, and Kosuke Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20 (1): 47–77.

Butler, Daniel M., and Jonathan Homola. 2017. "An Empirical Justification for the Use of Racially Distinctive Names to Signal Race in Experiments." *Political Analysis* 25 (1): 122–30.

Carnes, Nicholas, and Noam Lupu. 2016. "Do Voters Dislike Working-Class Candidates? Voter Biases and the Descriptive Underrepresentation of the Working Class." *American Political Science Review* 110 (4): 832–44.

Coppock, Alexander, and Oliver A. McClellan. 2019. "Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents." *Research & Politics* 6 (1). https://doi.org/10.1177/2053168018822174

Doherty, David, Conor M. Dowling, and Michael G. Miller. 2019. "Do Local Party Chairs Think Women and Minority Candidates Can Win? Evidence from a Conjoint Experiment."

*Journal of Politics* 81 (4): 1283–97.

Fryer, Roland G., Jr., and Steven D. Levitt. 2004. “The Causes and Consequences of Distinctively Black Names.” *Quarterly Journal of Economics* 119 (3): 767–805.

Gaddis, S. Michael. 2017a. "How Black are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies." *Sociological Science* 4: 469–89.

Gaddis, S. Michael. 2017b. "Racial/Ethnic Perceptions from Hispanic Names: Selecting Names to Test for Discrimination." *Socius* 3. https://doi.org/10.1177/2378023117737193

Goggin, Stephen N., John A. Henderson, and Alexander G. Theodoridis. 2020. “What Goes with Red and Blue? Mapping Partisan and Ideological Associations in the Minds of Voters.” *Political Behavior* 42: 985–1013.

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.

Horiuchi, Yusaku, Zachary Markovich, and Teppei Yamamoto. 2021. “Does Conjoint Analysis Mitigate Social Desirability Bias?” *Political Analysis*. Published ahead of print. https://doi.org/10.1017/pan.2021.30

Ono, Yoshikuni, and Barry C. Burden. 2019. “The Contingent Effects of Candidate Sex on Voter Choice.” *Political Behavior* 41: 583–607.

Zhirkov, Kirill. 2021. “Estimating and Using Individual Marginal Component Effects from Conjoint Experiments.” *Political Analysis*. Published ahead of print. https://doi.org/10.1017/pan.2021.4

**Conjoint Experiments and Social Desirability in a Sensitive Topic: The Case of Race and Welfare Stereotypes in the United States**

Supplementary Material

Kirill Zhirkov, Kristin Lunz Trujillo, and C. Daniel Myers

## Sample characteristics

Mean age is 49.5 years. Gender composition is 53.2% female. Median income is between $45,000 and $49,999. 45.8% of respondents have college degrees. 35.7% of respondents are Democrats, 38.4% Republicans, and 25.9% independents.

## Conjoint instructions

"In the following few screens, you will be shown profiles of hypothetical individuals. After reviewing each profile, you will be asked to rate how much you think this person is typical of people on welfare."

**Table S1.** Attributes for profiles in conjoint experiment other than race/ethnicity

| Attribute | Values |
|---|---|
| Gender | Male |
| | Female |
| Marital Status | Married |
| | Not Married |
| Number of Children | Zero |
| | One |
| | Two |
| | Three |
| Immigration Status | U.S. Citizen |
| | Green-Card Holder |
| | Undocumented/Illegal Immigrant |
| Employment Status | Employed |
| | Unemployed, Seeking Employment |
| | Unemployed, Not Seeking Employment |
| Criminal Record | No Criminal Record |
| | DUI |
| | Heroin Possession |
| | Drug Sales |
| | Aggravated Assault |
| | Robbery |
| | Threatening with a Weapon |