Should Researchers Signal Race Explicitly or Implicitly in Conjoint Experiments?

The Case of 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

Experiments explicitly including race—a socially sensitive topic in America—are potentially subject to biased effects, as respondents may hide socially undesirable views on racial issues. Thus, race is often signaled in experiments implicitly. Conjoint experiments are an exception because researchers assume they mitigate the social desirability bias associated with explicit racial cues. However, this assumption remains untested. To test it, we use an experiment-in-experiment that randomly assigns respondents to either explicit or implicit racial cues condition within a larger conjoint study measuring stereotypes of welfare recipients. In the explicit condition, race is signaled openly: profiles are described as white, black, or Hispanic. In the implicit condition, race is signaled through racially distinctive names. Across these two conditions we find no differences in the effects of the race attribute. Our results support the current practice of including the race attribute explicitly in conjoint experiments.

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

Race is a socially sensitive topic in America that is difficult to study using experimental methods. Social scientists conducting vignette or audit experiments worry that explicit racial labels can depress effects by priming the norm of race-neutrality. As a result, the convention has been to signal race implicitly, often by using distinctive names (Butler and Homola 2017). However, studies that employ conjoint experiments generally do not follow this convention and instead label the race attribute explicitly (Carnes and Lupu 2016; Hainmueller, Hopkins, and Yamamoto 2014; Ono and Burden 2019; Zhirkov 2021; but cf. Doherty, Dowling, and Miller 2019). This practice is likely informed by an assumption that the conjoint design mitigates social desirability bias because respondents are not asked to reveal preferences directly (Hainmueller, Hopkins, and Yamamoto 2014). Though recent work suggests that conjoint experiments reduce social desirability bias broadly (Horiuchi, Markovich, and Yamamoto 2021), it is unclear whether this holds for explicit racial labels specifically.

We explore potential benefits of signaling race in conjoint experiments implicitly via distinctive names using an experiment-in-experiment design. We find no differences in average marginal component effects (AMCEs) between the explicit (group labels) and implicit (distinctive names) conditions, suggesting that researchers can continue signaling race in conjoint experiments directly—which is currently the dominant practice.

Experimental Design

We recruited 1,281 non-Hispanic white U.S. adults in January 2021 using Lucid online panel.[1] Our design constituted an experiment within an experiment evaluating the content of Americans' stereotypes of welfare recipients.[2] Respondents were shown profiles of hypothetical persons and

---

[1] This number excludes 37 respondents who gave similar ratings to all conjoint profiles or failed attention checks. Sample demographics can be found in Online Appendix.

[2] It provides a useful case to evaluate the implicit vs. explicit signaling of race as non-conjoint experiments on welfare attitudes often use names instead of group labels (e.g., DeSante 2013).

asked to assess their typicality as welfare recipients (see Online Appendix for the exact instructions). Each respondent was presented with 30 profiles to maximize statistical power. Profiles were described in terms of seven attributes: race, gender, marital status, number of children, immigration status, employment status, and criminal record (order randomized across respondents).[3] One half of respondents were randomly assigned to rate profiles where race was signaled directly (white, black, or Hispanic; explicit condition). The other half rated profiles with racially distinctive names (implicit condition).[4] Sample profiles for the two conditions are shown in Figure 1.
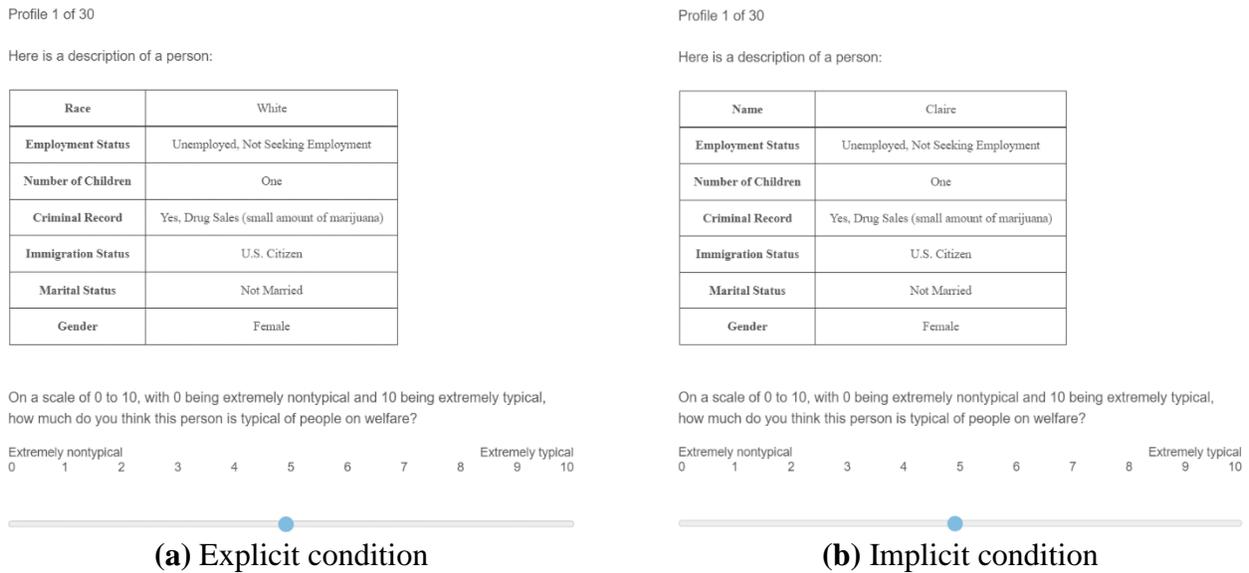

(a) Explicit condition
(b) Implicit condition

Figure 1. Conjoint design: sample profiles

Results

We estimate AMCEs for the two conditions using an OLS regression with standard errors clustered by respondent. Since all attribute values have been randomized independently, the

---

[3] See Table A1 in Online Appendix for the potential values of attributes other than race.

[4] Procedure of name selection is described in Online Appendix. See Table A2 for the list of names.

model includes only the race attribute categories with "White" as the reference. Results presented in Table 1 show no meaningful differences in AMCEs for the race attribute values across the two conditions. These null effects are estimated with a high degree of precision: for instance, standard errors of the estimated differences in AMCEs across the two conditions are as low as 0.08 (while the rating outcome is on a 0–10 scale with a standard deviation of 2.8).

Table 1. AMCEs of the race 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. Standard errors clustered by respondent. "White" is the reference category

Conclusion

We have compared the performance of explicit (group labels) and implicit (distinctive names) signaling of profiles' race in conjoint experiments. Applying an experiment within an experiment to the case of welfare stereotypes in the United States, we have demonstrated that the overall effects of race on welfare stereotypes are weak and that AMCE estimates do not differ across the two conditions.

The exact reasons for these findings are unclear. It is possible that social desirability bias depresses the effect of race in the explicit condition, whereas signaling race with names increases cognitive load on respondents (who need to recognize the names) thus depressing the same effect in the implicit condition. Our design cannot discriminate between these explanations. However, it points to some additional downsides of signaling race implicitly: for instance, it relies on

respondents correctly connecting names to racial groups.

Overall, our results show no benefits of signaling race in conjoint experiments via distinctive names. Therefore, scholars can continue the currently dominant practice of using racial group labels.

References

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.

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): 342–56.

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.

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. 2022. "Estimating and Using Individual Marginal Component Effects from

    Conjoint Experiments." *Political Analysis* 30 (2): 236–49.