The Impact of Customer Identity on Service Engagement
A Field Experiment in Union and Confederate States

Dylan Sivori
Jonathan Hernandez

December 12, 2025

Abstract

Does the race or gender of a potential customer affect service provision in U.S. states with historical ties to slavery? While the American Civil War ended over 160 years ago, it is often assumed that significant disparities persist in the experiences of racial minorities in the United States. To investigate this, we conducted a blocked randomized audit study of restaurants and caterers across states that comprised the Union and the Confederacy during the American Civil War. We made catering inquiries to restaurants in both regions, varying race and gender cues in the sender's name and email. Our results indicate complex regional heterogeneity. Minority applicants faced lower response rates in the historical-Union states. In historical-Confederate states, White applicants experienced lower engagement. These findings complicate common notions of regional discrimination and suggest that contemporary service disparities may be driven by unobserved covariates, or insufficient sample size to generate statistical power. Further research is encouraged to determine the extent to which local economic conditions like inflation and income, education levels, norms, and technological infrastructure influence these outcomes.

1 Introduction

In this audit study we investigate whether a potential customer's race or gender affects service engagement in U.S. states with historical ties to slavery. We use names to signal a specific race or gender to the service provider to see if that single variable changes the outcome, whether or not they respond - and how quickly they respond. We are auditing actual service providers in their normal business environment to see how they treat potential customers in real time. Our goal is to determine if identity characteristics affect engagement from service providers, specifically looking for a causal link between the customers' assumed demographic profile and the quality of service they receive, within the context of historical relationships to slavery.

2 Theoretical Framework

Restaurants and/or caterers are expected to treat all of our inquiries equivalently, regardless of the assumed race or gender. Related literature offers substantial support for this study's treatment design, assignment, and delivery, as well as appropriate techniques for sampling and statistical inference.

2.1 Audit Studies of Discrimination

Becker's (1957) model of taste-based discrimination, for example, finds that service providers will behave as if they incur non-pecuniary, psychic costs when interacting with members of a disfavored group, effectively treating transactions with those individuals as more costly than equivalent transactions with others. If Becker's analysis were true, then a restaurant operator who has such views might perceive an inquiry from a minority customer as carrying an implicit psychic cost, thereby reducing the perceived net gain of engaging with that customer. Under this framework, departures from the null of equal treatment could manifest as a lower probability of response.

Empirical research has successfully adapted this theoretical framework to detect *everyday* discrimination across various domains. In the labor market, Bertrand and Mullainathan (2004) demonstrated that fictitious resumes assigned White-sounding names received 50% more callbacks than identical resumes with African-American names, a gap that widened with candidate qualification. Subsequent studies have extended this "audit" methodology to the service and public sectors, operationalizing discrimination as a failure to respond to routine inquiries.

Block et al. (2021) characterize these ignored attempts as *paper cut* discrimination, finding that White senders requesting voluntary assistance received responses at a rate 15.5 percentage points higher than Black senders. Similarly, Hughes et al. (2020) utilized an email-based design to uncover persistent bias among local election officials, validating the use of correspondence audits to measure responsiveness in service-oriented roles. Within the hospitality sector specifically, ROC United (2015) has documented significant racial segregation and bias in fine-dining hiring, suggesting that the *tastes* Becker described remain active in this industry.

2.2 Evidence on Sample Sizes and Power

We also note the number of subjects required to gain statistical power. Bertrand and Mullainathan (2004) sent 4,890 fictitious resumes in response to help-wanted ads in Boston and Chicago newspapers, a sample size sufficient to identify a 50% racial gap in callbacks. Studies focused on detecting more subtle forms of discrimination often employ even larger scales; Block et al. (2021) contacted 250,000 individuals to more precisely estimate a 15.5% percentage point difference in response rates, favoring White senders compared with African-American names. Moreover, given that they asked subjects to complete a task, Black names were 9.5% less likely to receive a completed task.

However, robust results are also attainable with more targeted populations. Hughes et al.'s (2020) audit of local election officials found that Arab-Muslim senders faced the greatest barrier, receiving replies at a rate of 50.1% (SE 1.25), which was more than 11 percentage points lower than the 61.3% White baseline (SE 1.21). Latino senders also received replies nearly 2.9% less often (SE 1.23), while Black senders received responses .1% more often than White senders (SE 1.21). Hughes' findings, while small and within the standard error, offer interesting evidence in relation to this study's population, and the expected directional effects of identity-based implicit racism.

Meanwhile, design-adjacent and industry-specific audits like ROC United's (2015) investigation of labor conditions in fine-dining establishments often rely on smaller samples. They conducted over 400 matched-pair audit tests, finding that workers of color received 56% lower earnings than equally qualified White workers, with women of color in California earning just $10.13 per hour compared to $14.18 for White men.

2.3 Operationalization

By using a geographically-blocked randomization inference treatment assignment mechanism, we allow for the possibility that observable behaviors vary systematically across regions, and may vary in response to our intervention. Our use of web form-based inquiries situates the study within the literature on subtle, everyday forms of discrimination, which documents differential responses to otherwise comparable electronic communications based solely on racialized or gendered cues in sender identity. This body of work motivates our empirical test of whether we can statistically reject the null of no treatment effect in this particular service setting.

2.4 Conceptual and Terminological Clarifications

In this study, we refer to *gender* using binary categories (male and female) because the audit design relies on name cues that service providers typically read as signaling a binary gender. However, it's important for us to acknowledge that gender is expansive, nonbinary, and culturally diverse. Many individuals identify outside or across the categories of *male* and *female*, and our methodological simplification should not be interpreted as endorsing a binary view of gender.

Our use of these terms is strictly for the purpose of maintaining a controlled experimental environment. We are not making normative claims about the legitimacy or completeness of gender categories - only using a limited operationalization that allows us to test a very specific causal question in a constrained design.

When we use terms aligned with a particular *allegiance*, we are referring only to the historical Civil War designation of states as either part of the Union or the Confederacy. We use this classification because our study examines whether regional historical context moderates responses to identity-signaled emails.

This language describes states, not individuals. People living in these states today hold a wide range of values and beliefs, and no one should be assumed to share or reflect the politics or ideologies of a state's 19th-century government. Our framing is analytical, not judgmental: the historical 'allegiance' labels help us test whether long-term legacies correlate with contemporary service-sector behavior, but they are not meant to define, stereotype, or characterize anyone in our class or in our dataset.

3 Experiment Design

3.1 Hypotheses

Based on this theoretical framework, we test the following hypotheses:

H1: *Racial Differences*
H_0: There is no difference in response rate between profiles with African American names and White names, holding gender and region constant.
H_A: There is a measurable difference in the response rates observed between African American and White populations.

H2: *Gender Differences*
H_0: There is no difference in response rate between profiles with male names and female names.
H_A: There is a measurable difference observed in the response rates between male and female names.

H3: *Intersectional Differences*
H_0: There is no difference in response rate across subgroups (race, gender, and region).
H_A: There is a measurable difference in response rate across subgroups.

3.2 Potential Outcomes to Treatment

The key outcome in this audit study is the response behavior of restaurants and caterers in a 2×2 factorial design. For each subject, the potential outcome is defined as $Y = 1$ if the business responds to the catering request and $Y = 0$ otherwise, conditional on the treatment condition received: Y(white_male), Y(white_female), Y(black_male), and Y(black_female). By the fundamental reality of causally inferring treatment effects from assigning a single treatment, only one of these potential outcomes is observed for any given subject. Our specific treatment design was to fill out the restaurant or caterer's web based contact form with the same language, date, time, number of people, and phone number. The only change was in the name and email address of the requester. The white male treatment was Matthew Wilson (matthew.wilson0241@gmail.com), the white female treatment was Emily Smith (emily.smith4241@gmail.com), the African American male name was Andre Lamar (andre.lamar0241@gmail.com), and the African American female name Monique Jackson was (monique.jackson0241@gmail.com).

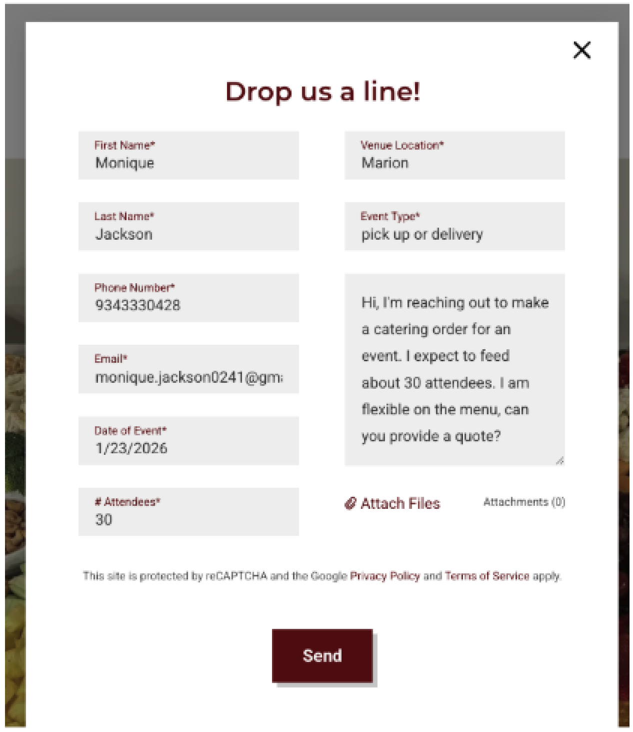


Figure 1: Black Female Treatment Example

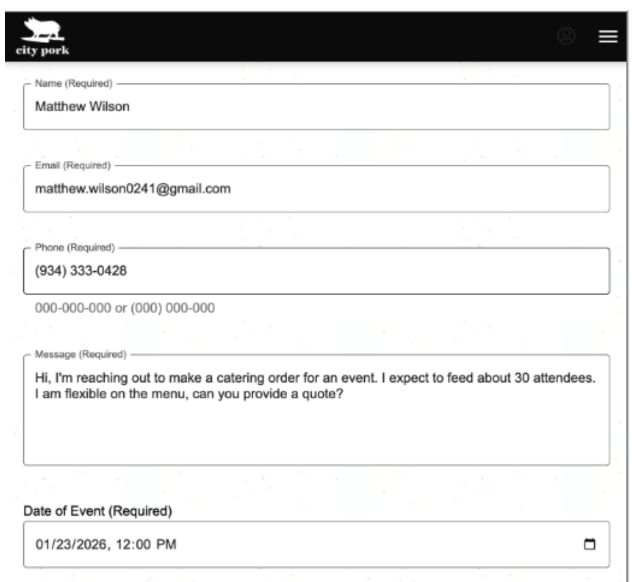


Figure 2: White Male Treatment Example

The realized outcomes we observe reflect only one potential response per business, and our inferences are further constrained by the modest scale and timing of the study. With a relatively small sample and a single wave of contacts, the experiment may have limited statistical power to detect modest effects and may be more sensitive to short-run economic or sector-specific shocks that shape how restaurants manage incoming requests. Consequently, any divergence from (or alignment with) the larger audit literature should be interpreted cautiously, as our operational constraints and the contemporary socioeconomic context may attenuate or amplify the patterns of discrimination that we are able to detect in this setting.

3.3 Power Analysis

To benchmark the statistical power of our design, we conducted a series of simulation-based power analyses using response rate patterns drawn from prior correspondence and audit studies of discrimination. Specifically, we considered three scenarios: (1) a moderate-effect scenario calibrated to Rosen's (2010) colloquially referred to as the *Jose vs. Colin* experiment, where a comparison of Hispanic and non-Hispanic legislators, names, and grammatically-oriented treatments were found to vary response rates between up to 38% between block cells; (2) a smaller-effect scenario based on the callback rates in Bertrand and Mullainathan (2004) where race gaps manifest at single-digit percentage levels; and (3) a very small-effect scenario reflecting the 1–5% response rates observed in Block et al. (2021).

For each scenario and a grid of total sample sizes from $n = 100$ to $n = 10{,}000$, we simulated datasets under the assumed cell means, fit linear probability models with robust standard errors, and recorded the proportion of simulations in which the race-by-region interaction term was statistically significant at the 5% level. Under the moderate-effect Jose/Colin scenario (29–58% cell means), power exceeded 80% once the total sample size reached approximately $n = 800$, and approached or exceeded 95% for $n \geq 1{,}200$. By contrast, in the Bertrand-like scenario (6–13% callback rates) and the Block-like scenario (1–5% response rates), power remained well below conventional thresholds even at $n = 2{,}000$, reflecting the difficulty of detecting small absolute differences in rare outcomes without very large samples.

Our realized design of 88 businesses (11 subjects per treatment block) lies below the level from which interaction effects can be realistically measured. As a result, any null findings in our study are consistent with a sample that is simply underpowered to detect statistically significant differences.

3.4 Blocked Randomization Design

We assembled an initial sampling frame of known states, that existed in 1861 and were clearly either a part of the Union or the Confederacy. We, therefore, excluded Union-Supporting Border States and Unorganized Territories. The Confederate States are composed of South Carolina, Mississippi, Florida, Alabama, Georgia, Louisiana, Texas, Virginia, Arkansas, Tennessee, North Carolina. The Union states are composed of California, Connecticut, Illinois, Indiana, Iowa, Kansas (admitted January 29, 1861), Maine, Massachusetts, Michigan, Minnesota, Nevada, New Hampshire, New Jersey, New York, Ohio, Oregon, Pennsylvania, Rhode Island, Vermont, Wisconsin.

We initially used the maps library for R to compose a geographically and population balanced list of 562 cities across Union states, and 261 cities in Confederate states, filtering for states with populations of less than 8,000 people.

We then used the Targetron API to obtain a list of 800 geographically-balanced businesses, initially intending to use email for treatment delivery. However, given that Targetron is a web-crawling company for targeted advertising, their data proved to be of poor quality for this study. Thus, the treatment design pivoted toward one that could be delivered by filling out web forms on restaurants' websites.

Then, the Google Places API was leveraged to obtain a sample (without email addresses). For each state in our set of Union and Confederate states, we submitted the search query "restaurants that offer catering for pickup in state". The API returns at most 60 results per state, which provided more than sufficient candidates for this study. For each region, samples were proportionally allocated across states, such that states were shuffled before each draw for treatment assignment. If the proposed subject passed a manual review (i.e. a website inspection by this study's researchers), then the business remained in its randomly assigned treatment cell. If a subject failed the website verification, then a new subject was selected based on sequentially examining the list of returned businesses, based only on whether or not their website would allow catering inquiries via a functioning web form. Treatment was randomly assigned to one of four treatment conditions, i.e. {`white_male`, `white_female`, `black_male`, `black_female`}, within region blocks {`Union`, `Confederacy`}. This ensured numerical balance by region.

In the interest of efficiency, additional covariates were considered, but not implemented. These included local socioeconomic conditions and trends, local demographic factors, and the average cost per restaurant.

4 Models

The models in this section estimate the causal effect of race, gender, and region on the response indicator Y_i, with response rate as the primary estimand. The analysis proceeds from main-effect specifications to models that include interactions capturing heterogeneous treatment effects. Robust standard errors are used throughout to avoid relying on homoskedasticity.

The first set of models estimates the main effects of race and gender, followed by specifications that add region (Union vs. Confederacy) and then interaction terms between race and region and between gender and region. These interaction terms represent heterogeneous treatment effects across blocks. The final, fully saturated model includes race, gender, region, all two-way interactions, and the three-way *race* × *gender* × *region* interaction, which captures whether the joint effect of race and gender varies across historical regions.

For each expansion of the model (e.g., adding block fixed effects or interaction terms), an F-test compares the enriched specification to its simpler counterpart to assess whether the additional terms explain statistically significant additional variation in response outcomes.

In these specifications, Y_i denotes a binary response indicator for business i, equal to 1 if the business replies to the catering inquiry and 0 otherwise. Race_i is a dummy variable coded 1 for profiles with Black names and 0 for profiles with White names, and Gender_i is a dummy coded 1 for female names and 0 for male names. Region_i is a dummy coded 1 for businesses in historically Confederate states and 0 for businesses in historically Union states, while $\phi_{\text{Region}[i]}$ represents a region-specific intercept (block fixed effect) that allows the baseline response level to differ between Union and Confederacy.

The interaction terms $(\text{Race} \times \text{Region})_i$, $(\text{Gender} \times \text{Region})_i$, and $(\text{Race} \times \text{Gender})_i$ are products of the corresponding dummies and capture how the effect of one characteristic depends on the level of another, and the three-way interaction $(\text{Race} \times \text{Gender} \times \text{Region})_i$ captures whether the joint effect of race and gender varies across regions. The coefficients $\beta_0, \ldots, \beta_7$ are linear probability model parameters, and ϵ_i is an idiosyncratic error term.

In all of these models, the omitted (baseline) category is profiles with White, male names located in historically Union states (i.e., $\text{Race}_i = 0$, $\text{Gender}_i = 0$, and $\text{Region}_i = 0$), so β_0 represents their average response rate and each coefficient on a dummy or interaction term captures the difference in response probability relative to this baseline group.

4.1

Race Main Effects:

$$Y_i = \beta_0 + \beta_1 \,\text{Race}_i + \epsilon_i \tag{1}$$

4.2

Race Main Effects + Blocking ($\phi_{\textbf{Region}}$):

$$Y_i = \beta_0 + \beta_1 \,\text{Race}_i + \phi_{\text{Region}[i]} + \epsilon_i \tag{2}$$

4.3

Race + Blocking + Race $\times$ Blocking:

$$Y_i = \beta_0 + \beta_1 \,\text{Race}_i + \beta_2 \,\text{Region}_i + \beta_3 \,(\text{Race} \times \text{Region})_i + \epsilon_i \tag{3}$$

4.4

Gender Main Effects:

$$Y_i = \beta_0 + \beta_1 \,\text{Gender}_i + \epsilon_i \tag{4}$$

4.5

Gender Main Effects + Blocking ($\phi_{\textbf{Region}}$):

$$Y_i = \beta_0 + \beta_1 \,\text{Gender}_i + \phi_{\text{Region}[i]} + \epsilon_i \tag{5}$$

4.6

Gender + Blocking + Gender $\times$ Blocking:

$$Y_i = \beta_0 + \beta_1 \,\text{Gender}_i + \beta_2 \,\text{Region}_i + \beta_3 \,(\text{Gender} \times \text{Region})_i + \epsilon_i \tag{6}$$

4.7

Fully Saturated Model:

$$\begin{aligned} Y_i = \beta_0 &+ \beta_1 \,\text{Race}_i + \beta_2 \,\text{Gender}_i + \beta_3 \,\text{Region}_i \\ &+ \beta_4 \,(\text{Race} \times \text{Gender})_i + \beta_5 \,(\text{Race} \times \text{Region})_i \\ &+ \beta_6 \,(\text{Gender} \times \text{Region})_i + \beta_7 \,(\text{Race} \times \text{Gender} \times \text{Region})_i \\ &+ \epsilon_i \end{aligned} \tag{7}$$

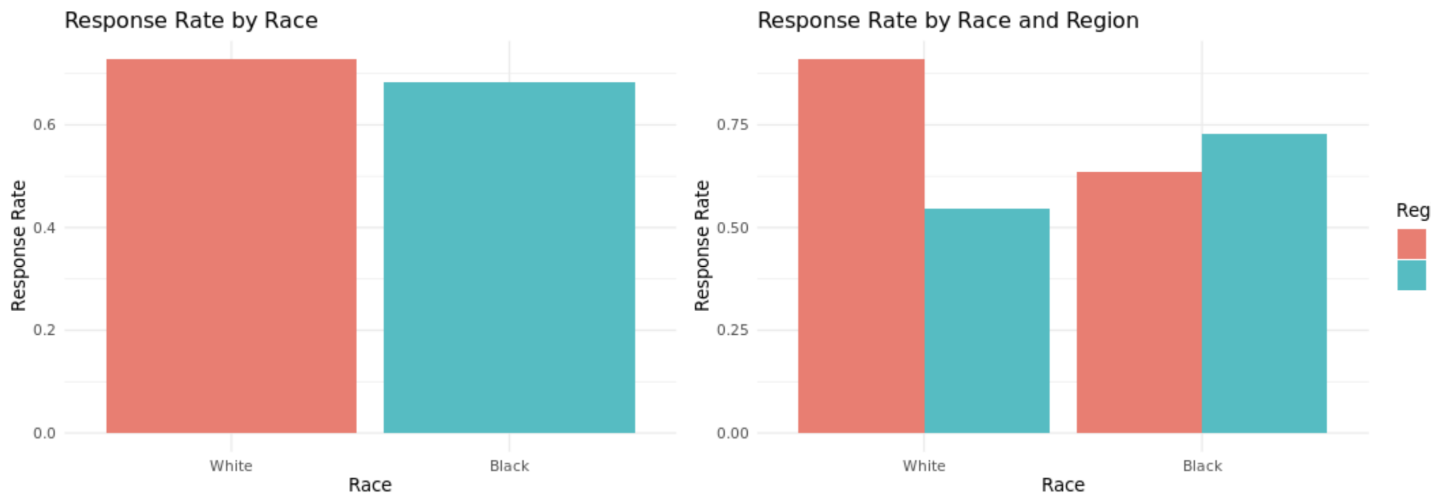


Figure 3: Response Rate By Race and Region

5 Results

Table 1: Response Rates by Treatment Block - Summary Statistics of Data Collected

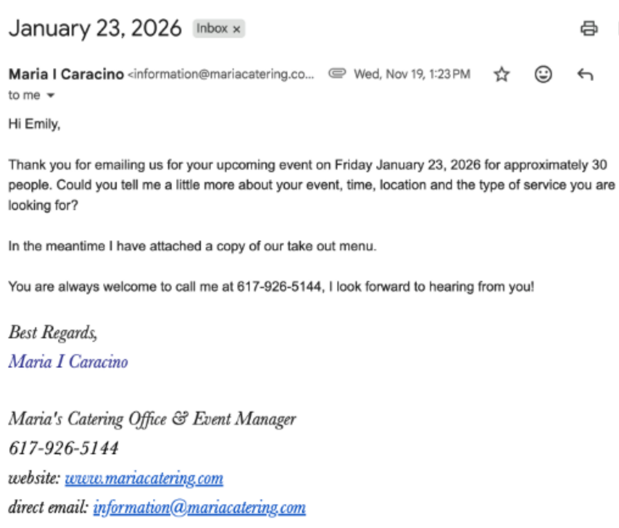
Gender	Region	Race	Response Rate	Response Count	Total Request
Male	Union	White	0.909	10	11
Male	Union	Black	0.727	8	11
Male	Confederacy	White	0.545	6	11
Male	Confederacy	Black	0.727	8	11
Female	Union	White	0.909	10	11
Female	Union	Black	0.545	6	11
Female	Confederacy	White	0.545	6	11
Female	Confederacy	Black	0.727	8	11

The results of the experiment appear in Table 1 and Figure 3. The outcome for each business is a binary indicator equal to 1 if it replied to the catering request within the observation window and 0 otherwise, and response rates are calculated within each race–gender–region cell. Overall, response rates are higher for profiles with White names than for profiles with Black names, but this pattern reverses in historically Confederate states, where businesses reply more often to Black than to White profiles. Response rates are also higher for male names than for female names, with the largest gender gap in historically Union states, and this gender difference is present in both regional blocks.

5.1 H1

The first hypothesis concerns the effect of race on the probability of receiving a response. In Table 2, Column (1) reports a linear probability model with the binary response indicator as the dependent variable and a race dummy variable used as the independent variable, such that Black = 1 and White = 0. The estimated intercept implies an average response rate of approximately 73% for White profiles, and the race coefficient indicates a 4.5 percentage point lower response rate for Black profiles relative to White profiles; this difference is not statistically distinguishable from zero. Column (2) adds region fixed effects, which reduces the standard errors slightly but leaves the estimated race effect substantively similar and still statistically insignificant. Column (3) introduces a race-by-region interaction term and provides evidence of heterogeneous treatment effects by region, where the Black–White difference in response rates is estimated to be 45.5 percentage points larger in historically Confederate states than in historically Union states.

Figure 4: Response Rate By Race and Gender



January 23, 2026 Inbox ×

Maria I Caracino <information@mariacatering.co... Wed, Nov 19, 1:23 PM
to me

Hi Emily,

Thank you for emailing us for your upcoming event on Friday January 23, 2026 for approximately 30 people. Could you tell me a little more about your event, time, location and the type of service you are looking for?

In the meantime I have attached a copy of our take out menu.

You are always welcome to call me at 617-926-5144, I look forward to hearing from you!

Best Regards,
Maria I Caracino

Maria's Catering Office & Event Manager
617-926-5144
website: www.mariacatering.com
direct email: information@mariacatering.com

Figure 5: Sample Response

An F-test comparing the interacted model in Column (3) to the simpler race-only specifications yields a p-value of 0.024, indicating that the additional interaction terms explain a statistically significant share of the variation in response outcomes.

5.2 H2

The second hypothesis concerns the effect of gender on the probability of receiving a response. In Table 3, Column (1) reports a model with the binary response indicator as the dependent variable and a gender dummy (female = 1, male = 0) as the independent variable. The estimated intercept again corresponds to an average response rate of roughly 73% for male profiles, and the gender coefficient implies a 4.5 percentage point lower response rate for female profiles relative to male profiles; this estimate is not statistically significant. Column (2) adds region fixed effects, which modestly increases precision but does not change the conclusion that the estimated gender effect is not statistically distinguishable from zero. Column (3) includes a gender-by-region interaction term to test for heterogeneous gender effects across Union and Confederate states; the interaction estimate is small and imprecise, and an F-test ($p = 0.347$) provides no evidence that the more complex model explains additional variance beyond the simpler gender specifications.

5.3 H3

The third hypothesis examines whether the joint effects of race and gender vary by region. This hypothesis is evaluated using the fully saturated model that includes race, gender, region, all two-way interactions, and the three-way race $\times$ gender $\times$ region interaction. In this specification, the coefficient on the three-way interaction term is close to zero and estimated with substantial uncertainty, indicating no statistically significant evidence that the combined effect of race and gender differs systematically between historically Union and Confederate states.

6 Conclusion

This study provides limited statistical evidence of treatment effects, which aligns with the low statistical power implied by the design's modest sample size and the power analysis. Nonetheless, the point estimates for race and gender effects are directionally consistent with prior audit studies of discrimination, and the sizable race-by-region interaction suggests that the Black–White gap in responsiveness may differ between historically Union and Confederate states. These patterns are not sufficient for strong causal claims in this sample but motivate a larger, more highly powered replication that can more precisely estimate main effects, interactions, and assess whether the apparent regional heterogeneity in race-based responsiveness is robust.

Table 2: Regression Results of Race on Response Rates: Statistically Signficant Heterogeneous Treatment Effects
Race (White=0, Black=1), Region (Union=0, Confederacy=1)

	Dependent variable:		
	Response Rate		
	Race Main Effects	Race Main Effects | Block	Race HTE
	(1)	(2)	(3)
race	-0.045	-0.045	-0.273^{**}
	(0.099)	(0.098)	(0.125)
region			-0.364^{***}
			(0.128)
race:region			0.455^{**}
			(0.195)
Constant	0.727^{***}		0.909^{***}
	(0.069)		(0.064)
Observations	88	88	88
R^2	0.002	0.025	0.087
Adjusted R^2	-0.009	0.002	0.054
Residual Std. Error	0.461 (df = 86)	0.458 (df = 85)	0.446 (df = 84)
F Statistic	0.214 (df = 1; 86)		2.663^{*} (df = 3; 84)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3: Regression Results of Gender on Response Rates: No Statistically Significant Effect of Gender Gender (Male=0, Female=1),Region (Union=0, Confederacy=1)

	Dependent variable:		
	Response Rate		
	Gender Main Effects	Gender Main Effects | Block	Gender HTE
	(1)	(2)	(3)
gender	-0.045	-0.045	-0.091
	(0.099)	(0.098)	(0.132)
region			-0.182
			(0.138)
gender:region			0.091
			(0.201)
Constant	0.727^{***}		0.818^{***}
	(0.069)		(0.086)
Observations	88	88	88
R^2	0.002	0.025	0.027
Adjusted R^2	-0.009	0.002	-0.007
Residual Std. Error	0.461 (df = 86)	0.458 (df = 85)	0.461 (df = 84)
F Statistic	0.214 (df = 1; 86)		0.786 (df = 3; 84)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Full Experimental Results including Fully Saturated Model: No Statistically Significant Heterogeneous Treatment Effects for Intersectional Subgroups
Race (White=0, Black=1),Gender (Male=0, Female=1),Region (Union=0, Confederacy=1)

	Dependent variable:						
				Response Rate			
	Race Main Effects	Race | Block	Race HTE	Gender	Gender | Block	Gender HTE	Fully Saturated HTE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
race	−0.045	−0.045	-0.273^{**}				−0.182
	(0.099)	(0.098)	(0.125)				(0.176)
region			-0.364^{***}			−0.182	-0.364^{*}
			(0.128)			(0.138)	(0.191)
race:gender							−0.182
							(0.259)
race:region			0.455^{**}				0.364
			(0.195)				(0.283)
gender:region						0.091	−0.000
						(0.201)	(0.270)
race:gender:region							0.182
							(0.407)
gender				−0.045	−0.045	−0.091	0.000
				(0.099)	(0.098)	(0.132)	(0.135)
Constant	0.727^{***}		0.909^{***}	0.727^{***}		0.818^{***}	0.909^{***}
	(0.069)		(0.064)	(0.069)		(0.086)	(0.095)
Observations	88	88	88	88	88	88	88
R^2	0.002	0.025	0.087	0.002	0.025	0.027	0.097
Adjusted R^2	−0.009	0.002	0.054	−0.009	0.002	−0.007	0.018
Residual Std. Error	0.461 (df = 86)	0.458 (df = 85)	0.446 (df = 84)	0.461 (df = 86)	0.458 (df = 85)	0.461 (df = 84)	0.455 (df = 80)
F Statistic	0.214 (df = 1; 86)		2.663^{*} (df = 3; 84)	0.214 (df = 1; 86)		0.786 (df = 3; 84)	1.224 (df = 7; 80)

Note: *p<0.1; **p<0.05; ***p<0.01

7 References

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago Press.

Bertrand, M., & Mullainathan, S. (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.

Block Jr, R., Crabtree, C., Holbein, J. B., & Monson, J. Q. (2021). Are Americans less likely to reply to emails from Black people relative to White people? *Proceedings of the National Academy of Sciences*, 118(51), e2110347118.

Hughes, D. A., Gell-Redman, M., Crabtree, C., Krishnaswami, N., Rodenberger, D., & Monogan, J. (2020). Persistent Bias Among Local Election Officials. *Journal of Experimental Political Science*, 7(3), 179–187.

Restaurant Opportunities Centers United. (2015). Ending Racial and Gender Occupational Segregation in the Restaurant Industry. *ROC United*.

Rosen, B. (2010). The effect of constituent identity and message quality on legislative responsiveness. *Unpublished manuscript*.

8 Appendix

Link to Dataset

Random Assignment in R

Stratified Random Sampling in R

Project Analysis in R