DOES RACIAL ANIMUS UNDERMINE SUPPORT FOR THE U.S. SOCIAL SAFETY NET?*

Jeffrey P. Carpenter†
Jakina Debnam Guzman
Peter Hans Matthews
Erin L. Wolcott

November 2025

Abstract

In two large, multistage experiments, we test whether beliefs about beneficiaries' race influence support for the U.S. social safety net. Participants who receive information that there are more (less) Black beneficiaries than expected supported welfare less (more). We find no evidence of an effect for the less racially stigmatized unemployment insurance program. Notions of deservingness help explain program differences in the effect of racial animus on support. Untreated beliefs remain stable over time, while our intervention persistently changes beliefs. In short, we find robust evidence of a causal relationship between racial animus and social safety net policy preferences.

Keywords: Experiment, Information Provision, Welfare, Unemployment Insurance, Social Safety Net, Discrimination
JEL codes: J08, J15, C9, D9

*For excellent research assistance, we thank Allison Lounsbury, Bishal Panthi, Blair Jia, and Audrey Wang. For assistance locating historical data, we thank Lisa Di Valentino. For helpful comments, we are grateful to Vicki Bogan, Leila Davis, Dania Francis, Erin Giffin, Heather McGhee, Martin Gilens, Sandra Goff, Remy Levin, Mike Shor, Steve Ross, Roberto Venziani, Kenicia Wright, seminar participants at Williams College, Fordham University, LACBEE, the New England Experimental Economics Workshop, the National Economics Association Meetings at the 2025 Meeting of the Allied Social Sciences Association, the 2025 Economics of Race, Racism, and Structural Inequality Conference at Duke University, the Sixth World Conference on Remedies to Racial and Social Inequality, and the University of Connecticut, the 2025 Canadian Economics Association conference, the 2025 Analytical Political Economy workshop, and the 2025 American Political Science Association conference.

Our intervention was pre-registered at AEARCTR-0013774 and approved by the Middlebury IRB. The authors have no relevant conflicts or financial interests to report.

†Corresponding author: 305 Warner Hall, Middlebury College, Middlebury VT 05753; jcarpent@middlebury.edu .

America's troubled race relations are clearly a major reason for the absence of an American welfare state.

- Alesina, Glaeser, and Sacerdote (2001)

I Introduction

In her book *The Sum of Us*, Heather McGhee (2021) recounts the story of the Fairground Park pool in St. Louis, in its time one of the largest and most elaborate public swimming facilities in the world. Efforts to integrate pool access in the 1950s – that is, to make it public in both name and spirit – produced such backlash, including violent mob attacks, that it was drained and forever closed. The provocative implication is that White St. Louisans were not averse to public goods provision, or the attendant tax implications, as long as access was restricted to other White St. Louisans. In broader terms, McGhee's book forces us to consider the notion that the differences between the American and, for example, European welfare states has less to do with individualism and self-reliance than it does with race. McGhee is not alone, of course: Kevin Kruse's *White Flight* (2007), for example, argues that in its present form, anti-tax sentiment owes something to conflicts over desegregation. And as Corey Robin (2018) reminds us, this was not a lesson lost on some politicians or their advisers: in a now infamous 1981 interview, Lee Atwater, then an adviser in the Reagan White House, observed that given its disproportionate effects, support for reduced public spending was an "abstract" expression of racial animus, more respectable than opposition to busing or, for that matter, the use of racial epithets. But while there is considerable correlational evidence consistent with the hypothesis that racial animus animates some opposition to redistributive or social insurance policies, what the literature lacks, and what we provide, is causal experimental corroboration.

We focus on the two largest income support programs in the United States social safety net: Temporary Assistance for Needy Families (TANF)—commonly known as welfare—and unemployment insurance (UI). Both programs were established by the Social Security Act of 1935 and provide income assistance in times of need.[1] The two programs differ along many dimensions, but most important in this context, they have different racialized histories and narratives.[2] Welfare has disproportionately served Black Americans. The earliest

[1]The Social Security Act of 1935 established Aid to Families with Dependent Children (AFDC), the precursor program to TANF.

[2]Across the set of national social safety net policies in the U.S., UI and TANF are the two programs which directly provide income replacement (rather than in-kind transfers). UI is not means-tested. Individuals are eligible for UI benefits if they demonstrate previous attachment to the labor market and lose a job "through no fault of their own". The amount of benefit depends on the individual's level of pre-termination

data available suggests nearly half of welfare recipients in the late 1960s were Black, and today the share of Black beneficiaries is still over double the rate of the population.[3] Unemployment insurance, on the other hand, disproportionately excluded Black Americans at its inception by not covering agricultural and domestic workers (Katznelson, 2005). Moreover, UI continues to require a stable work history, which is less common among marginalized workers. These histories, which unfolded against a backdrop of racial tension, contributed to the different narratives and notions of deservingness accompanying each program: welfare has been associated with supporting (Black) mothers "abusing the system," while unemployment insurance has been associated with (White) industrial workers "pulling themselves up by their bootstraps." To compare support for these two programs, we conduct two parallel information experiments (Akesson, Hahn, Metcalfe, and Rasooly (2022); Haaland and Roth (2023, 2020); Kuziemko, Norton, Saez, and Stantcheva (2015); Alesina, Ferroni, and Stantcheva (2021)) — one experiment exploring the role of racial beliefs for TANF policy support and one experiment exploring the role of racial beliefs for UI policy support.

Between June and August 2024, we ran two large ($n \approx 3000$), nationally representative, parallel, three-stage information provision experiments. The experiments were pre-registered (AEARCTR-0013774), and the tests of our primary hypotheses are well-powered. In the first stage, we elicited participants' demographic information as well as their implicit and explicit Black-White racial preferences. After a month had elapsed, we randomly assigned participants to either the UI or TANF experiment and conducted the second stage, in which we collected participants' prior beliefs about the share of program beneficiaries who are Black. We then randomly provided half of the participants in each experiment with the true racial composition of beneficiaries. Their anticipated response to learning this information differs depending on their taste or distaste for supporting Black Americans through the social safety net, in other words, their levels of anti-Black *racial animus*. Our main outcome is the extent to which participants think current TANF or UI benefit amounts should be changed. One month later, in the third stage, we asked participants to provide their current beliefs about usage in, and support for, the policy to which they had been initially randomized to test for persistence. At this stage, we also asked for information about the other program to examine any spillovers.

To preview our hypotheses and results, consider first two individuals with the same

earnings, and individuals receive UI benefits only for a defined maximum number of weeks (usually around 26). TANF is a means-adjusted cash-assistance program for very low- or no-income families with strict work requirements. Relative to UI recipients, TANF recipients are lower in the income distribution. TANF benefits are calculated by taking the maximal benefit (approximately 35% of the federal poverty level) and then reducing from this the amount of any other income.

[3]Authors' calculations using data from https://datacatalog.urban.org/dataset/afdc-data-archive and https://www.census.gov/library/visualizations/interactive/social-safety-net-benefits.html.

underestimate of the proportion of Black beneficiaries, one in the control group and the other in the treatment group. If racial animus is a motivation, then the latter, after learning the correct proportion is higher than they thought, will support welfare less than the former, whose beliefs were not "corrected." Similarly, now consider a participant in the treatment group who overestimated the proportion of Black recipients, and, after learning the true number is less than they thought, supports welfare more than their counterpart in the control group. This response is also consistent with racial animus.

Indeed, we find causal evidence that racial animus and misbeliefs about Black Americans affect policy support for TANF in the United States, as posited. First, we find that for every 10 more welfare recipients mistakenly thought to be Black (out of 100), experimental participants in the control preferred to reduce monthly welfare payments by 40 percent, on average. Second, our causal treatment effect estimates suggest that correcting beliefs about the number of Black TANF recipients significantly reduces support for those who find there are more Black recipients than expected and increases it for those who find there are fewer than expected. The casual treatment effects are driven by White participants and are strongest among participants who scored higher on implicit and explicit measures of anti-Black bias in the first stage.

Interestingly, in what we consider our quasi-placebo experiment, we do not find evidence that, on average, racial animus and misbeliefs about Black Americans affect policy support for UI in the United States. First, misbeliefs about Black participation in UI do not significantly predict the average participant's level of policy support in the control condition. Second, despite being well-powered to detect one, there is no significant effect of the information treatment on the average participants' level of support for UI. In fact, after being treated, participants who either under- or overestimated the number of Black recipients have, on average, the same level of UI support as do their counterparts in the control group. We present suggestive evidence that, in the case of UI, racial preferences are "crowded out" by perceptions of beneficiary worthiness. Overall, participants view UI beneficiaries as being more worthy of help than welfare beneficiaries. Further, in the UI experiment, those who rate UI beneficiary deservingness the highest exhibit neither a racial preference (in the control) nor any significant affect of the information (in the treatment); however, those participants who do not think UI beneficiaries are deserving react more similarly to those in the welfare experiment.

Finally, we summarize some other notable findings meant to test the robustness of our main results. First, in the second stage of both experiments, approximately equal numbers of respondents underestimated and overestimated the true share of Black Americans on each program, giving us statistical power to conclude that our results are broadly symmetric on

either side of the correct belief. Second, when participants were asked whether they wanted to donate to other experimental participants who had benefited from TANF or UI, they did so in ways that are consistent with their stated support for each program, suggesting our results are not due to experimenter demand or hypothetical biases. Lastly, the results from the third stage indicate that the information interventions persistently changed beliefs in both experiments, at least one month later. This suggests that the differential treatment effects we see between TANF and UI are not due to differences in our protocol's ability to update participant beliefs.

Our primary contribution is to the literature that highlights the important role that racial attitudes and beliefs play in shaping social safety net policy preferences in the United States (Gilens (1995, 1996, 2000); Lee, Roemer, and Van der Straeten (2006); Luttmer (2001); Alesina et al. (2021); Haaland and Roth (2023); Kruse (2005); Quadagno (1994)). Across this body of research, beliefs about Black Americans, in particular, play an important role in determining choices and outcomes. Our results extend this work in two directions: first, by introducing an exogenous source of variation in beliefs about beneficiaries' race; and second, by showing that a causal relationship between beliefs, the preferences they activate, and social safety net policy support does exist, but only for a program for which desert may be questioned – i.e., the one with the concomitant racialized history.

Within the experimental literature, we also add to work demonstrating that information provision can shift beliefs (Akesson et al. (2022); Alesina et al. (2021); Haaland and Roth (2023, 2020); Kuziemko et al. (2015)). While beliefs may be updated, there is no guarantee the effect will extend to policy support (Akesson et al. (2022); Haaland and Roth (2023)). For example, Haaland and Roth (2023) elicit a quantitative metric of beliefs about racial discrimination by asking what share of resumes with Black-sounding names receive callbacks (after being informed the callback rate of resumes with White-sounding names). After the treatment group is told the true callback rate, they are more likely to update their beliefs about the extent of racial discrimination but are not any more likely to support pro-Black policies.

We, however, contribute to the small set of studies finding that providing information can also change policy support. In an earlier paper, Haaland and Roth (2020) show that respondents do change their policy views and are more supportive of immigration if they are informed about research showing that immigration has no adverse labor market impacts. In another example, Alesina et al. (2021) find that providing information about the extent of racial inequality in the U.S. increases Democrats' favorability toward redistribution and decreases Republicans'. In their omnibus experiment, Kuziemko et al. (2015) draw a distinction between social safety net and transfer policies, finding that their information intervention

has an effect on estate tax preferences, a small effect on minimum wage preferences, but has no effect on preferences for food stamps. We likewise find a differential effect of information across policies, though in our case both relate to the social safety net and we collect data to explicitly test a hypothesis about why (and the direction in which) the effects should differ.

Our welfare treatment effect's persistence is also relatively novel. Among the information interventions which explore the possibility of persistence, we are able to identify two studies which also find lasting effects. First, Haaland and Roth (2020) find that the effect of their experiment on policy support for low-skilled immigration persisted one week later. Kuziemko et al. (2015) find that the effect of their personalized information interventions are able to change preferences for estate taxes for at least a month. We likewise find that our interventions change beliefs for at least a month.

In work closest to ours, Akesson et al. (2022) link racial beliefs to support for welfare, finding that White Americans support the program less when they are presented with information about a subsample of welfare recipients which suggests that a high proportion of them are Black (juxtaposed against a condition suggesting that a low proportion of welfare recipients are Black). Relative to not receiving any information at all, however, Akesson et al. find no significant effect of information provision on welfare support. This starkly contrasts with our main result - we find evidence of a causal effect of providing information about the share of welfare recipients who are Black on policy support. We conjecture that two protocol differences may explain why we find strong effects. Unlike Akesson et al. (2022), we provided participants with base rate information (that is, the number of Black Americans in the overall population, 13%) which generated a distribution of beliefs that included substantial numbers of both under- and overestimates about both policies. By contrast, Akesson et al. find that the vast majority of experimental participants (86%) overestimated the share of welfare recipients who are Black. Another important distinction between the two experiments is that we present racial composition information about the entire U.S. population of welfare recipients (as opposed to a selected subsample), a difference Akesson et al. themselves speculate may have mattered.

Lastly, because we test whether racial animus affects two policies with differing narratives of recipient deservingness, we contribute to the literature on altruism and racialized views of worthiness (Candelo, de Oliveira, and Eckel, 2019). In a phenomenon they term "sympathy for the diligent," Drenik and Perez-Truglia (2018) find that experimental support for redistribution is determined by recipients' level of pre-policy effort. Such perceptions of effort may be correlated with race. Fong and Luttmer (2011), for example, find that there is an effect of race on charitable giving *only* as moderated through its effect on the perceived deservingness of respondents. Alesina et al. (2021) find that people who believe that economic gaps

between Black Americans and White Americans are most likely due to racism (as opposed to the level of Black Americans' effort) are also more likely to support redistributive policy interventions. While we hypothesize that the less egregious racialization of the history of UI (compared to welfare) should (and does) imply that most beneficiaries of UI are considered "deserving" overall, we still find that for respondents who do not think that UI recipients are deserving, animus-based treatment effects do begin to emerge. By contrast, we find that the treatment effect of racial information on welfare support persists *whether or not* respondents believed recipients to be at fault. This result suggests that the effect of racial beliefs on redistributive policy preferences may extend beyond the extent to which race proxies for perceived deservingness. In the end, our results imply that information provisions can serve to move redistributive policy preferences in a more equitable direction - even in the presence of racial animus, a particularly entrenched form of in-group bias in the United States.

The remainder of the paper is organized as follows. Section II describes the design of our two parallel experiments. Section III describes the experimental results. Section IV considers the role of persistence and attenuation between the first and second stages of the experiment. Section V concludes.

II Experimental Design

We implemented two parallel three-stage information provision experiments on Connect between the middle of June and the beginning of August, 2024.[4] The details of the experimental protocol appear in Appendix A. For both experiments, we targeted a nationally representative sample of participants within the United States. In the experiments, we elicited participant beliefs about the number of Black people who benefited in 2021 from either Temporary Assistance for Needy Families (TANF, commonly known as "welfare") or Unemployment Insurance (UI).[5] Then, at random, half of the participants were informed of the correct number of Black program beneficiaries in that year. Said differently, half of the participants had their beliefs "corrected" (or confirmed in the case that their belief matched the actual number). All participants were then asked to indicate their level of support for their respective social safety net program. Hence, our participants were randomized into one of four experimental conditions: *TANF Control*, *TANF Corrected*, *UI Control* or *UI Corrected*. This design assures that, within either the TANF-focused or the UI-focused experiment, the comparison of control and treated responses provides a causal estimate of the

[4]Connect is an online experimental platform. Recent evidence suggests that Connect provides high data quality and is comparable to other online experimental platforms (Gupta, Rigotti, and Wilson (2021)).

[5]This is the most recent year for which data was available for both programs.

effect of correcting one's beliefs about the recent usage of these programs by Black Americans on one's support for the social safety net. Importantly, some participants in the treatment groups found that their belief about Black usage was too low (i.e., there are more Black program beneficiaries than they thought), and others realized that they thought that Black usage was too high (i.e., there are fewer Black program beneficiaries than they thought). We can therefore compare people with similar initial beliefs who did, and did not, have their beliefs corrected to learn how providing correct information about the racial composition of safety net programs affects program support.

Each experiment had three stages. Stage 1 ran during the second week of June and collected individual characteristics and racial preferences (both implicit and explicit) for Black relative to White people from a nationally representative sample of U.S. participants. Individual characteristics included basic demographics (e.g., race, age, sex, education, income), work status, and information about the respondent's political leanings (e.g., their ideology on a scale from liberal to conservative and who they voted for - or would have voted for - in the 2020 presidential election).

To observe explicit racial bias, we asked respondents directly: "Which statement best describes you?" At 0 was the statement, "I prefer Black people to White people," at 5 was the statement, "I like White people and Black people equally" and at 10 was, "I prefer White people to Black people." The respondent could move the resulting slider (which started at 5 but needed to be clicked to register a 5) in either direction and could see the resulting numeric response. To measure implicit racial bias, Stage 1 participants completed the Implicit Association Test (IAT) (Greenwald, McGhee, and Schwartz, 1998). The IAT is a tool developed by social psychologists to measure bias indirectly, thereby minimizing any social desirability bias that can confound self-reports and it is widely used in economics to measure implicit bias (Alesina, Carlana, La Ferrara, and Pinotti (2024); Corno, La Ferrara, and Burns (2022); Glover, Pallais, and Pariente (2017); Carlana (2019)). During an IAT, participants attempt to quickly and accurately sort pictures and words into groups. The more natural a pairing appears to a participant (or the closer the items' "implicit association" for the participant), the more easily (and quickly) that participants can place these notions together in an IAT. In our racial IAT, the pictures were either of Black or White people and the words were either positive (e.g., happy) or negative (e.g., hate). We can then measure a participant's implicit bias with respect to a given racial comparison by how quickly the person can associate the a photo of someone of a given race with either the positive or negative words. The resulting measure ranges between -2 (indicating a strong implicit bias against Black people) and 2 (indicating a strong implicit bias against White people). Zero indicates no bias in either direction. In the end, the median Stage 1 participant spent 6.9

minutes and earned a flat fee of $1.25 (the equivalent of $10.70 per hour).

To provide some separation and to minimize any experimenter demand effects or urge for consistency, we started Stage 2 of the experiment approximately a month after Stage 1 was completed. Without any cajoling, 94% of the Stage 1 respondents returned for Stage 2 and were randomly sorted into one of the four experimental conditions.[6] Stage 2 began by informing all the participants that they would be paid a $1 bonus if they correctly responded to the one of the following two belief questions to which they were randomized (a correct response was within 2 people of the true number):

- *Out of every 100 adults who received welfare from the U.S. government (sometimes referred to as TANF or Temporary Assistance for Needy Families) in 2021, how many do you think identified as Black?*

- *Out of every 100 adults who received unemployment benefits from the U.S. government in 2021, how many do you think identified as Black?*

At the time of the experiments, this information was not immediately available through search engine results and, as expected, the vast majority of participants answered these questions incorrectly. We frame the elicitation question specifically in terms of *Black* recipients, thereby directly engaging the core racial misbelief in which we are interested. While misbeliefs about the number of Black beneficiaries may be driven by misperceptions of Black population density or may be endogenous to racial animus itself, the effect of biased belief – regardless of the source of bias – is precisely what we seek to quantify. Eliciting beliefs about programs' racial composition without directly referring to Blacks risks measuring racial beliefs while leaving racialized misbeliefs about Blacks unobserved. To attenuate base rate neglect, just before being asked the question, all the participants were informed that during 2021, 13 out of every 100 adults in the U.S. identified as Black. Further, after answering, all the participants then revealed (on a scale from 0 to 10) how confident they were in their belief.

Those participants randomized into the treatment conditions were then told the correct information (29 people for TANF and 18 for UI) and, to prevent any confusion, were asked to compare their belief to the actual number by correctly answering the question: "How does your answer of [*belief inserted*] compare to the actual number?"

Directly after stating their beliefs and having them corrected or confirmed in the treatment conditions, we asked participants our two primary outcome questions. First, we told participants what the typical TANF or UI benefit was per month in 2021 and asked them

[6]As expected with 94% returning, there is no selection on observables that differentiates the Stage 1 participants that returned from the entire Stage 1 population.

how the benefit should change.[7] They could respond on a vertical slider anywhere from −100% (end the benefit altogether) to +100% (double the size of the benefit). We consider this our unincentivized policy support outcome measure.

We next elicited our second primary outcome measure - a version of the Dictator Game (Forsythe, Horowitz, Savin, and Sefton, 1994; Eckel and Grossman, 1996; Carpenter, Connolly, and Myers, 2008), in which participants were given a $1 bonus and told that they could donate any fraction of the dollar to another participant who, in the past five years, had received either TANF or UI, depending on the experiment.[8] After the experiments concluded, these donations were randomly assigned and bonused to other participants who had participated in the programs. Hence, our respondents incurred a direct monetary cost to show their support for someone who qualified and received program benefits recently. These donations are our incentivized policy support outcomes, which we designed to further reduce any experimenter demand effects.

Stage 2 concluded with four questions about how much participants believed that TANF and UI recipients were at fault for their situations and the extent to which they deserved support from taxpayers. In the end, Stage 2 was completed quickly – the median time was 4.1 minutes – for which participants received a flat fee of $1 and the potential to earn up to an additional $2 in bonuses (the equivalent of up to $43.90 per hour).

Stage 3 of the experiments was run a month after Stage 2 and was designed to test the persistence of our information interventions. Participants were given a $1 flat fee and provided additional incentives (like those in Stage 2) to provide their current beliefs about the number of Black people who benefited from *both* programs in 2021. Not only does this allow us to test the stability of uncorrected beliefs and the persistence of corrected beliefs, we can also examine any spillover effects from beliefs about one policy onto the other. Importantly, all bonus payments accruing in Stages 2 and 3 were made after Stage 3 was completed so that participants in the control groups were not inadvertently provided information about the actual rates of Black beneficiaries in the two social safety net programs.[9] We received responses from 82% of the Stage 2 participants and Stage 3 was also completed quickly by most participants (the median completion time was 3.8 minutes and pay per hour was up to

[7]According to the Center on Budget and Policy Priorities, in the median state in 2021 the maximum monthly TANF benefit received by a mother and two children was $498. According to the U.S. Department of Labor, in the median state in 2021 the maximum monthly UI benefit received was $1852. At the time of the experiments, 2021 was the most recent year for which the racial composition of TANF and UI recipients was reported by the Census Bureau at `www.census.gov/library/visualizations/interactive/social-safety-net-benefits.html`.

[8]As part of the demographic information collected in Stage 1, we asked recipients if they had received TANF or UI benefits within the previous five years.

[9]That is, if control participants received a belief bonus for Stage 2, they would know that they were within 2 of the correct response.

$23.68).[10]

III Results

III.A Sample Characteristics

Based on our pilot study (December, 2023; $N = 380$) and the rule of thumb developed in Haaland, Roth, and Wohlfart (2023), we planned to gather between 600 and 700 observations per treatment arm in Stage 2, enough to detect a 0.15 standard deviation effect (with power of 0.8 and significance equal to 0.05). With this goal in mind, we targeted 3,000 Stage 1 observations and ultimately recruited 3,029 participants, as seen in Table I. We also see in Table I that we collected Stage 2 data from 2,834 of the Stage 1 participants which surpasses our upper goal of 2,800 and represents a 94% return rate.[11] Lastly, 2,324 Stage 2 participants returned for Stage 3.

Also, notice in Table I that the demographics on which the Connect survey platform balanced our sample to make it representative of the U.S. population match to a great extent. The only difference of note between the U.S. Census data and our Stage 1 sample is that we were unable to get as many participants aged 65 or older as we would have liked. That said, our sample matches the population relatively well on other demographics. Further, as shown in Table II, randomization worked as intended and we achieved balance in our two experiments.

III.B Stage 1: Measures of Racial Animus

On average, both of the measures we collected in Stage 1 indicate some racial animus in our representative sample. Recall that the IAT measure is bounded between -2, which indicates a strong implicit bias against Black people, and 2, which shows a strong bias against White people. The IAT scores in our sample range from a low of -1.63 to a high of 1.74 and have a mean of -0.43, suggesting some bias against Black people across much of the sample.[12] Similarly, our explicit measure of racial preference ranged from 0 (a strong preference for Black people over White people) to 10 (a strong preference for White over Black people) and averaged 5.3 in our sample. In Figure I, we plot the individual level

[10] Considering selection into Stage 3, the only characteristic that differs from the entire Stage 2 population at the 5% level is age and the difference is small: Stage 3 participants are 1.4 years older, on average.

[11] The full Stage 2 sample is slightly larger because there are another 12 observations from people who did not participate in Stage 1 and for whom we do not have demographic information.

[12] As is common, the IAT test failed to compute scores for 2.6% of our sample because some respondents made too many errors in the sorting task.

pairs of these two measures along with the linear fit (and 95% confidence intervals) and find that in addition to there being considerable variation in each score, the two measures are highly correlated ($\rho = -0.256$, $p < 0.01$), as anticipated. Respondents who report a stronger explicit preference for White people over Black people also tend to be more implicitly biased in the same direction.

In Appendix Table B1, we report the demographic determinants of these two measures of animus. We find that both White and Black participants have significantly different scores from the omitted racial categories. While the IAT scores are lower and the explicit scores are higher for Whites, the opposite is true for Black participants; this provides some sense of the internal validity of these measures. We also find that politically conservative participants and those who report voting for Donald Trump in the 2020 presidential election both have higher explicit racism scores and that the Trump voters also have lower implicit scores. Lastly, older participants reveal more racial animus and women exhibit less on both scales.

III.C Stage 2: Beliefs

The prior beliefs that our participants held at the start of Stage 2 about how many Black people made use of TANF and UI in 2021 are summarized in Figure II. Welfare beliefs vary from 1 to 100 with a mean of 30, and are very close to accurate, on average. The participant mean is only significantly different from the true number (29) at the 5% level ($t = 2.03$, $p = 0.04$). We also find that participants reveal varying levels of confidence in these beliefs. While the mean confidence in the TANF experiment is 4.54, a quarter of people report confidence of 2 or lower and almost 10% are very confident (8 or higher). Interestingly, being confident in one's belief about TANF participation does not make our participants more accurate. Regressing the absolute difference between one's belief and the true number on confidence returns a positive coefficient ($\beta = 0.255$, $p = 0.04$).

Considering the determinants of participant TANF beliefs, Appendix Table B2 confirms that there is no significant difference in the prior beliefs held by the TANF experimental participants in the two treatment conditions. In terms of demographic determinants of beliefs, we only find that older participants have lower beliefs and that participants from the South of the U.S. have marginally significantly higher beliefs (relative to people in the West). Appendix Table B2 also confirms that there is no significant correlation between the level of one's belief and one's confidence in that belief.

Beliefs about the number of Black UI participants appear on the right of Figure II. As with TANF, there is considerable variation in UI beliefs (stated proportions range from 0 to 98) and the mean of 23 is significantly higher than the true value of 18 ($t = 10.22$,

$p < 0.01$). Similarly to beliefs about TANF, UI beliefs are balanced almost equally on either side of the actual number (the median UI belief is 17 and the median TANF belief is 26), which is important for our experimental design and inference. In other words, for both experiments there were roughly equal numbers of participants receiving news that their beliefs were too high as received the news that their beliefs were too low. Belief confidence in the UI experiment is also varied, averages to 4.33, and is also a poor predictor of accuracy: regressing absolute UI misbelief on confidence results in a very small coefficient that is far from statistically significant ($\beta = -0.017$, $p = 0.90$).

In Appendix Table B3, we show that UI beliefs do not differ by treatment assignment and that confidence is not a significant predictor of the level of one's UI belief. Here, the only demographic determinant of respondent beliefs is participant age. Older participants believe that fewer Black people received UI in 2021.

III.D Stage 2: Average Treatment Effects

Before reporting the average treatment effects for both experiments, we summarize our outcomes: the level of support for each program and donations to previous program beneficiaries. For the welfare experiment, overall support is summarized on the left of Figure III. Here, the responses vary from doubling the typical benefit (100% increase) to ending it altogether (-100%) and the most common responses are to level fund the program, to increase it by about a quarter, or to double it. The average response is to increase TANF benefits by 34%. On the right of Figure III, we see that there are fewer people who would like to increase UI benefits by a substantial amount than would like to substantially increase TANF benefits. The average response for UI is to increase benefits by 22% and the modal response is to increase benefits by about a quarter.

Our participants were also relatively generous with the donations they made to previous beneficiaries of the two programs. Figure IV illustrates the fraction of the one-dollar endowment provided that the respondents donated to another participant who received benefits in the past five years from either TANF (left) or UI (right). In both experiments, while the share of respondents who kept all the money is between 35% and 45%, most people donate some amount, often either 50 cents or the entire dollar. That said, there are some differences between the two programs. On average, donations are about 10 cents lower in the UI experiment (37 versus 27 cents) and this is driven mostly by the number of people giving the whole one-dollar endowment being halved in the UI experiment. Lastly, one possible explanation for the level differences in the support and donations is that the UI benefit quoted in the experiment ($1,852 for an individual) already seems substantially more generous than the

TANF benefit ($498 for a mother and two children). While these differences are interesting, it is the within experiment variation in these outcomes that is important for inference.

For the average treatment effects of correcting racial misbeliefs about Black participation in social safety net programs, we estimate, for each outcome, the following "information gradient", with and without controls:

$$Y_i^o = \beta_0 + \beta_1 T_i + \beta_2(Belief_i - True_i) + \beta_3 T_i(Belief_i - True_i) + \beta_4 X_i + \epsilon_i$$

where Y_i^o is policy support for participant i and outcome o, T_i is an indicator for equal to one if the participant received correct information about the number of Black program beneficiaries, the difference $(Belief_i - True_i)$ is the perception gap between a respondent's prior belief and the true statistic, and X_i is a vector of controls.

For participants whose prior beliefs match reality (i.e., where $Belief_i - True_i = 0$), the estimating equation reduces to $Y_i^o = \beta_0 + \beta_1 T_i + \beta_4 X_i + \epsilon_i$ and we expect β_1 to be equal to 0 because the respondent hasn't learned anything new from the information. When participants have some misperception of the true number of Black people receiving benefits, we expect the slope of the belief gradient (β_2) to depend on the participant's attitude toward Black people. Participants who are either implicitly or explicitly biased against Black people should be identified by $\beta_2 < 0$, those who hold no relative preference should result in $\beta_2 = 0$ (i.e., the race of program participants does not affect support), and for those who are biased in favor of Black people it should be the case that $\beta_2 > 0$.

Our main interest is in the coefficient β_3, which measures the causal effect of the information provided on a participant's support for each policy outcome. In particular, if participants are biased against Black program beneficiaries, we expect $\beta_3 > 0$ because when they learn that there are more (less) Black beneficiaries than they believed, they should support the policy less (more). That said, if participants have the opposite preference toward Black beneficiaries, the sign of β_3 should be reversed. We also consider the supplemental hypothesis $\beta_3 = -\beta_2$ that information is, in effect, the perfect antidote to misperception.

Our treatment effects are best represented graphically.[13] In Figure V, we present the estimated relationship between misbelief about participation and program support, separately for those participants in the control groups (solid lines) and those in the information provision treatment groups (dashed lines). In the TANF experiment (left), we see that β_2 is less than zero, indicating that, on average, beliefs about the number of welfare recipients who are Black do predict one's support for TANF funding in the control group. Specifically,

[13]To prevent the figures from becoming too cluttered, we represent the point estimates without confidence intervals; however, all the estimation details are also presented tabularly in either the main text or the appendices.

the more participants believe that the program serves the Black population, the less they support it. Further, we see in column (1) of Table III that this effect ($\beta_2 = -0.401$, $p < 0.01$) is significant. We also see that the gap between the lines at a misbelief of zero is about three percentage points but confirm in Table III that this small gap is not statistically significantly different than zero ($\beta_1 = -2.805$, $p = 0.22$), so indeed, those people who simply have their beliefs confirmed in the treatment group do not behave differently. Most importantly, we see that the slope of the information gradient is shallower for the treated respondents, indicating a positive treatment effect. Returning to Table III, we find that the estimate for $\beta_3 = 0.299$ ($p = 0.014$). Lastly, we see that these estimates change little in column (2) which includes demographic controls. This is no surprise, given we achieved random assignment to treatment condition.[14] Summing, we find causal evidence that correcting misperceptions about the participation of Black people in the TANF program affects their support for the program. On average, those people who incorrectly thought participation was higher (lower), support the program more (less) when they learn that fewer (more) Black people receive the benefits.

The results for UI are given on the right side of Figure V. Here we find that although the estimate of β_2 is negative, it is shallower than for the TANF program and column (3) of Table III indicates that the slope is not significantly different from zero ($\beta_2 = -0.095$, $p = 0.27$). Interestingly (and as hypothesized), beliefs about Black participation in the unemployment program do not correlate very strongly with how much our average participant supported the program. Like the TANF results, we find that the UI estimate of β_1 is also small and insignificant ($\beta_1 = -0.147$, $p = 0.94$), indicating, as posited, that simply confirming someone's prior has little affect on their support for the program. Considering β_3 in the UI experiment, the estimated effect of providing information on the actual number of Black beneficiaries is also negligible. This estimate is small, slightly negative, and not statistically different from zero ($\beta_3 = -0.066$, $p = 0.55$). All the estimates on the right of Figure V are effectively flat, consistent with our hypothesis about the (damped) racial politics of UI and, as a result, racial beliefs and information matter little for the level of UI program support. In this program, designed to help "deserving" workers who are trying to "pull themselves up by their bootstraps," our experiment indicates that racial animus appears to play a more muted role.

[14]The point estimates for all the controls are listed in Table B4 for the interested reader.

III.E Stage 2: Incentivized Outcomes and Experimenter Demand

Some may be concerned that the preceding estimates are based on hypothetical policy support outcomes which could be influenced by experimenter demand. Put differently, our participants may have sought cues or formed expectations about the purposes of our experiments, and sought to respond appropriately. To mitigate this concern, we also estimate the average treatment effects using incentivized donation outcomes. These treatment effects are presented in Figure VI. Notice that these estimates seem similar to the results from the unincentivized support measure shown in Figure V, especially when comparing across programs. On the left, for TANF recipient donations, we see that β_2 is again negative and significant according to the first column of Table IV ($\beta_2 = -0.002$, $p = 0.02$), indicating that overestimating the number of Black TANF recipients by 10 is associated with a 2 cent reduction in one's donation (compared to a 37 cent average). We also find that participants who have their beliefs confirmed instead of corrected donate almost 4 cents less, but the effect is only significant at the 10% level. What is most important, however, is that we estimate a positive and highly significant treatment effect ($\beta_3 = 0.003$, $p < 0.01$) that corroborates our policy support estimates. Again, we find causal evidence that racial animus affects our average (and mostly representative of the U.S. adult population) participant's support for TANF, this time with an incentivized measure of support.

As in Figure V, on the right of Figure VI and in Table IV, we see two results that are even more pronounced using donations to the UI recipients as the outcome. First, the information gradient is essentially flat ($\beta_2 = -0.00006$, $p = 0.93$), which suggests that for the average participant, racial preferences do not determine how much they donate to UI recipients. Second, correcting their misbeliefs changes this response very little ($\beta_3 = 0.0006$, $p = 0.51$). Further, none of the results summarized in Figure VI change when we control for demographics in columns (2) and (4) of Table IV.[15]

III.F Stage 2: Asymmetric Responses to Beliefs and Information

As another robustness test, we examined whether our estimates are symmetric on either side of the correct belief. Here we test two hypotheses. First, in our control conditions, does having a negative misbelief (i.e., $Belief < True$) correlate with support differently than having a positive one (i.e., $Belief > True$)? Second, in the treatment conditions, are respondents differentially motivated by one of the two kinds of surprises they can receive from the information provided? For participants with racial animus toward Black people, holding a low prior belief and being told there are actually more Black people benefiting

[15] Again, point estimates for the demographic variables appear in Appendix Table B5.

from the program might be considered "bad" news. Likewise, similar respondents who have high priors and find out that the true number is lower might consider the information to be "good" news. To this point, we have treated "good" and "bad" news symmetrically.

One way to test for asymmetric responses to prior beliefs and any surprises coming from the information provided is to create a "knot" where misbeliefs are equal to zero for each experiment and estimate linear spline functions on either side of this correct prior. In Table V, we report linear spline versions of all our average effect estimates (i.e., the estimates represented in Figures V and VI) and use F-tests to compare the slope of the two linear segments on either side of the knot. That is, as a first test of asymmetry, we simply ask whether any of these splines are kinked significantly as compared to the linear estimates (forced to be symmetrical) described above.

For the TANF support estimates in the first two columns of Table V, we see that both segments of the control estimates are significantly downward sloping and though the estimated slope in the treatment condition appears to be slightly upward sloping for respondents with negative misbeliefs while the slope on the other side of the knot in the treatment condition is downward sloping, neither estimate is significantly different from zero. Further, and most importantly for the asymmetry hypothesis, the F-tests return p-values of 0.75 and 0.39 for these two splines, indicating that the policy support responses for both control and treated participants in the TANF experiment do not seem to depend significancy on the sign of their prior misbeliefs or the type of surprise they receive. In addition, continuing to the right in Table V, we find that in just one of the eight cases (Column 3) do the slopes of the linear segments estimated change from one side of the zero misbelief knot to the other. Summarizing these regressions, treating positive and negative misbeliefs symmetrically seems to fit our data relatively well.

III.G Stage 2: Subgroup Effects

Given the context of our study, we expect the treatment effects we just discussed to be the average of a number of heterogeneous groups that respond differently to the information provided. Therefore, we split the sample to test whether the treatment effects differed (or were stronger) among various pre-registered subgroups for whom hypothesized differences were obvious: White participants, implicitly biased participants, explicitly biased participants, political conservatives and participants who were more confident in their prior beliefs (and for whom learning the truth might be more of a surprise).

Similar to the country as a whole, our sample is over 70% White. In the presence of in-group bias, we expect racial animus against Black people to be stronger among this

group compared to a population of non-Whites. In Figure VII, we illustrate the treatment effects for both programs and both outcomes split by race. The top two panels are for the welfare outcomes and the bottom two illustrate the unemployment subgroup treatment effects. Starting on the left, we see that the treatment effect is noticeably stronger than what was seen on average in Figure VI. Where β_3 was 0.299, on average, limiting the sample to just our White respondents, we find that β_3 rises to 0.501 ($p < 0.01$). This, and the other estimates associated with race as a subgroup, are compiled in Table VI. In addition to the fact that the belief gradient is noticeably steeper for the control participants in the White subgroup, we note that in the non-White sample β_3 is negative, meaning that among this group of respondents when someone finds out that there are fewer Black TANF recipients than they believed, their support actually falls and when they find out that there are more than they expected, their support rises – a reasonable pro-Black response. Lastly, in the second to last row of Table VI we see that a χ^2 test of equality among the two treatment effect estimates is strongly rejected: the treatment response of White participants is more consistent with racial animus.

In the top right panel of Figure VII, we find that when we change the outcome to incentivized donations to TANF recipients, the racial subgroup effects look similar. As columns (3) and (4) of Table VI indicate, while donations fall as beliefs rise among White participants, they are essentially flat for non-White participants (i.e., the racial composition does not affect donations) and the treatment effects are oppositely signed once more. Further, at the bottom of the table we confirm again that the treatment effects are significantly different ($\chi^2 = 5.61$, $p = 0.02$). Considering the welfare experiment as a whole, we find that our White participants react substantially more as if racial animus was a motivator than do the rest of our respondents.

Recall that for the UI experiment we expected null results and did not find average treatment effects of information provision in the previous section. The bottom two panels of Figure VII, indicate that this result was not because we were averaging over very different racial subgroup effects that canceled each other in the aggregate. For both outcomes, we see that the number of Black participants that one believes receive unemployment benefits has little affect on participant support for UI or on donations to the program's recipients. In addition, randomly telling some participants the true number does little to affect this support. And, crucially, these effects are common to both non-White and White participants, as confirmed in the last four columns of Table VI.

We now consider subgroups based on measures of racial animus. Recall that low (specifically negative) IAT scores identify participants who demonstrate an implicit bias against Black people. To binarize this characteristic, we split the sample at the median, -0.46, and

report the estimates of the subgroup treatment effects in Figure VIII and Table VII. Continuing the emerging pattern, we find implicit bias subgroup effects in the TANF experiment but not in the UI experiment. In the panel at the upper left of Figure VIII and in the first two columns of Table VII, we see that β_2 is significantly negative for both groups but twice as steep for the low IAT score group. We also find that the treatment effect (i.e., β_3) is zero for people with higher than median IAT scores but strongly positive for the low IAT group. Only those participants with a strong implicit bias against Black people react to being told there are more (less) of them receiving welfare benefits than they thought by reducing (increasing) their support for the program. Analyzing TANF recipient donations, in the upper right panel of Figure VIII and columns (3) and (4) of Table VII, we find similar results: donations in the control fall as misbeliefs increase but only significantly for those with a strong implicit bias and only among the same group of participants does finding out the true number of Black welfare recipients lead to a significant change in behavior.

The results of our UI experiment subgroup analysis based on implicit racism are similar to those based on respondent race. The bottom two panels of Figure VIII and the last four columns of Table VII indicate that beliefs about the racial composition of the beneficiaries of UI do not explain differences in how much individuals with lower than median IAT scores support the program, be it hypothetically, in terms of benefit levels (left), or directly by donating to program recipients (right). For both outcomes, our belief gradient estimates are close to zero and the treatment effects do not appear to be strong or change significantly for the subgroup of implicitly racist respondents.

As a second measure of racial animus against Black people in our sample, we evaluate subgroups based on whether a respondent explicitly told us that they had some preference for White people over Black people. It turns out that a median split of this measure overlaps with having a score larger than 5 (where 5 indicates having no preference). Despite some concern that experimenter demand might lead participants to not reveal a true preference for Whites over Blacks, we see that these results are even stronger than those we found among our race and implicit bias subgroups. The estimates are illustrated in Figure IX and listed in Table VIII. In the TANF experiment, we find that $\beta_2 < 0$ for all groups and both outcomes, but the slopes are steeper among explicitly racist participants. Most importantly, the treatment effects ($\beta_3 > 0$) are only strong and significant for the explicitly racist participants. As a result, the treatment effects are highly significantly different for both outcomes (see results of χ^2 tests shown in Table VIII). The last thing to notice about the explicit subgroup is that $\beta_3 > -\beta_2$: the information antidote is considerably stronger than the effect of the misperception.

In the lower panels of Figure IX, we find that there are no large differences in how

explicit racists support UI or react to information about the actual number of Black people who receive UI program benefits. This is confirmed in our estimates listed on the right half of Table VIII. We note that for UI, while not statistically significant, β_2 is actually *positive* for the subgroup of explicit racists. Among this group there is some (also racist) sense that Black people are more deserving of assistance, if they "work" for it.

As a final side note on subgroups constructed based on measures of implicit or explicit racial attitudes, in the Appendix we present analyses in which we look at the behavior of the participants at the intersection of these two measures – the 17% of our sample who responded above a 5 on the explicit scale *and* had an IAT score below -0.46. In Appendix Figure B1 and the accompanying Appendix Table B6, we confirm that, as is now anticipated, the subgroup effects are even more pronounced among these participants and we conclude that racial animus is a strong driver of the average treatment effects discussed in the previous section.

We also posited that our subgroup of politically conservative respondents might react to the prompts, policies and information differently. The estimated belief gradients and treatment effects are described in Figure X and Table IX. As above, we see stronger responses in the TANF experiment than in the UI experiment. Beginning with TANF support (upper left and first two columns), we see that beliefs matter in the control conditions ($\beta_2 < 0$), more so among the conservatives, and only for the conservatives is the treatment effect (β_3) both positive and significant. The reactions are similar when analyzing the TANF recipient donations, but they are not as strong as among the subgroups of White, implicitly biased or explicitly racist respondents. The belief gradients again slope downward in the control condition, the treatment effects are positive and this effect is stronger among conservatives than non-conservatives, but the differences between the groups are more muted.

When we consider the UI experiment in the bottom panels of Figure X and columns (5) - (8) in Table IX, the effects are very similar to what we have already seen. None of the estimates differ significantly from zero, reaffirming that race seems to play less of a role in support for the UI program. However, what is interesting is that for UI policy support, conservatives respond to the information as predicted ($\beta_3 > 0$) but non-conservatives behave oppositely, similar to what we witnessed among the non-White subgroup in the TANF experiment. Here when non-conservatives find out that fewer Black people are on the program than they thought, they support it less. While neither of these estimates of β_3 are significantly different from zero, they are significantly different from each other ($\chi^2 = 5.03$, $p = 0.02$). Finally, we note that on the left of Figure X, we see the largest difference between conservative and non-conservative respondents: conservatives simply support these two social safety net programs less than the other participants. We see this most clearly by

considering the constants in the four relevant regressions. Comparing them, we find that both the conservative subgroup constant in the TANF analysis and the one in the UI analysis are significantly lower ($\chi^2 = 47.77$, $p < 0.01$ and $\chi^2 = 39.88$, $p < 0.01$, respectively). As we might expect, politically conservative participants are more hesitant to spend taxpayer money on social safety net programs, regardless of the racial composition of the beneficiaries.

In accordance with our pre-analysis plan, we estimate two more subgroup analyses and report the results in the Appendix. Briefly, we created a subgroup of people who did or would have voted for Trump in the previous presidential election. As you can imagine, the correlation between supporting this candidate and revealing a conservative political ideology is strong ($\rho = 0.67$) and so the experimental results, reported in Appendix Figure B2 and Appendix Table B7 look very similar to those in Figure X. We also find some evidence in Appendix Figure B3 and Appendix Table B8, that participants who are more confident in their beliefs than the median are more likely to react to information as predicted (i.e., $\beta_3 > 0$) in the TANF experiment than participants who are less confident. This result is consistent with the evidence that "surprises" are more salient in one's decision-making (Kahneman, 2011; Bordalo, Gennaioli, and Shleifer, 2022).

III.H Stage 2: Beneficiary Worthiness and Differences in Support for Welfare and Unemployment Insurance

As part of Stage 2 of the experiment, we asked participants their views on the suitability of beneficiaries of welfare and unemployment insurance, in general. The purpose of these questions was to gather evidence on the hypothesized differences between the programs. Is it the case that people in the U.S. view UI recipients as more worthy of help than TANF recipients and does this difference account, to some extent, for the attenuated treatment effect we find in the UI experiment?

For each program, participants were asked whether program recipients deserved the support of taxpayers and whether the situation of program recipients was no fault of their own. The responses to these four worthiness questions were recorded on a 10-point likert scale (where 0 indicated "strongly disagree" and 10 indicated "strongly agree"). Note that we asked all four questions to each participant, regardless of their experiment and condition assignment, with the purpose of accounting for any possible treatment effects on these responses.

On the left of Figure XI, we summarize the data on deservingness. Here we see, as posited, that our participants view unemployment insurance recipients as more deserving than welfare recipients (pooled $t = 7.82$, $p < 0.01$) and that the views of the participants in

a given experiment are not different from those in the other experiment ($t = 0.65$, $p = 0.52$ for TANF recipients and $t = 0.17$, $p = 0.86$ for UI recipients). On the right of Figure XI, we see an even larger difference in the extent to which our participants think that receiving unemployment insurance is no fault of the recipients compared to receiving welfare (pooled $t = 17.48$, $p < 0.01$). Again, we also see that these views are shared by respondents in both experiments ($t = 1.15$, $p = 0.25$ for TANF recipients and $t = 0.27$, $p = 0.78$ for UI recipients). In sum, we do find evidence that respondents in the U.S. think of unemployment insurance beneficiaries as being more worthy of taxpayer help than welfare beneficiaries.[16]

While the data in Figure XI are consistent with our priors about the two programs, the simple differences in worthiness do not directly account for the absence of a treatment effect in the UI experiment. To test this hypothesis more directly, consider the regression point estimates illustrated in Figure XII. Here we conduct another analysis of potential heterogeneous treatment effects on the support for the two programs, this time the sample being split by how worthy the respondents find the beneficiaries. Here, the hypothesis is that racialized impulses interact with perceived worthiness in the UI experiment. Specifically, we test whether increasing worthiness in the UI experiment crowds out the racialized impulses of participants who think program recipients are not worthy.

To examine this hypothesis, we first use principle component analysis to create an index of "worthiness" from the two highly correlated individual measures described in Figure XI. We then split the samples in Figure XII between our participants who find program recipients to not be very worthy of support (i.e., the lowest quintile) and the rest. If worthiness crowds out racial preferences in just the UI experiment, we expect the treatment effect of correcting beliefs to be mostly stable in the welfare experiment (i.e., regardless of worthiness) and to change from being similar to the welfare experiment for participants who think UI beneficiaries are not very worthy to a much attenuated response for those who do find them worthy.

Perhaps the first thing to notice in both panels of Figure XII, is that the auxiliary hypothesis that worthiness and support are positively correlated is supported. In both experiments, respondents who find program recipients to be more worthy also support the program more. More to the point, however, on the left of Figure XII, we see that correcting prior beliefs in the welfare experiment results in the effect documented above both when respondents think that the program beneficiaries are not very worthy and when they do. In this case, the treatment effect for the top four quintiles is not different than the one for the

[16] As another test of whether the responses to these four prompts differ by experimental condition, we conducted all 24 pairwise Kolmogorov-Smirnov comparisons of the CDFs and none were significant at lower than the $p = 0.18$ level.

bottom quintile of the worthiness distribution ($\chi^2 = 1.84$, $p = 0.17$).[17]

On the right of Figure XII, the results are noticeably different, however. In the UI experiment, when respondents do not think the recipients are very worthy, we see an information treatment effect similar to the those found in the welfare experiment (i.e., $\beta_2 < 0$ and $\beta_3 > 0$). At the same time, when our respondents think that UI beneficiaries are worthy, this effect not only attenuates, it flips to some extent (though β_3 is not significantly negative). What is most important is that while the treatment effect was mostly stable in the welfare experiment (left of Figure XII), it changes in the hypothesized direction in the UI experiment (right of Figure XII) and this change is marginally significant ($\chi^2 = 3.24$, $p = 0.07$). These findings provide preliminary evidence of worthiness crowding out racial preferences for the unemployment insurance program only.

IV Stage 3: Persistence and Attenuation

The third stages of our experiments were designed to examine the extent to which the misbeliefs of our participants persist over time if uncorrected, or attenuate if corrected. We also examine whether levels of support for the safety net persist both for uncorrected and corrected participants. Specifically, for our findings to extend beyond the immediate time frame surrounding our interventions, we expected that after a month participants in the control treatments would hold similar beliefs about the number of Black people utilizing TANF or UI and their support for these policies would be roughly similar, while participants whose beliefs were explicitly corrected would update them in the direction of the true statistics while the effects of our experimental intervention on their support would persevere.[18]

IV.A Stage 3: Beliefs

To empirically assess the amount of persistence in uncorrected beliefs and attenuation in corrected misbeliefs, we estimate:

$$(Belief_i^{+1} - True_i) = \gamma_0 + \gamma_1 T_i + \gamma_2 (Belief_i - True_i) + \gamma_3 T_i (Belief_i - True_i) + \gamma_4 X_i + \epsilon_i$$

where the left side is now the Stage 3 misbelief for participant i in either the welfare or unemployment experiments and the right side is unchanged from Section 3.4. Here γ_2 mea-

[17] See Appendix Table B9 for the details of these estimates.

[18] We have previously commented on some of the results referenced in this subsection in Carpenter, Debnam Guzman, Matthews, and Wolcott (2025).

sures the persistence of misbeliefs among participants in the control group and γ_3 measures the degree to which Stage 2 information provision translates into updated Stage 3 beliefs. Here, our prior is that $\gamma_3 \leq 0$: if information provision leads participants to update based on this information, greater second stage misbeliefs will be associated with smaller third stage misbeliefs.

We see in Figure XIII that participants in both experiments (TANF on the left and UI on the right) are roughly Bayesian. Misbeliefs of participants in the control conditions, separated by a month, are positively correlated, and as shown in Table X, these associations are large and highly statistically significant ($\gamma_2 = 0.478$, $p < 0.01$ for TANF and $\gamma_2 = 0.413$, $p < 0.01$ for UI). Hence, the misbeliefs of control respondents persist, on average, for at least a month.

Most importantly, we also see in Figure XIII and Table X that the information we provided about the true usage of TANF and UI, made a lasting impact on the beliefs of participants in the treatment condition of both experiments. Here, we see that γ_3 is negative, as hypothesized, and the effects are sizeable. From Figure XIII, we see that the linear estimates intersect very near a Stage 3 misbelief of 0 (i.e., you held the correct prior, learned that and restated it a month later) and the relatively flat nature of the estimates for the corrected respondents indicates that regardless of the magnitude of their initial misbeliefs, after a month most treated participants respond with a posterior belief much closer to the true statistic. The magnitude and significance of the effects of providing this information on Stage 3 misbeliefs in both experiments are confirmed in Table X, where we find $\gamma_3 = -0.348$ ($p < 0.01$) for the TANF experiment and $\gamma_3 = -0.263$ ($p < 0.01$) for the UI experiment. Lastly, Table X also indicates that adding controls changes these estimates very little. Summing, we find that, for both experiments, the misbeliefs of respondents who did not receive any new information (in the control treatments) reported posteriors a month later that were similar and highly correlated with their priors. By contrast, respondents who did receive information to correct their prior misbeliefs, largely believed it, report as a result considerably smaller posterior misbeliefs (i.e., beliefs much closer to the true value) and this treatment effect lasts at least a month.

IV.B Stage 3: Spillovers

Because we asked all the returning participants in Stage 3 to report their beliefs about the number of Black people benefiting from *both* TANF and UI, we can test whether people hold consistent beliefs in the control treatments and whether being corrected in one information treatment "spills over" to affect one's Stage 3 posterior belief about the usage of the other

program by Black people. In this case, we regress one's Stage 3 misbelief about policy $p2$ on one's Stage 2 prior misbelief about policy $p1$ (along with a treatment indicator and the interaction) or:

$$(Belief^{+1}_{i,p2} - True_{i,p2}) = \kappa_0 + \kappa_1 T_i + \kappa_2 (Belief_{i,p1} - True_{i,p1}) + \kappa_3 T_i (Belief_{i'p1} - True_{i,p1}) + \kappa_4 X_i + \epsilon_i.$$

Notice in Figure XIV that people who participated in the control conditions of both experiments hold policy misbeliefs that are consistent: over- or under-estimating the number of Black recipients in one program is associated with holding similar misbeliefs about participation in the other program. We also see this in Table XI, where κ_2 is estimated to be positive and significant for both experiments.

Behavior among the treated participants is at least as interesting. Here, we see that having your misbelief corrected about participation in one policy attenuates your Stage 3 (posterior) misbelief about both that program and the other program. On the left of Figure XIV, we see that TANF experiment participants who were told that the correct number for TANF participation was 29 have smaller misbeliefs about the UI program as well. Further, notice that the intercept is close to 11 (as verified in Table XI), indicating the possible use of the heuristic of assuming that the true value is the same for both programs (but being wrong and resulting in an UI misbelief of $29 - 18 = 11$). Moving to the right of Figure XIV, notice that there is also attenuation of TANF posterior misbeliefs by UI experimental participants who learned that the correct UI number was 18. Consistent with the left side of the figure, Table XI reports the significance of this difference from the control participant responses (i.e., $\kappa_3 < 0$) and indicates that there is sizeable spillover on TANF posteriors from being corrected about one's UI belief. This time, however, while the intercept is negative, it is not quite -11, suggesting weaker support for the "apply the same statistic to the other program" updating heuristic.

IV.C Stage 3: Policy Support

To complete our analysis, we examine whether the TANF policy support treatment effect we found in Section 3.4 and Figure V persists for at least a month as well (along with the lack of an effect in our placebo, UI, experiment). As a first test, in Figure XV, we examine the extent to which measures of policy support, elicited a month apart from one another, are consistent within both experiments. As is obvious, the two panels indicate that policy support changed little during the intervening month. On the left of Figure XV, we see that respondents report very similar levels of support for TANF one month later, regardless of whether they were

assigned to the control or had their misbelief corrected. Here the raw (pooled) correlation between responses is $\rho = 0.71$, $p < 0.01$, and it should be clear that the correlation differs little between control and treatment. The same is true of the UI experiment, depicted on the right of Figure XV. Here the overall raw correlation is $\rho = 0.64$, $p < 0.01$, and, again, this association is essentially unchanged between control and treatment. Elicited policy support levels are persistent.

Given the persistence of policy support, it is no surprise that the estimated relationships between Stage 2 misbeliefs and Stage 3 policy support look remarkably similar to the original estimates of the effect of Stage 2 misbeliefs on Stage 2 policy support from Section 3.4. These coefficients are illustrated in Figure XVI and listed in Table XII. Consider first the TANF experiment (left). In the control treatment the new estimate of β_2, the association between Stage 2 misbeliefs and Stage 3 policy support, is between -0.43 and -0.45 (depending on whether controls are included); this is quite close to the estimate of -0.40 reported in Table III. Also similar to Table III, our estimate of the treatment effect of having one's (Stage 2) misbeliefs corrected on this association one month later is positive as hypothesized and between 0.18 and 0.19, about two-thirds the magnitude of the original estimate shown in Table III. While this new estimate of β_3 is only significant at the $p = 0.15$ level, it is importantly not significantly different from the estimate in Table III ($\chi^2 = 1.04$, $p = 0.31$). In fact, none of the TANF policy support coefficients differ significantly when switching focus from Stage 2 policy support to Stage 3 support.

Likewise, the estimates for Stage 3 policy support in the UI experiment on the right of Figure XVI (and listed in the second two columns of Table XII) resemble those based on Stage 2 policy support in Figure V. As above, the estimate of β_2 for the UI experiment is considerably shallower than for the TANF experiment and the effect of the information provision is minimal. Put differently, like before with Stage 2 in Section 3.4, misbeliefs and racial animus continue to have little explanatory power in Stage 3 of our (placebo) UI experiment. Further, like the TANF experiment, none of the Stage 3 UI policy support coefficients in Table XII differ significantly from those reported in the Stage 2 analysis of Table III. Taken together, these results provide some evidence that, like beliefs, policy support, measured a month after our interventions, is persistent in both experiments, as is the effect of correcting racial misperceptions, but only, again, in the TANF experiment.

V Discussion and Concluding Remarks

In this paper we present evidence of a direct causal link between the provision information about the racial composition of the beneficiaries and levels of support for social safety net

expenditures. We consider together the two largest social safety net programs which provide direct income support to beneficiaries in the U.S. – welfare and unemployment insurance. Within a large representative online sample of U.S. participants, we find casual evidence that participants' support of welfare (TANF) can be changed by correcting their misbeliefs about the number of Black welfare beneficiaries. Treatment effects are robust to considering either a stated measure of policy support or considering participants' costly donations to previous welfare beneficiaries. These effects are driven by racial animus among White participants and by those who report above-median levels of anti-Black preferences. In a parallel experiment, we fail to reject the null hypothesis that correcting racial misbeliefs about the number of Black Unemployment Insurance beneficiaries has no effect on participants' support of this program.

Regardless of which measure of support we consider or whether we aggregate the sample or drill down into subgroups of participants with a self-reported bias against Blacks, updating participants about the number of Black beneficiaries matters little for their support for the Unemployment Insurance program. From our analysis of misbelief persistence and attenuation (Figure XIII) we know that UI experiment participants updated their beliefs almost identically to participants in the welfare experiment, so racial beliefs do not themselves explain the lack of a treatment effect for UI. Instead, the results illustrated in Figures XI and XII suggest that participants view recipients of Unemployment Insurance as more deserving of help, and therefore racial information matters less for their support of this policy. Further, widely expanded UI benefits represented an important part of the federal policy response to the COVID-19 crisis, so recent policy events may have further increased the perception of program beneficiaries' deservingness. Extensions of our work could examine the role of deservingness in support of the social safety net with special attention to disentangling a true preference for giving support to people in need from a desire to appeal to notions of deservingness to provide "cover" for biased racial preferences (Bursztyn, Haaland, Rao, and Roth (2020)).

Our individual measures of bias allow us to differentiate between a TANF treatment effect driven by anti-Black racial animus and one in which behavior is driven by rational statistical discrimination. Suppose there exists some characteristic disproportionately prevalent among Black welfare beneficiaries, the presence of which decreases our participants' utility of spending on the welfare program. In this case, statistical discrimination could explain the significant negative relationship we observe between the perceived proportion of welfare recipients who are Black and patterns of stated and incentivized program support. Our main treatment effect, however, is concentrated among the subgroup of participants who also display high levels of anti-Black racial animus. If statistical discrimination is animating

our participants, it does so only among those with the highest distaste for Blacks.

We can likewise rule out simple in-group bias as an explanation for our results. Our separate consideration of treatment effects within White and non-White subgroups is consistent with a narrative in which in-group bias motivates support for welfare. In the case of Unemployment Insurance, however, subgroup analysis is inconsistent with in-group bias. Further research might directly explore the role of in-group bias within social safety net policies. In this case, the experiment could reveal information about the proportion of social safety net program beneficiaries who belong to the participant's own racial group, similar in spirit to the analysis of Luttmer (2001). Subsequently observed changes in policy support would provide direct evidence on the role of in-group bias (rather than exploring anti-Black racial bias as is our aim).

Our work has been motivated by the narrative that beliefs about Black Americans have eroded support for redistributive policy in the U.S. (Gilens (1995, 1996, 2000); Quadagno (1994)). As such, our information provision experiment provided participants in the treatment group with updated racial information about the number of *Black* respondents. Within our predominantly (87.4% in the first stage) non-Black sample, we therefore measure the effect of shocking racial beliefs about this one, particular, outgroup population of interest. Importantly, this treatment is distinct from updating participants' beliefs about the proportion of beneficiaries belonging to some other racial group (e.g., White, Hispanic). Had we instead done this, participants would have been forced to infer, perhaps through subtraction, the consequent change in their beliefs about the number of Black participants. These inferences would have been heterogeneous - we would not been able to directly treat or observe the change in beliefs about Black recipients.

Our information provision experiment finds evidence of an effect of information on policy support where other related experiments have not (e.g., Akesson et al. (2022); Haaland and Roth (2023)). We can speculate about why this is the case. First, it seems that racial information is a particularly salient intervention. We find that the effects of our information provision persist for at least a month. Second, we find that our first stage beliefs about the number of program beneficiaries who are Black within both income support program are well-distributed on either side of the truth. For the latter fact, the provision of a base rate of the proportion of the U.S. population which is Black seems to have been an important experimental design choice, reducing measurement error in participant beliefs.

In establishing the importance of racial beliefs (and preferences) for support for redistributive policy, we cast light on the relative sparseness of the modern social safety net in the United States relative to other similarly wealthy nations, and on the decline of redistributive policy in the U.S. after the Civil Rights Movement. We find support for the longstanding

idea that the United States' ethnic heterogeneity undermines support for a social safety net which benefits members of long perceived (and reinforced) marginalized groups. Baseline beliefs about the racial composition of welfare programs are almost certainly endogenous to racial animus itself. Nonetheless we find that shocking these beliefs – even for those with high levels of implicit and explicit racial bias – can disrupt the mapping between participants' racial animus and their policy preferences. We do not pretend that our intervention reduced racial animus, but our evidence suggests that it "short-circuited" one manifestation of such animus. Even so, a natural policy recommendation emerges, to wit, that the provision of accurate information about, in this case, welfare recipients serves as a guardrail, limiting the effects of racial animus on support.

References

Akesson, J., R. W. Hahn, R. D. Metcalfe, and I. Rasooly (2022). Race and redistribution in the United States: an experimental analysis. Technical report, National Bureau of Economic Research.

Alesina, A., M. Carlana, E. La Ferrara, and P. Pinotti (2024, July). Revealing Stereotypes: Evidence from Immigrants in Schools. *American Economic Review 114*(7), 1916–1948.

Alesina, A., M. F. Ferroni, and S. Stantcheva (2021). Perceptions of racial gaps, their causes, and ways to reduce them. Technical report, National Bureau of Economic Research.

Alesina, A., E. L. Glaeser, and B. Sacerdote (2001). Why doesn't the united states have a european-style welfare state? *Brookings Papers on Economic Activity 2001*(2), 187–277.

Bordalo, P., N. Gennaioli, and A. Shleifer (2022). Salience. *Annual Review of Economics 14*, 512–544.

Bursztyn, L., I. K. Haaland, A. Rao, and C. P. Roth (2020, May). Disguising prejudice: Popular rationales as excuses for intolerant expression. Working Paper 27288, National Bureau of Economic Research.

Candelo, N., A. C. de Oliveira, and C. Eckel (2019). Worthiness versus self-interest in charitable giving: Evidence from a low-income, minority neighborhood. *Southern Economic Journal 85*(4), 1196–1216.

Carlana, M. (2019, August). Implicit Stereotypes: Evidence from Teachers' Gender Bias*. *The Quarterly Journal of Economics 134*(3), 1163–1224.

Carpenter, J., C. Connolly, and C. K. Myers (2008). Altruistic behavior in a representative dictator experiment. *Experimental Economics 11*, 282–298.

Carpenter, J., J. Debnam Guzman, P. H. Matthews, and E. Wolcott (2025, May). Can incorrect beliefs about the racial composition of welfare and unemployment insurance beneficiaries be changed? *AEA Papers and Proceedings 115*, 445–50.

Corno, L., E. La Ferrara, and J. Burns (2022, December). Interaction, Stereotypes, and Performance: Evidence from South Africa. *American Economic Review 112*(12), 3848–3875.

Drenik, A. and R. Perez-Truglia (2018). Sympathy for the diligent and the demand for workfare. *Journal of Economic Behavior & Organization 153*, 77–102.

Eckel, C. C. and P. J. Grossman (1996). Altruism in anonymous dictator games. *Games and economic behavior 16*(2), 181–191.

Fong, C. M. and E. F. Luttmer (2011). Do fairness and race matter in generosity? evidence from a nationally representative charity experiment. *Journal of Public Economics 95*(5-6), 372–394.

Forsythe, R., J. L. Horowitz, N. E. Savin, and M. Sefton (1994). Fairness in simple bargaining experiments. *Games and Economic behavior 6*(3), 347–369.

Gilens, M. (1995). Racial attitudes and opposition to welfare. *The Journal of Politics 57*(4), 994–1014.

Gilens, M. (1996). "race coding" and white opposition to welfare. *American Political Science Review 90*(3), 593–604.

Gilens, M. (2000, October). *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*. Studies in Communication, Media, and Public Opinion. Chicago, IL: University of Chicago Press.

Glover, D., A. Pallais, and W. Pariente (2017, August). Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores*. *The Quarterly Journal of Economics 132*(3), 1219–1260.

Greenwald, A. G., D. E. McGhee, and J. L. Schwartz (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology 74*(6), 1464.

Gupta, N., L. Rigotti, and A. Wilson (2021). The experimenters' dilemma: inferential preferences over populations. *arXiv preprint arXiv:2107.05064*.

Haaland, I. and C. Roth (2020). Labor market concerns and support for immigration. *Journal of Public Economics 191*, 104256.

Haaland, I. and C. Roth (2023). Beliefs about racial discrimination and support for pro-black policies. *Review of Economics and Statistics 105*(1), 40–53.

Haaland, I., C. Roth, and J. Wohlfart (2023). Designing information provision experiments. *Journal of economic literature 61*(1), 3–40.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Katznelson, I. (2005). *When affirmative action was white: An untold history of racial inequality in twentieth-century America*. WW Norton & Company.

Kruse, K. M. (2005). *White Flight: Atlanta and the Making of Modern Conservatism*. Princeton University Press. Google-Books-ID: c5763Zgu4_oC.

Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review 105*(4), 1478–1508.

Lee, W., J. Roemer, and K. Van der Straeten (2006, May). Racism, Xenophobia, and Redistribution. *Journal of the European Economic Association 4*(2-3), 446–454.

Luttmer, E. F. (2001). Group loyalty and the taste for redistribution. *Journal of political Economy 109*(3), 500–528.

McGhee, H. (2021). *The Sum of Us: What Racism Costs Everyone and How We Can Prosper Together*. New York: Random House Publishing Group.

Quadagno, J. S. (1994). *The color of welfare: How racism undermined the war on poverty*. Oxford University Press.

Robin, C. (2018). *The reactionary mind: conservatism from Edmund Burke to Donald Trump*. Oxford University Press.

Table I: Participant characteristics.

	CPS	*Stage* 1		*Stage* 2		*Stage* 3	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
Age (18 - 64)	0.774	0.890	0.313	0.890	0.313	0.870	0.336
Female	0.505	0.511	0.500	0.512	0.500	0.516	0.500
White	0.753	0.717	0.450	0.719	0.450	0.740	0.439
Black	0.137	0.126	0.332	0.125	0.330	0.103	0.304
Observations	-	3029		2834		2324	

Notes. This table shows averages and standard deviations of selected participant demographic characteristics. Characteristics are shown separately for each of stage of the parallel experiments. The column "CPS" contains average values of these selected demographic characteristics as reported in the 20XX Census Bureau Current Population Survey Data. The Census age categories reflect that fact that participants must be at least 18.

Table II: Treatment balance on observables.

	TANF control		*TANF treatment*		*UI control*		*UI treatment*	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age	42.570	15.004	43.632	15.331	42.667	14.570	42.99	15.134
Female	0.523	0.500	0.502	0.500	0.520	0.500	0.505	0.500
White	0.700	0.459	0.752	0.432	0.715	0.452	0.710	0.454
Black	0.124	0.330	0.105	0.306	0.137	0.345	0.131	0.337
College	0.433	0.496	0.424	0.494	0.411	0.492	0.420	0.494
Masters or more	0.159	0.366	0.167	0.373	0.157	0.364	0.156	0.363
Income over $75k	0.408	0.492	0.436	0.496	0.442	0.497	0.394	0.489
Observations	709		707		706		703	

Notes. This table shows averages and standard deviations of selected demographic characteristics across the treatment and control arms of the TANF experiment (columns (1) through (4)), and the treatment and control arms of the UI experiment (columns (5) through (8)). For the TANF experiment, the F -statistic from regressing an indicator variable equal to one if the participant was randomized to the treatment group on these characteristics is 1.16 ($p = 0.32$). For the UI experiment, the F-statistic from regressing an indicator variable equal to one if the participant was randomized to the treatment group on these characteristics is 0.76 ($p = 0.62$).

Table III: Policy support treatment effects.

	(1) TANF	(2) TANF+	(3) UI	(4) UI+
β_1: Corrected Welfare Belief	-2.805	-2.451		
	(2.284)	(2.290)		
β_2: Misbelief (Welfare Belief - 29)	-0.401***	-0.390***		
	(0.085)	(0.085)		
β_3: Corrected × Misbelief (Welfare Belief - 29)	0.299**	0.298**		
	(0.121)	(0.121)		
β_1: Corrected UI Belief			-0.147	-0.078
			(1.809)	(1.804)
β_2: Misbelief (UI Belief - 18)			-0.095	-0.125
			(0.086)	(0.086)
β_3: Corrected × Misbelief (UI Belief - 18)			-0.066	-0.057
			(0.111)	(0.109)
β_0: Constant	35.605***	37.742***	22.473***	29.605***
	(1.584)	(4.255)	(1.337)	(3.513)
Observations	1415	1412	1417	1408

Notes. This table reports the estimated treatment effects on policy support. The outcome variable is desired percentage change to the average value of program benefit (measured in percentage points). Columns (1) and (2) consider TANF and columns (3) and (4) consider UI; + indicates that controls have been added (controls include age, sex, education, income and geographic region). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IV: Program recipient donation treatment effects.

	(1) TANF	(2) TANF+	(3) UI	(4) UI+
β_1: Corrected Welfare Belief	-0.037*	-0.037*		
	(0.019)	(0.019)		
β_2: Misbelief (Welfare Belief - 29)	-0.002**	-0.001*		
	(0.001)	(0.001)		
β_3: Corrected × Misbelief (Welfare Belief - 29)	0.003***	0.003***		
	(0.001)	(0.001)		
β_1: Corrected UI Belief			-0.013	-0.015
			(0.018)	(0.017)
β_2: Misbelief (UI Belief - 18)			-0.000	-0.000
			(0.001)	(0.001)
β_3: Corrected × Misbelief (UI Belief - 18)			0.001	0.001
			(0.001)	(0.001)
β_0: Constant	0.384***	0.236***	0.279***	0.155***
	(0.014)	(0.036)	(0.012)	(0.033)
Observations	1416	1413	1417	1408

Notes. This table reports the estimated treatment effects on participant donations to previous program beneficiaries. The outcome variable is measured in dollars. Columns (1) and (2) consider TANF and Columns (3) and (4) consider UI; + indicates that controls have been added (controls include age, sex, education, income and geographic region). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table V: Splines of the average effects.

	(1) TANF	(2) TANF	(3) TANF	(4) TANF	(5) UI	(6) UI	(7) UI	(8) UI
Misbelief (Welfare Belief - 29) ≤ 0	-0.461**	0.061	-0.008***	-0.001				
	(0.197)	(0.212)	(0.002)	(0.002)				
Misbelief (Welfare Belief - 29) > 0	-0.368***	-0.188	0.002*	0.003***				
	(0.141)	(0.131)	(0.001)	(0.001)				
Misbelief (UI Belief - 18) ≤ 0					-0.440	0.037	-0.003	-0.003
					(0.282)	(0.272)	(0.003)	(0.003)
Misbelief (UI Belief - 18) > 0					-0.013	-0.209**	0.001	0.001
					(0.124)	(0.096)	(0.001)	(0.001)
Constant	34.882***	34.680***	0.309***	0.320***	20.232***	23.584***	0.258***	0.246***
	(2.746)	(2.504)	(0.024)	(0.023)	(2.337)	(1.989)	(0.021)	(0.021)
Dependent Variable	TANF Support		TANF Donation		UI Support		UI Donation	
Experimental condition	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment
F-test of estimate equality	$p = 0.75$	$p = 0.39$	$p < 0.01$	$p = 0.16$	$p = 0.24$	$p = 0.45$	$p = 0.24$	$p = 0.25$
Observations	709	706	709	707	710	707	710	707

Notes. This table reports estimated effects of racial misbeliefs on policy support, using spline regressions to allow the effect of overestimating the number of Black beneficiaries to differ from the effect of underestimating the number of Black beneficiaries. Specifications are shown separately for TANF policy support measures (column (1) through column (4)) and for UI policy support measures (column (5) through column (8)). The dependent variable is the desired percentage change to the value of program benefit (measured in percentage points) in columns (1), (2), (5), and (6). The dependent variable is donation to previous program beneficiaries (measured in dollars) in columns (3), (4), (7), and (8). Robust standard errors are reported in parentheses. The F-tests report p-values from the hypothesis that the estimates are the same on either side of the spline knot. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table VI: Subgroup treatment effects for White participants.

	(1) Others	(2) Whites	(3) Others	(4) Whites	(5) Others	(6) Whites	(7) Others	(8) Whites
β_1: Corrected Welfare Belief	-4.716	-2.548	-0.045	-0.039*				
	(4.166)	(2.740)	(0.036)	(0.023)				
β_2: Misbelief (Welfare Belief - 29)	-0.195	-0.473***	0.000	-0.002***				
	(0.152)	(0.100)	(0.001)	(0.001)				
β_3: Corrected × Misbelief	-0.371	0.501***	-0.001	0.004***				
	(0.230)	(0.140)	(0.002)	(0.001)				
β_1: Corrected UI Belief					1.025	-0.594	-0.045	-0.007
					(3.587)	(2.106)	(0.033)	(0.021)
β_2: Misbelief (UI Belief - 18)					-0.252*	-0.039	-0.001	0.000
					(0.130)	(0.113)	(0.001)	(0.001)
β_3: Corrected × Misbelief					0.114	-0.139	0.001	0.001
					(0.168)	(0.143)	(0.002)	(0.001)
β_0: Constant	35.443***	35.751***	0.355***	0.397***	27.163***	20.791***	0.295***	0.276***
	(2.741)	(1.941)	(0.025)	(0.016)	(2.757)	(1.533)	(0.024)	(0.015)
Dependent Variable	TANF Support		TANF Donation		UI Support		UI Donation	
χ^2 test of β_3 equality	$\chi^2 = 10.56$, $p < 0.01$		$\chi^2 = 5.61$, $p = 0.02$		$\chi^2 = 1.32$, $p = 0.25$		$\chi^2 = 0.08$, $p = 0.78$	
Observations	385	1027	386	1027	404	1004	404	1004

Notes. This table reports, for both experiments, heterogeneous treatment effects estimated separately for White and non-White participants. The dependent variable in columns (1), (2), (5), and (6) is the desired percentage change to the average value of program benefit (measured in percentage points). The dependent variable in columns (3), (4), (7), and (8) is participant donations to previous program beneficiaries (measured in dollars). χ^2 tests report the p-values from the hypothesis that the main treatment effect, β_3, is equal between White and non-White participants for a given experiment and policy outcome. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table VII: Subgroup treatment effects for low IAT participants.

	(1) Others	(2) Low IAT	(3) Others	(4) Low IAT	(5) Others	(6) Low IAT	(7) Others	(8) Low IAT
β_1: Corrected Welfare Belief	-6.146*	-0.342	-0.024	-0.052*				
	(3.190)	(3.360)	(0.028)	(0.027)				
β_2: Misbelief (Welfare Belief - 29)	-0.242**	-0.484***	-0.001	-0.002*				
	(0.116)	(0.116)	(0.001)	(0.001)				
β_3: Corrected × Misbelief	-0.001	0.456***	0.001	0.005***				
	(0.180)	(0.164)	(0.001)	(0.001)				
β_1: Corrected UI Belief					-2.303	2.482	-0.024	-0.014
					(2.643)	(2.488)	(0.026)	(0.025)
β_2: Misbelief (UI Belief - 18)					-0.234*	-0.007	-0.001	0.000
					(0.124)	(0.132)	(0.001)	(0.001)
β_3: Corrected × Misbelief					0.048	-0.149	0.001	0.001
					(0.155)	(0.171)	(0.001)	(0.001)
β_0: Constant	40.069***	31.803***	0.406***	0.364***	27.735***	16.610***	0.296***	0.268***
	(2.042)	(2.449)	(0.020)	(0.020)	(1.931)	(1.843)	(0.019)	(0.017)
Dependent Variable	TANF Support		TANF Donation		UI Support		UI Donation	
χ^2 test of β_3 equality	$\chi^2 = 3.53$, $p = 0.06$		$\chi^2 = 2.93$, $p = 0.09$		$\chi^2 = 0.73$, $p = 0.39$		$\chi^2 = 0.07$, $p = 0.79$	
Observations	696	683	697	683	691	676	691	676

Notes. This table reports, for both experiments, heterogeneous treatment effects estimated separately for participants with below-median IAT ("Low IAT") and participants with above-median IAT ("Others"). The dependent variable in columns (1), (2), (5), and (6) is the desired percentage change to the average value of program benefit (measured in percentage points). The dependent variable in columns (3), (4), (7), and (8) is participant donations to previous program beneficiaries (measured in dollars). χ^2 tests report the p-values from the hypothesis that the main treatment effect, β_3, is equal between participants with above- and below-median IAT scores for a given experiment and policy outcome. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table VIII: Subgroup treatment effects for explicit participants.

	(1) Others	(2) Explicit	(3) Others	(4) Explicit	(5) Others	(6) Explicit	(7) Others	(8) Explicit
β_1: Corrected Welfare Belief	-2.303	-3.481	-0.027	-0.060*				
	(2.617)	(4.364)	(0.024)	(0.034)				
β_2: Misbelief (Welfare Belief - 29)	-0.243***	-0.669***	-0.001	-0.003**				
	(0.088)	(0.168)	(0.001)	(0.001)				
β_3: Corrected × Misbelief	-0.013	0.925***	0.001	0.007***				
	(0.135)	(0.229)	(0.001)	(0.002)				
β_1: Corrected UI Belief					-0.508	1.710	-0.019	-0.013
					(2.108)	(3.384)	(0.022)	(0.029)
β_2: Misbelief (UI Belief - 18)					-0.192*	0.176	-0.000	0.001
					(0.102)	(0.160)	(0.001)	(0.001)
β_3: Corrected × Misbelief					0.003	-0.237	0.001	0.001
					(0.131)	(0.205)	(0.001)	(0.002)
β_0: Constant	39.806***	25.884***	0.401***	0.348***	26.161***	12.750***	0.306***	0.215***
	(1.784)	(3.146)	(0.016)	(0.025)	(1.531)	(2.592)	(0.015)	(0.022)
Dependent Variable	TANF Support		TANF Donation		UI Support		UI Donation	
χ^2 test of β_3 equality	$\chi^2 = 12.51$, $p < 0.01$		$\chi^2 = 7.21$, $p < 0.01$		$\chi^2 = 0.98$, $p = 0.32$		$\chi^2 = 0.01$, $p = 0.93$	
Observations	992	420	993	420	990	418	990	418

Notes. This table reports, for both experiments, heterogeneous treatment effects estimated separately for participants with a stated preference for White people over Black people ("Explicit") and for participants who do not state such a preference ("Others"). The dependent variable in columns (1), (2), (5), and (6) is the desired percentage change to the average value of program benefit (measured in percentage points). The dependent variable in columns (3), (4), (7), and (8) is participant donations to previous program beneficiaries (measured in dollars). χ^2 tests report the p-values from the hypothesis that the main treatment effect, β_3, is equal between participants with who do and do not report an explicit preference for White people for a given experiment and policy outcome. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IX: Subgroup treatment effects for participants who are politically conservative.

	(1) Others	(2) Conservative	(3) Others	(4) Conservative	(5) Others	(6) Conservative	(7) Others	(8) Conservative
β_1: Corrected Welfare Belief	-0.828	-6.930	-0.019	-0.093**				
	(2.419)	(4.936)	(0.023)	(0.036)				
β_2: Misbelief (Welfare Belief - 29)	-0.295***	-0.319**	-0.002**	-0.001				
	(0.093)	(0.160)	(0.001)	(0.001)				
β_3: Corrected × Misbelief	0.167	0.524**	0.003**	0.005***				
	(0.135)	(0.219)	(0.001)	(0.002)				
β_1: Corrected UI Belief					-0.423	1.097	-0.039*	0.044
					(2.129)	(3.104)	(0.021)	(0.032)
β_2: Misbelief (UI Belief - 18)					-0.021	-0.276	-0.000	0.000
					(0.094)	(0.169)	(0.001)	(0.001)
β_3: Corrected × Misbelief					-0.205	0.335	0.001	0.000
					(0.128)	(0.205)	(0.001)	(0.002)
β_0: Constant	42.118***	14.998***	0.389***	0.365***	26.831***	9.942***	0.306***	0.207***
	(1.679)	(3.566)	(0.016)	(0.028)	(1.603)	(2.153)	(0.015)	(0.022)
Dependent Variable	TANF Support		TANF Donation		UI Support		UI Donation	
χ^2 test of β_3 equality	$\chi^2 = 1.94$, $p = 0.16$		$\chi^2 = 1.14$, $p = 0.28$		$\chi^2 = 5.03$, $p = 0.02$		$\chi^2 = 0.13$, $p = 0.72$	
Observations	1054	358	1055	358	1033	375	1033	375

Notes. This table reports, for both experiments, heterogeneous treatment effects estimated separately for participants who are politically conservative ("Conservative") and for participants who are not ("Others"). The dependent variable in columns (1), (2), (5), and (6) is the desired percentage change to the average value of program benefit (measured in percentage points). The dependent variable in columns (3), (4), (7), and (8) is participant donations to previous program beneficiaries (measured in dollars). χ^2 tests report the p-values from the hypothesis that the main treatment effect, β_3, is equal between participants who are and are not politically conservative for a given experiment and policy outcome. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table X: Persistence of misbeliefs in Stage 3.

	(1) TANF	(2) TANF+	(3) UI	(4) UI+
γ_1: Corrected Welfare Belief	-0.783	-0.798		
	(0.990)	(0.987)		
γ_2: Misbelief (Welfare Belief - 29)	0.478***	0.482***		
	(0.040)	(0.040)		
γ_3: Corrected × Misbelief (Welfare Belief - 29)	-0.348***	-0.354***		
	(0.058)	(0.058)		
γ_1: Corrected UI Belief			-0.256	-0.264
			(0.932)	(0.918)
γ_2: Misbelief (UI Belief - 18)			0.413***	0.405***
			(0.047)	(0.047)
γ_3: Corrected × Misbelief (UI Belief - 18)			-0.263***	-0.256***
			(0.062)	(0.062)
γ_0: Constant	2.070***	3.560*	5.131***	6.302***
	(0.697)	(1.849)	(0.662)	(1.877)
Observations	1161	1159	1163	1155

Notes. This table reports the estimated treatment effects of the Stage 2 information provision on Stage 3 racial misbelief. The dependent variable in all specifications is the difference between participants' incentivized reports in Stage 3 of the proportion of the beneficiaries of their respective social safety net program who are Black, and the true proportion. The independent variable "Misbelief" refers to the same difference, elicited from participants approximately a month earlier during Stage 2 of the experiment. Columns (1) and (2) consider TANF and Columns (3) and (4) consider UI; + indicates that controls have been added (controls include age, sex, education, income and geographic region). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table XI: Testing for policy belief spillover

	(1) UI	(2) UI+	(3) TANF	(4) TANF+
κ_1: Corrected Welfare Belief	-1.243	-1.081		
	(0.944)	(0.942)		
κ_2: Misbelief (Welfare Belief - 29)	0.401***	0.396***		
	(0.038)	(0.038)		
κ_3: Corrected × Misbelief (Welfare Belief - 29)	-0.291***	-0.292***		
	(0.053)	(0.053)		
κ_1: Corrected UI Belief			-1.404	-1.301
			(1.040)	(1.038)
κ_2: Misbelief (UI Belief - 18)			0.374***	0.376***
			(0.049)	(0.050)
κ_3: Corrected × Misbelief (UI Belief - 18)			-0.245***	-0.248***
			(0.066)	(0.066)
κ_0: Constant	9.519***	11.994***	-0.568	-0.364
	(0.682)	(1.757)	(0.758)	(2.031)
Observations	1161	1159	1163	1155

Notes. This table reports the estimated treatment effects of the Stage 2 information provision on Stage 3 racial misbelief about the social safety net program featured in the *other* experiment. The independent variable "Misbelief" is the difference between participants' incentivized reports in Stage 2 of the proportion of the beneficiaries of their respective social safety net program who are Black, and the true proportion. The dependent variable "Misbelief" refers to the same difference, now elicited a month later about the beneficiaries of the social safety net program featured in the other experiment. Columns (1) and (2) consider the treatment effects of updating TANF beliefs on Stage 3 UI misbeliefs. Columns (3) and (4) consider treatment effects of updating UI beliefs on Stage 3 UI misbeliefs; + indicates that controls have been added (controls include age, sex, education, income and geographic region). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table XII: Stage 3 policy support treatment effects.

	(1) TANF	(2) TANF+	(3) UI	(4) UI+
β_1: Corrected Welfare Belief	-4.465*	-4.352*		
	(2.454)	(2.449)		
β_2: Misbelief (Welfare Belief - 29)	-0.448***	-0.432***		
	(0.086)	(0.085)		
β_3: Corrected × Misbelief (Welfare Belief - 29)	0.181	0.187		
	(0.132)	(0.131)		
β_1: Corrected UI Belief			2.496	2.649
			(2.082)	(2.083)
β_2: Misbelief (UI Belief - 18)			-0.182*	-0.211**
			(0.097)	(0.097)
β_3: Corrected × Misbelief (UI Belief - 18)			-0.018	-0.000
			(0.125)	(0.125)
β_0: Constant	40.867***	39.228***	20.329***	24.388***
	(1.730)	(4.600)	(1.565)	(4.211)
Observations	1161	1159	1163	1155

Notes. This table reports the estimated treatment effects on Stage 3 policy support (measured approximately one month after the completion of Stage 2). The outcome variable is desired percentage change to the average value of program benefit (measured in percentage points), as reported in Stage 3 of the experiment. Columns (1) and (2) consider TANF and Columns (3) and (4) consider UI; + indicates that controls have been added (controls include age, sex, education, income and geographic region). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

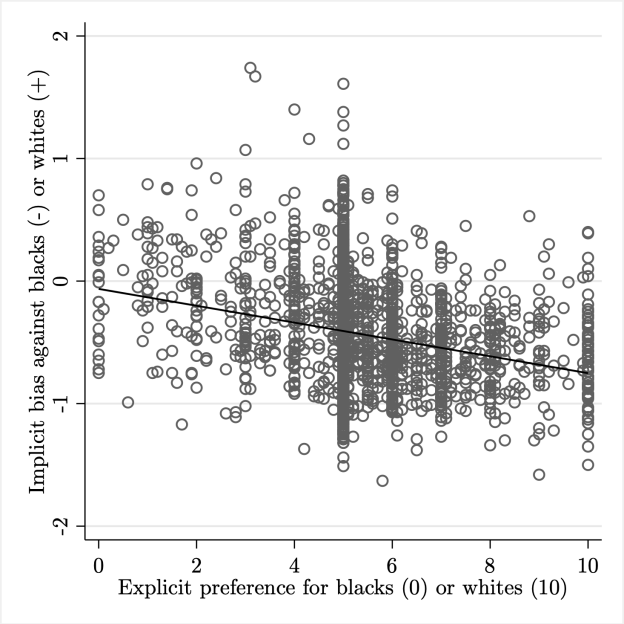


Figure I: Individual measures of explicit and implicit racial animus.

Notes. A scatterplot of participants' implicit and explicit biases against Blacks ($\rho = -0.26, p < 0.01$). Implicit bias is measured using an Implicit Association Test (IAT), which ranges from -2 (indicating a strong implicit bias against Black people) and 2 (indicating a strong implicit bias against White people). A score of 0 on the IAT indicates no implicit racial bias in either direction. Explicit bias is measured on a scale which indicates participants' self-assessed location on a scale between "I prefer Black people to White people" (0), to "I prefer White people to Black people" (10). A response of 5 indicates agreement with the statement "I like White people and Black people equally".

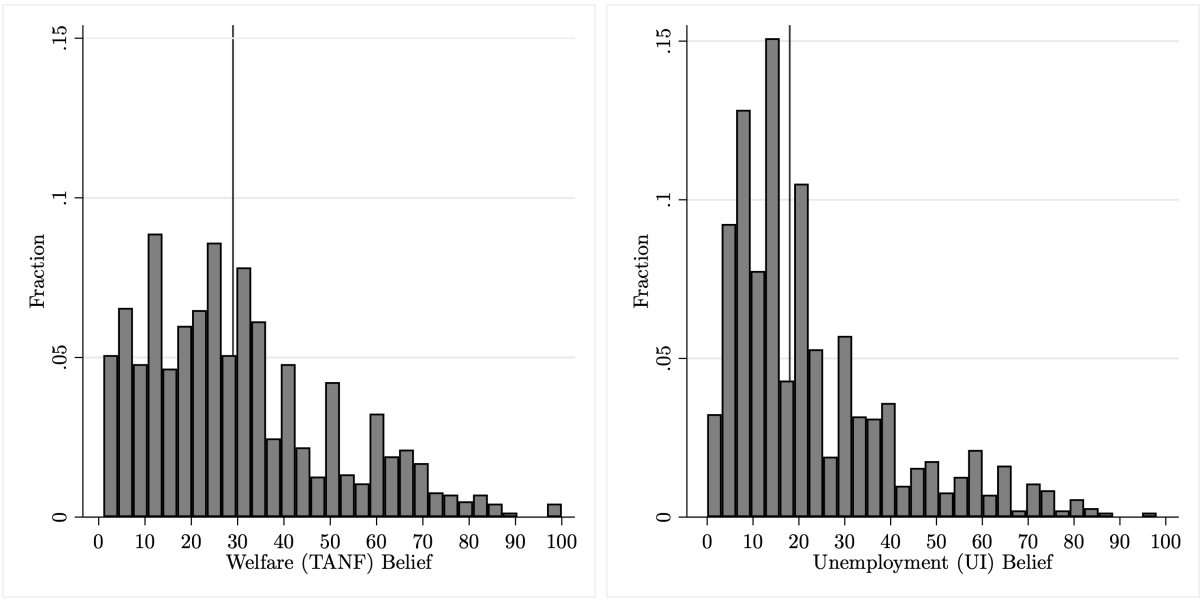


Figure II: Participant second-stage beliefs about the number of number of TANF, UI beneficiaries (out of 100) who are Black

Notes. This figure shows the density of participants' initial second-stage beliefs about the number of beneficiaries of TANF (left) and UI (right) out of 100 who are Black. Vertical lines indicate the true proportions.

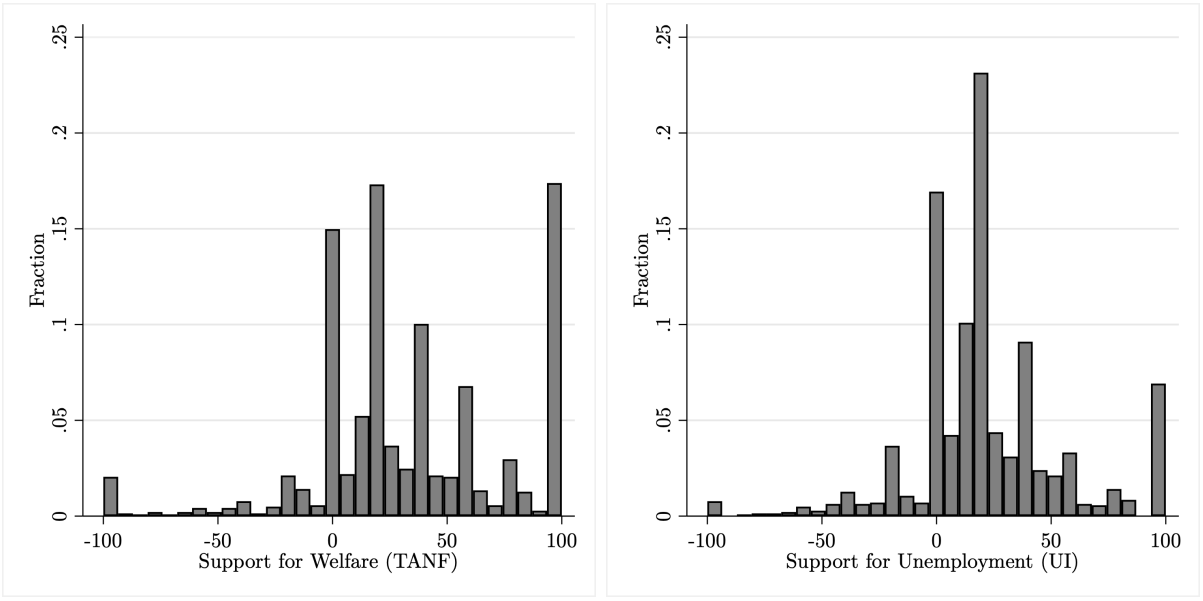


Figure III: Participant support for welfare (TANF) and unemployment insurance (UI).

Notes. This figure shows the density of participants' desired changes to the level of average of benefits for TANF (left) and UI (right). The figures include outcomes of participants in both the treatment and control groups. The x-axis is desired change to the average value of program benefit measured in percentage points.

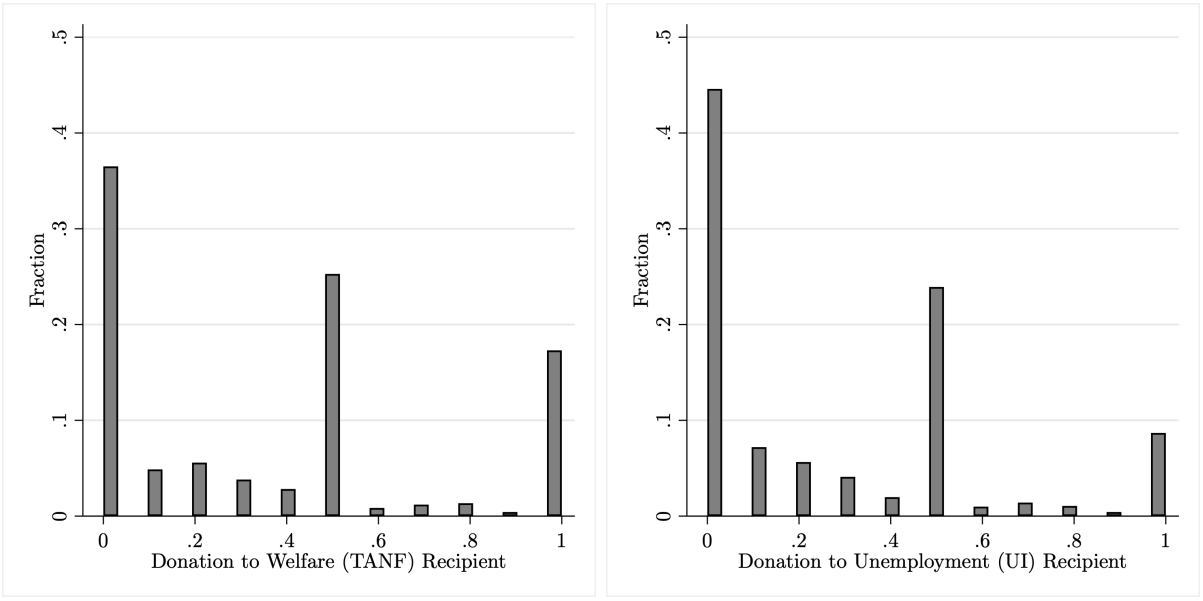


Figure IV: Participant donations to previous welfare (TANF) and unemployment (UI) recipients.

Notes. This figure shows the density of participant donations to a randomly selected beneficiary of TANF in the preceding 5 years (left) or beneficiary of UI in the preceding 5 years (right). Amounts shown on the horizontal axis indicate the fraction of a one-dollar bonus that the participant chose to donate.

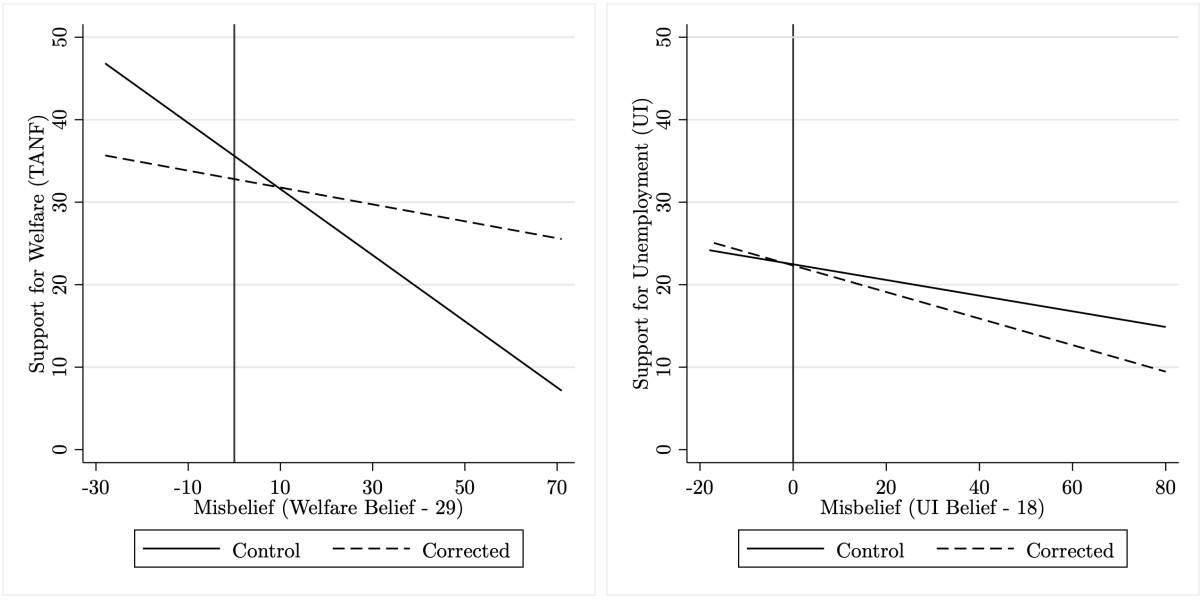


Figure V: Policy support average treatment effects.

Notes. This figure shows a linear prediction of the effect of racial misbelief on second-stage policy support, shown separately for control groups (solid) and for treatment participants whose beliefs have been updated (dashed). Policy support is measured by the desired percentage change to the average value of program benefit (in percentage points). Linear predictions reported separately for TANF (left) and for UI (right). The regression coefficients upon which these linear predictions are based are shown in Table III .

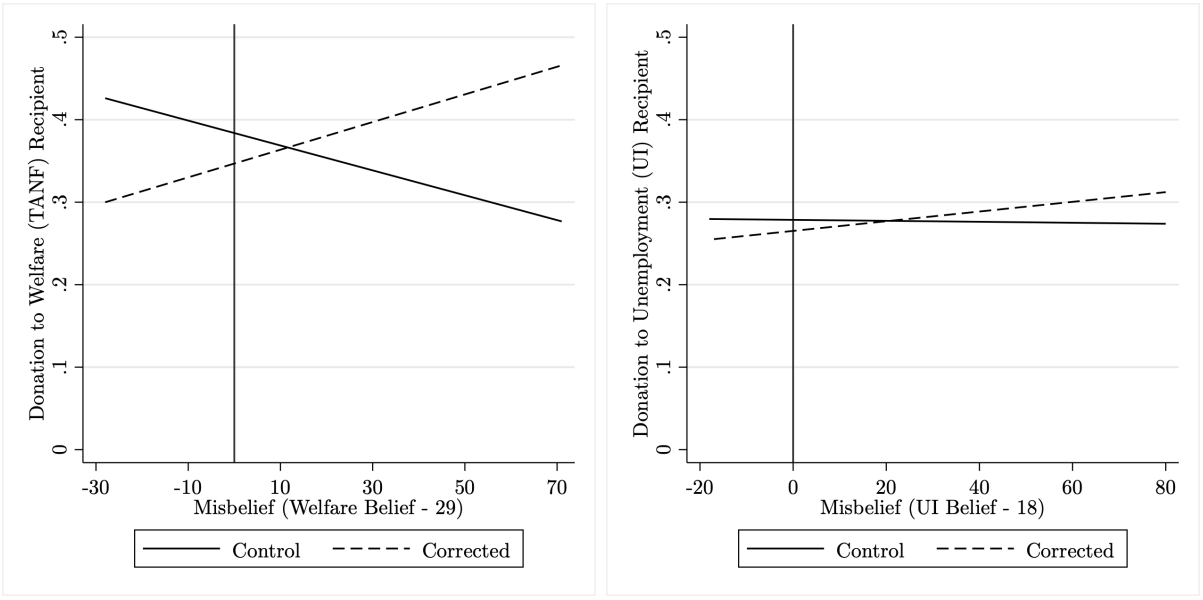


Figure VI: Program participant donation average treatment effects.

Notes. This figure shows a linear prediction of the effect of racial misbelief on second-stage policy support, shown separately for control groups (solid) and for treatment participants whose beliefs have been updated (dashed). In this case, policy support is measured by participant donations to a randomly selected previous beneficiary of TANF (left) or UI (right). Linear predictions reported separately for TANF (left) and for UI (right). The regression coefficients upon which these linear predictions are based are shown in Table IV .

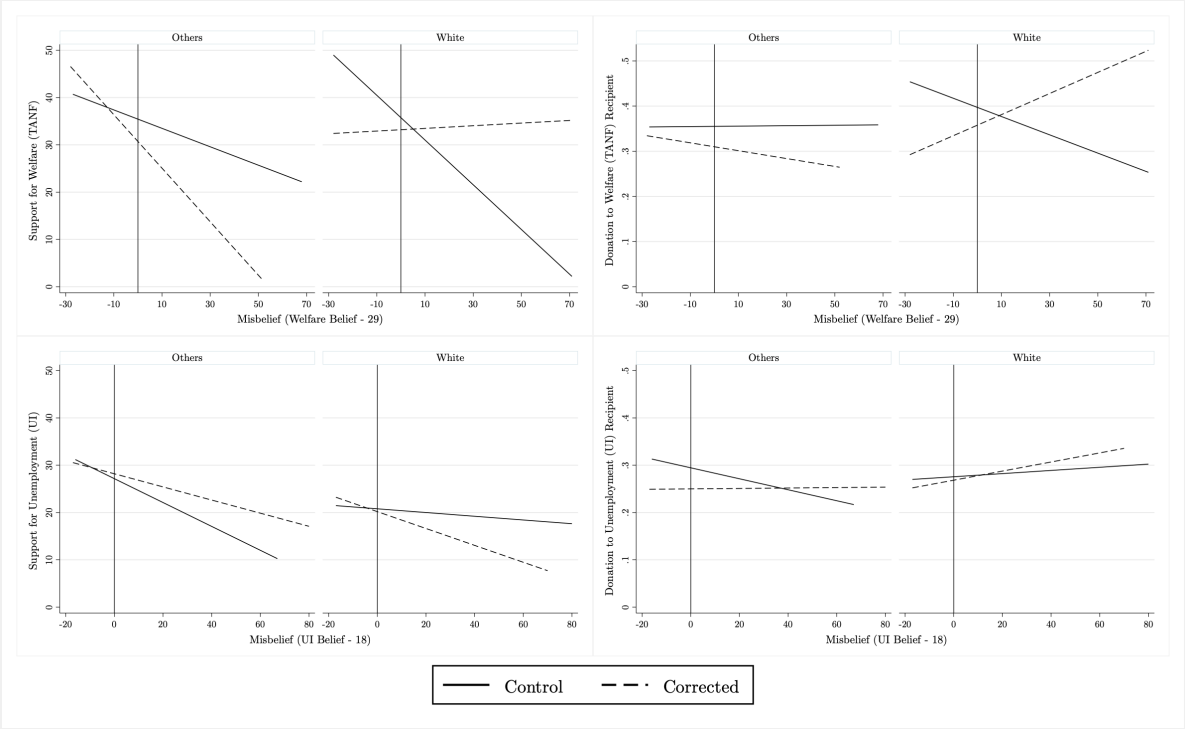


Figure VII: Comparing treatment effects between White and other participants.

Notes. This figure shows heterogeneous treatment effects between White participants and participants of other ethnicities within the UI experiment (panels C and D) and the TANF experiment (panels A and B). Each panel shows, separately for participants who are White (right) and who are not White (left), linear predictions of the effect of racial misbelief on second-stage policy support, shown separately for control groups (solid) and for treatment participants whose beliefs have been updated (dashed). In panels (A) and (C), policy support is measured by the desired percentage change to the average value of program benefit (measured in percentage points). In panels (B) and (D), policy support is measured by donations to previous program beneficiaries (measured in dollars). The regression coefficients upon which these linear predictions are based are shown in Table VI .

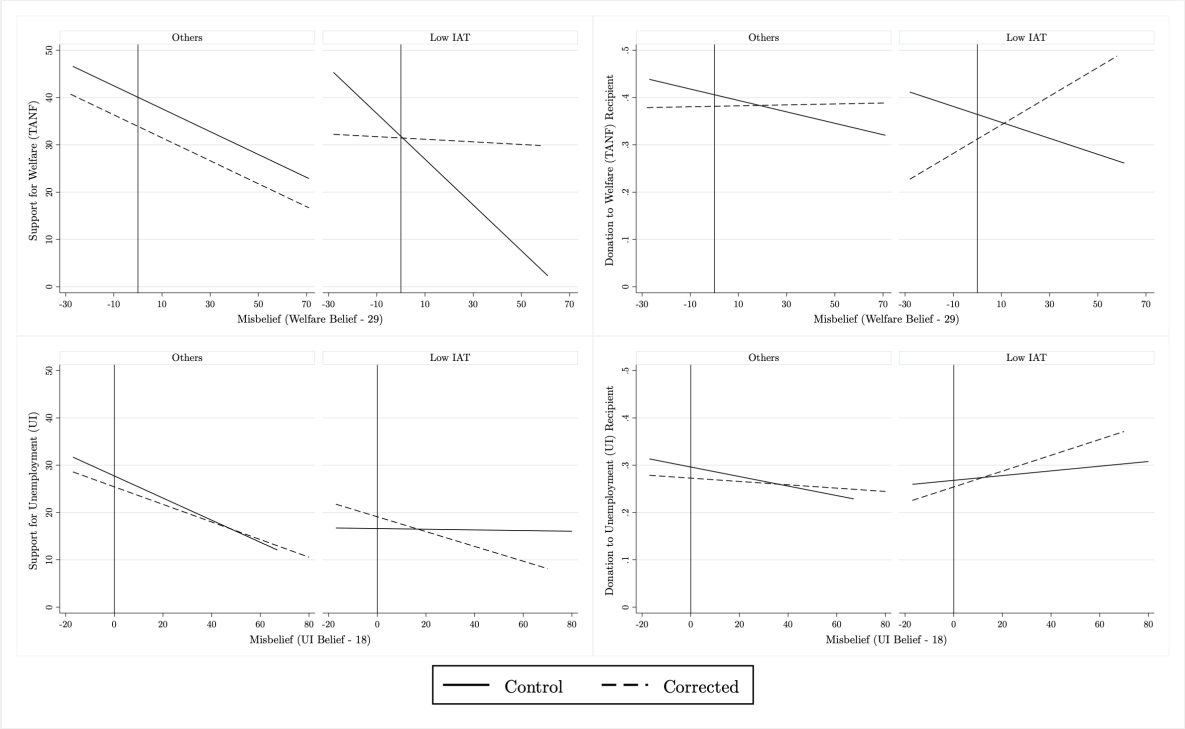


Figure VIII: Comparing treatment effects between implicitly biased (low IAT) and other participants.

Notes. This figure shows heterogeneous treatment effects between implicitly biased participants ("Low IAT", participants with below-median IAT) and participants with above-median IAT ("Others") within the UI experiment (panels C and D) and the TANF experiment (panels A and B). Each panel shows, separately for participants who are low IAT (right) and who are not low IAT (left), linear predictions of the effect of racial misbelief on second-stage policy support, shown separately for control groups (solid) and for treatment participants whose beliefs have been updated (dashed). In panels (A) and (C), policy support is measured by the desired percentage change to the average value of program benefit (measured in percentage points). In panels (B) and (D), policy support is measured by donations to previous program beneficiaries (measured in dollars). The regression coefficients upon which these linear predictions are based are shown in Table VII.

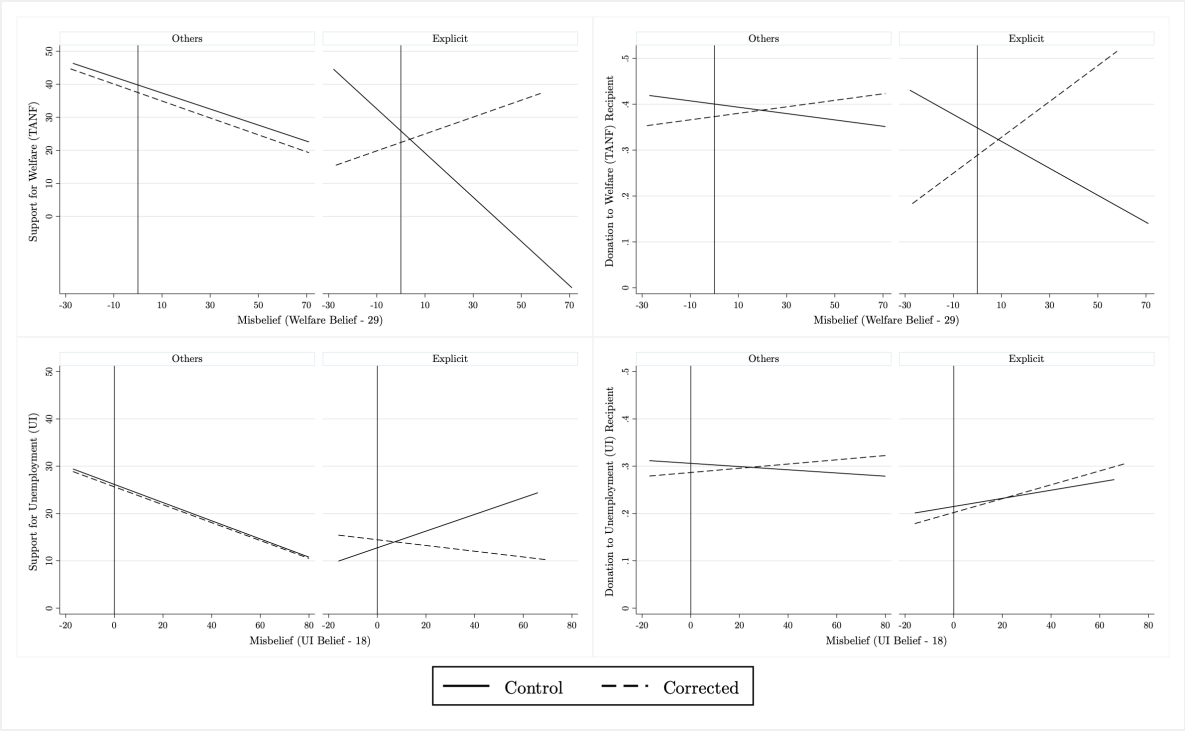


Figure IX: Comparing treatment effects between explicitly racist and other participants

Notes. This figure shows heterogeneous treatment effects between participants who stated that they have an explicit preference for White people over Black people ("Explicit") and participants who do not report such a preference ("Others") within the UI experiment (panels C and D) and the TANF experiment (panels A and B). Each panel shows, separately for participants who explicitly prefer White people to Black people (right) and who do not (left), linear predictions of the effect of racial misbelief on second-stage policy support, shown separately for control groups (solid) and for treatment participants whose beliefs have been updated (dashed). In panels (A) and (C), policy support is measured by the desired percentage change to the average value of program benefit (measured in percentage points). In panels (B) and (D), policy support is measured by donations to previous program beneficiaries (measured in dollars). The regression coefficients upon which these linear predictions are based are shown in Table VIII .

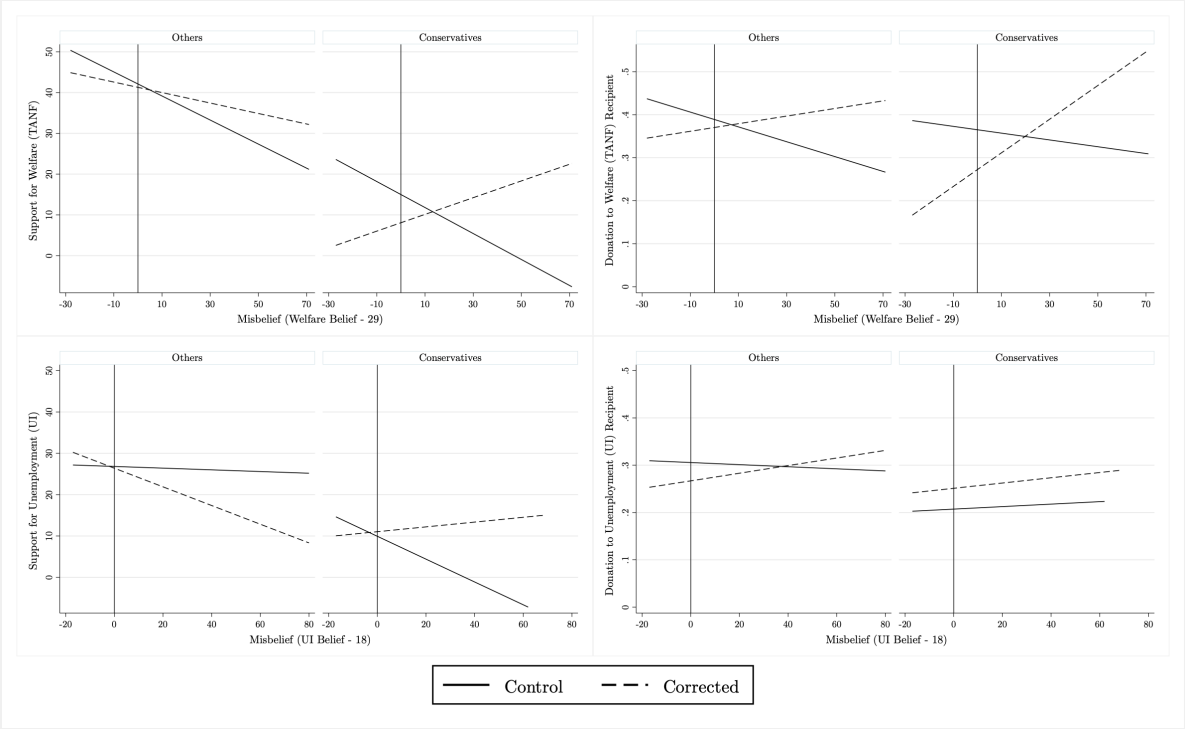


Figure X: Comparing treatment effects between politically conservative and other participants.

Notes. This figure shows heterogeneous treatment effects between politically conservative participants ("Conservatives") and participants who are not politically conservative ("Others") within the UI experiment (panels C and D) and the TANF experiment (panels A and B). Each panel shows, separately for political conservatives (right) and for those who are not politically conservative (left), linear predictions of the effect of racial misbelief on second-stage policy support, shown separately for control groups (solid) and for treatment participants whose beliefs have been updated (dashed). In panels (A) and (C), policy support is measured by the desired percentage change to the average value of program benefit (measured in percentage points). In panels (B) and (D), policy support is measured by donations to previous program beneficiaries (measured in dollars). The regression coefficients upon which these linear predictions are based are shown in Table IX .

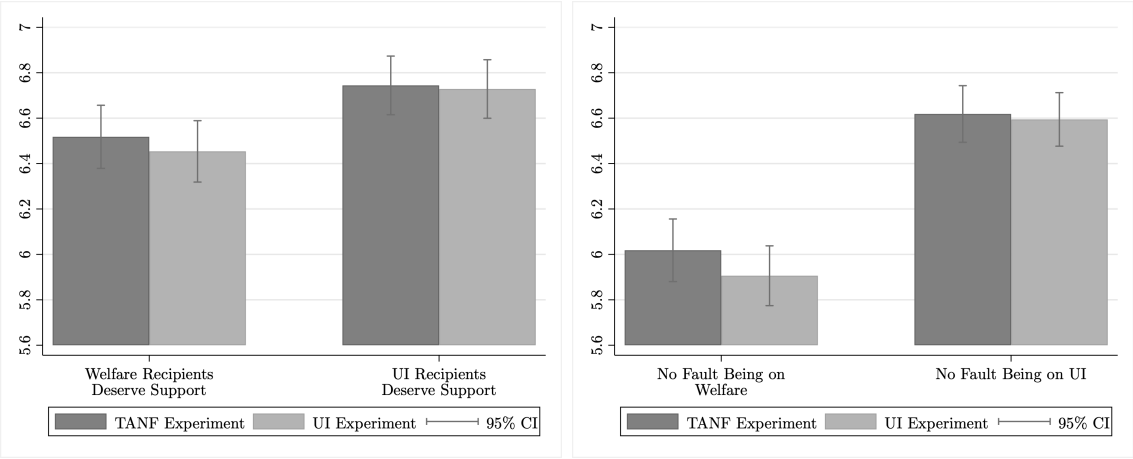


Figure XI: Differences in beneficiary worthiness by program and experimental assignment.

Notes. This figure illustrates differences by program and by experimental assignment in reports of beneficiaries' levels of fault for their own situations and their deservingness of taxpayer support. In the left panel, bars report mean second-stage responses to the question "To what extent do you agree or disagree with the following statement about [welfare/unemployment insurance]? *The situation of most [welfare/unemployment benefit] recipients is no fault of their own.*" where 0 indicates "Strongly disagree" and 10 indicates "Strongly agree". In the right panel, bars report mean second-stage responses to the question "To what extent do you agree or disagree with the following statement about [welfare/unemployment insurance]? *[Welfare/Unemployment benefit] recipients deserve support from taxpayers.*" where 0 indicates "Strongly disagree" and 10 indicates "Strongly agree". 95% confidence intervals shown.

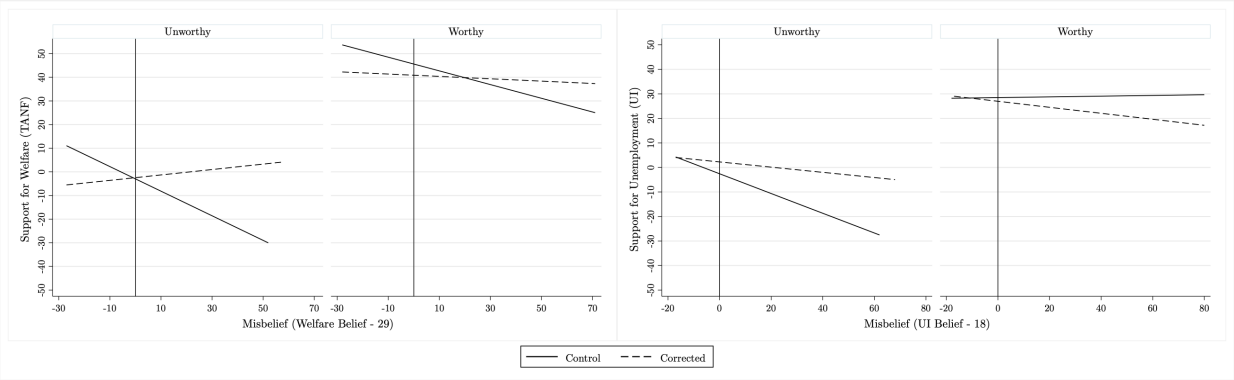


Figure XII: The interaction of worthiness and the treatment effect of information in the Unemployment Insurance experiment.

Notes. This figure shows heterogeneous treatment effects between participants who view the beneficiaries of their respective income support program as worthy of help ("Worthy") and participants who do not view the beneficiaries of their respective income support program as worthy of help ("Unworthy") within the UI experiment (panels C and D) and the TANF experiment (panels A and B). Participants are categorized according to their views of worthiness by: 1) collapsing their measure of beneficiaries' fault and their measure of beneficiaries' deservingness of support into a single index using principal components analysis, then 2) categorizing those within the bottom quintile of this index into the "Unworthy" subgroup. Remaining participants are categorized into the "Worthy" subgroup. Each panel shows, separately for the "Worthy" (right) and "Unworthy" (left) subgroups, linear predictions of the effect of racial misbelief on second-stage policy support, shown separately for control groups (solid) and for treatment participants whose beliefs have been updated (dashed). Policy support is measured by the desired percentage change to the average value of program benefit (measured in percentage points). The regression coefficients upon which these linear predictions are based are shown in Appendix Table B9.

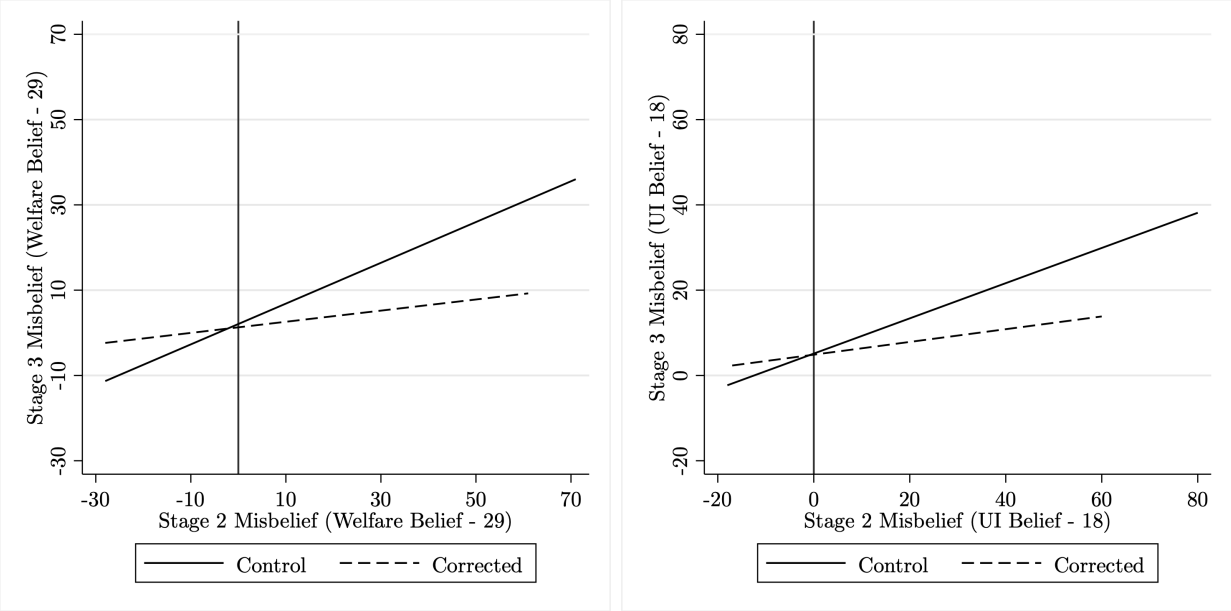


Figure XIII: Persistence and attenuation of Stage 3 beliefs.

Notes. Each panel shows linear predictions of the effect of Stage 2 racial misbelief on Stage 3 racial misbelief, shown separately for control groups (solid) and for treatment participants whose beliefs have been updated (dashed). In each stage, "Misbelief" is defined as the difference between participants' incentivized reports of the proportion of the beneficiaries of their respective social safety net program who are Black, and the true proportion. Shown separately from the UI experiment (left) and the TANF experiment (right). The regression coefficients upon which these linear predictions are based are shown in Table X.

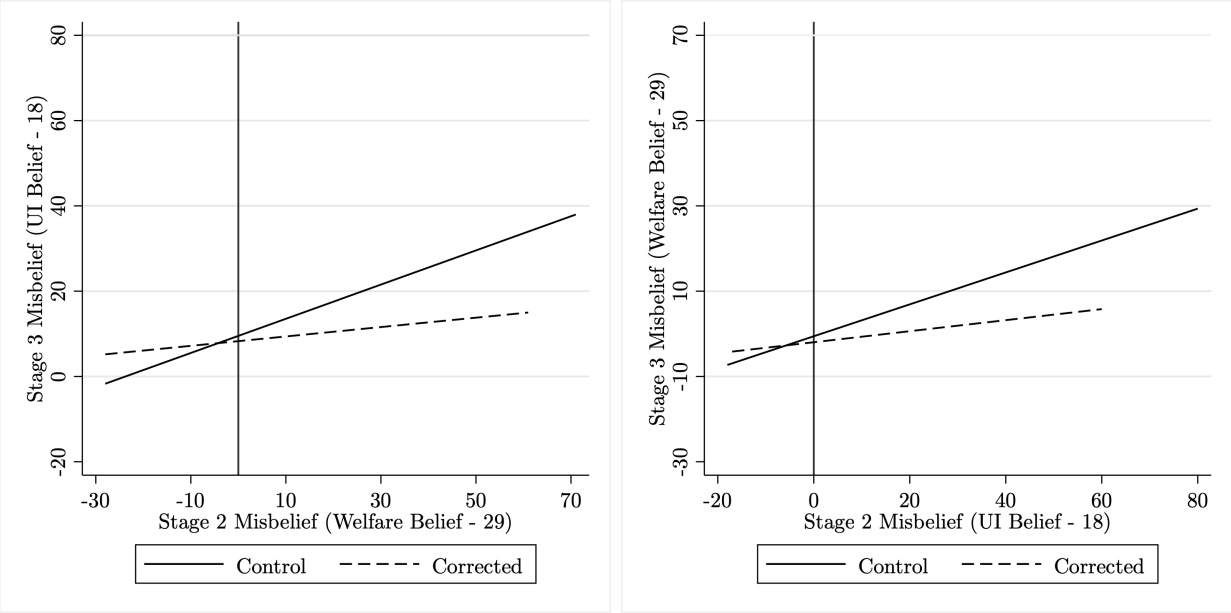


Figure XIV: Do beliefs spillover from one experiment to the other?

Notes. Each panel shows linear predictions of the effect of Stage 2 racial misbelief about the social safety net program featured in the participant's own experiment on Stage 3 racial misbelief about the social safety net program featured in the *other* experiment, shown separately for control groups (solid) and for treatment participants whose beliefs have been updated (dashed). In each stage, "Misbelief" is defined as the difference between participants' incentivized reports of the proportion of the beneficiaries of their respective social safety net program who are Black, and the true proportion. Shown separately for participants randomly assigned to complete the UI experiment (right) and those randomly assigned to complete the TANF experiment (left). The regression coefficients upon which these linear predictions are based are shown in Table XI.

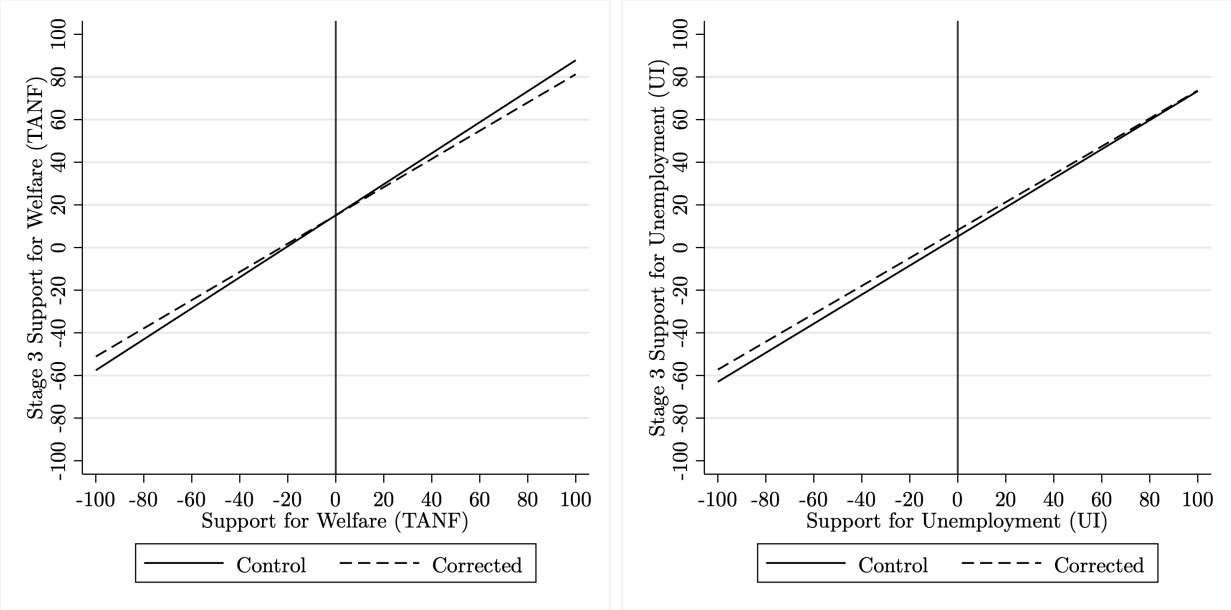


Figure XV: Persistence of policy support from Stage 2 to Stage 3.

Notes. This table shows the relationship between Stage 2 and Stage 3 policy support, where policy support is measured by the desired percentage change to the average value of program benefit (measured in percentage points). Each panel shows linear predictions of the effect of Stage 2 program support on Stage 3 program support, shown separately for control groups (solid) and for treatment participants whose beliefs have been updated (dashed). In each stage, "Misbelief" is defined as the difference between participants' incentivized reports of the proportion of the beneficiaries of their respective social safety net program who are Black, and the true proportion. Linear predictions reported separately for TANF (left) and for UI (right).

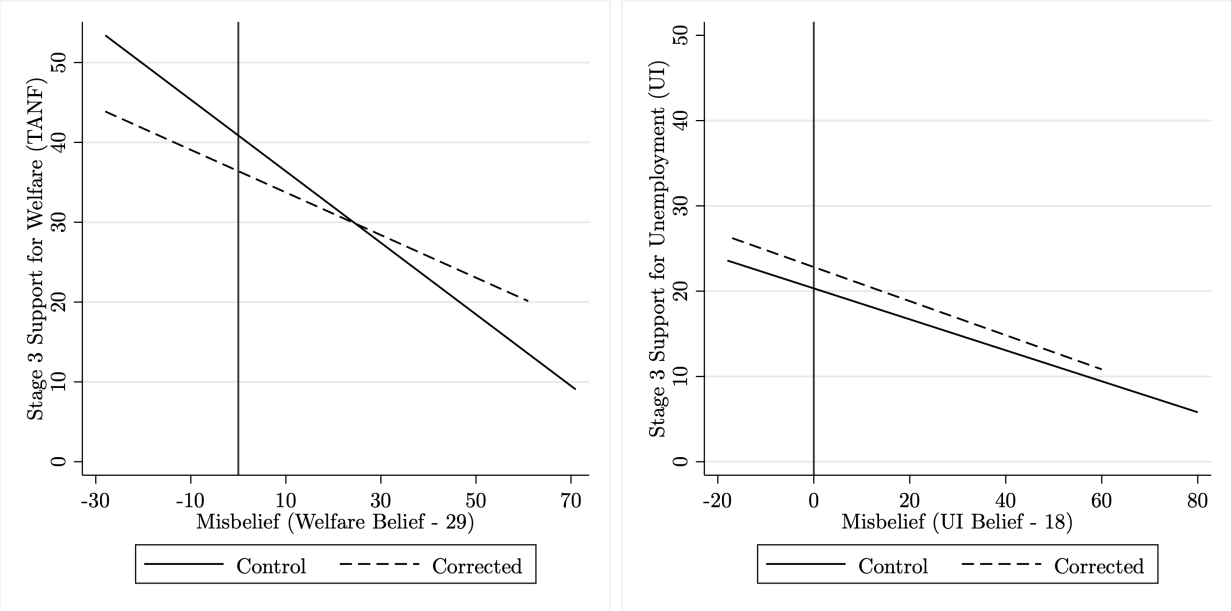


Figure XVI: Policy support average treatment effects after one month.

Notes. This figure shows a linear prediction of the effect of racial misbelief on Stage 3 policy support, shown separately for control groups (solid) and for treatment participants whose beliefs have been updated (dashed). Policy support is measured by the desired percentage change to the average value of program benefit (measured in percentage points). Linear predictions reported separately for TANF (left) and for UI (right). The regression coefficients upon which these linear predictions are based are shown in Table XII.

Supplementary Material for *Does Racial Animus Undermine Support for the U.S. Social Safety Net?*

Jeffrey P. Carpenter
Jakina Debnam Guzman
Peter Hans Matthews
Erin L. Wolcott

Appendix A: Experimental Protocol

Stage 1

The experiment was coded in Qualtrics and conducted on Connect with a representative sample from the U.S. In Stage 1 of the experiment (as in all three stages), participants first gave consent, checked to make sure their Connect user ID number auto-filled correctly, and completed a "recaptcha" question. They were then given instructions on how to complete the Implicit Association Task (IAT), told the current part of the study would last about 5 minutes, that they would receive a flat payment of $1.25 for completing the current part of the study and would be eligible for further payments in Stages 2 and 3.

After completing the IAT, which compared the sorting of Black and White people along with positive and negative words, participants were asked about their explicit racial preferences on a $0-10$ scale where 0 was "I prefer Black people to White people", 5 was labelled "I like White people and Black people equally" and 10 indicated "I prefer White people to Black people".

In the last part of Stage 1, participants answered questions about the state in which they lived, their ethnicity, age, gender, education, income, zip code, work status, whether they had received TANF and/or UI benefits in the previous five years, how politically liberal or conservative they were and whether they voted in the last presidential election.

Stage 1 ended by reminding participants that they would be eligible to participate in the next two stages of the experiment (without specifying what would occur in those stages) and then participants were redirected back to Connect for payment.

Stage 2

After the same start as Stage 1, in Stage 2, all participants were shown the following prompt informing them that the data they would be asked about came from the BLS and if they answer correctly they would receive $1.

All participants were then asked for their incentivized prior beliefs, with the exact wording depending on the experiment to which they were randomized. On the left below is the prompt used for the welfare experiment. The prompt used for the unemployment experiment is on the right of the figure. In each case, they were shown the base rate information first and then the graphic simultaneously darkened the same number of stick figures as the slider reported. The welfare prompt is slightly longer because it has to indicate that the welfare prior we are asking for has to do with TANF specifically. We did this to prevent the conflation of the other various programs that are often called "welfare", such as Aid to Families with Dependent Children (AFDC) or the Supplemental Nutrition Assistance Program (SNAP).

Once the participant had settled on a prior, they then were asked how confident they were

in their belief (as shown below) and clicked a submit button to move on to the assessment of their policy support.

Those participants randomly assigned to the information provision treatment saw some version of the screen below before being asked about their policy preferences. In this example, the participant was sorted into the unemployment treatment condition, where they were first told that the correct statistic is 18 Black people out of 100. To assure that all participants in the treatment were "treated", participants were required to compare their prior to the correct number and could not proceed until they selected the correct response.

We used versions of the following prompts to solicit unincentivized policy support in the two experiments. The prompt for the welfare control treatment is on the left and the unemployment treatment prompt is on the right of the figure. The prompts all started with a reminder of the belief that a participant submitted (and how that prior compared to the true number in the treatment conditions). In all cases, participants were informed of the "typical" benefit provided by the program, which was averaged across all the states for the year 2021, and then asked how much they would prefer to change this benefit. Participants responded using a vertical slider. The slider was initialized at 0%, which level-funded the program. Pushing the slider all the way up to +100% doubled the benefit and pulling it all the way down to −100% zeroed-out or ended the benefit. Before hitting submit, the participant could see the exact percentage change they were about to submit at the bottom of the screen.

For the incentivized policy response, participants were shown the following prompts (again, welfare on the left and unemployment on the right). These prompts always followed the unincentivized policy preference instruments. To be clear, every prompt began by informing the respondent that they would be dividing a dollar between themselves and a person who had been a welfare or unemployment insurance recipient at some point in the previous five years. Here the slider was initialized at $0 and the participants could donate to another participant who had been a benefit recipient in the the past in 10 cent increments by moving the slider to the right. Whatever the participant did not donate was paid to them as a bonus.

Before finishing Stage 2, participants were asked to respond (on a 10-point likert scale) to four statements having to do with the "fault" and "deservingness" of welfare and unemployment recipients. After Stage 2 ended, participants were thanked, asked to keep an eye open for Stage 3 and sent back to Connect for payment.

Stage 3

In Stage 3 of the experiment, returning participants were paid a flat wage of $1 and asked to submit posterior beliefs about the racial composition of benefit recipients for *both* safety net programs using elicitation instruments that were identical to those used in Stage 2. Like before, these beliefs were incentivized, this time with a 25 cent bonus for submitting a belief within two people of the correct answer. Here, the order of the belief questions was randomized (TANF then UI or UI then TANF).

Lastly, after all three stages of the experiment were complete, participants were paid any bonuses due. Participants knew this was how bonuses would accrue and be paid at the outset of the experiment and we made this design choice to not inadvertently convey the belief information provided in the treatments to control participants (i.e., if a participant got a bonus after Stage two they should infer that their belief was correct and if they did not that must have been because their belief was wrong).

Appendix B: Additional Results

Table B1: Determinants of racial animus.

	(1) IAT	(2) Explicit
White	-0.071***	0.256***
	(0.021)	(0.071)
Black	0.250***	-1.070***
	(0.029)	(0.109)
Politically Conservative	-0.028	0.455***
	(0.021)	(0.090)
Vote for Trump	-0.053***	0.417***
	(0.020)	(0.083)
Age	-0.001***	0.005***
	(0.000)	(0.002)
Female	0.026*	-0.297***
	(0.014)	(0.051)
College Degree	0.002	0.067
	(0.016)	(0.057)
Masters Degree or more	0.006	0.050
	(0.021)	(0.073)
Income over $75k	0.003	-0.036
	(0.015)	(0.052)
New England	-0.020	0.276**
	(0.035)	(0.125)
Mid Atlantic	-0.065***	0.093
	(0.021)	(0.079)
South	-0.019	0.098
	(0.018)	(0.066)
Mid West	-0.032	0.119*
	(0.021)	(0.072)
Constant	-0.320***	4.830***
	(0.030)	(0.101)
Observations	2946	3025

Dependent variable is IAT or Explicit score.
OLS with robust standard errors reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B2: Determinants of welfare beliefs.

	(1) None	(2) Controls
Corrected Welfare Belief	0.923	1.027
	(1.065)	(1.068)
Confidence in Welfare Belief	-0.077	-0.119
	(0.200)	(0.202)
Age		-0.087**
		(0.035)
Female		-0.291
		(1.076)
College Degree		1.556
		(1.205)
Masters Degree or more		-0.088
		(1.566)
Income over $75k		0.983
		(1.140)
New England		-1.808
		(2.692)
Mid Atlantic		0.431
		(1.613)
South		2.695*
		(1.400)
Mid West		1.601
		(1.489)
Constant	29.966***	31.837***
	(1.104)	(2.118)
Observations	1419	1416

Dependent variable is welfare (TANF) belief.
OLS with robust standard errors reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B3: Determinants of unemployment beliefs.

	(1) None	(2) Controls
Corrected UI Belief	-0.255	-0.288
	(0.967)	(0.964)
Confidence in UI Belief	-0.140	-0.160
	(0.171)	(0.172)
Age		-0.161***
		(0.032)
Female		0.391
		(0.975)
College Degree		-0.441
		(1.075)
Masters Degree or more		1.993
		(1.545)
Income over $75k		-1.231
		(1.026)
New England		2.367
		(2.543)
Mid Atlantic		0.414
		(1.460)
South		1.339
		(1.241)
Mid West		1.296
		(1.379)
Constant	23.684***	30.111***
	(0.958)	(1.993)
Observations	1418	1409

Dependent variable is unemployment belief.
OLS with robust standard errors reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B4: Policy support treatment effects.

	(1) TANF	(2) TANF+	(3) UI	(4) UI+
β_1: Corrected Welfare Belief	-2.805	-2.451		
	(2.284)	(2.290)		
β_2: Misbelief (Welfare Belief - 29)	-0.401***	-0.390***		
	(0.085)	(0.085)		
β_3: Corrected × Misbelief (Welfare Belief - 29)	0.299**	0.298**		
	(0.121)	(0.121)		
Age		0.016		-0.216***
		(0.076)		(0.056)
Female		4.165*		2.158
		(2.286)		(1.819)
College Degree		-0.461		1.805
		(2.603)		(2.012)
Masters Degree or more		3.264		7.288***
		(3.418)		(2.652)
Income over $75k		-5.398**		-3.326*
		(2.407)		(1.932)
New England		-8.137		5.536
		(6.186)		(4.855)
Mid Atlantic		-1.525		4.576*
		(3.367)		(2.656)
South		-4.935*		0.047
		(2.988)		(2.271)
Mid West		-5.539*		-1.779
		(3.298)		(2.671)
β_1: Corrected UI Belief			-0.147	-0.078
			(1.809)	(1.804)
β_2: Misbelief (UI Belief - 18)			-0.095	-0.125
			(0.086)	(0.086)
β_3: Corrected × Misbelief (UI Belief - 18)			-0.066	-0.057
			(0.111)	(0.109)
β_0: Constant	35.605***	37.742***	22.473***	29.605***
	(1.584)	(4.255)	(1.337)	(3.513)
Observations	1415	1412	1417	1408

Dependent variable is TANF or UI support.
Controls include age, sex, education, income and geographic region.
OLS with robust standard errors reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B5: Program recipient donation treatment effects.

	(1) TANF	(2) TANF+	(3) UI	(4) UI+
β_1: Corrected Welfare Belief	-0.037*	-0.037*		
	(0.019)	(0.019)		
β_2: Misbelief (Welfare Belief - 29)	-0.002**	-0.001*		
	(0.001)	(0.001)		
β_3: Corrected × Misbelief (Welfare Belief - 29)	0.003***	0.003***		
	(0.001)	(0.001)		
Age		0.003***		0.001**
		(0.001)		(0.001)
Female		0.074***		0.072***
		(0.019)		(0.017)
College Degree		-0.017		-0.002
		(0.021)		(0.019)
Masters Degree or more		0.007		0.020
		(0.030)		(0.026)
Income over $75k		0.013		0.046**
		(0.021)		(0.019)
New England		-0.066		-0.046
		(0.052)		(0.048)
Mid Atlantic		0.004		-0.025
		(0.030)		(0.026)
South		0.007		0.018
		(0.026)		(0.022)
Mid West		-0.056**		0.016
		(0.026)		(0.024)
β_1: Corrected UI Belief			-0.013	-0.015
			(0.018)	(0.017)
β_2: Misbelief (UI Belief - 18)			-0.000	-0.000
			(0.001)	(0.001)
β_3: Corrected × Misbelief (UI Belief - 18)			0.001	0.001
			(0.001)	(0.001)
β_0: Constant	0.384***	0.236***	0.279***	0.155***
	(0.014)	(0.036)	(0.012)	(0.033)
Observations	1416	1413	1417	1408

Dependent variable is donation to a TANF or UI recipient.
Controls include age, sex, education, income and geographic region.
OLS with robust standard errors reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B6: Subgroup treatment effects for participants who are both implicitly and explicitly racist.

	(1) Others	(2) Both	(3) Others	(4) Both	(5) Others	(6) Both	(7) Others	(8) Both
β_1: Corrected Welfare Belief	-3.274	-3.641	-0.034	-0.070*				
	(2.472)	(5.897)	(0.022)	(0.042)				
β_2: Misbelief (Welfare Belief - 29)	-0.246***	-0.912***	-0.001	-0.005***				
	(0.083)	(0.231)	(0.001)	(0.001)				
β_3: Corrected × Misbelief	-0.002	1.317***	0.001	0.010***				
	(0.130)	(0.302)	(0.001)	(0.002)				
β_1: Corrected UI Belief					-0.266	3.272	-0.019	-0.011
					(1.997)	(4.348)	(0.020)	(0.038)
β_2: Misbelief (UI Belief - 18)					-0.190*	0.289	-0.000	0.001
					(0.100)	(0.204)	(0.001)	(0.002)
β_3: Corrected × Misbelief					0.003	-0.431	0.001	0.001
					(0.125)	(0.277)	(0.001)	(0.002)
β_0: Constant	38.693***	23.492***	0.395***	0.345***	24.868***	8.510**	0.298***	0.204***
	(1.640)	(4.505)	(0.015)	(0.032)	(1.442)	(3.514)	(0.014)	(0.029)
Dependent Variable	TANF Support		TANF Donation		UI Support		UI Donation	
χ^2 test of β_3 equality	$\chi^2 = 12.51$, $p < 0.01$		$\chi^2 = 7.21$, $p < 0.01$		$\chi^2 = 0.98$, $p = 0.32$		$\chi^2 = 0.01$, $p = 0.93$	
Observations	1129	250	1130	250	1134	233	1134	233

Dependent variable is policy support or policy recipient donation. OLS with robust standard errors reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B7: Subgroup treatment effects for Trump voters.

	(1) Others	(2) Trump	(3) Others	(4) Trump	(5) Others	(6) Trump	(7) Others	(8) Trump
β_1: Corrected Welfare Belief	-1.665	-5.587	-0.012	-0.092***				
	(2.531)	(4.364)	(0.023)	(0.034)				
β_2: Misbelief (Welfare Belief - 29)	-0.315***	-0.291**	-0.002**	0.000				
	(0.103)	(0.137)	(0.001)	(0.001)				
β_3: Corrected × Misbelief	0.256*	0.357*	0.004***	0.003*				
	(0.142)	(0.204)	(0.001)	(0.002)				
β_1: Corrected UI Belief					-1.198	2.717	-0.029	0.013
					(2.108)	(3.198)	(0.021)	(0.032)
β_2: Misbelief (UI Belief - 18)					-0.012	-0.282*	0.000	-0.001
					(0.092)	(0.162)	(0.001)	(0.001)
β_3: Corrected × Misbelief					-0.121	0.178	0.001	0.001
					(0.127)	(0.195)	(0.001)	(0.001)
β_0: Constant	43.214***	16.722***	0.396***	0.346***	28.118***	8.571***	0.301***	0.228***
	(1.786)	(2.983)	(0.016)	(0.026)	(1.558)	(2.342)	(0.015)	(0.022)
Dependent Variable	TANF Support		TANF Donation		UI Support		UI Donation	
χ^2 test of β_3 equality	$\chi^2 = 0.17$, $p = 0.68$		$\chi^2 = 0.21$, $p = 0.65$		$\chi^2 = 1.66$, $p = 0.20$		$\chi^2 = 0.01$, $p = 0.90$	
Observations	1000	412	1001	412	994	414	994	414

Dependent variable is policy support or policy recipient donation. OLS with robust standard errors reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B8: Subgroup treatment effects for participants confident in their beliefs.

	(1) Others	(2) Confident	(3) Others	(4) Confident	(5) Others	(6) Confident	(7) Others	(8) Confident
β_1: Corrected Welfare Belief	0.807	-5.614*	-0.002	-0.063**				
	(3.440)	(3.038)	(0.030)	(0.026)				
β_2: Misbelief (Welfare Belief - 29)	-0.376***	-0.418***	-0.001	-0.002**				
	(0.128)	(0.113)	(0.001)	(0.001)				
β_3: Corrected × Misbelief	0.239	0.317**	0.002	0.004***				
	(0.195)	(0.153)	(0.002)	(0.001)				
β_1: Corrected UI Belief					0.060	-0.550	-0.030	-0.000
					(2.631)	(2.505)	(0.026)	(0.024)
β_2: Misbelief (UI Belief - 18)					-0.009	-0.147	-0.001	0.000
					(0.136)	(0.110)	(0.001)	(0.001)
β_3: Corrected × Misbelief					0.047	-0.195	0.002	-0.000
					(0.166)	(0.145)	(0.002)	(0.001)
β_0: Constant	36.369***	34.933***	0.376***	0.391***	21.273***	23.433***	0.275***	0.282***
	(2.368)	(2.134)	(0.020)	(0.018)	(1.963)	(1.829)	(0.018)	(0.017)
Dependent Variable	TANF Support		TANF Donation		UI Support		UI Donation	
χ^2 test of β_3 equality	$\chi^2 = 0.10$, $p = 0.75$		$\chi^2 = 1.15$, $p = 0.28$		$\chi^2 = 1.22$, $p = 0.27$		$\chi^2 = 1.51$, $p = 0.22$	
Observations	648	767	648	768	664	753	664	753

Dependent variable is policy support or policy recipient donation. OLS with robust standard errors reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B9: Subgroup treatment effects based on percieved worthiness of beneficiaries.

	(1) Unworthy	(2) Worthy	(3) Unworthy	(4) Worthy
β_1: Corrected Welfare Belief	0.527	-4.753**		
	(5.439)	(2.208)		
β_2: Misbelief (Welfare Belief - 29)	-0.520***	-0.290***		
	(0.172)	(0.085)		
β_3: Corrected × Misbelief (Welfare Belief - 29)	0.635**	0.240**		
	(0.268)	(0.118)		
β_1: Corrected UI Belief			4.826	-1.527
			(4.026)	(1.887)
β_2: Misbelief (UI Belief - 18)			-0.403**	0.015
			(0.160)	(0.086)
β_3: Corrected × Misbelief (UI Belief - 18)			0.297	-0.137
			(0.214)	(0.114)
β_0: Constant	-2.961	45.621***	-2.564	28.499***
	(3.526)	(1.513)	(3.075)	(1.372)
Dependent Variable	TANF Support		UI Support	
χ^2 test of β_3 equality	$\chi^2 = 1.84$, $p = 0.17$		$\chi^2 = 3.24$, $p = 0.07$	
Observations	281	1134	278	1139

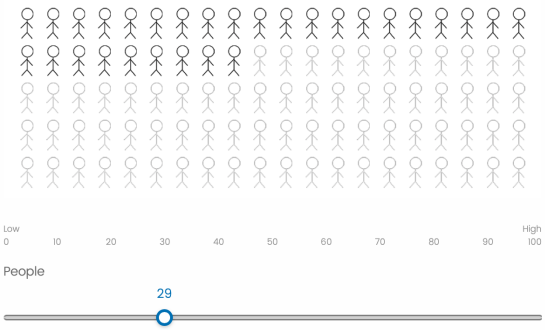
Dependent variable is policy support. OLS with robust standard errors reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The following questions are based on data from the United States Census Bureau and the Bureau of Labor Statistics. **A correct answer to the next question will be rewarded with an additional bonus of $1.00.**

→

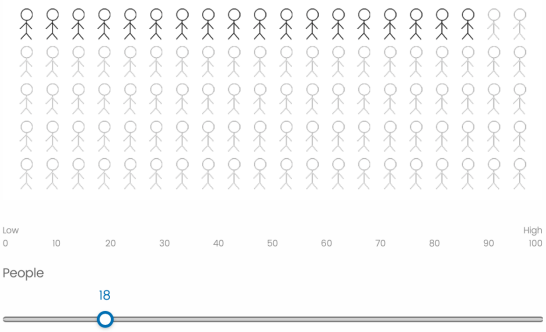
During 2021, 13 out of every 100 adults in the U.S. identified as Black.

Out of every 100 adults who received welfare from the U.S. government (sometimes referred to as TANF or Temporary Assistance for Needy Families) in 2021, how many do you think identified as Black? A correct answer is within 2 people on either side of the actual number.



During 2021, 13 out of every 100 adults in the U.S. identified as Black.

Out of every 100 adults who received unemployment benefits from the U.S. government in 2021, how many do you think identified as Black? A correct answer is within 2 people on either side of the actual number.



How confident are you in this number? *(0 means that this number is a guess, 10 means that you are certain of your response)*

In the previous question you answered that in 2021, for every 100 adults receiving unemployment benefits in the U.S., 18 identified as Black. The actual answer is 18.

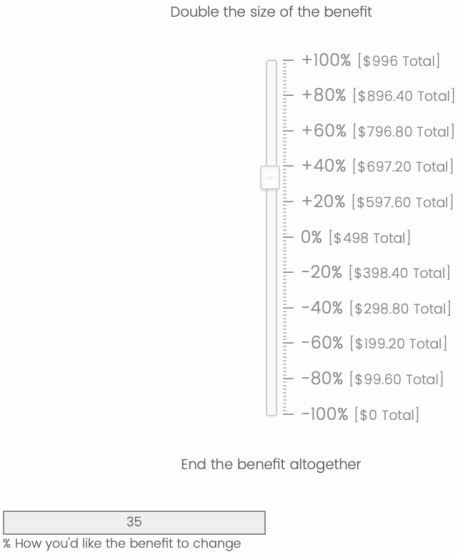
How does your answer of 18 compare to the actual number?

- ○ My answer is higher (by more than 2 people) than the actual number
- ◉ My answer is around the same (± 2 people) as the actual number
- ○ My answer is lower (by more than 2 people) than the actual number

As a reminder, you previously estimated that for every 100 adults receiving welfare in the U.S., 29 identified as Black.

The typical state paid a mother and two children on welfare (sometimes referred to as TANF or Temporary Assistance for Needy Families) a maximum of $498 a month at the end of 2021.

How should this benefit be changed?



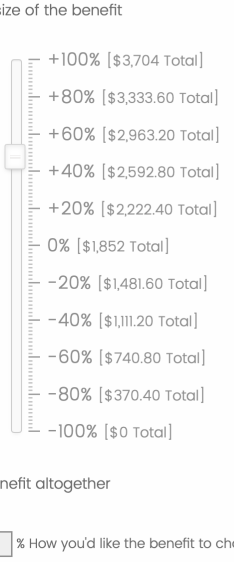
35

% How you'd like the benefit to change

As a reminder, you previously estimated that for every 100 adults receiving unemployment benefits in the U.S., 18 identified as Black. Upon reflection you acknowledged that "My answer is around the same (± 2 people) as the actual number."

The typical state paid a person filing for unemployment insurance a maximum of $1,852 a month at the end of 2021.

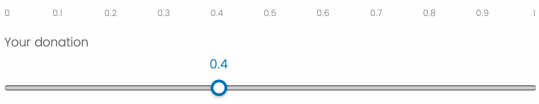
How should this benefit be changed?



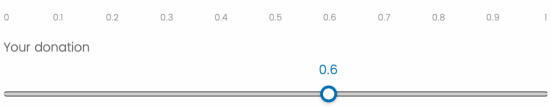
48

% How you'd like the benefit to change

Divide a dollar: We will give you a second bonus of $1, any portion of which you can donate to another participant in this study and keep the rest. The person who will receive your donation will be selected at random from the group of participants who have received welfare (i.e. TANF) in the past five years. Please use the slider to indicate how much you want to donate to this person. You will receive whatever amount you do not donate as a bonus.



Divide a dollar: We will give you a second bonus of $1, any portion of which you can donate to another participant in this study and keep the rest. The person who will receive your donation will be selected at random from the group of participants who have received unemployment benefits in the past five years. Please use the slider to indicate how much you want to donate to this person. You will receive whatever amount you do not donate as a bonus.



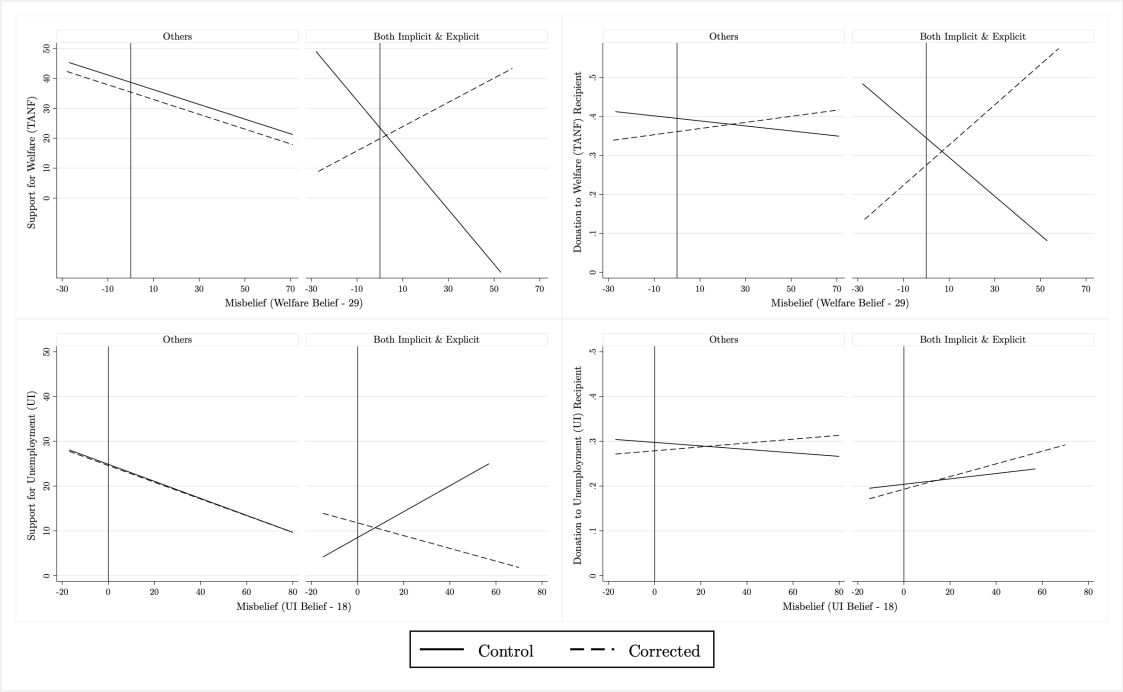


Figure B1: Comparing treatment effects between participants who are both implicitly and explicitly racist and other participants.

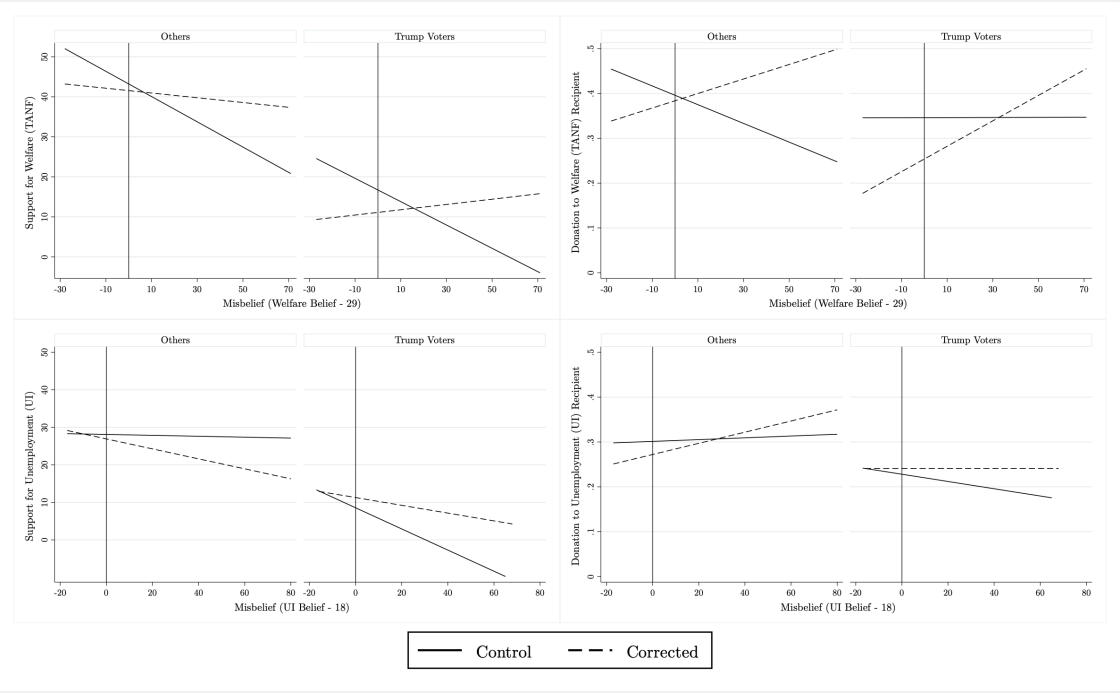


Figure B2: Comparing treatment effects between Trump voters and other participants.

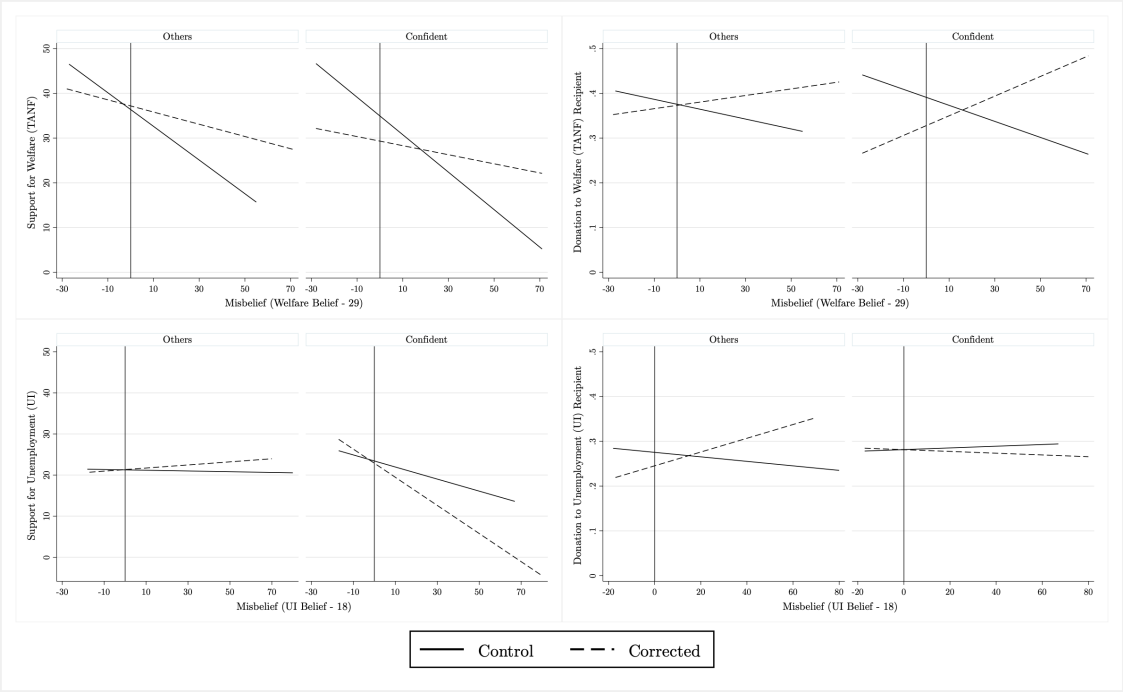


Figure B3: Comparing treatment effects between participants who are confident in their beliefs and others.