Does racial animus determine support for redistributive policies in the United States?*

Jeffrey Carpenter† Jakina Debnam Guzman‡
Peter Hans Matthews§ Erin Wolcott¶

March 2025

Abstract

It has been hypothesized that support for redistributive policy in the United States is influenced by beliefs about beneficiaries' racial composition. We report findings from two large parallel, three-stage information provision experiments that leverage the different racialized histories of welfare and unemployment insurance. We first explore the extent to which racial beliefs predict policy preferences and then test whether these beliefs can be *persistently* changed through an information intervention. We find that the untreated beliefs are more predictive of support for the racially stigmatized welfare program than for the less stigmatized unemployment insurance program, but also that participants who received information that there are more (less) Black recipients than expected supported welfare less (more). Along with various robustness tests and extensions, we also find that untreated beliefs are stable over time but that our intervention generates persistent changes. In short, we find robust evidence of a causal relationship between racial animus and redistributive policy preferences.

DRAFT: DO NOT QUOTE OR CITE WITHOUT PERMISSION.

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

*For excellent research assistance, we thank Allison Lounsbury, Bishal Panthi, and Blair Jia. For helpful comments, we are grateful to seminar participants at Williams College, Fordham University, and the Sixth World Conference on Remedies to Racial and Social Inequality. 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.

†Economics, Middlebury College and IZA; jcarpent@middlebury.edu.
‡Economics, Amherst College; jguzman@amherst.edu
§Middlebury College and Aalto University; pmatthew@middlebury.edu
¶Middlebury College; wolcott@middlebury.edu

1 Introduction

Differences between the American and European welfare states are often attributed to differing beliefs about the sources of inequalities, such as effort versus luck (Almås, Cappelen, and Tungodden, 2020). As Alesina, Glaeser, and Sacerdote (2001) claimed more than two decades ago, however, "America's troubled race relations are clearly a major reason for the absence of an American welfare state." The two are, of course, not inconsistent – racial animus likely contaminates the perception of effort. Our question is whether we can build on these insights to explain the dynamic patterns in Figure 1, which documents the rise in Aid to Families with Dependent Children (AFDC) payments coterminously with the Civil Rights Movement, followed by a steady decline until 1996, when it was replaced with the Temporary Assistance to Needy Families (TANF) program and the much less pronounced decline in unemployment insurance (UI) benefits beginning in the 1970s. One explanation is that as the number of Black program beneficiaries increased, and as beliefs or, sometimes, misbeliefs about the racial composition of TANF and UI beneficiaries evolved, the support of voters and their elected representatives weakened. There is a nascent literature exploring the possible nexus between racial animus and program benefits (Quadagno, 1994; Massey, 2009; McGhee, 2022; Wolcott, 2024), but unlike these studies, we seek to establish a causal link between support for TANF and UI and beliefs about the racial composition of beneficiaries. In particular, we report the results of a series of experiments that draws on the different racialized histories of these two programs to estimate and contextualize the causal role of misbeliefs about recipients on Americans' support for the social safety net.[1]

Welfare, in particular, has a long history of being racialized in the U.S., famously typified by Ronald Reagan's "welfare queen" trope about an undeserving (Black) mother using multiple aliases to abuse government assistance. The racialized narra-

[1]Over time, the term "welfare" has been ambiguously used in the literature to refer to several different programs including Medicaid, Housing Assistance, Aid to Families with Dependent Children (AFDC), Temporary Assistance to Needy Families (TANF), the Supplemental Nutrition Assistance Program (SNAP), and Supplemental Security Income (SSI).

tive around unemployment insurance, on the other hand, had focused on supporting deserving (White) industrial workers during temporary spells of bad luck, and helping them "pull themselves up by their bootstraps." It is with these differences in mind that our experimental design provides causal evidence of racial (mis)beliefs on policy preferences for these two particular programs. With regard to welfare, we find that policy preferences — which are often viewed as stubborn and deep-seated — can be changed through the exogenous and persistent manipulation of beliefs about the composition of beneficiaries, results that stand in contrast with some of the information provision literature. In the case of unemployment insurance, we are able to manipulate beliefs but find scant evidence of an effect on support - as expected given the different program histories.

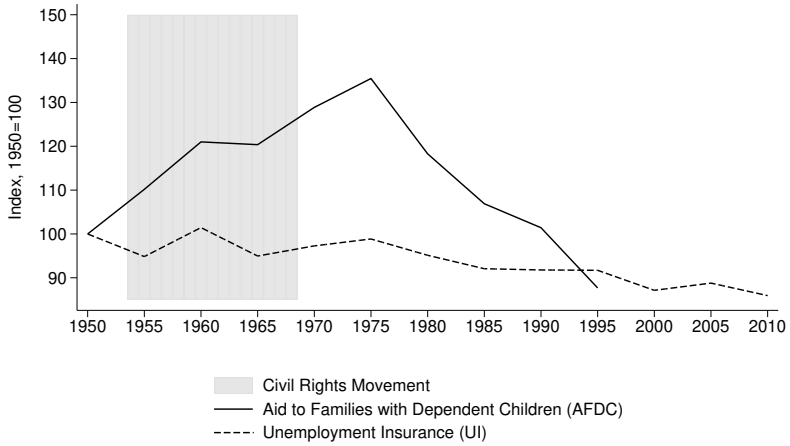


Figure 1: Declining Redistributive Policy Post-Civil Rights (Notes: Aid to Families with Dependent Children is average per recipient benefit in 1995 dollars from Wexler and Engel (1999). Unemployment Insurance is average weekly benefits as a share of average wages from Wolcott (2024). The Civil Rights Movement spans from 1954 through 1968.

Considering the details, in this paper we describe the results from two large ($n \approx 3000$), nationally representative, parallel, three-stage information provision experiments we conducted between June and August of 2024 in the U.S. In the first stage, we elicited participants' implicit and explicit Black-White racial preferences. After a month had elapsed, we conducted the second stage, in which we collected partic-

ipants' prior beliefs about the racial composition of beneficiaries of either TANF or UI. We then randomized half of the participants into a treatment group and provided them with the correct information. We focus on two outcomes. The first is the extent to which participants think TANF or UI benefit amounts should be changed, based on an uncentivized sliding scale measure. Second, we also asked what fraction of a bonus payment they would like to give to another participant who, in the past five years, received either TANF or UI, an incentivized measure that serves to attenuate possible (experimenter demand, hypothetical) biases. The results are consistent. In the third stage one month later, we asked participants to provide their current beliefs about the number of Black people who used both programs.

We find that participants who in the second stage received information that there are more (less) Black welfare recipients than expected had less (more) support for welfare. This result is driven disproportionately by White participants and by participants who in the first stage held above median (implicit or explicit) anti-Black views. We also find that perceived notions of worthiness matter: Participants who think that beneficiaries of either program are more deserving or less at fault are more willing to increase the level of program benefits. In contrast to our welfare results, we do not find a significant relationship between corrected misbeliefs and policy support for unemployment insurance in aggregate; however, we do find some evidence that perceptions of worthiness may be crowding out racial preferences for the UI experiment, an important amendment to conventional views. Last, we find that uncorrected beliefs about race are temporally stable while corrected beliefs can be persistently changed, which suggests that our manipulation was externally valid.

Our primary contribution is to the literature which explores the role that racial beliefs play in shaping redistributive policy preferences. Consider, as in Alesina et al. (2001), an electorate whose support for redistribution is increasing in their perception that members of their own group benefit from it. To the extent that in-group and out-group status is defined by race, members of majority groups will decrease their support for redistributive policies if they perceive that the beneficiaries of such policies are primarily from outside of their group. This mechanism, coupled with the relatively

high level of ethnic heterogeneity in the U.S., has been applied to explain the relatively small size of the social safety net in the U.S. relative to Europe (Alesina et al., 2001; Lipset, 1996). Consistent with this, Luttmer (2001) uses geographic variation in data from the General Social Survey to find that if individuals perceive that a higher share of welfare recipients are from their own racial group, they view welfare spending more favorably.

Beginning with Gilens (1995, 1996), there is evidence that respondents with negative perceptions of Black people are also less supportive of welfare. Alesina, Ferroni, and Stantcheva (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. Our results are consistent with these claims. We find that people who are (upwardly) surprised by the share of Black people receiving benefits are less supportive of welfare, and this result is driven by White respondents. Our results extend this literature in two directions, by introducing an exogenous source of variation in beliefs about beneficiaries, and by showing that the same relationship does not extend to unemployment insurance.

We also contribute to the experimental literature on whether information provisions affect preferences for redistribution by documenting that for some information provisions and for some polices, respondents update their views. Much of this literature finds that preferences about redistributive policy are difficult to move. Kuziemko, Norton, Saez, and Stantcheva (2015) find that despite their information intervention affecting perceptions of inequality, it has almost no effect on preferences about minimum wages and food stamps. The exception they document is for estate taxes: After respondents learn how little inheritance is taxed, they are more supportive of estate tax. Alesina et al. (2021) find that providing information about the extent of racial inequality in the U.S. affects perceptions about racial inequality but does not garner support for redistribution. Only for Democrats, does explaining some of the causes and consequences of systemic racism increase favorability towards redistribution; for Republicans, doing so decreases favorability. Luttmer and Singhal (2011) conclude

redistributive preferences are stable over time and determined by culture because immigrants' policy preferences closely align with the average preferences of their origin country.

The literature is mixed on whether information interventions about racial attitudes shift policy views. Haaland, Roth, and Wohlfart (2023) find that their information provision does not change support for pro-Black policies. They 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 informed the true callback rate, they are more likely to update their beliefs about the extent of racial discrimination but not any more likely to support pro-Black policies. In an earlier paper, Haaland and Roth (2020) find respondents do change their policy views and are more supportive of immigration if they are informed about research showing no adverse labor market impacts of immigration. Similar to Haaland and Roth (2023), we conclude that whether information can change policy preference is context specific. Relative to this literature, we document a novel information provision that alters support for welfare: the racial composition of beneficiaries. Because racial attitudes are very stable and formed by the time people are teenagers, our results indicate why little else can move policy preferences (Gilens, 1995; Alesina et al., 2021). That said, this novel information provision only effects support for welfare and not support for unemployment insurance.

In work closest to ours, Akesson, Hahn, Metcalfe, and Rasooly (2022) link racial beliefs to support for welfare, finding that when White Americans are presented with either a stylized high proportion of welfare recipients who are Black or a stylized low proportion, viewing the higher proportion causes them to support the program less. Unlike the information intervention of Akesson et al. (2022) which finds a null effect of *accurate* information, we find causal evidence of providing the accurate share of welfare recipients who are Black on policy support. We conjecture that one reason for this is that unlike Akesson et al. (2022), we provided a base rate (that is, the number of Black Americans in the overall population, 13%) which generated a distribution of

beliefs that, if not symmetric, included substantial numbers of positive and negative misbeliefs about both policies. In their work, the vast majority of survey participants (86%) overestimated the share of welfare recipients who are Black. Another important distinction between the two papers is that we survey a representative sample of the U.S. population, while Akesson et al. (2022) restrict their sample to White Americans. Our broader sample allows us to test whether the causal effect of information provision also varies by race.

Lastly, because we test whether racial animus affects two policies with differing narratives of recipient fault, we contribute to the literature on altruism and racialized views of worthiness (Fong and Luttmer, 2011; Drenik and Perez-Truglia, 2018; 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. The potential for perception of fault to moderate the effect of racial beliefs on altruism is noted by Alesina et al. (2001) who suggest that the widespread perception in the U.S. that the poverty of the poor is due to their own laziness (while in Europe poverty is more likely to be perceived as due to misfortune), compounds the majority groups' lack of support for redistribution. 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. Fong and Luttmer (2011) find that there is an effect of Black race on charitable giving *only* as moderated through its effect on the perceived deservingness of respondents.

2 Experimental Design

We implemented two parallel three-stage information provision experiments on Connect[2] between the middle of June and the beginning of August, 2024. For both

[2]Connect is an online experimental platform. We chose to this platform because evidence suggests that this platform provides higher data quality than other comparable online experimental platformsGupta, Rigotti, and Wilson (2021).

experiments, we target a nationally representative sample of participants within the United States. For experimental details, see Appendix A. In the experiments, we elicit participant beliefs about the number of Black people who benefited in 2021[3] from either Temporary Assistance for Needy Families (TANF, commonly known as "welfare") or Unemployment Insurance (UI). Then, at random, half of the participants were informed of the correct number of Black program beneficiaries in 2021. 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 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., more Black people use the program than they thought), and others realized that they thought that Black usage was too high (i.e., fewer Black people actually use the program). 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 measure implicit racial bias, we had our Stage 1 participants complete the

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

Implicit Association Test (IAT) in which they were asked to quickly and accurately sort pictures and words into groups. The pictures were either of Black or White people and the words were either positive (e.g., happy) or negative (e.g., hate). Depending on how quickly (and therefore, naturally) people can associate the racial groups of people with either the positive or negative words, the resulting measure (between -2 indicating a strong implicit bias against Black people and 2 indicating a bias against White people, with 0 indicating no bias in either direction) gives us a sense of any implicit bias a respondent carries with respect to a given racial comparison. For the explicit racial bias information, we asked respondents directly: "Which statement best describes you?" Here, 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. 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 urge for consistency or experimenter demand effects, 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.[4] Stage 2 began by informing all the participants that they would be paid a $1 bonus if they correctly responded to one of the two following 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. gov-*

[4] 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.

ernment in 2021, how many do you think identified as Black?

To attenuate any 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 how the benefit should change.[5] They could respond on a vertical slider anywhere from −100% (end the benefit altogether) to +100% (double the size of the benefit). We consider these our (unincentivized) policy support outcomes.

For each participant, the next question (our second primary outcome) was a new 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, either received TANF or UI, depending on the experiment.[6] 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

[5]According to the Center on Budget and Policy Priorities, the maximum TANF benefit paid to a mother (with two children) in the median state was $498 a month at the end of 2021. According to the U.S. Department of Labor, the maximum UI benefit paid to a claimant in the median state was $1852 a month at the end of 2021. We chose 2021 because, at the time of the experiment, it is the most recent year the Census Bureau reported the racial composition of TANF and UI recipients at `www.census.gov/library/visualizations/interactive/social-safety-net-benefits.html`.

[6]One of the questions we asked in Stage 1 was whether people had received TANF or UI in the previous five years.

our incentivized policy support outcomes, which we designed to further reduce any experimenter demand effects.

Stage 2 ended 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 used *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 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 participants in the two redistributive programs.[7] 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 $23.68).[8]

3 Results

3.1 Sample characteristics

Based on our pilot study (December, 2023; $N = 380$) and the rule of thumb developed in Haaland et al. (2023), we planned to gather between 600 and 700 observations per

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

[8] 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.

Table 1: 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: The Census age categories reflect that fact that participants must be at least 18.

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 ended up with 3,029, as seen in Table 1. We also see in Table 1 that we collected Stage 2 data from 2,834 of the Stage 1 participants which surpasses our upper goal of 2800 and represents a 94% return rate.[9] Lastly, we had 2324 Stage 2 participants return for Stage 3.

Also, notice in Table 1 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 2, randomization worked as intended and we achieved balance in our two experiments.

[9]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.

Table 2: 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: The F statistic from: (1) regressing a TANF indicator on the observables is 1.16 ($p = 0.32$), (2) regressing an UI indicator on the observables is 0.76 ($p = 0.62$).

3.2 Stage 1: measures of racial animus

On average, both of the measures we collected in Stage 1 show some racial animus. 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.[10] 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 2, we plot the individual level 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. Put differently, respondents who report a stronger explicit preference for White people over Black people also tend to be more implicitly biased in the same direction.

[10] 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.

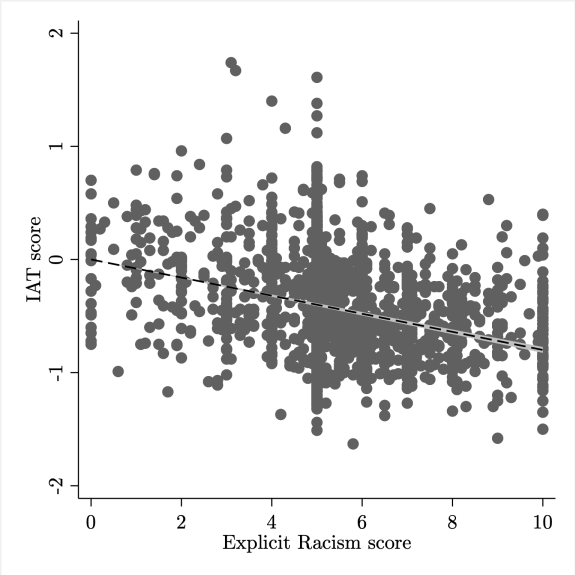


Figure 2: Individual measures of explicit and implicit racial animus.

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; however, while the IAT scores are lower and the explicit scores are higher for Whites, the correlations are the opposite for Black participants, which provides some sense of internal validity in these measures. We also find that politically conservative participants and Trump voters 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.

3.3 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 3. Starting with welfare on the left, we see that these beliefs vary from 1 to 100 with a mean of 30, which is only significantly different from the true number (29) at the 5% level

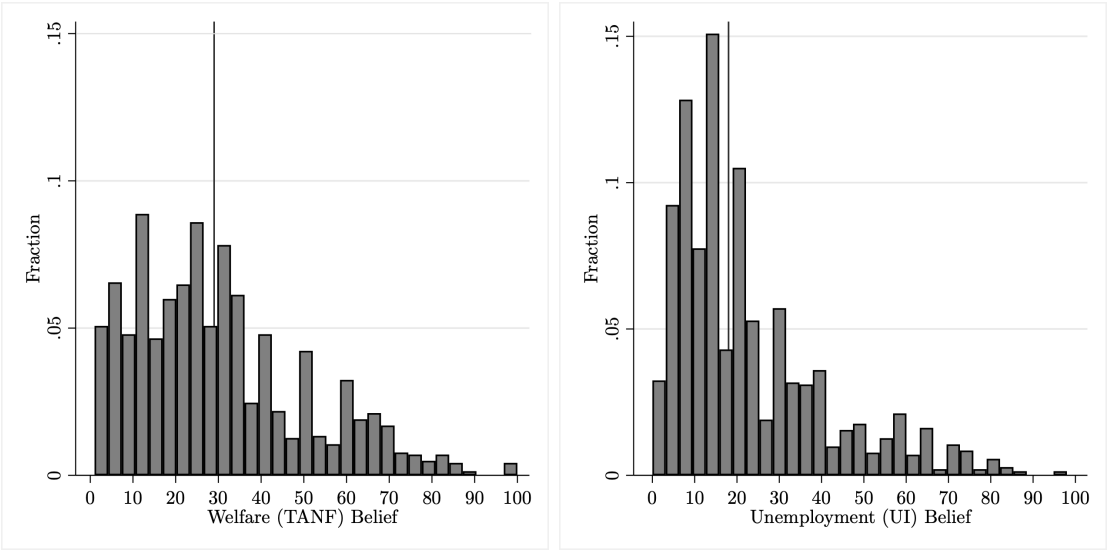


Figure 3: Individual beliefs about the number of Black people on TANF and UI (vertical lines indicate the actual numbers).

($t = 2.03$, $p = 0.04$). Put differently, the beliefs about TANF participation of our participants are very close to accurate, on average. 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 also confirms that there is no significant correlation between the level of one's belief and one's confidence in that belief. Importantly, there is no significant difference in the prior beliefs held by the TANF experimental participants in the two treatment conditions. In terms of demographics, we only find that older participants have lower beliefs and that participants from the South of the U.S. have marginally significantly higher beliefs (than people in the West).

Beliefs about the number of Black UI participants appear on the right of Figure 3. 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 ($t = 10.22$, $p < 0.01$). That said, and similar to the TANF beliefs, we see that 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, in this case, is 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 within the UI experiment 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.

3.4 Stage 2: Average Treatment Effects

Before reporting the average treatment effects for both experiments, we summarize our outcomes: 1) support for each program and 2) donations to previous program participants. For the welfare experiment, overall support is summarized on the left of Figure 4. 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 a third (34%). On the right of Figure 4, we see that there are fewer people who would like to dramatically increase UI benefits. Here the modal response is to increase UI by about a quarter and the average increase is 22%.

Our participants were also relatively generous with the donations they made to previous beneficiaries of the two programs. Figure 5 illustrates the fraction of the one-

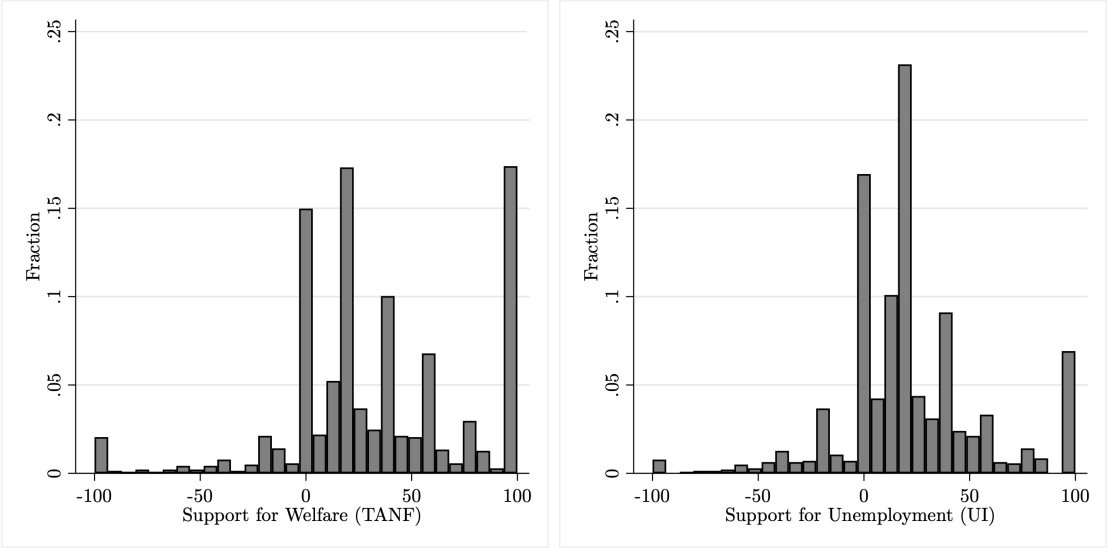


Figure 4: Participant support for welfare (TANF) and unemployment insurance (UI). The x-axis is stated preference for increasing or decreasing program benefit levels measured in percentage points.

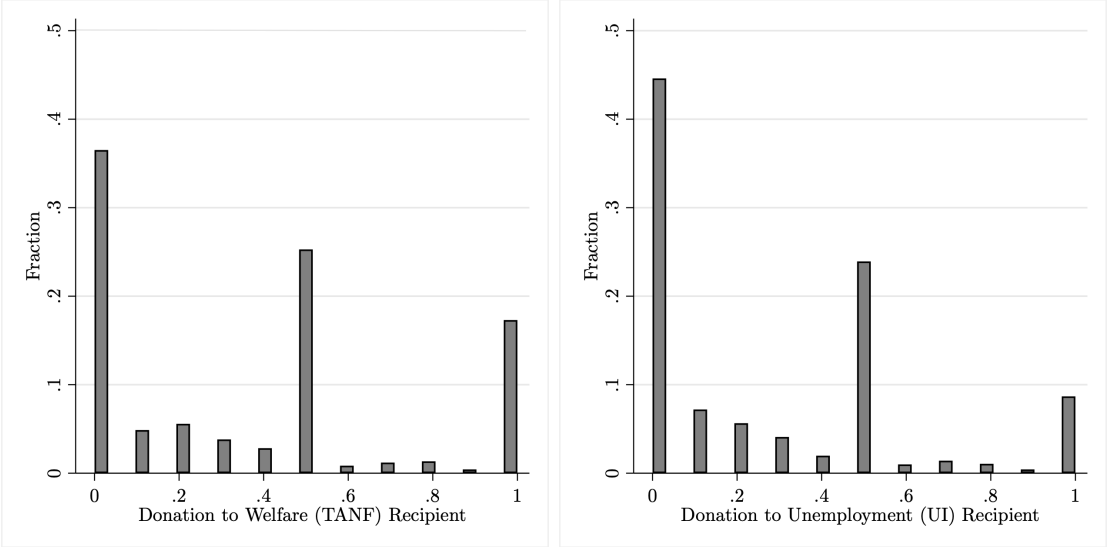


Figure 5: Participant donations to previous welfare (TANF) and unemployment (UI) recipients (transfer from a one-dollar endowment).

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 ($1852 for an individual) already seems substantially more generous than the TANF benefit ($498 for a mother and two children). That said, while these differences are interesting, it is the within experiment variation in these outcomes that is important for inference.

Turning to the average treatment effects of correcting people's beliefs about Black participation in redistributive programs, we estimate, for each outcome, the following "information gradient", with and without controls:

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

where Y_i^k is policy support for participant i and outcome k, T_i is an indicator for the participants who receive the information about the correct number of Black people benefiting from a policy, 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 $Y_i^k = \beta_0 + \beta_1 T_i + \beta_4 X_i + \epsilon_i$ 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 the 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 participants, we expect $\beta_3 > 0$ because when they learn that there are more (less) Black participants than they believed, they should support the policy less (more). That said, if participants have the opposite preference toward Black policy participants, 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.[11] In Figure 6, 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). Starting on the left with the TANF experiment, 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, the more participants believe that the program serves the Black population, the less they support it. Further, we see in column (1) of Table 3 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 3 that this small gap isn't significant ($\beta_1 = -2.805$, $p = 0.22$), 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 3, we find that the estimate for $\beta_3 = 0.299$ ($p = 0.014$), which is positive as expected. Lastly, we see that these estimates change little in column (2) which includes demographic controls. This is no surprise, given

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

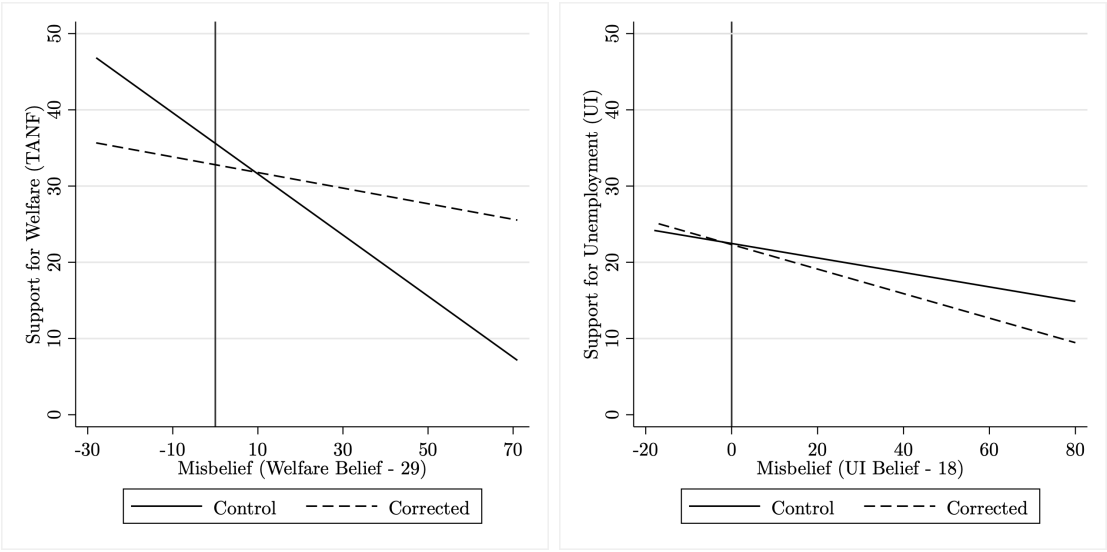


Figure 6: Policy support average treatment effects.

we achieved random assignment to treatment condition.[12] 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 6. 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 3 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 in the estimate of UI support 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 participants is

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

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 6 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 workers trying to "pull themselves up by their bootstraps," our experiment indicates that racial animus appears to play a more muted role.

Table 3: 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

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

3.5 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 our incentivized donation outcomes. These treatment effects are presented in Figure 7. Notice that these estimates seem similar to the results from the unincentivized support measure shown in Figure 6, 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 4 ($\beta_2 = -0.002$, $p = 0.02$), 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 6, on the right of Figure 7 and in Table 4, 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 7 change when we control for demographics in Columns (2) and (4) of Table 4.[13]

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

Table 4: 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

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

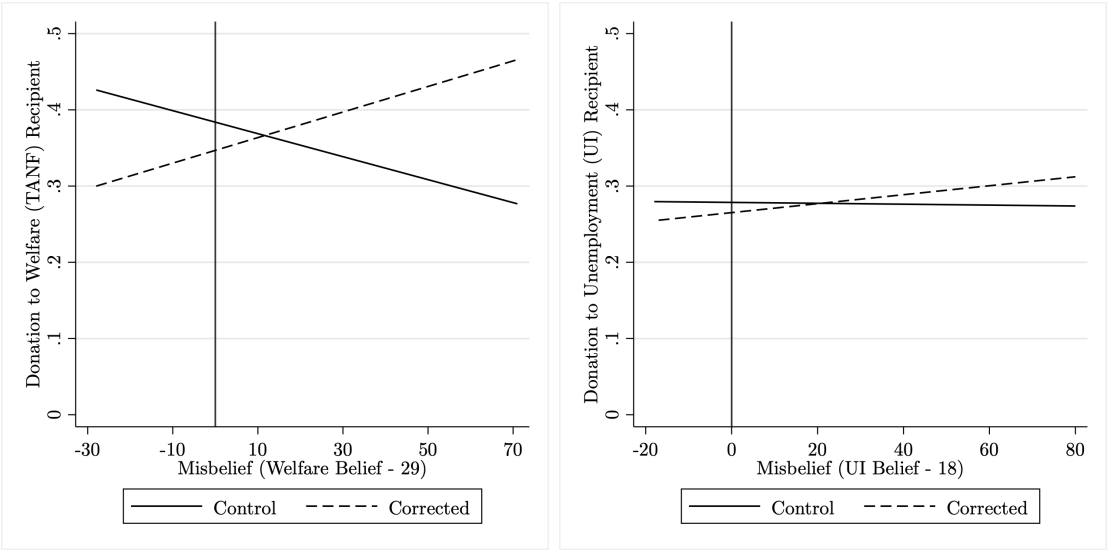


Figure 7: Program participant donation average treatment effects.

3.6 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 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 5, we report linear spline versions of all our average effect estimates (i.e., the estimates represented in Figures 6 and 7) and use F-tests to compare the slope of the two linear segments on either side of the knot. Put differently, 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.

Table 5: Splines of the average effects.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	TANF	TANF	TANF	TANF	UI	UI	UI	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

Dependent variable is policy support or donation to a former recipient. OLS with robust standard errors reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The F-tests report p-values from the hypothesis that the estimates are the same on either side of the spline knot.

For the TANF support estimates in the first two columns of Table 5, 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

5, 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. Put differently, treating positive and negative misbeliefs symmetrically seems to fit our data relatively well.

3.7 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 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 8, 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 7. 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 6. 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 6 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

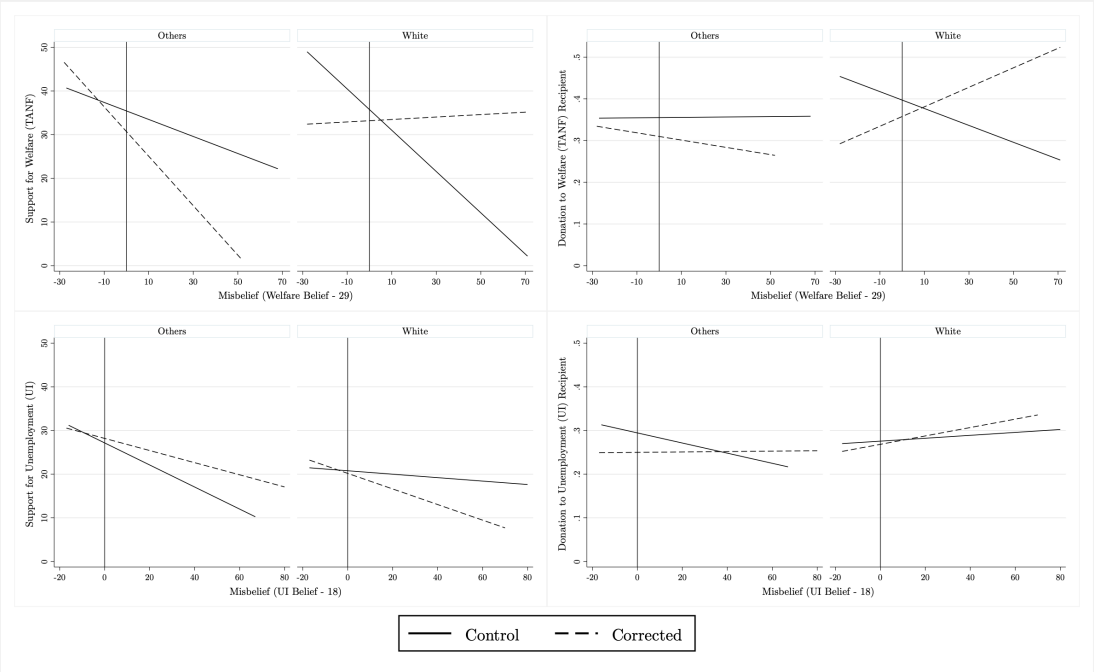


Figure 8: Comparing treatment effects between White and other participants.

consistent with racial animus.

In the top right panel of Figure 8, 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 6 indicate, while donations fall as beliefs rise among White participants, they are essentially flat for non-White participants (i.e., the racial composition doesn't 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 8, indicate that this result was not because we were averaging over very different racial subgroup effects that canceled each other in the aggregate. For

Table 6: 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

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

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 6.

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 9 and Table 7. 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 9 and in the first two columns of Table 7, 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

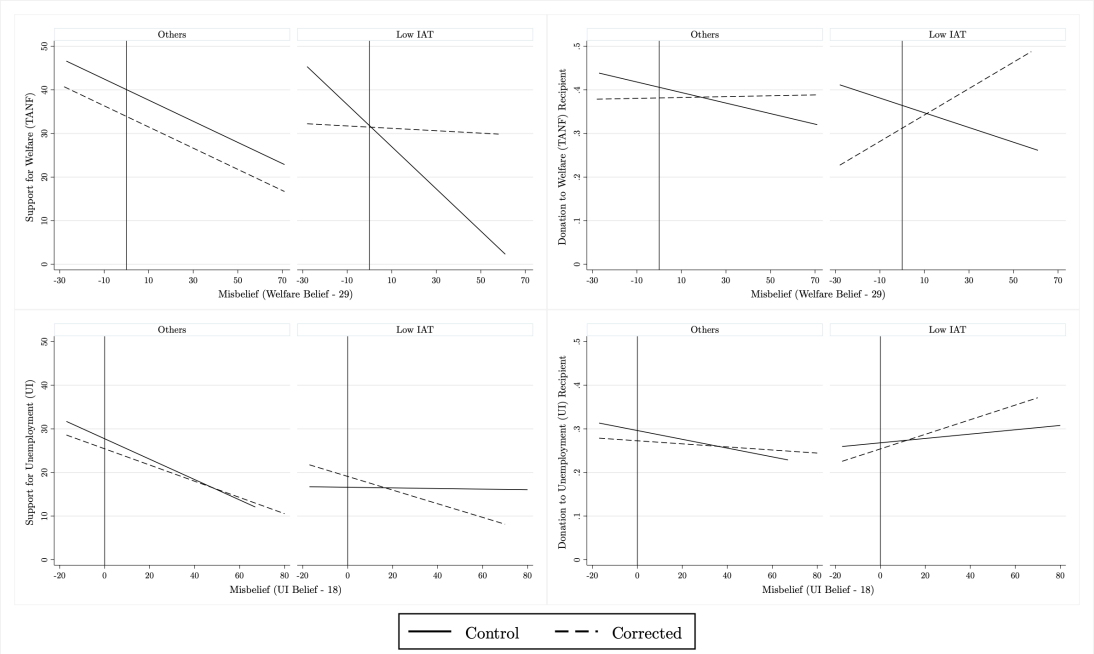


Figure 9: Comparing treatment effects between implicitly biased (low IAT) and other participants.

Table 7: 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

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

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 9 and Columns (3) and (4) of Table 7, 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 race. The bottom two panels of Figure 9 and the last four columns of Table 7 indicate that beliefs about the racial composition of the beneficiaries of UI do not explain differences in how much individuals 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

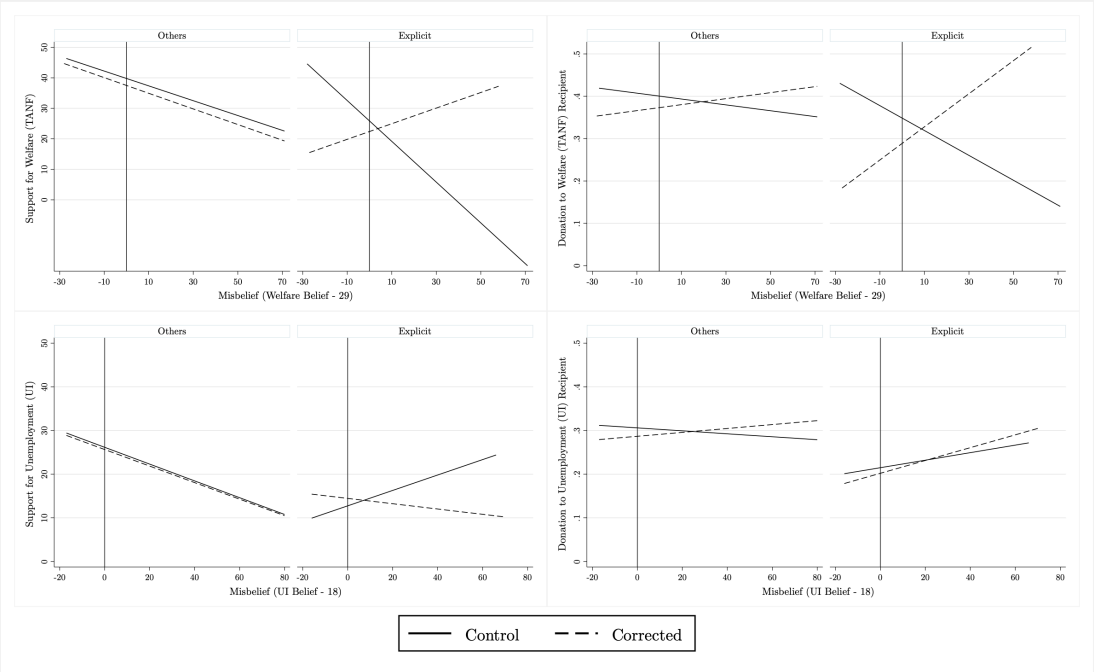


Figure 10: Comparing treatment effects between explicitly racist and other participants.

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 10 and listed in Table 8. 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 Table 8,

Table 8: 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

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

second from the bottom row). 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.

On the bottom of Figure 10, 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 8. The one thing that is slightly more pronounced here, but not significant, is the fact that 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 an analysis 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 Figure B1 and the accompanying Table B6, we confirm that, as

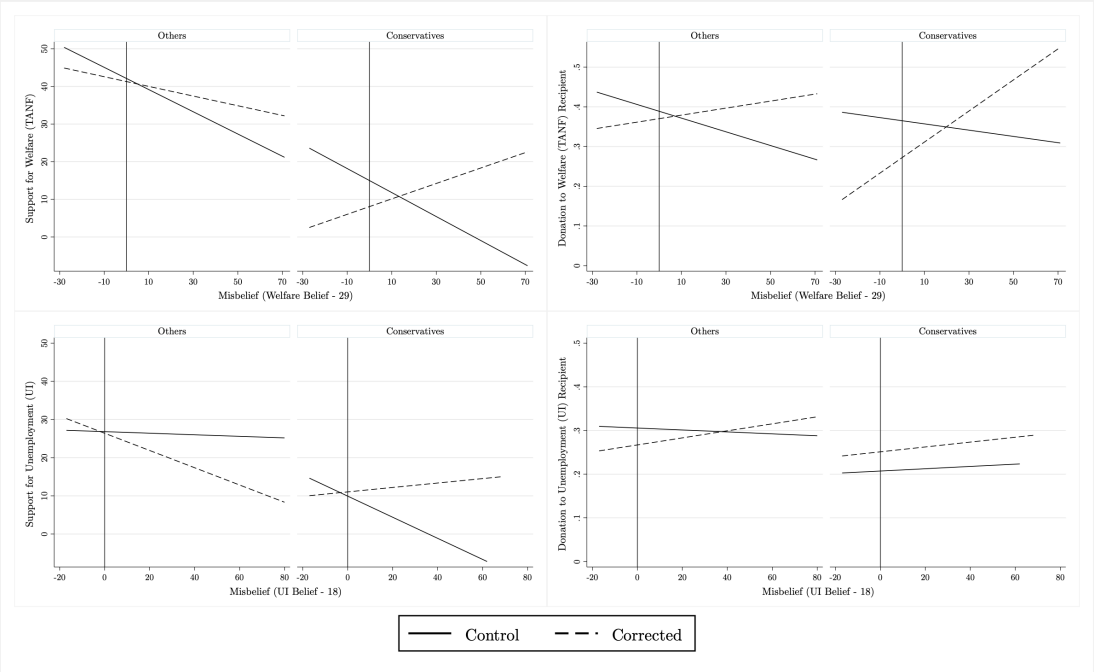


Figure 11: Comparing treatment effects between politically conservative and other participants.

is now anticipated, the subgroup effects are even more pronounced and we conclude that racial animus is a strong driver of our 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 11 and Table 9. As above, we see stronger responses in the TANF experiment than in the UI experiment, but comparing across programs also reveals estimates that jibe with what one would expect from conservative participants. 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 when we looked at the subgroups

of White, implicitly biased or explicitly racist respondents. The belief gradients slope downward again in the control condition, the treatment effects are positive and this effect is stronger among the conservatives but the differences between the groups are more muted.

Table 9: Subgroup treatment effects for participants who are politically conservative.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Others	Conservative	Others	Conservative	Others	Conservative	Others	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

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

When we consider the UI experiment on the bottom of Figure 11 and Columns (5) - (8) in Table 9, the effects are very similar to what we have already seen. None of the estimates are significantly different from zero, reaffirming that race seems to play less of a role in one's support for the UI program. However, what is interesting is that in the UI policy support cell, conservatives respond to the information as predicted ($\beta_3 > 0$) but non-conservatives behave oppositely, similar to what we witnessed using the 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 ($\chi^2 = 5.03$, $p = 0.02$). The last thing to notice about the conservative subgroup is on the left side of Figure 11 where we find that

conservatives simply support these two social safety net programs considerably 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 redistributive programs, regardless of the racial composition of the beneficiaries.

In accordance with our pre-analysis plan, we ran 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 Figure B2 and Table B7 look very similar to those in Figure 11. We also find some evidence in Figure B3 and Table B8, that participants who are more confident in their beliefs than the median are more likely to react 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).

3.8 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 unemployment recipients as more worthy than welfare 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

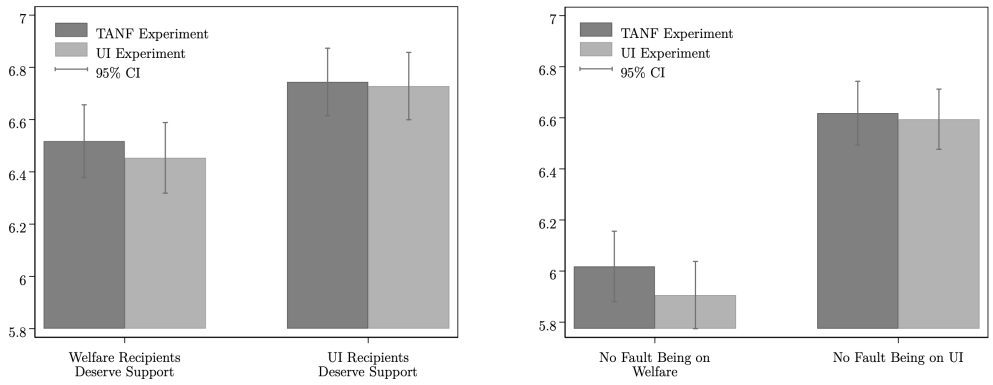


Figure 12: Differences in beneficiary worthiness by program and experimental assignment.

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 12, 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 12, we see an even larger difference in the extent to which our participants think that being on unemployment insurance is no fault of the recipients compared to being on 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 are welfare beneficiaries.[14]

[14]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

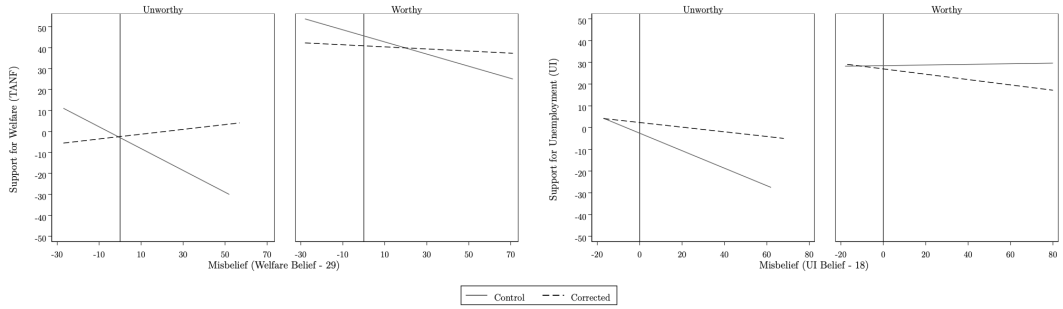


Figure 13: The interaction of worthiness and the treatment effect of information in the Unemployment Insurance experiment.

While the data in Figure 12 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 13. 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 12. We then split the samples in Figure 13 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 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.

significant at lower than the $p = 0.18$ level.

Perhaps the first thing to notice in both panels of Figure 13, 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 13, 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 bottom quintile of the worthiness distribution ($\chi^2 = 1.84$, $p = 0.17$).[15]

On the right of Figure 13, 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 one's we have 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 (on the left), it changes in the hypothesized direction in the UI experiment (on the right) and that 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.

3.9 Stage 3: The Persistence of Beliefs

The third stage of our experiments was designed to examine the extent to which the misbeliefs of our participants persist over time, if uncorrected, or attenuate, if corrected. Specifically, for our findings to extend beyond the immediate time frame surrounding our interventions, we expected that after a month the beliefs about the number of Black people utilizing TANF or UI of our participants in the control treatments would be roughly similar, while participants whose beliefs were explicitly corrected would update them in the direction of the true statistics. We have previously

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

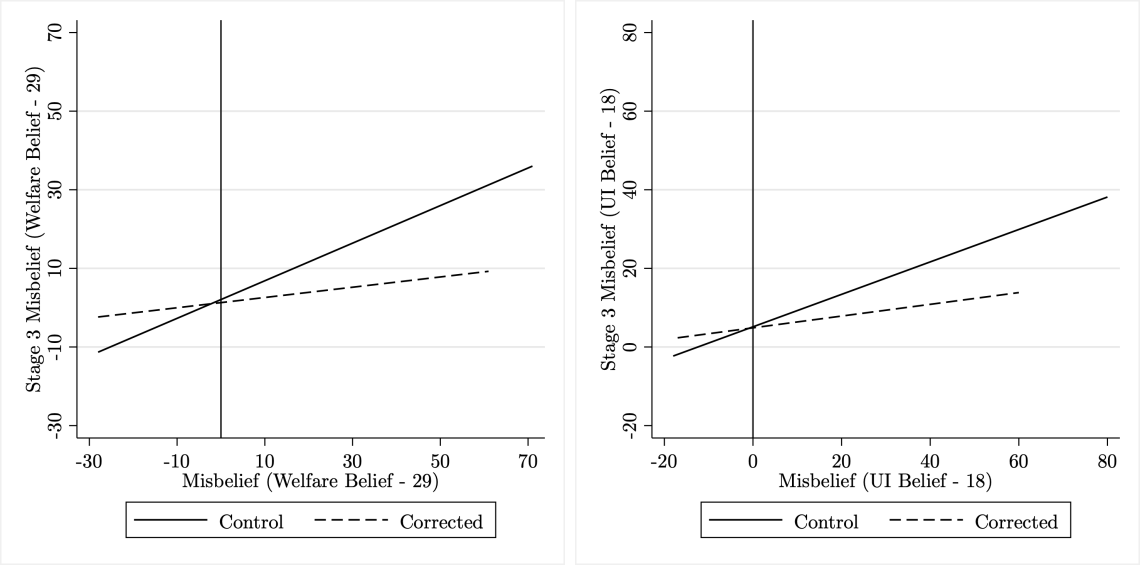


Figure 14: Persistence and attenuation of Stage 3 beliefs.

commented on some of the results referenced in this subsection in Carpenter, Debnam Guzman, Matthews, and Wolcott (2025).

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. In this case, γ_2 measures the stability or persistence of misbeliefs in the control population and the coefficient γ_3 measures the "pass through" of information provision. Here, our prior is that $\gamma_3 \leq 0$. That is, the larger the misbelief gap in the second stage, the smaller it will be in the third if corrected participants update their beliefs based on the new information.

Figure 14 illustrates these estimates and reveals that our participants in both

experiments (TANF on the left and UI on the right) are roughly Bayesian. For participants in the control conditions, misbeliefs separated by a month are positively correlated and as one can see in Table 10, these associations are large and highly 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.

Table 10: 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

Dependent variable is Stage 3 misbelief in TANF or UI experiment. OLS with robust standard errors reported. + indicates controls added. Controls include age, sex, education, income and geographic region.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Most importantly, we also see in Figure 14 and Table 10 that the information we provided about the true usage of TANF and UI, made a lasting impact on the beliefs of our 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 14, 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 midbeliefs are confirmed in Table 10, 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 10 also indicates that adding controls changes these estimates very little. Summing, we find that 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 considerably smaller posterior misbeliefs (i.e., beliefs much closer to the true value) as a result and this treatment effect lasts at least a month.

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 j on one's Stage 2 prior misbelief about policy i (along with a treatment indicator and the interaction) or

$$(Belief_j^{+1} - True_j) = \kappa_0 + \kappa_1 T_i + \kappa_2 (Belief_i - True_i) + \kappa_3 T_i (Belief_i - True_i) + \kappa_4 X_i + \epsilon_i.$$

The first thing to notice in Figure 15, is that people who participated in the control conditions of our two experiments hold policy misbeliefs that are consistent: over- or under-estimating the number of Black recipients in one program leads these people to hold similar misbeliefs about participation in the other program. We also see this in Table 11, 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

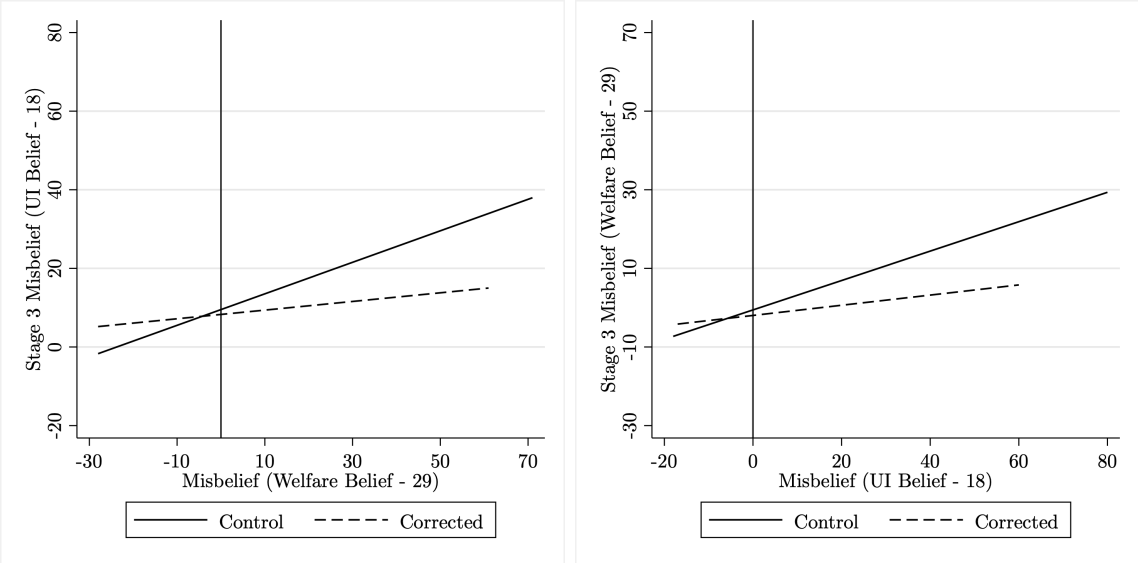


Figure 15: Do beliefs spillover from one experiment to the other?

Stage 3 (posterior) misbelief about both that program and the other program. On the left of Figure 15, 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 too. Further, notice that the intercept is close to 11 (as verified in Table 11), 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 15, 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 11 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 of the "apply the same statistic to the other program" updating heuristic.

Table 11: 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

Dependent variable is Stage 3 misbelief in the other experiment. OLS with robust standard errors reported. + indicates controlls added. Controls include age, sex, education, income and geographic region.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4 Discussion and Concluding Remarks

We are the first to provide a direct causal link between the provision of racial information and updated policy beliefs. Within a large representative 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 driven by racial animus among White participants and by those who report above-median levels of anti-Black preferences. Our main result immediately suggests a naive policy recommendation - racist beliefs in the United States (with their accompanying policy preferences) can be ameliorated through simply informing people of the truth. This conclusion, however, exists in an environment where information receipt is increasingly endogenous, polarized, and correlated with race itself.

We are also the first to formulate and test a hypothesis comparing two policy programs with differing racial histories in this setting. While most of the related literature focuses on welfare alone, we use the Unemployment Insurace program to highlight how equally important perceptions of "worthiness" accentuate or attenuate racialized reactions to various safety net programs and emphasize the role of narrative in constructing economic policy preferences. In our case, differential perceptions of deservingness between the two programs we consider derive from historical events, rather than from policy differences.

We find null effects when correcting beliefs about the racial composition of beneficiaries of Unemployment Insurance. Using two measures, in the aggregate or within any number of subgroups, even among participants with a self-reported high level of 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 14) we know that UI experiment participants updated their beliefs almost identically to participants in the welfare experiment, so this does not explain the lack of a treatment effect for UI. Instead, the results illustrated in Figures 12 and 13 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.

Extensions of our work could examine the role of deservingness in support of redistributive policy 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)). Also, when taken in conjunction with the work of Akesson et al. (2022) and others who find null results of providing true racial information on policy preferences, our results suggest that providing participants with information about base rates regarding racial information is important for identification.

In establishing the importance of racial beliefs 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 includes members of long perceived (and reinforced) out-groups.

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. 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.

Almås, I., A. W. Cappelen, and B. Tungodden (2020). Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than scandinavians? *Journal of Political Economy 128*(5), 1753–1788.

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.

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*.

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.

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.

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.

Lipset, S. M. (1996). *American Exceptionalism: A Double-Edged Sword*. New York, NY: W.W. Norton Company.

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

Luttmer, E. F. P. and M. Singhal (2011). Culture, context, and the taste for redistribution. *American Economic Journal: Economic Policy 3*(1), 157–179.

Massey, D. S. (2009). Globalization and inequality: Explaining american exceptionalism. *European Sociological Review 25*(1), 9–23.

McGhee, H. (2022). *The sum of us: What racism costs everyone and how we can prosper together*. One World.

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

Wexler, S. and R. J. Engel (1999). Historical trends in state-level adc/afdc benefits: Living on less and less. *J. Soc. & Soc. Welfare 26*, 37.

Wolcott, E. L. (2024). Did racially motivated labor policy reverse equality gains for everyone? Technical Report 90, Federal Reserve Bank of Minneapolis.

5 Appendix A: Experimental protocol

5.1 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.

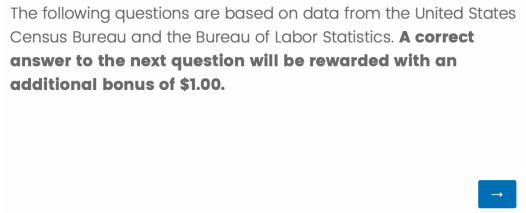
After completing the IAT, which compared the sorting of Black and White people along with positive and negative words, participants were asked about their explicit racial preferences on a $0-10$ scale where 0 was "I prefer Black people to White people", 5 was labelled "I like White people and Black people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

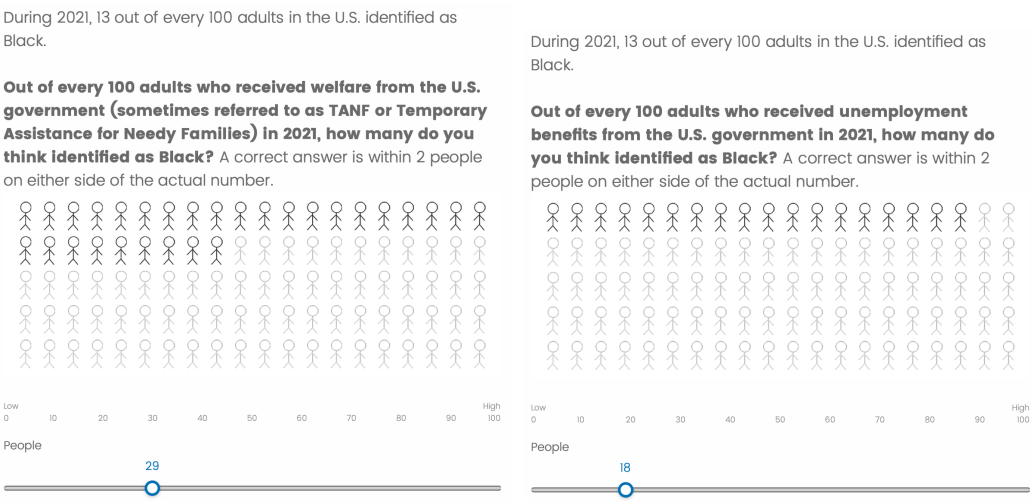
Stage 1 ended by reminding participants that they would be eligible to participate in the next two stages of the experiment (without specifying what would occur in those stages) and then participants were redirected back to Connect for payment.

5.2 Stage 2

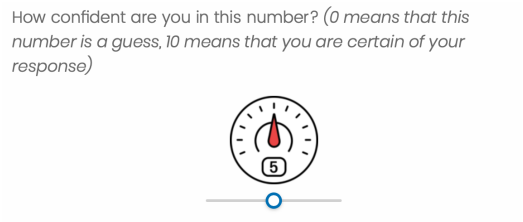
After the same start as Stage 1, in Stage 2, all participants were shown the following prompt informing them that the data they would be asked about came from the BLS and if they answer correctly they would receive $1.



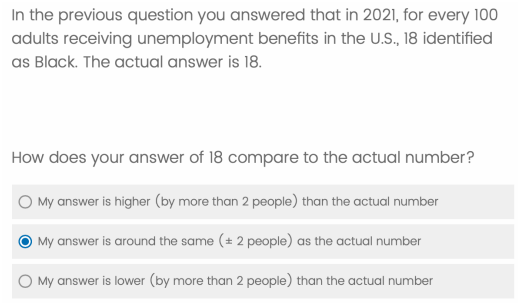
All participants were then asked for their incentivized prior beliefs, with the exact wording depending on the experiment to which they were randomized. On the left below is the prompt used for the welfare experiment. The prompt used for the unemployment experiment is on the right of the figure. In each 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.

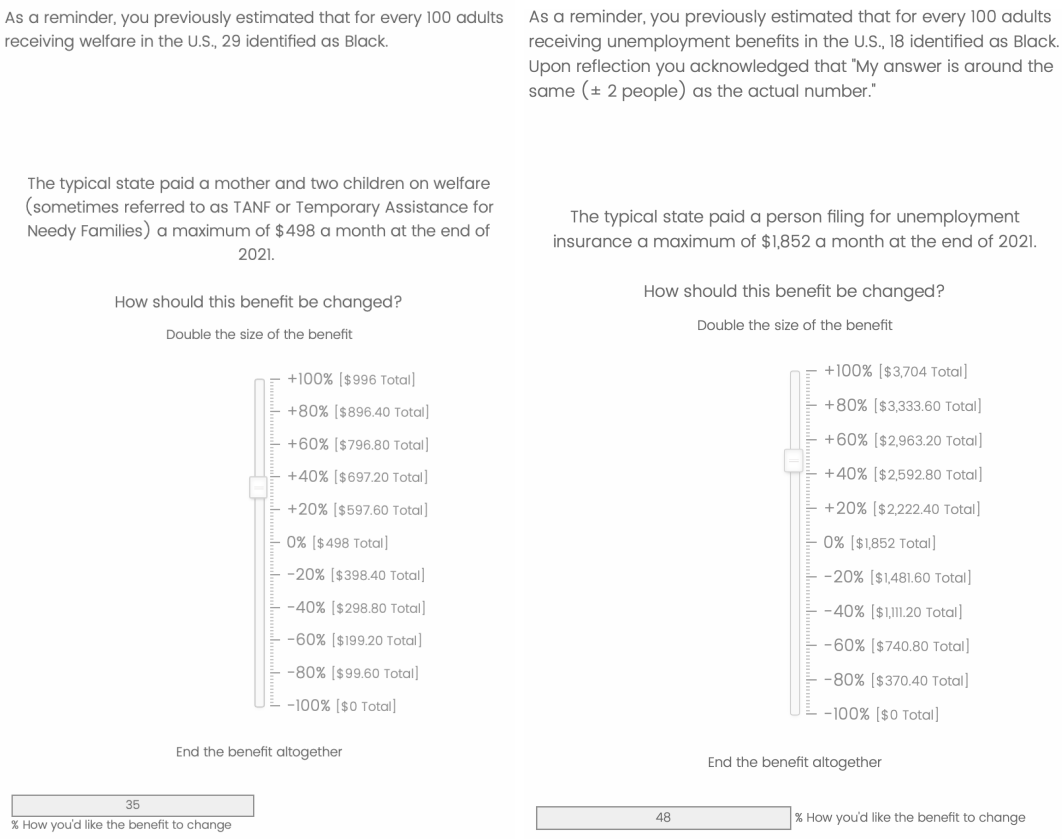


Those participants randomly assigned to the information provision treatment saw some version of the screen below before being asked about their policy preferences. In this example, the participant was sorted into the unemployment treatment condition, where they were first told that the correct statistic 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.



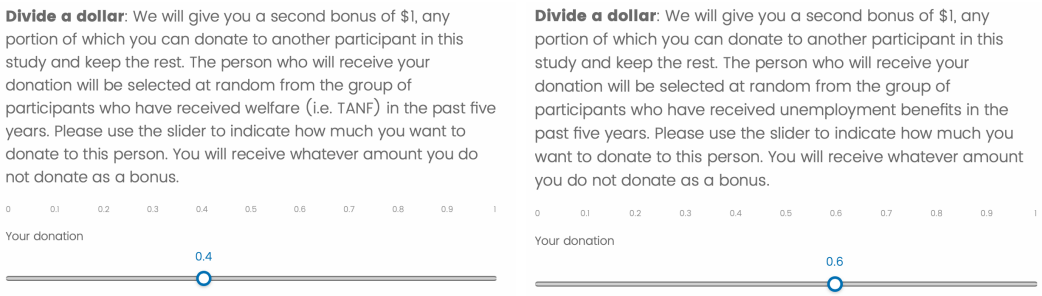
We used versions of the following prompts to solicit unincentivized policy support in the two experiments. The prompt for the welfare control treatment is on the left and the unemployment treatment prompt is on the right of the figure. The prompts all started with a reminder of the belief that a participant 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.

5.3 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).

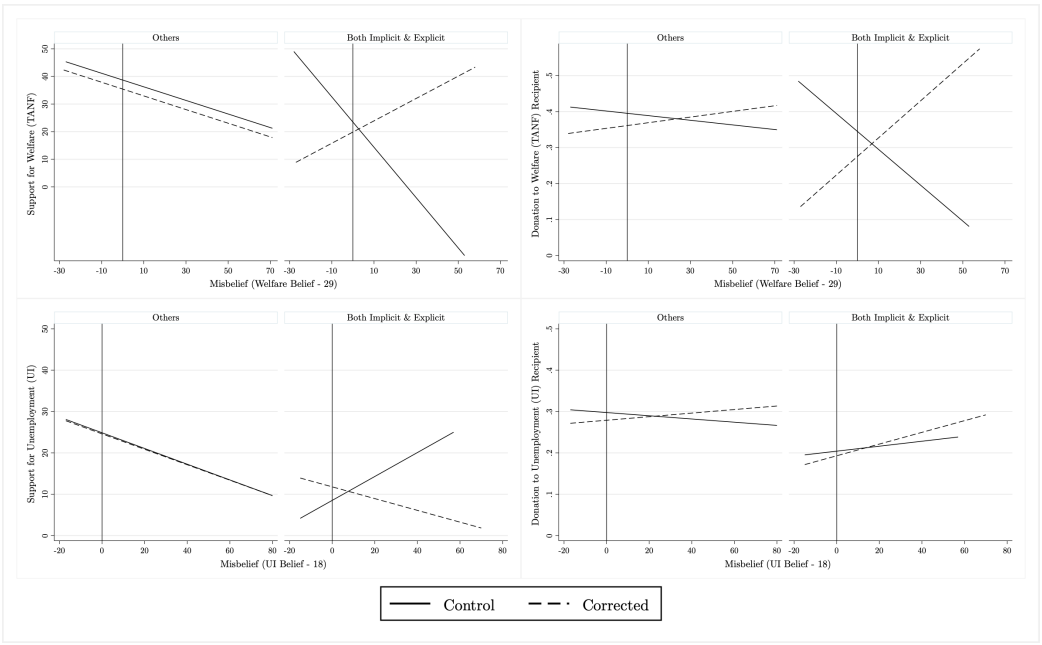


Figure B1: Comparing treatment effects between participants who are both implicitly and explicitly racist and other participants.

6 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

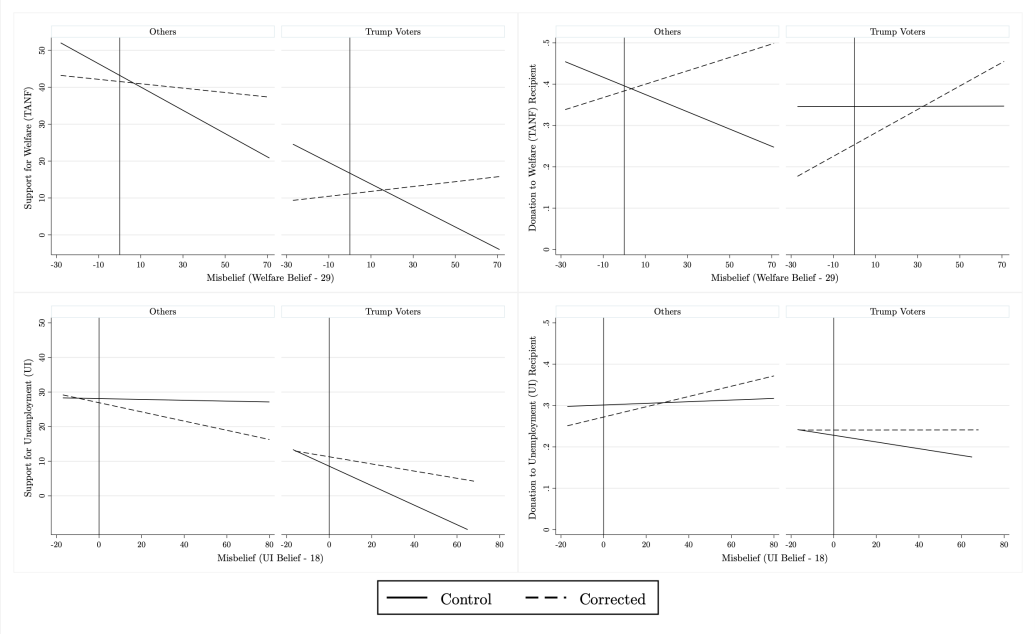


Figure B2: Comparing treatment effects between Trump voters and other participants.

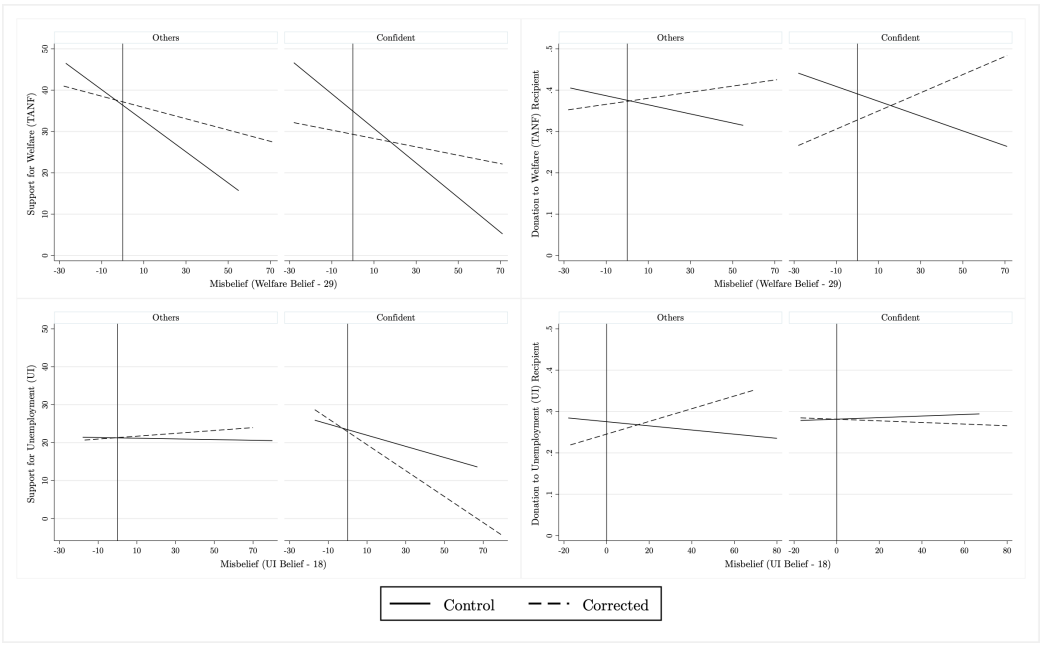


Figure B3: Comparing treatment effects between participants who are confident in their beliefs and others.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Others	Confident	Others	Confident	Others	Confident	Others	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)	(2)	(3)	(4)
	Unworthy	Worthy	Unworthy	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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.