

Do Survey Measures of Racial Prejudice Predict Racial Discrimination?
Experimental Evidence on Anti-Black Discrimination in Resource Allocations.

Scholars regularly measure whites' racial attitudes using symbolic racism (racial resentment) and, more rarely, overt prejudice batteries. Symbolic racism is a strong predictor of race-related policy attitudes and other political outcomes, but has been criticized for being tightly tied to conservative social value orientations and other (non-race related) factors that might explain conservative attitudes toward race-related policies. In this paper, we examine the predictive power of both overt and symbolic racism measures in explaining anti-black discrimination by white Americans. In Study 1 we obtain a behavioral measure of racial discrimination using the Ultimatum Game (UG). In Study 2 we ask white third-party observers to evaluate the fairness of interactions among black and white players in the UG. In Study 1, we find that white responders in the UG were more likely to engage in costly discrimination against black proposers by rejecting offers they would otherwise accept from whites. Further, we demonstrate that overt racism predicts which whites discriminate whereas the symbolic racism measure does not. In Study 2 we also find that overt racism predicts racially biased evaluations of the fairness of resource distributions made by black proposers to white responders but symbolic racism does not, suggesting that racially biased conceptions of fairness are a plausible explanation for the discriminatory behavior of prejudiced individuals. These results have important implications for how we measure whites' racial attitudes and the theoretical constructs underlying both symbolic and overt racism measures.

Kyle Peyton
Yale University, PhD candidate
Department of Political Science
Institution for Social and Policy Studies
77 Prospect Street, PO Box 208209
New Haven, CT 06520-8209
Kyle.peyton@yale.edu

Gregory A. Huber*
Yale University, Professor
Department of Political Science
Institution for Social and Policy Studies
77 Prospect Street, PO Box 208209
New Haven, CT 06520-8209
gregory.huber@yale.edu

December 16, 2017

*Corresponding author.

Intergroup conflict and persistent racial inequality are distinguishing features of American politics. Enduring anti-black attitudes, in particular, are a prominent feature in explanations for America's many institutional exceptionalisms (e.g., Smith 1993), including the comparative absence of social welfare institutions (e.g., Gilens 1999) and the "hyper-segregation" of urban areas (e.g., Massey and Denton 1993). Many of these accounts, either explicitly or implicitly, assume strong connections between racial prejudice and both preferences for racially discriminatory policies and the discriminatory behaviors of white Americans. Consequently, the study of racial prejudice in public opinion is of crucial importance to advancing our understanding of how racial attitudes shape politics and policy in the United States.

Despite this enduring interest in racial attitudes, important questions remain about the theoretical underpinnings of different measures of those attitudes. Two important survey measures of explicit¹ (outwardly expressed) racial attitudes are overt racial prejudice—the degree to which whites believe in the innate racial inferiority of black Americans—and "symbolic racism" (also called "racial resentment" or "modern racism," see Hutchings and Valentino 2004 for a review). Symbolic racism has been operationalized as a blend of both anti-black prejudice (perhaps based in perceptions of group conflict) and traditional (Protestant) moral values potentially at odds with government efforts to rectify the consequences of historical racial discrimination (Sears and Kinder 1971; Kinder and Sears 1981; Sears and Henry 2003). Emphasizing the importance of symbolic racism, Kinder and Sanders (1996) argued that while overt prejudice had declined, "a new form of racial prejudice has appeared" (p. 97–98). Moreover, unlike the overt measures of racial prejudice, whites continued to express conservative views on this measure long after changes in social norms rendered outward expressions of overt prejudice

¹ Other scholars measure implicit racial attitudes, which are unconscious racial biases (e.g. Dovidio, Kawakami and Gaertner 2002), and link those measures to racially discriminatory behavior (e.g., Kubota et al. 2013). Unlike implicit measures, which often require lengthy indirect measurement, explicit measures are quickly and easily fielded on conventional telephone, paper, face-to-face, and internet surveys.

unpopular. Accordingly, work in political science using symbolic racism measures has largely displaced work using measures of overt racism.²

In this paper, we examine whether these survey measures of explicit racial prejudice are prognostic of racially discriminatory behavior. Allport (1954) described prejudice as “an antipathy based on faulty and inflexible generalization.” (p. 9). Discrimination, by contrast, is the act of “[denying] individuals or groups of people equality of treatment which they may wish” because of their group membership (pp. 51-52). Prejudice is therefore a negative evaluation of another person based on their group membership, whereas discrimination is a negative behavior toward that person (Dovidio & Gaertner 2010). Although prejudice is often used to predict discrimination, the two are not synonymous. Indeed, a classic result in social psychology is that prejudicial attitudes expressed in surveys may poorly predict discriminatory behavior (e.g. La Pierre 1934). Historically, overt racial prejudice in the United States was the justification for both personal and political (state sponsored) racial discrimination against blacks, and whites endorsed negative stereotypes of blacks at high levels.³ It follows naturally therefore that overt prejudice should explain an individual willingness to discriminate, but contemporary behavioral tests of the survey measure do not appear to exist. Alternatively, however, social desirability biases may cause individuals who are willing to discriminate to underreport their true level of overt prejudice (see Sears and Kinder 1971; Schuman et al. 1997; Sears and Henry 2003), which would reduce the predictive power of the measure.

² Underscoring this shift, the ANES has included a 4-item symbolic racism scale in presidential election year surveys since 1986. Measures of overt racism have sometimes been included in the ANES, but never consistently. Although questions about differences in work ethic have been asked since 1992, the question about group differences in violence that appears in the 2016 ANES is the first time it has been asked in a presidential election year survey since 1992. Questions about differences in intelligence were asked from 1992 to 2012 and questions about differences in trustworthiness were asked in the 1996, 2000 and 2004 versions.

³ For example, in 1942 nearly 60% of whites believed “Negroes” were less intelligent than whites, but by 1956 only 20% expressed support for racial intelligence differences (see Hyman and Sheatsley 1956; Hyman and Sheatsley 1964). These trends occurred alongside the Civil Rights movement and policy changes such as the desegregation of public schools, suggesting that the decline of overt anti-black prejudice may have helped sustain more racially inclusive policies. Despite these hopes, however, negative racial stereotypes have not disappeared. According to the 2016 American National Election Survey (ANES), for example, approximately 54% of white Americans rated blacks as more violent than whites, and 49% rated blacks as lazier than whites

By contrast, symbolic racism is argued to have emerged after the Civil Rights movement had rendered hostile expressions of bigotry and white supremacy unpopular. This new form of racism was “not the racism of the red-neck bigots of old who spewed forth hatred, doctrines of racial inferiority, and support for de jure segregation” (McConahay and Hough 1976, p. 23). Rather, it reflected an evolution in white racism that affected a broader segment of American society. As McConahay and Hough explained, “As with other forms of American racism, symbolic racism consists of both prejudicial or negative attitudes, opinions, and feelings, and of discriminatory acts or behaviors on the part of white Americans directed toward black Americans” (1976, p. 24). Symbolic racism has since been widely used to predict whites’ attitudes about policies addressing racial inequality (e.g., school busing, affirmative action, etc.) as well as their racial preferences in candidate choice (e.g. voting against black candidates [Pasek et al. 2009; Payne et al. 2009]). Yet the link between symbolic racism and discriminatory behavior is less clear, and whether symbolic racism is prognostic of individual-level discrimination against blacks is largely untested. Further, the correlation between conservative political choices and symbolic racism has been challenged as evidence of construct validity because the measure is also correlated with conservative ideology and related social values. Such ambiguity is heightened by the fact that the survey items used to measure symbolic racism invoke a variety of conservative social value orientations (see Huddy and Feldman 2009 for a review). Those values, rather than racial prejudice, may be the reason individuals who score high on measures of symbolic racism oppose liberal policies and liberal (black) candidates (e.g., Sniderman and Tetlock 1986).

Our reading of this debate about the validity of the symbolic racism measure is that its critics are arguing that one needs evidence other than the measure’s predictive power in racialized policy domains to show it is a measure of prejudice that is prognostic of racial discrimination.⁴ Thus, we argue that if symbolic racism explains a willingness to discriminate, it is natural to understand contemporary symbolic

⁴ Indeed, Carmines, Sniderman, and Easter (2011) make the point that although the two measures (overt prejudice and symbolic racism) both predict racial policy attitudes, this is not evidence that they do so for the same reason. What is left unstated in that work is the specific evidence would show the two measures are theoretically overlapping or distinct.

racism as sharing a racially prejudicial and discriminatory impetus, whereas if it does not (but survey measures of overt racism do), then overt and symbolic racism are conceptually distinct in at least one important way. In this case, symbolic racism may still offer important insight into whites' policy views, but not because it captures a willingness to engaged in racial discrimination against black individuals.

Our research approach is therefore to examine whether these two explicit measures of racial prejudice—overt and symbolic racism—explain anti-black discrimination by white Americans. In our first experiment, we obtain behavioral measures of racial discrimination using a non-anonymous version of the ultimatum game (UG), a strategy previously employed to study the link between implicit racism measured using the Implicit Association Test (IAT) and anti-black discrimination (Kubota et al., 2013). In the UG, a player in the role of proposer, the first mover, offers a division of a fixed endowment between herself and a responder, the second mover. If the responder accepts the offer, each side receives their portion of the division, but if the responder rejects the offer, both players receive nothing. Rejecting a non-zero offer therefore requires that a responder give up real money in order to deny the first mover her proposed share of the endowment. Consistent with Kubota et al. (2013), we find that white responders in the UG were more likely to engage in costly discrimination against randomly assigned black proposers by rejecting offers they would otherwise accept from whites. More importantly, we demonstrate that overt racism reliably distinguishes which whites discriminate whereas the symbolic racism measure does not.

In our second experiment, we explore whether this discrimination is motivated by fairness perceptions using an evaluation experiment in which white Americans in a panel survey act in the role of a third-party observer to rounds of UG play between randomly assigned pairings of proposers and responders. We find that overt racism predicts racially biased evaluations of resource distributions made by black proposers to white responders but symbolic racism does not, suggesting that racially biased conceptions of fairness are a plausible explanation for the discriminatory behavior of prejudiced individuals.

Research Overview

We investigate the link between attitudes and discriminatory behavior in two experiments. Study 1 measures costly racial discrimination against black and white proposers in non-anonymous one-shot Ultimatum Game (UG) play by 738 white responders. Responders were randomly assigned to either a 4 second (time pressure) or 10 second (time delay) decision constraint for the entire experiment and then answered an opinion survey that included measures of symbolic and overt racism. In the experiment, responders played 30 rounds of the UG against 30 unique proposers, 25 of whom were white or black. The race of the proposer and the amount offered were randomly assigned so that they were statistically independent by design. This allows us to test the predictive power of survey measures of racial prejudice for explaining discriminatory behavior, and whether these relationships are affected by the amount of time respondents had to make their decisions. Differences in behavior across time to decision are relevant for understanding whether the frequency of prejudiced behavior is affected by dual process cognition (e.g. Kahneman 2011), which we manipulate by altering the speed with which respondents must make choices.

The standard behavioral prediction for the anonymous one-shot UG is that the responder will play the sub-game perfect Nash equilibrium and accept any positive (non-zero) offer, and the proposer will offer the smallest possible division (Güth, Schmittberger, and Schwarze 1982). Experimental regularities across a variety of human societies show, however, that most proposers offer between 30% and 50% of the endowment and that responders frequently reject offers of less than 20% of the endowment (see Henrich et al. 2004), even in anonymous one-shot interactions (Güth and Tietz 1990). A common interpretation of this pattern is that responder behavior is motivated by a social preference to punish proposers who make offers they perceive as unfair, and that proposers anticipate this and make offers to avoid rejection (Henrich et al. 2001).

The responder's desire to punish unfair offers, a social preference called "negative reciprocity," may be even greater in non-anonymous one-shot UGs when the proposer is a member of an out-group.⁵

⁵ In contrast to in-groups, groups to which the person belongs, an out-group is one that the person does not belong to (see Brewer 1979).

In-group bias is stronger when the groups are made salient with realistic categorizations such as race and ethnicity (see Mullen, Brown, and Smith 1992), but group identity can even be established on the basis of arbitrary differences such as preferences for paintings by artists Klee or Kandinsky (e.g. Tajfel et al. 1971). Numerous studies in economics and psychology have shown evidence of in-group favoritism in resource allocations, often using this “minimal group paradigm” setup (see Lane 2016 and Balliet, Wu, and De Dreu 2014 for meta-analyses). Kubota et al. (2013) provide evidence of costly racial discrimination by white responders against black proposers in the UG and show that this discrimination (rejecting comparable offers from blacks but accepting them from whites) is increasing in respondents’ implicit racial prejudice as measured using an Implicit Association Test (IAT). They note that this anti-black discrimination may be driven by (implicitly) racially prejudiced responders perceiving offers from disliked out-group (black) proposers as less fair than monetarily equivalent offers from in-group (white) proposers.

Although Study 1 allows us to examine the relationship between anti-black stereotypes and behavior, it does not directly examine the perceived fairness of offers across intergroup bargaining scenarios. Study 2 therefore examines how 753 white third party observers perceive allocations in rounds of UG play between white and black proposers. In this experiment, each evaluator rated the fairness of the proposer’s offer in 36 rounds of the UG between 76 unique responders and proposers. The racial pair of the responder and proposer (e.g. black proposer and white responder) and the offer amounts were independently randomly assigned so that the racial pairing and the proposed distribution of the endowment in any particular round were statistically independent by design. This allows us to test how the racial pair affects perceived offer fairness, and whether the observer’s racial attitudes predict racially biased fairness evaluations.

Study 1

In Study 1 we recruited 931 unique workers from Amazon’s Mechanical Turk marketplace (see Berinsky, Huber, and Lenz 2012) and achieved a sample of 738 treated white workers (see Supplementary

Appendix for additional details on sample recruitment and exclusion criteria). The instructions for the UG were explained and subjects were required to answer comprehension questions about at least one practice round of the UG before proceeding.⁶ Subjects were told they would always be in the role of the responder and that they would play 30 rounds of the UG with 30 different proposers. Subjects were paid a flat fee of \$1.25 plus an additional bonus based on the decisions they made in 5 randomly selected rounds of play.⁷ For each randomly selected round, subjects were told they would be paid the amount of any offer they accepted (while the proposer would get the remainder) and \$0 for every offer they rejected (in this case, the proposer would also get nothing).

Subjects were informed they would always be in the role of the responder and that they would play the game 30 times (rounds) with 30 unique proposers. Subjects were randomly assigned to a time pressure (“Fast”) or time delay (“Slow”) decision constraint for the entire experiment. In the Fast condition, subjects were allowed a maximum of 4 seconds to decide whether to accept or reject the proposer’s offer.⁸ In the Slow condition, subjects were required to spend a minimum of 10 seconds before deciding whether to accept or reject the proposer’s offer. The 30 putative proposers were in fact a randomly selected subset of the 160 male faces with neutral expressions (15 white, 10 black and 5 non-white/non-black faces) used by Kubota et al. (2013) in their study of racial bias in the UG. All subjects were exposed to the same set of 30 faces, presented in a random order, with an offer amount (out of an initial endowment of \$1.00) randomly drawn from the set $\{0, 0.05, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50,$

⁶ Subjects answered questions about a second practice round only if they failed to pass the initial comprehension questions. Sixty subjects required a second comprehension check and all of them passed. The Supplementary Appendix provides additional details.

⁷ Average total earnings were \$2.68 (inclusive of the \$1.25 base rate) and the median time to complete the task was 13 minutes (11 minutes for those assigned to the fast condition and 15 minutes for those assigned to the slow condition)

⁸ In the “Fast” condition, if a respondent did not make a decision in the allocated time, their response is missing. 1.4% of responses (N=132) are missing in this condition, compared to 1.0% (N=90) in the 10-second condition. We discuss in the Supplementary Appendix the robustness of our findings to our treatment of these missing observations. We find that restricting our sample to those who provided evaluations of all faces, as well as assuming that offers not responded to were all accepted, or were all rejected, or that all offers to black proposers were accepted while those by white proposer were rejected, does not alter our conclusions. The last approach is an extreme bounds analysis that would bias against the findings we report below.

0.55, 0.60} with equal probability. After the subject completed 30 rounds of play with 30 unique proposers they completed an opinion survey that captured demographic covariates and racial attitudes.⁹ In all analyses presented in this paper, we restrict attention to the 738 white responders in 25 rounds of play against black or white proposers (738 x 25 rounds = 18,450 potential observations without missing data).

Like Kubota et al. (2013), we led subjects to believe the offers from the proposers that appeared on their screens were from real players, but they were instead randomly assigned faces and offer amounts (see Supplementary Appendix for example rounds). After all subjects completed the study they were all simultaneously debriefed and informed of the deception. Subjects were then paid *as if* they had accepted all offers in the five rounds that were selected for determining their bonus payment, thereby maximizing the bonus they could have received.

Although the design of Study 1 is similar to Kubota et al. (2013), it has important differences. First, and most central for our purposes, the measures of racial prejudice included in our study are different. Kubota et al. (2013) measured implicit racial bias using the Implicit Association Test (IAT), whereas we employ explicit measures of racial attitudes that can and are used widely in many different survey settings. Second, unlike Kubota et al. (2013), we did not instruct responders to accept offers they considered to be fair.¹⁰ Third, the larger number of subjects recruited for this study came from an online labor market and played 30 rounds of the UG, whereas Kubota et al. (2013) recruited 49 subjects from the New York University campus and surrounding area to each play 160 rounds of the UG. Finally, this study randomly assigned subjects to two different decision scenarios whereas Kubota et al. (2013) only investigated behavior under a 4 second decision constraint (what we label the Fast condition).

⁹ We measure racial attitudes post-treatment, like Kubota et al. (2013), to avoid priming subjects' racial attitudes before playing the UG. Prior research finds that racial attitudes are socialized early and remarkably stable over time, reducing concerns that playing the UG affected measured racial attitudes (see Harding et al. 1969; Sears 1988). Because the offer amounts and race of proposer were randomly assigned over 30 rounds, there are small differences in the average offer by race, reducing concerns that playing the UG would induce differences in racial attitudes.

¹⁰ Compare the instructions of the UG described in their Supplementary Materials with ours to see this difference.

Study 2

In Study 2 we designed an evaluation task to explore how whites in the role of a third party observer *perceived* intergroup resource allocations in the UG. If whites impose different fairness standards on black proposers' behavior toward white responders, this should lead to differences in their evaluations of the fairness of proposed monetary splits across different racial pairs. Given that whites in Study 1 were more likely to reject comparable offers from black than white proposers, we therefore expected that whites would evaluate offers from whites to whites more positively than offers from blacks to whites (even though they do not have a personal stake in the outcome). Additionally, to rule out the possibility that all intergroup interactions are simply evaluated differently, we also compare fairness evaluations of offers to black responders from both white and black proposers.

The evaluation task was administered in the second wave of an online panel study. In Wave 1, a target population of white Americans was recruited from Survey Sampling International (SSI) for a public opinion survey that included demographics and measures of overt and symbolic racism. To minimize potential demand effects associated with measuring racial attitudes and evaluations in the same survey, approximately 10 days later all subjects who completed the Wave 1 survey (N=1715) were invited to Wave 2 (60%, N=1029, completed), a putatively unrelated study on decision making (see Supplementary Appendix for additional details).

In Wave 2, subjects were shown images from 41 rounds of play in the UG and led to believe the rounds had already occurred between real humans. Subjects were asked to evaluate the fairness of the proposer's offer and the likelihood the responder accepted. Again we used a random sample of the neutral male faces in Kubota et al. (2013) for all UG pairs. Evaluators were told the proposer saw the responder prior to deciding on the offer to make, and the responder likewise saw the proposer and the offer amount before making their decision. Offer amounts were randomly drawn from the set $\{0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50\}$. In order to increase statistical power in the region of the offer distribution where there was the most variation in acceptance behavior in Study 1, offers in $\{0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45\}$ were assigned with probability 0.10 and offers in $\{0, 0.05, 0.45, 0.50\}$

were assigned with probability 0.05. The racial pairs, 82 unique faces sampled without replacement, were randomly assigned across the 41 unique UG rounds.

For each round of play, subjects were first asked “How fair was the Proposer’s offer?” Responses were captured using a continuous sliding scale from “completely unfair” (0) to “completely fair” (100), where “neither fair nor unfair” (50) served as a neutral midpoint. Next, subjects were asked (on the same screen while the faces and offer amounts were still displayed) “How likely is it that the Responder accepted the offer?” Responses were captured using a continuous sliding scale from “certainly did not accept” (0) to “certainly accepted” (100) where “neither likely nor unlikely” (50) served as the neutral midpoint. As with the individual decision making experiment (Study 1), the UG was explained and subjects were required to answer comprehension questions about at least one practice round of the UG. 267 of the 1029 subjects who returned to complete Wave 2 of Study 2 failed the comprehension test twice in a row and were excluded from analysis (see Supplementary Appendix for details).¹¹ In all analyses presented in this paper we restrict attention to 741 white evaluators who passed a comprehension test and completed at least 75% of the evaluations assigned to them.¹² We focus on their evaluations in the 36 instances of UG play involving black or white proposers and responders (741 x 36 rounds = 26,676 potential observations without missing data).

Measures of Overt and Symbolic Racism

The measure of overt racism we used is based on 4 questions that have previously been included on the General Social Survey and other surveys of public opinion (see Huddy and Feldman 2009). Each item in the overt racism scale provides a measure of a subject’s willingness to endorse group-level black

¹¹ As with Study 1, any subject who failed the first set of comprehension questions was presented with a detailed explanation of the correct answer and given a second opportunity to pass a different version of the same test. Subjects who failed the test a second time were not told that they failed the test a second time and were allowed to continue with the study. The significant difference in pass rates between the SSI sample (74%, Study 2) and MTurk samples (100%, Study 1) is consistent with prior research finding MTurk workers are more attentive to instructions than research subjects drawn from other populations (see Hauser and Schwarz, 2016).

¹² We exclude 14 respondents who completed fewer than 75% of their assigned evaluations. These respondents all stopped answering evaluations at some point prior to the end of the experiment. See Supplementary Appendix Table A1 for robustness of results to the treatment of missing data.

inferiority in work ethic, trustworthiness, intelligence, and propensity for violence using a sliding scale (see Feldman and Huddy 2005). Each responder provided a measure on each of the four dimensions using a 7-point scale for each of four racial groups — Asians, blacks, Hispanics and whites.¹³ A positive difference in a responder's score for "whites" versus "blacks" indicates a belief in group-level white superiority. The white-black differences for each of the questions are summed and divided by 4 (the number of questions) to provide our *Overt Racism Scale* with range [-6,6].

The measure of symbolic racism is identical to the 4-item battery of questions asked on various waves of the ANES since 1986. Respondents were asked to agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly with each of these four statements:

1. Irish, Italians, Jewish and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors.
2. Generations of slavery and discrimination have created conditions that make it difficult for blacks to work their way out of the lower class.
3. Over the past few years, blacks have gotten less than they deserve.
4. It's really a matter of some people not trying hard enough; if blacks would only try harder they could be just as well off as whites.

We note that these items potentially invoke a larger and more ambiguous set of considerations than the overt racism measures. For example, the first question asks agreement with 1) a statement about the historical behavior of certain groups, 2) a comparison of their behavior to blacks, and 3) "special favors" for blacks. We construct the *Symbolic Racism Index* with the canonical transformation used in prior empirical studies (e.g. Payne et al. 2010; Tesler 2012). Responses to each of the questions are coded from 0 to 1 by .25 increments (1 being the most resentful response and 0.5 a neutral midpoint) then summed and divided by 4 (the number of questions) to make a scale with range [0,1].

Prior empirical work on whether symbolic racism predicts individual-level discrimination is ambiguous. In a 2x2 factorial experiment with 81 white undergraduate subjects, McConahay (1983)

¹³ In Study 1, responses on this scale were recorded as discrete integers, while in Study 2 responses were continuous on the range [1,7]. In both studies, respondents evaluated all four groups simultaneously so that relative differences expressed among groups were both visually and numerically apparent.

found those scoring high on racial resentment, measured with the Modern Racism Scale¹⁴, evaluated a resume with a photo of a black (versus a white) job applicant less favorably when this resume was presented first, but more favorably after viewing a mix of more and less qualified resumes without photos. This finding suggests symbolic racism does not always lead to a willingness to discriminate against a black individual. In a related study, Brief et al. (2000) found that baseline levels of racial resentment did not predict discrimination in resume evaluations because both low and high racism individuals recommended hiring an equal number of black candidates. However, directives from a business superior that explained white candidates work better with white co-workers led both high and low racism subjects to recommend fewer black candidates, but the effect was larger for those that scored higher on survey measures of racism. What is unclear is whether this difference reflects a greater desire to discriminate or greater deference to the leader's directive.

One argument that led researchers away from using measures like the overt racism scale is that social desirability concerns may motivate respondents to rate all groups equally (Huddy and Feldman 2009). Indeed, the modal value on the overt racism measure in both of our studies was 0, indicating that many respondents rated blacks and whites equally. These critiques, which would tend to depress the predictive power of overt prejudice measures (if those who are prejudiced and would discriminate pooled with those who are not prejudiced and would not discriminate), are generally directed at data collected in conventional face-to-face or telephone interviews. Web administered surveys, by contrast, have been shown to increase respondents' reporting accuracy and willingness to disclose sensitive information (Kreuter, Presser, and Tourangeau 2008). Consistent with this prior work, we find that a substantial proportion of respondents are willing to endorse negative stereotypes of blacks (compared to whites).

¹⁴ The Modern Racism Scale used in McConahay (1983) and Brief et al. (2000) is a different question than the 8-item "Symbolic Racism 2000 Scale (SR2K)" introduced by Henry and Sears (2002). Four of these eight items make up the widely used 4-item ANES measure. Modern Racism is derivative of Symbolic Racism and, in the past, these were viewed as the same concept and measure (Henry, 2010). The scale used in McConahay (1983, p. 557) was the same scale used in the seminal study of symbolic racism by Kinder and Sears (1981).

Figure 1 plots the univariate distributions of the overt (Panel A) and symbolic measures (Panel B) for the analysis samples of white subjects from both Study 1 (MTurk, N=738) and Study 2 (SSI sample, N=741), as well as their bivariate relationship (Panel C). In Study 1, 42% of the white responders were willing to endorse views of group-level black inferiority and therefore have positive scores on the overt racism measure. In Study 2, 59% of the white evaluators scored above 0 on the overt racism measure. The mean level on the -6 to 6 overt racism scale is 0.44 among white responders in the MTurk sample and 0.45 among the white evaluators in the SSI sample. On face value, these patterns undercut the concern that survey respondents are unwilling to express overt anti-black views.

<<Figure 1 about here>>

The mean symbolic racism score among white responders in the MTurk sample was 0.50 and 46% of respondents were coded as symbolic racists (Symbolic Racism Index > .5), while in the SSI sample the figures were 0.56 and 52%.¹⁵ The latter figure is comparable to the average score among white Americans according to the 2016 wave of the American National Election Study (ANES) survey (unweighted mean = 0.58). The bivariate correlation between the overt and symbolic scale measures is 0.38 in the MTurk sample and 0.35 in the SSI sample.

Results: Study 1 Racial prejudice and discriminatory behavior

In *anonymous* one-shot instances of the Ultimatum Game, most proposers offer between 30 and 50% of the endowment and offers below 20% are frequently rejected across a variety of populations, including respondents recruited using MTurk (see Camerer and Fehr 2004; Rand et al. 2013). In non-anonymous games, however, there may be reasons to expect differences in acceptance rates when the proposer is a member of an out-group, although the expected direction of this effect is less clear. The in-group favoritism hypothesis suggests that responders will react more favorably to unfair offers from in-group proposers because the responder cares about the well-being of the proposer, which is diminished by

¹⁵ That the MTurk sample scores lower on the symbolic racism measure is unsurprising given the correlation between scoring high on that measure and conservative ideology and the fact that prior research finds the MTurk pool is more liberal than the general population (Berinsky, Huber, and Lenz 2012).

rejecting offers. By contrast, the expectancy violation hypothesis suggests that responders will react *less favorably* to unfair offers from in-group proposers because they expect in-group members to consider their well-being by making more generous offers and react negatively to those who do not (see Fiske and Taylor 1991; Valenzuela and Srivastava 2012). For interactions between white and black Americans, prior empirical evidence finds white responders will discriminate against black proposers by rejecting their offers at higher rates (Kubota et al. 2013). No prior work that we are aware of examines how white proposers behave toward black responders, or how third party evaluators judge the fairness of interactions between white and black proposers (the focus of Study 2).

In Study 1, 69% of offers were accepted by the 738 white responders in the 25 rounds of play with black or white proposers (N=18,228 subject-rounds with a measured acceptance decision). There is strong evidence that white responders were less likely to accept offers from black proposers. Figure 2 shows the proportion of offers accepted (the vertical axis) from black and white proposers (white proposers are the solid line, black proposers are the dashed line) at each of the ten randomly assigned offer amounts. The likelihood an offer is accepted is increasing in the offer amount, but marginally unfair offers (20-30% of the endowment) from black proposers were about 4 percentage points ($p < 0.01$, t-test difference of means) less likely to be accepted than offers from white proposers.

<<Figure 2 about here>>

To leverage the independent random assignment of faces and offer amounts across rounds, and to account for the within-subjects design, we regress accept/reject decisions (coded 1/0 respectively) on an indicator for the (randomly assigned) proposer's race (coded 1 for black, 0 for white), the offer amount (entered as indicators for each offer amount), and the round of play (entered as indicators). In a second model, we also leverage the random assignment of the 4 or 10 second decision constraint and estimate an interaction between the randomly assigned decision condition and the race of the proposer. Table 1

presents estimates for the likelihood of acceptance from linear regressions with robust standard errors clustered at the responder level for both models.¹⁶

<<Table 1 about here>>

Several patterns are apparent from these results. First, confirming the graphical presentation, larger offers are more likely to be accepted than smaller ones (the excluded category is a \$0.00 offer). Per the results from Column (1), an offer of \$0.60 was 64 percentage points more likely to be accepted than an offer of \$0.05 (.892-.249, $p < 0.01$). Second, controlling for offer amount and round of play (coefficients not reported) does not materially alter the finding that offers from black proposers were less likely to be accepted. On average, there was an approximately 1.3% *decrease* ($p < 0.01$) in the probability a white responder accepted an offer when the proposer was black. To put this result in substantive terms, a 1.3 point difference is about 33% of the decrease in the average likelihood an offer from a black proposer is accepted when an offer is \$0.40 rather than \$.50 (4=99-95%). Third, per the results from Column (2), there is no evidence of an average effect of the 10 second (“Slow”) treatment, relative to the 4 second (“Fast”) treatment, on the likelihood of offer acceptance. Further, the coefficient on the interaction between the decision treatment and the proposer’s race provides no evidence that out-group bias operated differently across the Slow and Fast decision constraints, a finding consistent with previous research designs that utilize arbitrary group distinctions (e.g. Everett et al. 2017).¹⁷

Next, we turn to our core question, which is whether the overt and symbolic racism survey measures predict observed bias against black proposers in the UG. We classify “overt racists,” who score 1 on the *overt racism indicator*, as those who endorse group-level black inferiority (overt scale score > 0) and “symbolic racists,” who score 1 on the *symbolic racism indicator*, as those who exhibit a level of

¹⁶ Parallel analysis using Probit appears in the Supplementary Appendix and yields substantively similar conclusions.

¹⁷ This analysis helps rule out the possibility that discrimination against black proposers is driven solely by the need to make quick decisions across offers from black and white proposers. We find the greater propensity to reject proposals from black proposers is no different across the amount of time responders have to make their decisions.

symbolic racism above the (neutral) midpoint of the symbolic racism scale (symbolic scale score > .50).¹⁸ If symbolic and overt racism predict racial discrimination then there should be a negative effect on the interaction between each indicator of racial prejudice and the indicator that denotes a black proposer. To conduct this analysis, we regress the binary accept/reject decision on indicators for offer amount, round of play, the proposer's race, the appropriate survey measure (symbolic or overt), and an interaction between the survey measure and the proposer's race.

Results from three different specifications are shown in columns (3) through (5) of Table 1. The first (column 3) includes the symbolic indicator and its interaction with the proposer's race (coded 1 for black, 0 for white), the second (column 4) includes the overt indicator and its interaction with the proposer's race, and the third (column 5) estimates both interactions with the proposer's race in the same specification. The estimates shown in column (3) reveal that non-symbolic racists were about 1.4% less likely to accept offers from black than white proposers ($p < 0.05$). Symbolic racists, however, are no more likely to discriminate against offers from black proposers (the coefficient on the interaction is positive, rather than negative, but imprecisely estimated). Among symbolic racists, the estimated likelihood of accepting an offer from a black proposer was about 1.1 percentage points smaller than for a white proposer ($-1.4 + 0.3$), a statistically indistinguishable difference of .3 percentage points relative to non-symbolic racists ($p = 0.76$). Overall, this result shows that the symbolic racism measure was a poor predictor of discrimination against black proposers.

By contrast, the estimates from column (4) show that non-overt racists were about 0.5 percentage points less likely to accept an offer from a black proposer, a difference that was not statistically distinguishable from zero ($p = 0.41$). Overt racists, however, were about 2.5% less likely to accept offers from black proposers ($-.5 - 2.0$, $p < 0.001$), a statistically distinguishable difference of 2.0 points relative to non-overt racists ($p = 0.03$). The survey measure of overt racism is therefore a good predictor of racial discrimination, and there is little evidence that non-overt racists discriminated. In substantive terms, this

¹⁸ This is the dichotomization used in prior work in political science, see Pasek et al. 2009. We assess the robustness of this result to alternative cutpoints in the Supplementary Appendix, as well as to tripartite measures of overt and symbolic racism.

2.5 point effect represents a decrease in acceptance rates for offers by black proposers of about 4% relative to the average 63% acceptance rate for an offer of 15 cents, and is about 28% of the decrease in the likelihood of acceptance associated with an offer being 15 cents rather than 20 cents (9%).

Finally, the estimates in column (5) confirm this inference is essentially unchanged in a model that includes both interactions simultaneously.¹⁹ Per this specification, overt racists are about 2.3 points less likely than non-overt racists to accept offers from black proposers than white proposers ($p = 0.02$). For symbolic racists, as well as those who are neither overt nor symbolic racists, there is no evidence of discrimination against black relative to white proposers (for these groups, the effect of a black proposer is indistinguishable from zero, and for symbolic racists the point estimate is positive). We can reject the hypothesis that the two interaction effects (overt x black proposer and symbolic x black proposer) are equal ($p = 0.04$), and the 95% confidence interval for the interaction term symbolic x black proposer is -0.009 to 0.028. In other words, overt racism is a good predictor of discrimination against black proposers, even after accounting for the possibility of different levels of discrimination among symbolic racists.²⁰ Moreover, whether one accounts for overt racism or not, symbolic racism does not predict anti-black discrimination in the UG. Overall, this is the first evidence we are aware of to directly compare the predictive power of overt and symbolic racism survey measures in explaining racial discrimination, and this test reveals that overt racism, but not symbolic racism, identifies which individuals engage in this behavior.

¹⁹ Only 196/728 (27%) of responders were classified as both Overt and Symbolic Racists.

²⁰ A related concern is that if the symbolic racism scale measures both a willingness to engage in racial discrimination and ideology, failing to account for ideology may mask the measure's ability to predict racial discrimination. For this reason, we also estimated models after including a measure of respondent ideology (7 point scale, 1=very liberal to 7=very conservative) and the interaction between that measure and a black proposer. These results appear in Appendix Table A4. After including both terms in the regression specification, we continue to find no evidence that the symbolic racism measure predicts anti-black discrimination, implying that the correlation between the symbolic racism measure and ideology is not obscuring its predictive power. Overt racism, by contrast, continues to predict discrimination against black proposers.

Results: Study 2 Do fairness expectations explain why overt racists are more likely to discriminate?

The results from the previous section established two important facts. First, whites were less likely to accept offers from black than white proposers in the UG. This is a costly behavioral measure of racial discrimination that cannot be explained by the offer amount or the round of play. Second, this discriminatory behavior was strongest among whites who endorsed group-based black inferiority on survey measures of overt racism. One potential explanation for the observed behavior, suggested by prior work, is that white responders impose higher fairness standards on black proposers than white proposers. Study 2 allows us to further examine if the imposition of different fairness standards is a potential explanation for the observed pattern of racial discrimination found in Study 1.

In Study 2 we restrict attention to 741 white (non-Hispanic) evaluators and 36 rounds of the UG played involving only black and white players ($N = 26,651$ subject-rounds with non-missing outcome measures). While offers by white and black proposers to white responders simulate the interactions from Study 1, Study 2 also includes offers by white and black proposers to black responders. This allows us to examine whether there is something in particular about whites' evaluations of the fairness of black behavior that extends beyond intergroup negotiations that might explain the pattern seen in Study 1. For example, perhaps all negotiations involving black players, regardless of their role as a proposer or responder, are viewed differently by whites, even those that do not involve other whites.

As expected, the results from Study 2 show a strong association between the perceived fairness of an offer and the predicted likelihood it was accepted: higher offers were on average perceived as fairer and more likely to be accepted by the responder.²¹ There was, however, clear evidence that white evaluators, on average, viewed intergroup rounds between black and white players, as well as rounds between two black players, as systematically less fair than rounds of play between whites only. Differences in fairness evaluations in offers to whites by proposer race were strongest among overt racists, who gave systematically lower fairness evaluations when offers were made by black proposers to white responders.

²¹ The correlation between perceived fairness and predicted probability the offer was accepted was .86.

Figure 3 plots the relationship (averages by offer amounts) between the randomly assigned moderate offer amounts (\$.15 to \$.40 on the horizontal axis) and perceived fairness (0 to 100, vertical axis) for each of the 4 race pair conditions: White Proposer and White Responder, Black Proposer and White Responder, White Proposer and Black Responder or Black Proposer and Black Responder. The top two panels partition the data by the race of the responder while the bottom two panels partition by the race of the proposer. Across all panels, the perceived fairness of an offer is increasing in the offer amount.

<<Figure 3 about here>>

Focusing next on racial differences, the top left panel shows that offers to white responders by black proposers (the dotted line) are perceived as less fair than offers by white proposers (the solid line) for offer amounts greater than \$0.30. At the same time, the upper right panel shows that offers to black responders by white proposers (solid line) are perceived as less fair than offers by black proposers (dotted line) over a similar range of offers. In other words, intergroup offers are, on average, perceived as less fair than within-group interactions. Comparing the bottom two panels, which just partition the data shown in the top two panels differently, the lower left panel shows that offers by whites to blacks are perceived as less fair than offers by whites to whites. In the lower right panel, which presents evaluations of offers by black proposers, there is also more muted evidence that offers from black proposers to white responders are perceived as less fair than offers to black responders.

We more formally investigate differences in fairness evaluations with regression analyses that appear in Table 2. Specifically, we predict fairness evaluations using indicators for offer amount, round of play, and racial pairs (*black proposer white responder* (BW), *black proposer black responder* (BB), *white proposer black responder* (WB), with the omitted reference category *white proposer white responder*).²² As with Study 1, all specifications use robust standard errors clustered at the evaluator level.

<<Table 2 about here>>

²² We also examined evaluators' perceived likelihood that the responder will accept the proposers offer. Findings are similar to those for fairness evaluations. Results available upon request.

The coefficients shown in column (1) confirm the earlier graphical presentation that intergroup interactions are perceived as less fair. The -0.65 ($p = 0.01$) coefficient for BW means offers to white responders by black proposers were, all else equal, perceived as less fair than identical offers by white proposers to white responders. The difference WB – BB, is a nearly identical -0.61 ($p = 0.03$) and means that when the responder is black offers from white proposers are also perceived as less fair than offers from black proposers.²³

In columns (2) and (3) we separately examine the predictive power of the symbolic and overt racism survey measures, respectively, in explaining these differences in fairness evaluations, while in column (4) we estimate a model with both sets of interactions included simultaneously. We do so by including indicators for the binary symbolic and overt indicators interacted with each racial pair combination (because we include all racial pair combinations for the interaction terms, we do not include the symbolic and overt indicators separately). Substantively, our results are somewhat affected by whether one includes the two sets of interactions separately or simultaneously, and so we focus on the column (4) specification that accounts for the correlation between the two racial attitude measures and which we therefore think of as more conservative.

To understand the meaning of the results presented in that specification, we begin by examining predicted differences in evaluations of offers to white responders by white, rather than black, proposers. For ease of interpretation, these comparisons are summarized in Panel A of Table 3. First, among non-symbolic/non-overt racists, the average offer by a white proposer is perceived as .36 units *less* fair than a comparable offer by a black proposer. This estimate is indistinguishable from 0. Among symbolic racists, the estimate is a similar .06 units less fair, which is also indistinguishable from 0 or the estimate for non-symbolic/non-overt racists (difference in differences [DID] = .30 $p=.55$). Both symbolic racists and non-symbolic/non-overt racists therefore do not seem to distinguish the fairness of offers to whites by whites

²³ While we focus in this analysis on differences in proposer race holding responder race constant because it is the most clearly linked to our theory, we can also make other comparisons fixing the proposer's race.

from comparable offers by blacks, just as in Study 1 they did not appear to be more likely to reject offers from blacks than whites.

<<Table 3 about here>>

But among overt racists, the results are starkly different. Per these results, overt-racists evaluate offers to whites by whites as 1.08 ($p=.02$) units more fair than equivalent offers by black proposers, a DID of 1.44 units that is statistically distinguishable from the difference in evaluations by race of proposer for non-overt/non-symbolic racists ($p<.01$). Overt racists therefore appear to hold black proposers to a different standard than white proposers. These fairness evaluations may help explain their willingness to engage in costly discrimination in Study 1—overt racists appear to believe black proposers need to give more of their share in order for the black proposer's behavior to be perceived as fair as the white proposer's behavior.

This analysis reveals overt racists—but not symbolic racists or non-overt/non-symbolic racists—discriminate against blacks in cases in which blacks, rather than whites, propose resource allocations to whites. Does this preference among overt racists for same-race resource allocations extend to cases in which the recipient of a proposed division is black rather than white? In short, we find no evidence that it does. Panel B of Table 3 displays difference in fairness evaluations for offers to black responders by black, rather than white, responders.

Among non-overt/non-symbolic individuals we find that proposals to black responders by black proposers are perceived as .98 units more fair than equivalent offers by white proposers ($p=.03$). Among this population of whites, therefore, whites have to give more to blacks than blacks do for an offer to be perceived as equally fair.²⁴ Symbolic racists also display a similar preference for in-race interactions: they rate offers by blacks to blacks as 1.18 units fairer than offers by whites to blacks ($p=.03$), although this difference is substantively and statistically indistinguishable from the behavior of non-symbolic non-overt racists (DID=.22, $p=.69$).

²⁴ There are many potential explanations for this latter effect, including empathy, a belief blacks are worse off than whites on average and therefore have less money to give, etc. We do not have data to test or distinguish among these explanations.

Overt racists, however, do not distinguish by the race of the proposer in interactions with black responders, rating offers by black proposers only .17 units fairer than offers by white proposers ($p=.74$). This -.79 difference in differences compared to non-overt/non-symbolic racists is not statistically significant ($p=.16$), but is suggestive that overt racists are distinct in two ways. First, they do not expect whites to give more to blacks to be as fair as when blacks interact with other blacks. Second, and more important theoretically, they do not regard all intergroup racial interactions as intrinsically unfair. Instead, overt racists regard offers *to whites from blacks*, rather than offers *from whites to blacks*, as less fair. This fairness penalty is specific to interactions in which a white person, rather than a black person, is the recipient of a proposed resource allocation.

Overall, comparing the performance of the symbolic and overt racism measures, symbolic racists and non-symbolic/non-overt racists appear to view interactions where blacks receive resources from whites as less fair than when they receive resources from blacks, while viewing interactions in which whites are the recipients of resources as unaffected by the race of the proposer. The fact that symbolic racists did not penalize resource allocations that were favorable to blacks is surprising because high levels of symbolic racism are correlated with opposition to policies, like affirmative action, that are designed to give preferential treatment to black Americans as a group. One interpretation of this opposition is that symbolic racists reject policies designed to reduce racial inequality because they believe they are unfair because the perceived beneficiaries are black (Rabinowitz et al. 2009). But our results suggest symbolic racists may view group targeted policies in general as unfair, which may explain why the measure does not predict fairness evaluations in situations where racial inequality is a product of individual behavior.

Overt racists, on the other hand, only view resource allocations to whites as less fair when the person who decides on the allocation is black, and view resource allocations to blacks as unaffected by the race of the proposer. Turning back to the results from Study 1, we found that whites on average, and overt racists specifically, were distinct in being willing to engage in costly discrimination against black proposers by rejecting offers they would accept from whites. The results of Study 2 provide supportive evidence that this pattern originates in differences in fairness expectations among (overtly racist) whites

when blacks, rather than whites, propose resource divisions toward whites.²⁵ The unique penalty that overt racists apply to black proposers in the evaluation experiment (when the responder is white) suggests they expected black proposers to give up more than white proposers when whites had the power to reject unfair offers, yet they did not expect white proposers to reciprocate when blacks were in the same position of power.

Discussion and Conclusion

This research provides new evidence about the nature and correlates of white discrimination against black Americans. In Study 1, we find that white Americans on average discriminate against black proposers in the Ultimatum Game. This study is, to our knowledge, the first to link this behavior to explicit attitudes about group-level black inferiority and symbolic racism, measures that can readily be included on standard public opinion surveys and are widely used as predictors of racial policy preferences. We find that overt racists—those who endorse overt negative stereotypes of blacks—were about 2% more likely to discriminate against black proposers than non-overt racists compared to the 58% of white responders who did not endorse views of group-level black inferiority and did not appear to discriminate. By contrast, the symbolic racism measure was a poor predictor of discriminatory behavior. Responders who endorsed anti-black symbolic racism were no more likely to engage in discriminatory behavior than responders who did not. Overt racism, like the implicit racism measure using the IAT (Kubota et al. 2013), can therefore be used to identify white individuals who are willing to engage in costly discrimination against black individuals, but symbolic racism cannot.

Study 2 examined whether this discriminatory behavior might be partly motivated by different views about the fairness of racial interactions by measuring how white third party observers evaluated interactions between black and white players in the Ultimatum Game. We find that whites on average

²⁵ In general, those in the role of proposer in the UG should expect rejection of unfair offers, so there is a strong incentive for proposers to be just as fair to out-group responders as they would be to in-group responders. Indeed, there is a substantial reduction of proposer bias in the UG relative to the Dictator Game because in the latter context responders can penalize proposers by rejecting unfair offers (e.g. Stagnaro, Dunham, and Rand, 2017).

evaluate negotiations over resource distributions between racial groups as less fair than racially homogeneous interactions. Among overt racists, we find that racial bias in evaluations is strongest when black proposers make offers to white responders. Overt racists therefore expect black proposers to give up more of their share when interacting with white responders, a higher standard than they impose when whites propose resource allocations to blacks. These third-party evaluations are consistent with the individual decisions of overt racists in Study 1. By contrast, those who score high on symbolic racism do not appear to impose disproportionately higher fairness standards on black proposers (vis-à-vis white proposers) in interactions with white responders, which is consistent with the lack of racial discrimination among this group of white research subjects in Study 1.

These results have several important implications. First, per Study 1, they document a widespread willingness by whites to discriminate, in a costly manner, against blacks. Whites demanded that black proposers give up more of their share than they expected of other whites, and systematically acted against their narrow self-interest in order to discriminate. Despite hopes for social norms of racial equality and equal treatment in the United States, at least some whites expect more from black individuals than they do from other white individuals, and are even willing to take personally costly decisions that punish black individuals when this expectation is violated.

Second, theoretically, this work documents fairness as a potential mechanism for the discriminatory behavior observed in Study 1. In Study 2, we provide direct evidence that, on average, whites expected blacks to offer more to whites than they would expect other whites to offer in order for a resource division to be perceived as equally fair. Modern day white racial prejudice and discrimination in the United States may be rooted in whites' differential expectations about the fairness of black-white versus white-white resource divisions. These evaluators had no financial stake in the outcome of the intergroup interactions, but the results suggest it would be worthy in future research to study whether whites are willing to engage in third party punishment of black individuals who violate expectations when interacting with other whites.

Third, we provide a means to identify using survey data those who engage in this costly discriminatory behavior. We show that overt racism, but not symbolic racism, predicts racial discrimination. This validates the usefulness of the overt racism measure as a predictor of whites' willingness to engage in costly discriminate against blacks. More generally, however, the fact that symbolic racism does not predict discriminatory behavior (or racially biased fairness evaluations) raises important questions about the theoretical constructs tapped by the measure. If symbolic racism were a measure of prejudice leading to a willingness to discriminate (or impose different fairness expectations on blacks when they interacted with whites over resource allocations), it would be natural to understand the correlation between symbolic racism and racially relevant policy views as sharing a discriminatory impetus with the association between overt racism and the history of black political exclusion and discriminatory policies. But it does not, implying that our understanding of the meaning of the symbolic racism scale cannot be that it proxies a willingness to engage in racial discrimination. Whatever symbolic racism measures, and despite its predictive power, it is theoretically distinct from overt racism.

Stepping back, researchers have shown that symbolic racism is strongly correlated with voters' racially-relevant policy preferences and voting behavior. Recent studies, for example, suggest that symbolic racism is correlated with support for Barack Obama and policies associated with Obama, such as the Affordable Care Act (e.g. Tesler and Sears 2010; Tesler 2012). The results presented here, coupled with the measure's strong association with conservatism, partisanship, and voting behavior, however, suggest extreme caution about taking these correlations as evidence that racial prejudice leading to anti-black discrimination explains the predictive power of the symbolic racism measure. In a direct test, symbolic racism does not explain which whites engage in costly discrimination toward blacks, while overt measures of prejudice do.

Having demonstrated the explanatory power of an alternative survey measure that has been used much less in contemporary political science research, we close by considering whether this measure may also be subject to some of the same shortcomings and criticisms raised against the symbolic racism measure. Symbolic racism is strongly associated with political conservatism and this makes it difficult to

distinguish between racial prejudice and conservative social value orientations (for example, see Sniderman and Tetlock 1986; Sniderman and Piazza 1995; Schuman 2000). Although it has been suggested that overt racism provides an unambiguous measure of racial prejudice (Huddy and Feldman 2009), we know less about the relationship between overt racism and political ideology. If overt racism is also deeply tied to conservative political orientations then it may also be an unsatisfactory measure and isolating its effects vis-à-vis factors like partisanship and ideology is likely to be difficult in the absence of direct random manipulation of overt prejudice. (Just as it is hard to identify the effect of symbolic racism given its tight correlation with many other important political values and orientations.²⁶)

It is therefore interesting to note that pooling the data across both studies the correlation between conservative ideology (a 7 point scale running from “Very Liberal” to “Very Conservative”) and the symbolic racism scale is much larger ($\rho = 0.56$) than the correlation between ideology and overt racism ($\rho = 0.22$). Similarly, partisanship (measured on a 7-point scale from “Strong Democrat” to “Strong Republican” that includes leaners but excludes pure independents) is more strongly correlated with symbolic racism ($\rho = 0.53$) than overt racism ($\rho = 0.20$), as is 2012 Republican voting ($\rho = 0.49$ for symbolic versus 0.19 for overt racism). Altogether, these patterns show that overt prejudice is not simply a proxy for political conservatism or Republican Party orientation.

In conclusion, we have shown that symbolic racism and overt racism are theoretically distinct because the latter predicts white racial discrimination while the former does not (Study 1). This behavior is consistent with observed differences in whites’ third party fairness evaluations of resource allocations by black rather than white proposers toward whites in the Ultimatum Game (Study 2). Our work therefore provides an empirical underpinning for ongoing discussions about the relationship between symbolic racism and racial prejudice and discrimination. We show that both symbolic and overt racism remain prominent among members of the white American public, and while the former is associated with many important political views, it is analytically distinct from overt racism because it does not predict a

²⁶ Furthermore, the possibility of measurement error means that “controlling for” partisanship and ideology when assessing the effects of another survey item may still produce biased estimates of the other factor included in the predictive model.

willingness to discriminate against blacks vis-à-vis whites when making costly decisions or providing judgements about the fairness of unequal resource allocations between blacks and whites. In light of these findings, it is our view that more research should measure the prevalence and political importance of overt racism as defined here – the explicit endorsement of group-level white superiority. Additionally, scholars who use or critique the symbolic racism measures now have a solid evidentiary basis for understanding that symbolic racism is not based in a personal willingness to discriminate against blacks, despite its predictive power for many contemporary attitudes about public policies and political choices.

References

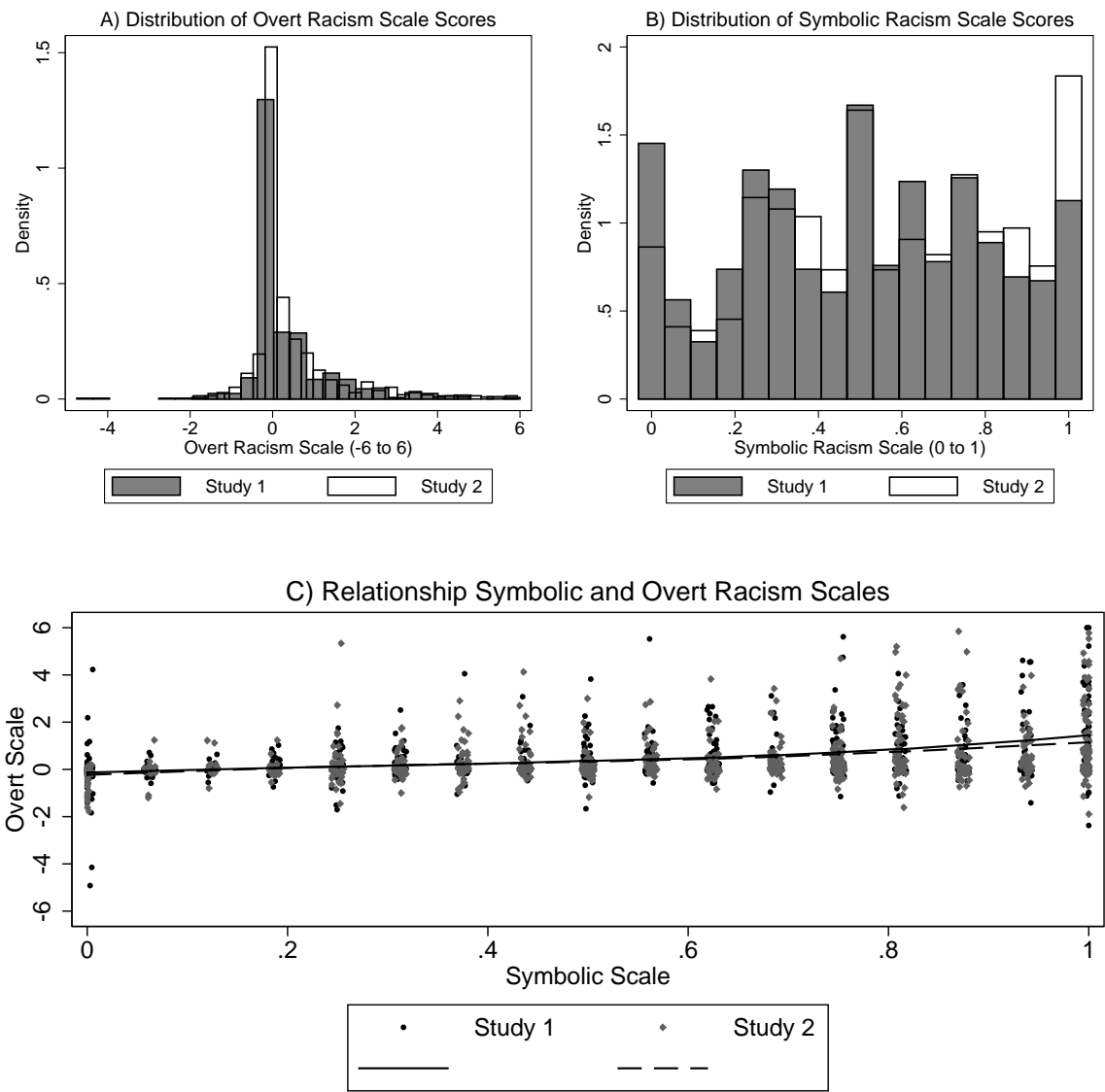
- Balliet, Daniel, Junhui Wu, and Carsten KW De Dreu. 2014. "Ingroup Favoritism in Cooperation: A Meta-Analysis." *Psychological Bulletin* 140(6):1556-1581.
- Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20: 351–68.
- Brewer, Marilynn B. 1979. "In-group bias in the minimal intergroup situation: A cognitive-motivational analysis." *Psychological bulletin* 86.2: 307.
- Brief, Arthur P., Joerg Dietz, Robin Reizenstein Cohen, S. Douglas Pugh, and Joel B. Vaslow. "Just doing business: Modern racism and obedience to authority as explanations for employment discrimination." *Organizational behavior and human decision processes* 81, no. 1 (2000): 72-97.
- Camerer, Colin F, and Ernst Fehr. 2004. "Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists." *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* 97. Oxford: Oxford University Press: 55–95.
- Carmines, Edward G., Paul M. Sniderman, and Beth C. Easter. "On the meaning, measurement, and implications of racial resentment." *The Annals of the American Academy of Political and Social Science* 634.1 (2011): 98-116.
- Dovidio, John F., Kerry Kawakami, and Samuel L. Gaertner. "Implicit and explicit prejudice and interracial interaction." *Journal of personality and social psychology* 82.1 (2002): 62.
- Everett, Jim, Zachary Ingbretsen, Fiery Cushman, and Mina Cikara. 2017. "Deliberation erodes cooperative behaviour—even towards competitive outgroups, even when using a control condition, and even when controlling for sample bias." *Journal of Experimental Social Psychology* (73): 76-81.
- Feldman, Stanley, and Leonie Huddy. 2005. "Racial Resentment and White Opposition to Race-Conscious Programs: Principles or Prejudice?" *American Journal of Political Science* 49 (1). Wiley Online Library: 168–83.
- Fiske, Susan T, and Shelley E Taylor. 1991. "Social Cognition, 2nd." NY: McGraw-Hill, 16–15.
- Gilens, Martin. 1999. *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*. University of Chicago Press.
- Güth, Werner, and Reinhard Tietz. 1990. "Ultimatum Bargaining Behavior: A Survey and Comparison of Experimental Results." *Journal of Economic Psychology* 11 (3). Elsevier: 417–49.
- Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior & Organization* 3 (4). Elsevier: 367–88.
- Harding, J, H Proshansky, B Kutner, and I Chein. 1969. "Prejudice and Ethnic Relations. in G. Lindzey & E. Aronson (Eds.), the Handbook of Social Psychology (Vol. 5."

- Hauser, David J, and Norbert Schwarz. 2016. "Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks Than Do Subject Pool Participants." *Behavior Research Methods* 48 (1). Springer: 400–407.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis. 2004. "Foundations of Human Sociality-Economic Experiments and Ethnographic: Evidence from Fifteen Small-Scale Societies."
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *The American Economic Review* 91 (2). JSTOR: 73–78.
- Henry, P. J. (2010). Modern racism. In J. Levine & M. Hogg (Eds.), *The encyclopedia of group processes and intergroup relations, Vol 2* (pp. 575-577). Thousand Oaks, CA: Sage Publications.
- Henry, P. J., & Sears, David O. (2002). "The symbolic racism 2000 scale." *Political Psychology*, 23, 253-283.
- Huddy, Leonie, and Stanley Feldman. 2009. "On Assessing the Political Effects of Racial Prejudice." *Annual Review of Political Science* 12. Annual Reviews: 423–47.
- Hutchings, Vincent L, and Nicholas A Valentino. 2004. "The Centrality of Race in American Politics." *Annu. Rev. Polit. Sci.* 7. Annual Reviews: 383–408.
- Hyman, Herbert H, and Paul B Sheatsley. 1956. "Attitudes Toward Desegregation." *Scientific American*. Scientific American, Inc.
- Hyman, Herbert H, and Paul B Sheatsley. 1964. "Attitudes Toward Desegregation." *Scientific American* 211: 16–23.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Macmillan.
- Kinder, Donald R, and David O Sears. 1981. "Prejudice and Politics: Symbolic Racism Versus Racial Threats to the Good Life." *Journal of Personality and Social Psychology* 40 (3). American Psychological Association: 414.
- Kinder, Donald R, and Lynn M Sanders. 1996. *Divided by Color: Racial Politics and Democratic Ideals*. University of Chicago Press.
- Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. "Social Desirability Bias in Cati, Ivr, and Web Surveys the Effects of Mode and Question Sensitivity." *Public Opinion Quarterly* 72 (5). AAPOR: 847–65.
- Kubota, Jennifer T, Jian Li, Eyal Bar-David, Mahzarin R Banaji, and Elizabeth A Phelps. 2013. "The Price of Racial Bias Intergroup Negotiations in the Ultimatum Game." *Psychological Science* 24 (12). Sage Publications: 2498–2504.
- Lane, Tom. 2016. "Discrimination in the Laboratory: A Meta-Analysis of Economics Experiments." *European Economic Review* 90. Elsevier: 375–402.
- LaPiere, Richard. "Attitudes and actions,." *Social Forces* 13 (1934): 230-237.

- Massey, Douglas S, and Nancy A Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press.
- McConahay, John B, and Joseph C Hough. 1976. "Symbolic Racism." *Journal of Social Issues* 32 (2). Wiley Online Library: 23–45.
- McConahay, John B. 1983. "Modern racism and modern discrimination: The effects of race, racial attitudes, and context on simulated hiring decisions." *Personality and Social Psychology Bulletin* 9.4: 551-558.
- Mullen, Brian, Rupert Brown, and Colleen Smith. 1992. "Ingroup Bias as a Function of Salience, Relevance, and Status: An Integration." *European Journal of Social Psychology* 22 (2). Wiley Online Library: 103–22.
- Pasek, Josh, Alexander Tahk, Yphtach Lelkes, Jon A Krosnick, B Keith Payne, Omair Akhtar, and Trevor Thompson. 2009. "Determinants of Turnout and Candidate Choice in the 2008 Us Presidential Election Illuminating the Impact of Racial Prejudice and Other Considerations." *Public Opinion Quarterly* 73 (5). AAPOR: 943–94.
- Payne, B Keith, Jon A Krosnick, Josh Pasek, Yphtach Lelkes, Omair Akhtar, and Trevor Thompson. 2010. "Implicit and Explicit Prejudice in the 2008 American Presidential Election." *Journal of Experimental Social Psychology* 46 (2). Elsevier: 367–74.
- Rabinowitz, J. L., Sears, D. O., Sidanius, J., & Krosnick, J. A. 2009." Why Do White Americans Oppose Race-Targeted Policies? Clarifying the Impact of Symbolic Racism." *Political Psychology*, 30(5), 805–828.
- Schuman, Howard, Charlotte Steeh, Lawrence Bobo, and Maria Krysan. 1997. "Racial Attitudes in America: Trends and Interpretations, Rev." Harvard University Press.
- Schuman, Howard. 2000. "The Perils of Correlation, the Lure of Labels, and the Beauty of Negative Results." *Racialized Politics: The Debate About Racism in America*. University of Chicago Press, 302.
- Sears, David O, and Harris M Allen Jr. 1984. "The Trajectory of Local Desegregation Controversies and Whites' Opposition to Busing." *Groups in Contact: The Psychology of Desegregation*. Academic Press Orlando, FL, 123–51.
- Sears, David O, and Patrick J Henry. 2003. "The Origins of Symbolic Racism." *Journal of Personality and Social Psychology* 85 (2). American Psychological Association: 259.
- Sears, David O. 1988. "Symbolic Racism." In *Eliminating Racism*, 53–84. Springer.
- Sears, David O., and Donald R Kinder. 1971. *Racial Tension and Voting in Los Angeles*. Vol. 156. Institute of Government; Public Affairs, University of California.
- Sears, David O., Carl P. Hensler, and Leslie K. Speer. 1979. "Whites' opposition to "busing": Self-interest or symbolic politics?." *American Political Science Review* 73.2: 369-384.
- Smith, Rogers M. 1993. "Beyond Tocqueville, Myrdal, and Hartz: The Multiple Traditions in America." *American Political Science Review* 87 (03). Cambridge Univ Press: 549–66.

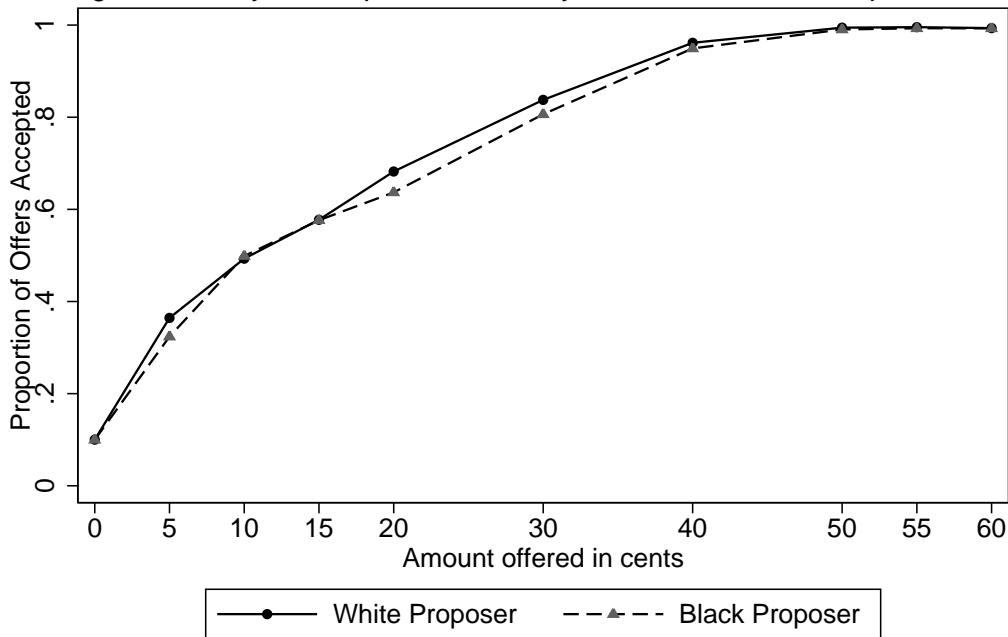
- Sniderman, Paul M, and Philip E Tetlock. 1986. "Symbolic Racism: Problems of Motive Attribution in Political Analysis." *Journal of Social Issues* 42 (2). Wiley Online Library: 129–50.
- Sniderman, Paul M, and Thomas Leonard Piazza. 1995. *The Scar of Race*. Harvard University Press.
- Stagnaro, Michael N., Yarrow Dunham, and David G. Rand. "Profit versus prejudice: harnessing self-interest to reduce in-group bias." *Social Psychological and Personality Science* (2017): 1948550617699254.
- Tajfel, H., Billig, M.G., Bundy, R.P. and Flament, C., 1971. "Social categorization and intergroup behaviour." *European Journal of Social Psychology*, 1(2), pp.149-178.
- Tesler, Michael, and David O Sears. 2010. *Obama's Race: The 2008 Election and the Dream of a Post-Racial America*. University of Chicago Press.
- Tesler, Michael. 2012. "The Spillover of Racialization into Health Care: How President Obama Polarized Public Opinion by Racial Attitudes and Race." *American Journal of Political Science* 56 (3). Wiley Online Library: 690–704.
- Valenzuela, A. and Srivastava, J., 2012. "Role of information asymmetry and situational salience in reducing intergroup bias: The case of ultimatum games." *Personality and Social Psychology Bulletin*, 38(12), pp.1671-1683.

Figure 1: Survey measures of racial attitudes by study



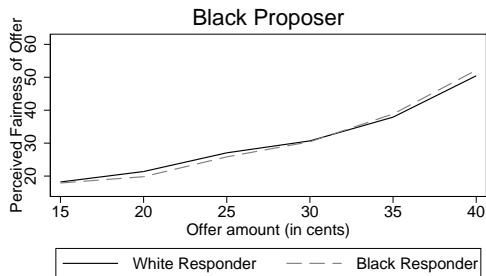
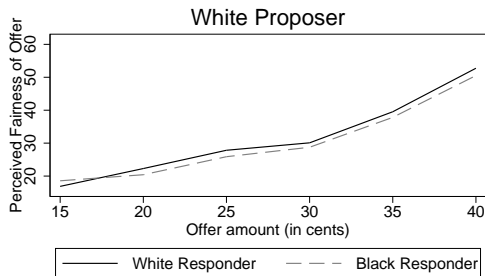
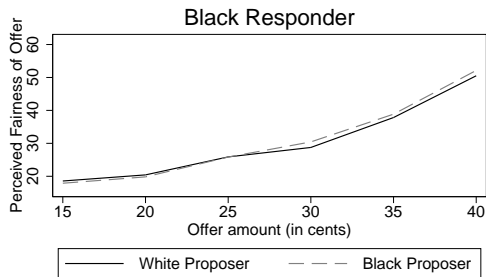
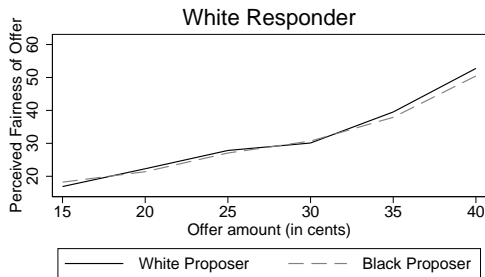
Note: Lines in panel C are lowess smoothers.

Figure 2: Study 1 Acceptance Rates by Offer Amount and Proposer Race



Total N=18233

Figure 3: Study 2 Fairness Evaluations
by race of proposer and responder



See text for details.

Data are partitioned by responder race in top two panels and by proposer race in bottom two panels

Table 1: Study 1 Acceptance of Ultimatum Game Proposals by White Respondents

	(1)	(2)	(3)	(4)	(5)
	Offer accepted (1=Yes) (0=No)				
Amount offered in cents					
\$0.05	0.248 [0.019]***	0.248 [0.019]***	0.248 [0.019]***	0.249 [0.019]***	0.249 [0.019]***
\$0.10	0.396 [0.020]***	0.396 [0.020]***	0.395 [0.020]***	0.394 [0.020]***	0.395 [0.020]***
\$0.15	0.476 [0.021]***	0.476 [0.021]***	0.476 [0.021]***	0.476 [0.021]***	0.476 [0.021]***
\$0.20	0.564 [0.020]***	0.564 [0.020]***	0.564 [0.020]***	0.562 [0.020]***	0.562 [0.020]***
\$0.30	0.725 [0.018]***	0.725 [0.018]***	0.725 [0.018]***	0.725 [0.018]***	0.724 [0.018]***
\$0.40	0.857 [0.013]***	0.857 [0.013]***	0.857 [0.013]***	0.856 [0.013]***	0.855 [0.013]***
\$0.50	0.893 [0.012]***	0.893 [0.012]***	0.893 [0.012]***	0.894 [0.012]***	0.894 [0.012]***
\$0.55	0.894 [0.012]***	0.894 [0.012]***	0.894 [0.012]***	0.894 [0.012]***	0.893 [0.012]***
\$0.60	0.893 [0.012]***	0.893 [0.012]***	0.893 [0.012]***	0.892 [0.012]***	0.892 [0.012]***
Proposer Black (1=yes)	-0.013 [0.004]***	-0.012 [0.006]*	-0.014 [0.006]**	-0.005 [0.006]	-0.008 [0.007]
Fast condition (4 seconds)		0.004 [0.016]			
Proposer Black x Fast condition		-0.003 [0.009]			
Symbolic Racist (1=yes)			-0.010 [0.016]		0.005 [0.016]
Proposer Black x Symbolic Racist			0.003 [0.009]		0.009 [0.009]
Overt Racist (1=yes)				-0.049 [0.016]***	-0.051 [0.017]***
Proposer Black x Overt Racist				-0.020 [0.009]**	-0.023 [0.010]**
Constant	0.132 [0.016]***	0.130 [0.018]***	0.137 [0.018]***	0.153 [0.018]***	0.151 [0.019]***
Observations	18233	18233	18233	18233	18233
R-squared	0.414	0.414	0.414	0.418	0.418

OLS coefficient with robust standard errors in brackets, clustered at the respondent level.

* significant at 10%; ** significant at 5%; *** significant at 1%.

Note: Coefficients (indicators) for rounds of play not reported to save space.

Table 2: Study 2 Third Party Fairness Evaluations of UG Proposals

	(1)	(2)	(3)	(4)
	Perceived Offer Fairness (0-100)			
Treatment: Black Proposer, White Respondent	-0.649 [0.238]***	-0.302 [0.325]	0.251 [0.334]	0.358 [0.365]
Treatment: Black Proposer, Black Respondent	-0.546 [0.242]**	-0.304 [0.346]	-0.030 [0.341]	0.063 [0.380]
Treatment: White Proposer, Black Respondent	-1.151 [0.255]***	-0.898 [0.345]***	-1.068 [0.365]***	-0.894 [0.395]**
Symbolic Racist x White Proposer, White Respondent		1.956 [1.020]*		1.172 [1.055]
Symbolic Racist x Black Proposer, White Respondent		1.284 [1.038]		0.871 [1.067]
Symbolic Racist x Black Proposer, Black Respondent		1.484 [1.023]		0.906 [1.054]
Symbolic Racist x White Proposer, Black Respondent		1.464 [1.007]		0.684 [1.037]
Overt Racist x White Proposer, White Respondent			3.350 [1.001]***	3.037 [1.041]***
Overt Racist x Black Proposer, White Respondent			1.833 [1.017]*	1.601 [1.047]
Overt Racist x Black Proposer, Black Respondent			2.480 [1.010]**	2.239 [1.043]**
Overt Racist x White Proposer, Black Respondent			3.211 [0.987]***	3.029 [1.022]***
Constant	11.858 [0.966]***	10.823 [1.052]***	9.875 [1.106]***	9.436 [1.138]***
Observations	26651	26651	26651	26651
R-squared	0.559	0.560	0.562	0.562

OLS coefficient with robust standard errors in brackets, clustered at the respondent level.

* significant at 10%; ** significant at 5%; *** significant at 1%.

Note: Coefficients (indicators) for offer amounts and rounds of play not reported to save space.

See Table 3 for relevant linear combination of coefficient tests.

Table 3: Study 2 Differences in fairness evaluations by
race of proposer, race of responder, and racial attitudes measures

Panel A: Difference in Fairness of offers to White Responder by White rather than Black proposers
(Positive values indicate proposals by whites to white are evaluated more favorably than proposals by blacks to whites)

	Estimate	p-value of difference	Difference in Differences compared to baseline (non-overt non- symbolic racists)	p-value of difference in differences
Baseline (Non-overt non-symbolic racists)	-0.36	0.33	N/A	N/A
Among symbolic racists (but not overt racist)	-0.06	0.91	0.30	0.55
Among overt racists (but not symbolic racist)	1.08	0.02	1.44	0.00

Panel B: Difference in Fairness of offers to Black Responder by Black rather than White proposers
(Positive values indicate proposals by blacks to blacks are evaluated more favorably than proposals by whites to blacks)

	Estimate	p-value of difference	Difference in Differences compared to baseline (non-overt non- symbolic racists)	p-value of difference in differences
Baseline (Non-overt Non-symbolic racists)	0.96	0.03	N/A	N/A
Among symbolic racists (but not overt racist)	1.18	0.03	0.22	0.69
Among overt racists (but not symbolic racist)	0.17	0.74	-0.79	0.16

Note: Results are from linear combination of coefficient tests for regression model reported in column (3) of Table 2.