Who Benefits? Experimental Evidence of Gender Differences in Evaluations of the Deserving Poor*

June 8, 2023

Abstract

This study examines how Americans evaluate the deservingness of men and women applying for government aid using a survey experiment. I find that, on average, male applicants earn less than female applicants with identical needs. Further, I find the amount awarded to women varies conditional on being rated a "Poor" or "Excellent" worker, while there is no difference in the amount awarded to men who are rated "Poor" workers compared to men who are rated "Excellent." Thus, women's advantage over men applicants extends only so far as women are perceived as deserving and hard workers. The totality of these results suggest that people are more inclined to help poor women rather than poor men, but only when the quality of women applicants is validated by an external source. I bolster these findings by providing evidence that neither respondent-level ideology nor gender drives these results.

*The experiment reported in this paper is preregistered at EGAP Registry (redacted). This research was conducted while the author was a graduate student at (redacted). This research was reviewed and approved by the (redacted) Institutional Review Board. All errors remain the author's own responsibility.

The question of "who should get what, when, and how?" is central to the study of politics (Lasswell 1936). Understanding the answer in a democratic society requires a robust understanding of the public's answers to these questions and how their values shape their responses. Strategic politicians will not advance redistributive policies for which they believe the public will punish them (Fearon 1999). Beyond the electoral connection, street level bureaucrats are swayed by public opinion. Across many policy areas, including Medicaid (Weissert 1994), labor (Schmidt 2002), disability insurance (Keiser 1999), food stamps (Kogan 2021), and immigration (Lewis, Provine, Varsanyi and Decker 2013), street-level bureaucrats use their discretion to align realized policy outcomes with personal and community values, even when their formal responsibility is to implement otherwise identical statutory language. Thus, it is critical that scholars understand how the American public thinks about government aid and its recipients, as scholarship has shown that attitudes about public policies are closely related to attitudes about the beneficiaries of those policies (Nelson and Kinder 1996; Rabinowitz, Sears, Sidanius and Krosnick 2009; Fossati 2018).

The dimension that has characterized past work and public debate on welfare is deservingness (Schneider and Ingram 1993; Ingram and Schneider 1993; Schneider and Ingram 2005; van Oorschot, Roosma, Meuleman and Reeskens 2017; Gilens 2000; Petersen 2012; Petersen, Sznycer, Cosmides and Tooby 2012; Aarøe and Petersen 2014). The so-called "deserving poor" are those whose personal financial circumstances have been devastated by structural or macroeconomic forces beyond the individual's control, rather than personal qualities of dependence and laziness. For decades, Americans have cared a great deal that aid go only to the most deserving, especially those they feel have earned federal assistance or are particularly helpfuless (Bobocel, Son Hing, Davey, Stanley and Zanna 1998; Katz and Hass 1988; Sniderman and Tetlock 1986; Sniderman, Carmines, Layman and Carter 1996; McClosky and Zaller 1984). But who do Americans think most deserves aid from the government? Past work has shown that these deservingness criteria are channels through which Ameri-

cans express racially biased attitudes (Gilens 2000; Cassese and Barnes 2019; Desante 2013; Gilliam 1999; Gilens 2000; Cassese and Barnes 2019) and programs seen as disproportionately servicing black Americans are viewed significantly less favorably than those servicing white Americans, contributing to negative racial stereotypes such as the "welfare queen."

Yet remarkably little is known about how gender attitudes influence and affect attitudes toward aid recipients. Indeed, much work on gendered attitudes toward welfare recipients considers only attitudes toward men (Petersen et al. 2012; Aarøe and Petersen 2014; Willrich 2000) or only attitudes toward women (Monnat 2010*a*; Desante 2013; Hayes and Mitchell 2020; Cassese and Barnes 2019), diminishing scholars' ability to make clear, causally identified statements about the role of gender in welfare attitudes.

In this paper, I address these critical but heretofore unanswered questions by leveraging an original survey experiment to causally identify the effect of gender on perceived deservingness. My analyses show that, when men and women are put in direct competition for scarce public resources, gender-based prejudice interacts with American perceptions of work ethic to amplify existing stereotypes about men and women. I find that, on average, male applicants earn less than female applicants with identical needs. Further, I find the amount awarded to women varies conditional on being rated a "Poor" or "Excellent" worker by a third party, while there is no difference in the amount awarded to men who are rated "Poor" workers compared to men who are rated "Excellent" workers. Thus, women's advantage over men applicants extends only so far as women are perceived as "deserving." The totality of these results suggest that people are more inclined to help poor women rather than poor men, but only deserving women, when the quality of women applicants is validated by an external source. This research provides evidence toward benevolent sexism, in which women are viewed as weak, in need of extra care, and are rewarded for conforming with traditional, feminine stereotypes (Glick and Fiske 1997, 2001).

The rest of this paper proceeds as follows. Section I outlines my theory and preregistered

hypotheses. Section II details the experimental design. Section III presents the main results. Section IV presents exploratory analyses to confirm that the mechanism driving my results is gender differences in the assigned treatment, rather than respondent covariates. Section V concludes.

Theory

Scholars have found that the perceived beneficiaries of public policies can have profound affect on their support for the policy. For example, Mettler (2011) finds that Americans' support for popular tax policies depends on which Americans are perceived to benefit most from the policy. When Americans learn, for example, that the mortgage interest deduction provides the greatest benefit to more affluent Americans, support for that policy drops significantly. These effects also persist in work on welfare support, which has revealed widespread bias induced by the target beneficiaries of welfare support (Gilens 2000; Cassese and Barnes 2019; Winter 2006; Desante 2013; Hayes and Mitchell 2020).

This work follows from a tradition of scholarship in political science suggesting "group interest" thinking drives Americans' political attitudes (Converse 1964). That is, Americans tend to think in general terms about how policies affect recognizable groups, and then form their support for such policies based on their affect toward the group and the intended effect of the policy on the group (Winter 2008; Theiss-Morse 2009; Huddy 2004; Iyengar, Sood and Lelkes 2012). Accordingly, if the target population associated with a policy shifts, individual Americans will shift their evaluations. For example, Hayes, Fortunato and Hibbing (2021) find that white Americans favor much more firearm availability when they perceive the target population of such policies to be white women than when they believe the target population is black men.

In the context of federal aid, Americans evaluate target populations on the basis of their perceived deservingness. Deservingness tends to operate along five key dimensions: control, need, identity, attitude, and reciprocity (van Oorschot 2000). Table 1 provides an overview of these components of deservingness. Different combinations of these criteria often underlie the constructed target populations for federal aid.

Table 1: Dimensions of Deservingness from van Oorschot (2000)

Dimension	**Definition**
Control	poor people's control over their neediness, or their responsibility for it
Need	the level of need
Identity	the identity of the poor, their proximity to the rich or their 'pleasantness'
Attitude	poor people's attitude towards support, or their docility or gratefulness
Reciprocity	the degree of reciprocation by the poor, or having earned support

Target populations are a concept from the policy design literature that draw attention to the fact that policy is purposeful and attempts to achieve its goals by altering the behavior of specific groups. The social construction of target populations refers to the valence-oriented, cultural characterizations of people and groups whose behavior are affected by public policy (Schneider and Ingram 1993, 2005). They tell us who *should* benefit from public policy. These constructs may emerge organically or be developed strategically to influence public opinion of welfare programs or recipients (Ingram and Schneider 1993).

One prominent example of strategic construction of a target population is Ronald Reagan's 1976 presidential campaign, which repeatedly referenced Linda Taylor, a black woman who committed welfare fraud, in order to characterize undeserving welfare recipients (Gilman 2013). This, and the publicity surrounding the Taylor affair, led to the construction of the "welfare queen" stereotype of black women defrauding the government by benefiting from welfare while secretly living in luxury. This stereotype is decidedly racialized. Consequently, it is challenging to develop priors on how we should expect it to inform public opinion on women, in general, receiving aid. Indeed, scholarship has consistently found that black

women seeking government assistance are evaluated less favorably than otherwise identical white women (Gilliam 1999; Desante 2013; Hayes and Mitchell 2020; Cassese and Barnes 2019; Gilens 2000).

While women receive different treatment than comparable men in everything from applying to jobs (Neumark, Bank and Van Nort 1996; Goldin and Rouse 2000; Quadlin 2018*a*) to running for office (Hassell and Visalvanich 2019; Clayton, Robinson, Johnson and Muriaas 2020) to the price they pay for basic household goods in some countries (Betz, Fortunato and O'brien 2021) to their salaries (Castillo, Petrie, Torero and Vesterlund 2013; Mandel 2013), scholarly intuition is unclear about how gender differences in these domains translates into perceptions of deservingness. The racialization of welfare attitudes makes it particularly challenging to evaluate the effect of gender attitudes on evaluations of deservingness (Winter 2006), because racial and gender attitudes can be difficult to disentangle and may not be additively separable (Hayes, Fortunato and Hibbing 2021). The prominence and persistence of the welfare queen stereotype suggests that we should perhaps expect women to be evaluated as less deserving, but the racialization of this stereotype suggests that we should not expect that it would apply to non-black women.

A limited body of empirical work considers how gender might affect perceptions of deservingness. Women, especially mothers, seeking aid are more likely to be seen as victimized, needy, pleasant, and docile (Glick and Fiske 1997, 2001; Schneider and Ingram 1993; Monnat 2010*a*,*b*; Cassese and Barnes 2019), whereas men seeking aid have historically been perceived as failed breadwinners and "home slackers" (Willrich 2000), in control of their poverty and having not earned aid. Compounding this, Americans have long believed people "ought to take care of their personal problems by themselves" without relying on the government for aid (Sniderman and Brody 1977), yet these values tend to skew toward valuing men's hard work and devaluing men who cannot or do not work (Willrich 2000). Indeed, modern social welfare is often considered a "women's issue" and welfare policies are favored by women in

politics (Krook and O'Brien 2012; Greene and O'Brien 2016). The combined weight of these attitudes may contribute to public hostility for men receiving public benefits: such policies help those who should be helping themselves (Bobocel et al. 1998; Katz and Hass 1988; Sniderman and Tetlock 1986; Sniderman et al. 1996; McClosky and Zaller 1984).

Given this, I hypothesize:[1]

Hypothesis 1 *On average, male applicants will be awarded less than female applicants.*

A key dimension on the deservingness literature is applicant quality (Petersen et al. 2012; Petersen 2012). Applicants perceived as lazy or in control of their poverty are considered less deserving of aid than those who are hardworking and whose financial circumstances are out of their control (Aarøe and Petersen 2014). Because notions of deservingness and undeservingness are frequently characterized by ability and willingness to work, we might think that factors like individual competence will affect perceived deservingness. Specifically, conditional on gender, high competence workers should receive more than their low-competence counterparts.

Hypothesis 2 *On average, high-competence applicants will be awarded more than low-competence applicants.*

Deservingness might also be signaled by external factors, such as a college degree, a robust employment record, or participation in a certificate or job training program. If individuals truly prioritize giving aid to good workers whose personal financial circumstances were devastated by forces outside of their control and who they feel have "earned" assistance, then high-quality workers should be awarded more than low-quality workers. However, scholarship suggests that these kinds of quality evaluations are avenues through which discriminatory attitudes are expressed. For example, Gilens (2000) and Desante (2013) show

[1]All hypotheses and components of the research design were preregistered. My preregistration can be found at: https://osf.io/u5wxg/

that worker quality-based perceptions of deservingness are racialized (e.g. avenues through which respondents express anti-black bias).

The literature on quality evaluations and gender has focused primarily on gender gap reduction in pay, hiring, and promotions. While women in the workforce tend to be penalized for self advocacy (Quadlin 2018*a*; Exley, Niederle and Vesterlund 2020), they tend to benefit disproportionately from external, objective signals of their quality such as GPA (Quadlin 2018*b*), letters of recommendation (Abel, Burger and Piraino 2020), sub-baccalaureate certificates (Dadgar and Trimble 2015), and college degrees (Jepsen, Troske and Coomes 2014). I therefore hypothesize:

Hypothesis 3 *For male applicants, there will be no significant difference between amounts awarded to "Excellent" workers as compared to "Poor" workers.*

Hypothesis 4 *For female applicants, there will be a significant difference between amounts awarded to "Excellent" workers as compared to "Poor" workers.*

Experimental Design and Data

To test how gender, worker quality, and perceived work ethic shape attitudes toward welfare recipients and the amounts they are awarded, I conducted a survey experiment in which participants were asked to budget money to different pairs of applicants for state assistance subject to a budget constraint. I diverge from past work in which welfare attitudes are elicited in isolation or without budgetary constraints (Rabinowitz et al. 2009; Aarøe and Petersen 2014; Sniderman and Piazza 1993; Cassese and Barnes 2019). Asking participants to make a costly decision is methodologically preferable, because it mitigates concerns about acquiescence bias and is more analogous to real-world decisions in which achieving benefits in one domain may come with costs in another. For example, the 2020 ANES asks respondents

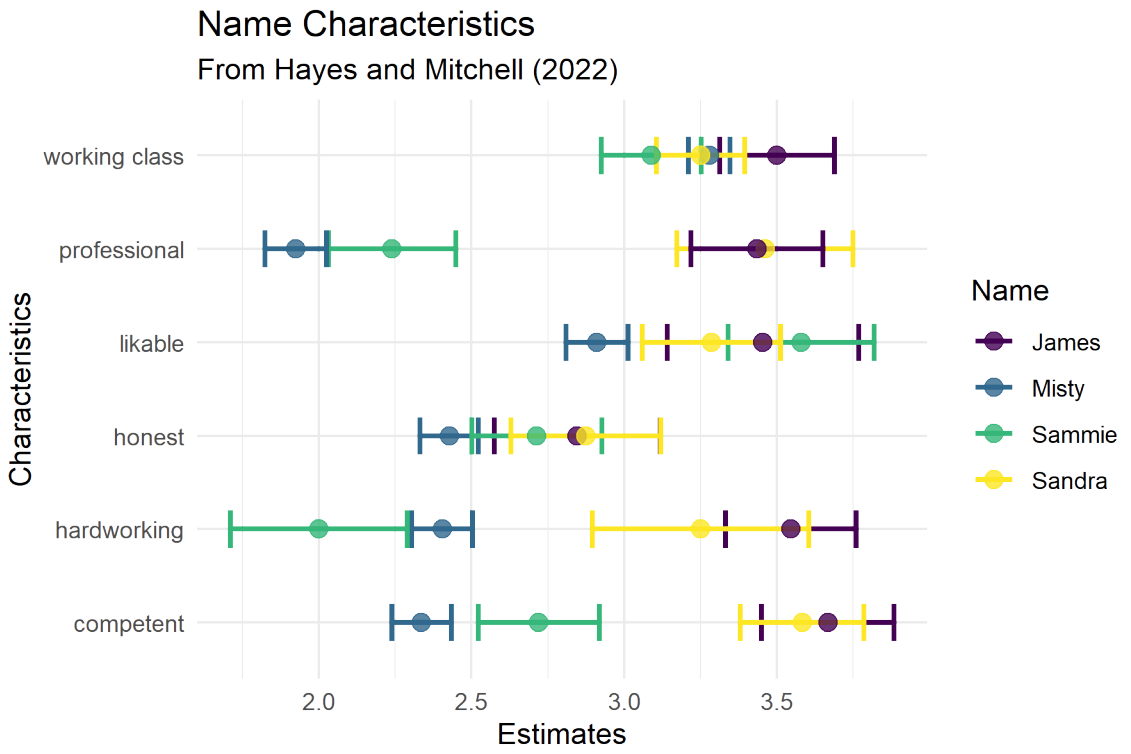
whether they think the government should increase spending on services like healthcare and education "even if it means an increase in spending." That is, given that the government cannot satisfy every need, which groups deserve to be taken care of first?

I use hand-redacted welfare applications to manipulate targets' for assistance sex, perceived competence, and objective work-quality rating (Figure 1). To manipulate quality, each applicant is randomly assigned a worker quality rating of "Excellent" or "Poor." To manipulate sex and perceived competence, I use two male and two female names from the Hayes and Mitchell (2020) name-characteristics dataset: Sandra, James, Misty, and Sammie.[2] These names were specifically selected to minimize the likelihood that factors other than sex, objective, and perceived work ethic would affect treatment. Names are used rather than more overt cues to minimize demand effects (Quidt, Vesterlund and Wilson 2019; Mummolo and Peterson 2019). To mitigate concerns about the effects of race and the racialization of welfare confounding results, all four names chosen were coded as racially distinct white names.

These names are matched on characteristics that Americans report as relevant considerations when evaluating deservingness (Bobocel et al. 1998; Katz and Hass 1988; Sniderman and Tetlock 1986; Sniderman et al. 1996; McClosky and Zaller 1984). Sandra and James are rated highly in professionalism (3.46, 3.43), competence (3.58, 3.66), and work ethic (3.25, 3.54), while Misty and Sammie are rated lower in all three characteristics (1.92, 2.24; 2.34, 2.72; 2.4, 2). Figure 1 below shows the complete breakdown of name-characteristics.

Subjects are given a budgeting task in which they are asked to allocate $1,500 to two applicants for federal assistance, each of whom has a state-determined need of $900. Respondents may also choose to give some or all of the funds to "offset the state deficit." Participants were given significant discretion in how aid was awarded, constrained only by a budgetary

[2]To create the dataset, the authors obtained a list of given names among people born in the United States between the years 1955 and 1990 from the U.S. Social Security Administration (SSA). To identify the gender of the names, they used information on sex included in the SSA data.

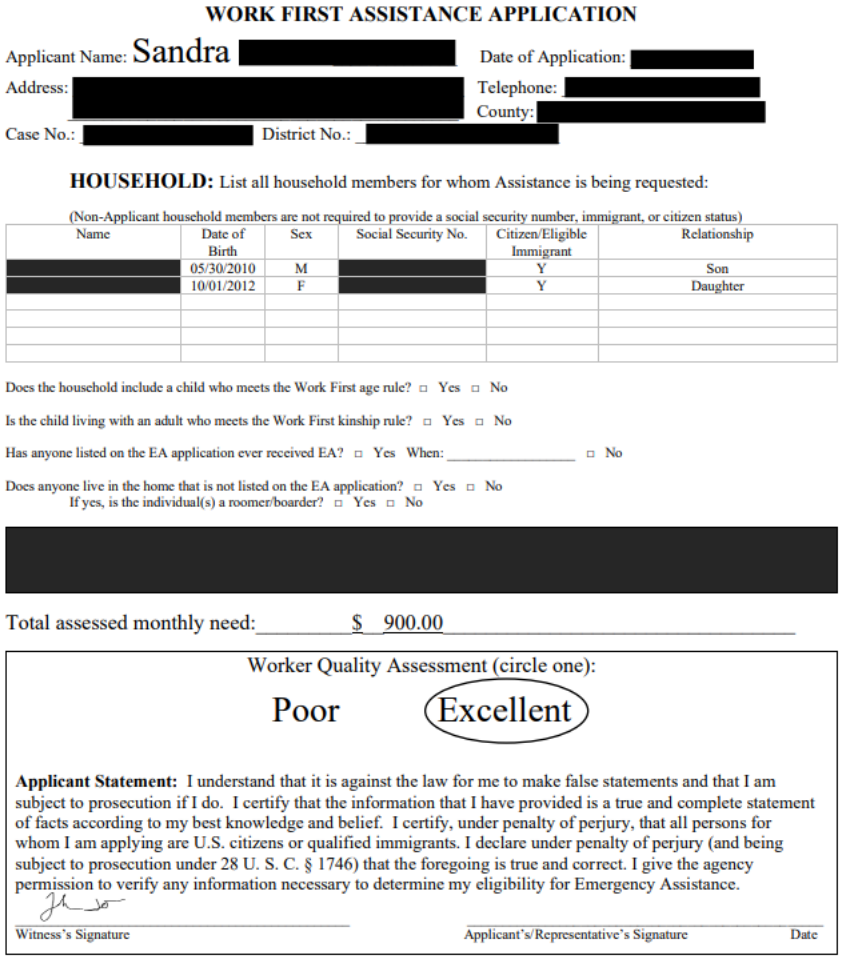


Takeaway: This figure shows the point estimates and confidence intervals of the estimated valence characteristics of the four chosen treatment names chosen from Hayes and Mitchell (2020). Higher point estimates correspond to higher levels of that value (e.g. a score of 5 in "professional" corresponds to greater perceived professionalism than a score of 4).

Figure 1: Differences in Valence Characteristics Between Treatment Names

limit and the inability to give an applicant more than their assessed need. Because both applicants' full needs cannot be met, I use the amount awarded to Applicant 1, Applicant 2, and the State as a direct estimate of an applicant's deservingness. These allocations are my main variables of interest. Everything else about the applicants remains identical, except for a worker quality assessment of Excellent or Poor and their name, which cues both sex (male or female) and competence (high or low). If an applicant receives a different allocation across treatments, we can infer that this difference is due to experimental manipulation. Further, as the applicant's characteristics were manipulated via random assignment, *any* difference in the relative importance respondents place on fiscal responsibility—illustrated by giving more to offset the budget deficit—can also be traced back to the experimental treatment.

Having the option to allocate dollars to a non-aid resource (in this case, the budget deficit) allows me to separate gender-neutral values from gendered discrimination. This deficit option also allows for individuals to take a principled position, a socially desirable and available option, to decide that the money would be better spent in some other way. While it may be the case that some respondents chose to allocate dollars to the state out of principled objection, whereas others believe that deficit reduction is simply a higher priority, both are a function of gender-neutral policy positions and should have no bearing on the *difference between* the amounts awarded to the baseline and treatment applicants.



WORK FIRST ASSISTANCE APPLICATION

Applicant Name: Sandra [redacted] Date of Application: [redacted]

Address: [redacted] Telephone: [redacted]

County: [redacted]

Case No.: [redacted] District No.: [redacted]

HOUSEHOLD: List all household members for whom Assistance is being requested:

(Non-Applicant household members are not required to provide a social security number, immigrant, or citizen status)

Name	Date of Birth	Sex	Social Security No.	Citizen/Eligible Immigrant	Relationship
[redacted]	05/30/2010	M	[redacted]	Y	Son
[redacted]	10/01/2012	F	[redacted]	Y	Daughter

Does the household include a child who meets the Work First age rule? □ Yes □ No

Is the child living with an adult who meets the Work First kinship rule? □ Yes □ No

Has anyone listed on the EA application ever received EA? □ Yes When: _____ □ No

Does anyone live in the home that is not listed on the EA application? □ Yes □ No
If yes, is the individual(s) a roomer/boarder? □ Yes □ No

[redacted]

Total assessed monthly need: $ 900.00

Worker Quality Assessment (circle one):

Poor Excellent

Applicant Statement: I understand that it is against the law for me to make false statements and that I am subject to prosecution if I do. I certify that the information that I have provided is a true and complete statement of facts according to my best knowledge and belief. I certify, under penalty of perjury, that all persons for whom I am applying are U.S. citizens or qualified immigrants. I declare under penalty of perjury (and being subject to prosecution under 28 U. S. C. § 1746) that the foregoing is true and correct. I give the agency permission to verify any information necessary to determine my eligibility for Emergency Assistance.

Witness's Signature Applicant's/Representative's Signature Date

Figure 2: Example Aid Application

In order to isolate the effects of sex versus the traits people ascribe to different names, I fielded a survey experiment with YouGov (n=1880) in April 2022. The sample is a mixture of a nationally representative sample with a low socio-economic status oversample. Due to this, in interpreting my results, I focus on effect direction, rather than magnitude, though I report point estimates in all cases (Horton, Rand and Zeckhauser 2011).[3]

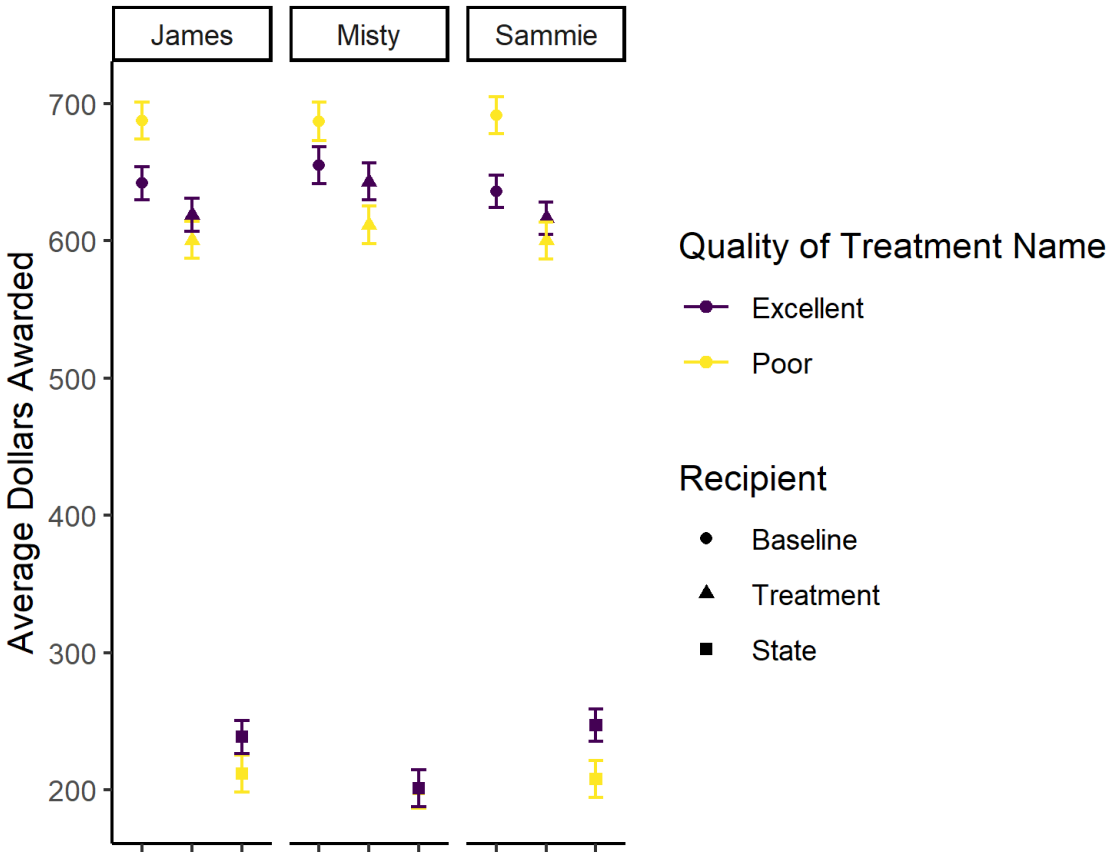
Respondents viewed two applications identical in appearance except for the realized treatment conditions. Rather than randomizing both applications, all respondents viewed the same baseline application of "Sandra" who was rated as "Excellent" compared to a second application. I used a 2x2x2 factorial design for this second application, randomizing sex (male/female), competence (high/low), and quality assessment (excellent/poor) of the second application, using the names James, Misty, and Sammie as my cue for sex and competence. Respondents were then asked to allocate funding to the two applicants or to offset the state budgetary deficit.

Experimental Results

When men and women are put in direct competition for scarce resources, how do women fare compared to similarly-situated men? How does perceived competence and external quality ratings affect this relationship? Figure 3 presents the main results of the experiment by treatment name, with color denoting the quality rating (Excellent/Poor) of the treatment name, and point-shape denoting the amount given to each recipient. Recall that the baseline condition is an applicant named Sandra, a high-competence name, who is rated as an "Excellent" quality worker.

H1 predicts that, on average, male applicants will be awarded less than female applicants. The first panel of Figure 3 demonstrates that, when she is paired with a high-competence,

[3]I present the same analysis conducted on the nationally representative and low-SES subsamples and find substantively similar results.



Takeaway: This figure shows the results for each treatment condition. The y-axis is the average dollars awarded to each possible recipient (the baseline, the treatment, or the state). The x-axis is each possible recipient (the baseline, the treatment, or the state). This allows us to compare the treatment to the baseline in each condition, as well as the treatment name to itself when the quality rating is manipulated. Error bars are 95% confidence intervals.

Figure 3: Gender Differences in Aid Allocated by Treatment Name

"Excellent" male name, James, Sandra is awarded significantly more on average ($642.06 vs. $619.20, p = 0.014). [4] However, there is also no significant difference between Excellent Sandra and Excellent Misty ($655.31 vs $643.20, p=0.141). These results are consistent with H1, in that Excellent Misty earns no less than Excellent Sandra, while Excellent James and Excellent Sammie earn substantively and significantly less than the baseline.

Turning to competence, I consider the difference between name valence characteristics

[4]Excellent Sandra also earns more on average than Sammie, a low-competence name, when he is rated "Excellent" ($635.98 vs $616.60, p=0.000).

and aid amounts awarded. Recall that competence is determined by valence-characteristic scores in the Names Dataset (Hayes and Mitchell 2020). Hayes and Mitchel's findings suggest that name-characteristics have the potential to minimize differences in aid allotment due to racial prejudice; however, the mitigating effects of name-characteristics do not appear to extend to gender bias. Sandra and James are both high-competence names, while Misty and Sammie are relatively low-competence. I predicted in H2 that high competence names should receive more, on average, than low competence names. Thus, holding applicant quality constant, Sandra should receive more than Misty and James should receive more than Sammie, relative to the baseline. I do not find empirical support for H2. There is no statistically or substantively significant difference between the amounts awarded to Excellent Misty and Excellent Sandra ($655.31 vs $643.20, p=0.141), nor is there between Excellent James and Excellent Sammie ($619.20 and $616.60, p=0.88), and Poor James and Poor Sammie ($600.60 and $600.32, p=0.986). This suggests that, among whites, perceived competence may matter less than when making inter-racial comparisons.

Finally, I predicted that, for male applicants, there will be no significant difference between amounts awarded to workers rated "Excellent" as compared to workers rated "Poor." Conversely, I predicted that the opposite will be true for female applicants; that is, worker quality rating will result in a significant difference in the amount awarded. To evaluate H3 and H4, I compare the difference in means for each treatment name to (James, Sammie, and Misty) when they are "Excellent" rated workers to when they are rated "Poor." Table 2 shows the results of this test, which support H3 and H4. There is a statistically significant difference in the amount of aid awarded to Excellent Misty compared to Poor Misty; however, there is no such difference between Excellent Sammie or James compared to Poor Sammie or James.

The totality of these results suggest that identically-situated men and women are evaluated differently when put in competition for scarce resources. Though implicit competence

Table 2: The Effect of Quality on Deservingness

Condition	Mean 1	Mean 2	Upper CI	Lower CI	p-value
Excellent Misty vs Poor Misty	643.198	611.797	0.728	62.073	0.045**
Excellent James vs Poor James	619.198	600.589	-13.564	50.781	0.256
Excellent Sammie vs Poor Sammie	616.598	600.317	-18.015	50.576	0.352

does not appear to play a role in aid evaluations when comparing white men and women to each other, positive third-party quality ratings affect women's earnings and do not alter men's.

Evaluation of the Mechanism[5]

I theorized that perturbing the gender of applicants is the sole cause of changes in the dollar amounts awarded to applicants and took special care to allow participants to have a socially desirable channel to express gender-neutral, value-based sentiment. Still, public aid is a partisan policy position. Historically, liberal voters and politicians favor more robust public aid programs, where as conservatives tend to oppose them. It could be the case that partisanship is driving both gender-based attitudes *and* attitudes towards welfare.

Further, scholarship shows that perceiving one's own group as having a stake in the outcomes of policy decisions can shape policy attitudes (Wolpert and Gimpel 1998; Smith 1980). It may therefore be the case that survey participants' responses are a function of in-group attitudes. That is, if female respondents award more to female applicants and male respondents award more to male applicants, my results may not be a function of gender attitudes, but simply a function of receiving a treatment condition in one's in-group compared to one's out-group.

To bolster confidence that the mechanism driving my results is varying applicants' gender and not other factors, like respondents' ideology or in-group favoritism, in this section I

[5]Exploratory analysis. The following analysis was not part of my pre-registration plan.

perform a series of tests for heterogeneous treatment effects by respondent ideology and gender.

Ideology

To mitigate concerns about respondent ideology driving my results, I compare the difference in the amount given to the baseline and the treatment name across self-reported ideology. First, I check whether there is a difference in the differences between the baseline and excellent-rated Misty between ideologies. If ideology, rather than gender, is driving my results, we should see statistically significant differences in the differences between the treatment and baseline, regardless in the overall magnitude of the results. In other words, even if conservative respondents give less, on average, to both applicants due to a principled objection to public aid, the *difference* in the amount given to Excellent Misty and the amount given to Excellent Sandra should not vary *unless* ideological differences are driving my results. Table 3 shows the results of unpaired t-tests across ideological self-identifications. The differences in the differences is statistically indistinguishable from zero across ideologies.

Table 3: Ideological Differences in Excellent Misty Treatment

Group 1	Group 2	n_{Group1}	n_{Group2}	p
Very Conservative	Conservative	42	57	0.867
Very Conservative	Moderate	42	102	0.506
Very Conservative	Liberal	42	49	0.82
Very Conservative	Very Liberal	42	35	0.607
Conservative	Moderate	57	102	0.446
Conservative	Liberal	57	49	0.922
Conservative	Very Liberal	57	35	0.417
Moderate	Liberal	102	49	0.509
Moderate	Very Liberal	102	35	0.207
Liberal	Very Liberal	49	35	0.388

Next, I examine whether there is a difference in the differences between the baseline

and excellent-rated James between ideologies. As before, if ideology is driving my results, there should be statistically significant differences in the differences from baseline between ideologies, because respondents with different ideologies are behaving differently. Table 4 shows the results of unpaired t-tests across ideological self-identifications. The differences in the differences is once again indistinguishable from zero across ideologies.

Table 4: Ideological Differences in Excellent James Treatment

Group 1	Group 2	n_{Group1}	n_{Group2}	p
Very Conservative	Conservative	36	43	0.89
Very Conservative	Moderate	36	101	0.917
Very Conservative	Liberal	36	52	0.987
Very Conservative	Very Liberal	36	41	0.621
Conservative	Moderate	43	101	0.95
Conservative	Liberal	43	52	0.876
Conservative	Very Liberal	43	41	0.722
Moderate	Liberal	101	52	0.9
Moderate	Very Liberal	101	41	0.568
Liberal	Very Liberal	52	41	0.485

In-Group Favoritism

Finally, I consider whether respondent's gender could be accounting for the results I observe. Extant work shows that individuals are more generous to members of their in-group (Chen and Li 2009). If respondents respond heterogeneously to the treatment based on whether the treatment they receive matches their gender, we should expect to see differences in the amounts male and female respondents allocate to Misty and James. Table 5 shows the results of an unpaired t-test between male and female respondents on the differences from baseline in the Excellent Misty and Excellent James condition. There are no statistically significant differences.

Given these results, we can be quite confident that the results are not simply a func-

Table 5: Gender Differences in Differences from Baseline

Group 1	Group 2	n_{Group1}	n_{Group2}	p	Treatment Name
Female	Male	187	126	0.723	Excellent Misty
Female	Male	164	134	0.742	Excellent James

tion of respondent's ideology or in-group favoritism, but are in fact due to experimental manipulation eliciting attitudes about male and female aid applicants.

Sub-Sample Analysis

The main analysis above reports results from a pooled sample made up of a nationally representative sample and a low-socioeconomic status (SES) oversample. However, there is reason to believe that socioeconomic status may correlate with differences in the type or level of sexism captured in this experiment. For example, perhaps those who are higher income express less sexism overall, but are more sensitive to the quality of those for whom their tax dollars are providing public aid. In order to address these concerns and provide higher quality inference, this section replicates the results from the pooled sample reported above for each of the two subsamples.

Nationally Representative Sample

Restricting the sample to the nationally representative subsample, I find that male names are still awarded less dollars relative to the baseline than female names, but this difference is no longer statistically significant at conventional levels. Table 6 shows the results of a difference in means test evaluating H1.

Table 7 shows the comparison between applicant names of different qualities. I find that Excellent Misty is awarded more than Poor Misty and that these differences are statistically

Table 6: Testing H1 with Nationally Representative Sample

Condition	Mean Difference	Upper CI	Lower CI	p-value
Excellent Misty vs Baseline	3.537	-11.467	18.542	0.643
Excellent James vs Baseline	16.065	-8.571	40.7	0.2
Excellent Sammie vs Baseline	10.041	-7.728	27.81	0.266

significant at conventional levels, whereas the difference in means between Poor and Excellent James and Poor and Excellent Sammie remain statistically insignificant. These findings conform with my predictions (H3 and H4) and the results from the pooled sample.

Table 7: Testing H3ab with Nationally Representative Sample

Condition	Mean 1	Mean 2	Upper CI	Lower CI	p-value
Excellent Misty vs Poor Misty	648.09	601.396	7.527	85.86	0.02**
Excellent James vs Poor James	624.393	590.662	-7.489	74.952	0.108
Excellent Sammie vs Poor Sammie	614.041	599.09	-28.278	58.181	0.497

Low-SES Sample

Restricting the analysis to the low-SES sample, I find that many results from the full sample hold. As shown in Table 8, the amount awarded to Excellent Misty is statistically indistinguishable from the amount awarded to the baseline. Excellent James earns significantly less than the baseline, as does Excellent Sammie. This is consistent with my predictions (H1).

Table 8: Testing H1 with Low SES Subsample

Condition	Mean Difference	Upper CI	Lower CI	p-value
Excellent Misty vs Baseline	27.491	-9.06	64.042	0.139
Excellent James vs Baseline	36.948	12.719	61.178	0.003***
Excellent Sammie vs Baseline	36.318	8.942	63.694	0.01**

However, Excellent Misty earns only $4 more on average than Poor Misty and this difference is not statistically significant. Excellent James earns *less* than Poor James on average, but this difference is not statistically significant either. Excellent Sammie earns more than

Poor Sammie, but this difference is also not statistically significant. These results are shown in Table 9.

Table 9: Testing H3a-b with Low SES Subsample

Condition	Mean 1	Mean 2	Upper CI	Lower CI	p-value
Excellent Misty vs Poor Misty	634.42	630.119	-45.183	53.783	0.864
Excellent James vs Poor James	608.433	618.682	-61.091	40.592	0.691
Excellent Sammie vs Poor Sammie	621.234	602.602	-38.23	75.494	0.519

Implications

The differences between subsamples suggest one of two realities. First, it is possible that the gender treatments effect the two subsamples differently. It may be the case that the general population cares more about applicant quality and less about gender, whereas low income populations bring stronger gender stereotypes into their deservingness evaluations. Alternatively, those who are *a priori* lower income may simply be less confident in the legitimacy of quality-based evaluations, perhaps having had experience with such evaluations themselves. While a thorough evaluation of these mechanisms is beyond the scope of this study, future work would benefit from a disentangling of *expressions* of sexism and deservingness from socioeconomic class.

However, another important possibility is that these differences in results are an artifact of the randomization process. Because treatments were randomized over the full, pooled sample, not over the two sub-samples individually, imbalance on pre-treatment covariates or a lack of observations for one treatment in a subsample may drive these differences. Thus, differences between samples should be interpreted conservatively, as suggestive of a possible underlying mechanism and an avenue for future work.

The Caustic Effect of Gender Stereotypes

This article addresses a fundamental question concerning American's political attitudes toward government redistribution: Does gender matter? Qualitative and historical scholars posit a strong link between gender and deservingness, and empirical research exploring the consequences of perceived deservingness often draws heavily on this normative literature. At the same time, gender and politics scholarship largely focuses on differences in perceived deservingness within men and women or within racial groups. Less attention has been dedicated to examining how gender attitudes influence American's attitudes toward welfare recipients. My work fills this gap by providing a theoretical framework and experimental tests that illuminate the link between gender, American values, and deservingness.

I find that gender and American values of hard work jointly influence how individuals evaluate the deserving poor. My results suggest that, on average, women are perceived as more deserving than similarly situated men. This is in line with historical and qualitative scholarship Willrich (2000) and broadly suggestive of ambivalently sexist attitudes in the public (Glick and Fiske 2001, 1997). Moreover, I find that third-party quality evaluations yield higher dividends for women, whereas they yield no effects for men. This result comports with a large body of scholarship suggesting that, absent external interventions, women's abilities and contributions are taken less seriously than their male counterparts. One solution identified by scholars across disciplines for helping disadvantaged women overcome these hurdles is additional job training and certificate programs that provide an external, reliable evaluation of their quality as a worker (Abel, Burger and Piraino 2020; Dadgar and Trimble 2015; Jepsen, Troske and Coomes 2014). My results suggest that these kinds of interventions are likely to improve the outcomes of women applying for aid, but not men. This is concerning, because it implies that those with the most external qualifications - those least likely to fall on hard times and require government assistance - should be the most likely to receive it.

Further, as a point of departure from past scholarship on the realization of welfare attitudes (Hayes and Mitchell 2020), I find that the valence characteristics of treatment names has no effect on the amount awarded applicants.

Moreover, my experimental methodology serves as a point of departure for a much broader research agenda. By imposing a budgetary constraint, I am able to avoid acquiescence bias and put respondents in a decision-making environment which more closely mirrors the real world. In so doing, I am better able to answer the question: given that the government cannot satisfy every need, who should it take care of first? Protecting women, especially those who are good workers, appears to be the foremost priority. One implication of this result is that policymakers are most likely to garner support for government aid programs when their messaging about these programs is focused on how they will support hardworking women, such as single mothers. Further work is needed to understand how the gender attitudes reported here extend into political messaging and policy evaluations.

Future work would also benefit from a more robust understanding of why good female workers, in particular, are favored, as this seems to cut against notions of ambivalent sexism, in which women fulfilling traditional, home-bound duties are praised and protected and those failing to do so are punished. Additionally, future work should consider what can be done in the face of gender differences in deservingness between otherwise identical applicants. If the normative goal is an equitable society, then presumably the normatively best result is one in which there are no differences between a "Sandra" or a "James", holding all else constant. This work raises clear questions as to what additional measures will reduce the gender differences in perceived deservingness of applicants. In this way, this work lays the foundation for additional research on the American welfare state.

References

Aarøe, Lene and Michael Bang Petersen. 2014. "Crowding out culture: Scandinavians and Americans agree on social welfare in the face of deservingness cues." *The Journal of Politics* 76(3):684–697.

Abel, Martin, Rulof Burger and Patrizio Piraino. 2020. "The Value of Reference Letters: Experimental Evidence from South Africa." *American Economic Journal: Applied Economics* 12(3):40–71.
URL: *https://www.aeaweb.org/articles?id=10.1257/app.20180666*

Betz, Timm, David Fortunato and Diana Z. O'brien. 2021. "Women's Descriptive Representation and Gendered Import Tax Discrimination." *American Political Science Review* 115(1):307–315.

Bobocel, D. Ramona, Leanne S. Son Hing, Liane M. Davey, David J. Stanley and Mark P. Zanna. 1998. "Justice-based opposition to social policies: Is it genuine?" *Journal of Personality and Social Psychology* 75:653–669.

Cassese, Erin C and Tiffany D Barnes. 2019. "Intersectional motherhood: Investigating public support for child care subsidies." *Politics, Groups, and Identities* 7(4):775–793.

Castillo, Marco, Ragan Petrie, Maximo Torero and Lise Vesterlund. 2013. "Gender differences in bargaining outcomes: A field experiment on discrimination." *Journal of Public Economics* 99:35–48.

Chen, Yan and Sherry Xin Li. 2009. "Group identity and social preferences." *American Economic Review* 99(1):431–457.

Clayton, Amanda, Amanda Lea Robinson, Martha C. Johnson and Ragnhild Muriaas. 2020. "(How) Do Voters Discriminate Against Women Candidates? Experimental and Qualitative Evidence From Malawi." *Comparative Political Studies* 53(3-4):601–630.

Converse, Phillip. 1964. *The Nature of Belief Systems in Mass Publics.* London: Free Press of Glencoe.

Dadgar, Mina and Madeline Joy Trimble. 2015. "Labor Market Returns to Sub-Baccalaureate Credentials: How Much Does a Community College Degree or Certificate Pay?" *Educational Evaluation and Policy Analysis* 37(4):399–418.

Desante, Christopher. 2013. "Working Twice as Hard to Get Half as Far: Race, Work Ethic, and America's Deserving Poor." *American Journal of Political Science* 57:342–356.

Exley, Christine L, Muriel Niederle and Lise Vesterlund. 2020. "Knowing when to ask: The cost of leaning in." *Journal of Political Economy* 128(3):816–854.

Fearon, James D. 1999. "Electoral accountability and the control of politicians: selecting good types versus sanctioning poor performance." *Democracy, accountability, and representation* 55:61.

Fossati, Flavia. 2018. "Who wants demanding active labour market policies? Public attitudes towards policies that put pressure on the unemployed." *Journal of Social Policy* 47(1):77–97.

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

Gilliam, Franklin D. 1999. "The "Welfare Queen" Experiment: How Viewers React to Images of African-American Mothers on Welfare." *Nieman Reports* 53.
URL: *https://escholarship.org/uc/item/17m7r1rq*

Gilman, Michele Estrin. 2013. "The return of the welfare queen." *Am. UJ Gender Soc. Pol'y & L.* 22:247.

Glick, Peter and Susan T. Fiske. 1997. "Hostile and Benevolent Sexism." *Psychology of Women Quarterly* 21(1):119–135.

Glick, Peter and Susan T. Fiske. 2001. "An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality." *American Psychologist* 56:109–118.

Goldin, Claudia and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians." *American Economic Review* 90(4):715–741.

Greene, Zachary and Diana O'Brien. 2016. "Diverse parties, diverse agendas? Female politicians and the parliamentary party's role in platform formation." *European Journal of Political Research* 55(3):435–453.

Hassell, Hans J.G. and Neil Visalvanich. 2019. "The Party's Primary Preferences: Race, Gender, and Party Support of Congressional Primary Candidates." *American Journal of Political Science* 63(4):905–919.

Hayes, Matthew, David Fortunato and Matthew V Hibbing. 2021. "Race–gender bias in white Americans' preferences for gun availability." *Journal of public policy* 41(4):818–834.

Hayes, Matthew and Elizabeth Mitchell. 2020. "Name Characteristics Dataset." *Harvard Dataverse* .
URL: *https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/HEOSVW*

Horton, John J, David G Rand and Richard J Zeckhauser. 2011. "The online laboratory: Conducting experiments in a real labor market." *Experimental economics* 14:399–425.

Huddy, Leonie. 2004. "Contrasting theoretical approaches to intergroup relations." *Political Psychology* 25(6):947–967.

Ingram, Helen and Anne Schneider. 1993. "Constructing citizenship: The subtle messages of policy design." *Public policy for democracy* pp. 68–94.

Iyengar, Shanto, Gaurav Sood and Yphtach Lelkes. 2012. "Affect, not ideologya social identity perspective on polarization." *Public opinion quarterly* 76(3):405–431.

Jepsen, Christopher, Kenneth Troske and Paul Coomes. 2014. "The Labor-Market Returns to Community College Degrees, Diplomas, and Certificates." *Journal of Labor Economics* 32(1):95–121.
URL: *https://www.journals.uchicago.edu/doi/full/10.1086/671809*

Katz, Irwin and R. Glen Hass. 1988. "Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures." *Journal of Personality and Social Psychology* 55:893–905.

Keiser, Lael R. 1999. "State bureaucratic discretion and the administration of social welfare programs: The case of social security disability." *Journal of public administration research and theory* 9(1):87–106.

Kogan, Vladimir. 2021. "Do Welfare Benefits Pay Electoral Dividends? Evidence from the National Food Stamp Program Rollout." *The Journal of Politics* 83(1):58–70.

Krook, Mona Lena and Diana Z O'Brien. 2012. "All the president's men? The appointment of female cabinet ministers worldwide." *The Journal of Politics* 74(3):840–855.

Lasswell, Harold D. 1936. *Politics: Who gets what, when, how*. Pickle Partners Publishing.

Lewis, Paul G, Doris Marie Provine, Monica W Varsanyi and Scott H Decker. 2013. "Why do (some) city police departments enforce federal immigration law? Political, demographic, and organizational influences on local choices." *Journal of Public Administration Research and Theory* 23(1):1–25.

Mandel, Hadas. 2013. "Up the Down Staircase: Women's Upward Mobility and the Wage Penalty for Occupational Feminization, 1970-2007." *Social Forces* 91(4):1183–1207.
URL: *https://doi.org/10.1093/sf/sot018*

McClosky, Herbert and John Zaller. 1984. "The American Ethos.".

Mettler, Suzanne. 2011. *The submerged state: How invisible government policies undermine American democracy*. University of chicago Press.

Monnat, Shannon M. 2010*a*. "The color of welfare sanctioning: Exploring the individual and contextual roles of race on TANF case closures and benefit reductions." *The Sociological Quarterly* 51(4):678–707.

Monnat, Shannon M. 2010*b*. "Toward a critical understanding of gendered color-blind racism within the US welfare institution." *Journal of Black Studies* 40(4):637–652.

Mummolo, Jonathan and Erik Peterson. 2019. "Demand effects in survey experiments: An empirical assessment." *American Political Science Review* 113(2):517–529.

Nelson, Thomas E and Donald R Kinder. 1996. "Issue frames and group-centrism in American public opinion." *The Journal of Politics* 58(4):1055–1078.

Neumark, David, Roy J. Bank and Kyle D. Van Nort. 1996. "Sex Discrimination in Restaurant Hiring: An Audit Study." *The Quarterly Journal of Economics* 111(3):915–941. **URL:** *https://doi.org/10.2307/2946676*

Petersen, Michael Bang. 2012. "Social welfare as small-scale help: evolutionary psychology and the deservingness heuristic." *American Journal of Political Science* 56(1):1–16.

Petersen, Michael Bang, Daniel Sznycer, Leda Cosmides and John Tooby. 2012. "Who deserves help? Evolutionary psychology, social emotions, and public opinion about welfare." *Political psychology* 33(3):395–418.

Quadlin, Natasha. 2018*a*. "The Mark of a Woman's Record: Gender and Academic Performance in Hiring." *American Sociological Review* 83:331–360.

Quadlin, Natasha. 2018*b*. "The mark of a woman's record: Gender and academic performance in hiring." *American Sociological Review* 83(2):331–360.

Quidt, Jonathan de, Lise Vesterlund and Alistair J. Wilson. 2019. "Experimenter demand effects." *Handbook of Research Methods and Applications in Experimental Economics* pp. 384–400.

Rabinowitz, Joshua L, David O Sears, Jim Sidanius and Jon A Krosnick. 2009. "Why Do White Americans oppose race-targeted policies? Clarifying the impact of symbolic racism." *Political psychology* 30(5):805–828.

Schmidt, Diane E. 2002. "Politicization and responsiveness in the regional offices of the NLRB." *The American Review of Public Administration* 32(2):188–215.

Schneider, Anne and Helen Ingram. 1993. "Social Construction of Target Populations: Implications for Politics and Policy." *American Political Science Review* 87(2):334–347.

Schneider, Anne L and Helen M Ingram. 2005. *Deserving and entitled: Social constructions and public policy*. Suny Press.

Smith, Tom W. 1980. "The 75% solution: An analysis of the structure of attitudes on gun control, 1959-1977." *J. Crim. L. & Criminology* 71:300.

Sniderman, Paul M., Edward G. Carmines, Geoffrey C. Layman and Michael Carter. 1996. "Beyond Race: Social Justice as a Race Neutral Ideal." *American Journal of Political Science* 40(1):33–55.
URL: *https://www.jstor.org/stable/2111693*

Sniderman, Paul M. and Philip E. Tetlock. 1986. "Symbolic Racism: Problems of Motive Attribution in Political Analysis." *Journal of Social Issues* 42(2):129–150.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-4560.1986.tb00229.x*

Sniderman, Paul M. and Richard A. Brody. 1977. "Coping: The Ethic of Self-Reliance." *American Journal of Political Science* 21(3):501–521.
URL: *https://www.jstor.org/stable/2110579*

Sniderman, Paul M and Thomas Piazza. 1993. *The scar of race*. Harvard University Press.

Theiss-Morse, Elizabeth. 2009. *Who counts as an American?: The boundaries of national identity*. Cambridge University Press.

van Oorschot, Wim. 2000. "Who should get what, and why? On deservingness criteria and the conditionality of solidarity among the public in." *Policy & Politics* 28.

van Oorschot, Wim, Femke Roosma, Bart Meuleman and Tim Reeskens. 2017. *The social legitimacy of targeted welfare: Attitudes to welfare deservingness*. Edward Elgar Publishing.

Weissert, Carol S. 1994. "Beyond the organization: The influence of community and personal values on street-level bureaucrats' responsiveness." *Journal of Public Administration Research and Theory* 4(2):225–254.

Willrich, Michael. 2000. "Home slackers: Men, the state, and welfare in modern America." *The Journal of American History* 87(2):460–489.

Winter, Nicholas J. G. 2006. "Beyond Welfare: Framing and the Racialization of White Opinion on Social Security." *American Journal of Political Science* 50(2):400–420.

Winter, Nicholas JG. 2008. *Dangerous frames: How ideas about race and gender shape public opinion*. University of Chicago Press.

Wolpert, Robin M and James G Gimpel. 1998. "Self-interest, symbolic politics, and public attitudes toward gun control." *Political Behavior* 20:241–262.