

Incarceration and Deservingness: Experimental Evidence on Less Eligibility*

Christopher D. Berk[†] Lucas Núñez[‡] Brett Burkhardt[§]

August 17, 2026

The principle of less eligibility holds that no person subject to punishment should be better off than the poorest free laborer. Originally a utilitarian deterrence claim, it survives as a folk standard of moral accounting structuring public resistance to rehabilitative prison programming—yet it has received no direct experimental test. Using a conjoint survey experiment with $N = 1,929$ respondents, we estimate the causal penalty incarcerated applicants face in public eligibility evaluations and decompose its mechanisms. Incarcerated applicants are substantially less likely to receive program support than equally poor but free applicants; a penalty that holds across offense types and survives favorable evidence that the incarcerated applicant benefits from the program more than a comparable free applicant. Testing three mechanisms, we find that portraying the program as imposing greater or lesser hardship does not move opinion, community membership raises support for all applicants but does not narrow the incarceration penalty, and portraying incarcerated applicants as hard-working paradoxically widens the gap rather than narrowing it. The incarceration penalty, in short, is moral rather than instrumental: a judgment about desert that evidence can narrow but not dislodge.

Keywords: incarceration, public opinion, moral reasoning, criminal justice, conjoint experiment

*This study is pre-registered at the Open Science Framework (OSF). Pre-analysis plan: [URL blinded for peer review], registered June 22, 2026.

[†]Assistant Professor, Schar School of Policy and Government, George Mason University

[‡]Associate Professor, Schar School of Policy and Government, George Mason University

[§]Associate Professor, School of Public Policy, Oregon State University

1 Introduction

Americans are, by most measures, a punitive public: they favor mandatory minimums and harsh sentences and view courts as chronically too lenient (Whitman, 2003). Yet most also support what Lisa Miller terms “punishment plus”—strict sanctions *and* effective programming to reduce recidivism and prepare individuals for release (Miller, 2016). This is not a minor inconsistency. The United States operates an expansive carceral system with no parallel among industrial democracies, holding nearly 2 million people in prisons and jails at any given time—roughly 700 of every 100,000 adult residents—and the lifetime incidence of imprisonment is far higher still, touching roughly 1 in 18 Americans over the course of their lives.¹ Punishment’s reach also extends well beyond the sentence itself: a conviction triggers a wide array of civil disabilities and collateral consequences—felony disenfranchisement chief among them—that persist long after release (Zimring, 2020; Lerman and Weaver, 2014). Disenfranchisement alone affects several million Americans nationwide (Manza and Uggen, 2006). Whether the state can make good on the “plus” in punishment-plus—education, job training, therapy—for a population this large depends on how the public resolves the tension between wanting punishment and wanting rehabilitation. How it does so is a question existing accounts have not satisfactorily answered.

The punitiveness and punishment-plus literatures document these two commitments separately, but neither specifies the boundary between them: what determines when support for a program shifts from a willingness to help to an insistence that help is undeserved? We suggest that support for rehabilitative programming is conditional on a principle of moral accounting familiar from the broader politics of the welfare state: deservingness. A substantial literature shows that support for redistribution turns on markers like need, reciprocity, and effort that sort recipients into deserving and undeserving (Jensen and Petersen, 2017; Petersen, 2012; Oorschot, 2000; Gilens, 1999). Incarceration status functions as an unusually powerful marker of this kind, one whose logic the politics of punishment has crystallized into an explicit and durable principle: “less eligibility.”

First articulated in the 1790s by Jeremy Bentham, the principle holds that no person subject to punishment should be better off than the poorest free laborer. Today it survives as a powerful, largely unquestioned folk standard, expressed in the familiar refrain: “Why should people who break the law enjoy comforts denied to those who follow it?” The two commitments—to punishment and to rehabilitation—can coexist, but only so long as the punished are adequately deprived of things enjoyed by their free counterparts. When the punished are seen to receive benefits or programs not available to workers on the outside, less eligibility looms as a potential popular veto.²

Despite its prominence, whether and how less eligibility actually structures public support for prison programming has received little direct empirical scrutiny. We address that shortcoming by providing a direct measurement of the principle as it operates in public opinion, decomposing it into component parts. To do this, we recruited a panel of $N = 1,929$ U.S. adults to complete a pre-registered conjoint experiment online. Respondents evaluated hypothetical applicants to a government social program—one year of college courses and job counseling—across profiles that independently varied applicant social position (incarcerated, near the poverty line, or on welfare), offense severity, moral-cultural components of less eligibility (applicant effort, applicant belonging, and program design), instrumental cost-benefit evidence, and applicant race. Randomizing social position within the experiment yields causally identified estimates of the penalty incarcerated appli-

¹The lifetime risk of imprisonment for all U.S. residents rose from 3.6% in 1986 to 10.3% in 2003 before falling to 5.7% in 2016. On the basis of 2016 first-time prison admission rates, 1 in 11 males (9.2%) and 1 in 49 females (2.1%) could expect to be imprisoned over the course of their lives. See Roehrkasse and Wildeman (2022) and Robey et al. (2023).

²The same categorical logic of desert extends beyond programming into civic status itself. For example, philosophical defenses of felony disenfranchisement invoke much the same reasoning (Cholbi, 2002; Dilts, 2014), and empirical work documents its scope and persistence (Manza and Uggen, 2006; Morris, 2021). Less eligibility, in other words, is not a quirk of prison policy; it is what a general deservingness heuristic looks like when the marker sorting the deserving from the undeserving is legal status rather than employment, income, or health.

cants face relative to their free counterparts as well as the attributes that exacerbate or diminish it. We find that the incarceration penalty is substantial; incarcerated applicants are viewed as significantly less worthy of the government social program than either welfare recipients (4.9 percentage points) or free-poor applicants (7.7 percentage points). Moreover, the incarceration penalty is only partly answerable to evidence. Telling respondents that the incarcerated applicant benefits three times as much from the identical program as the comparable free-poor applicant shrinks the incarceration penalty by only about half. The incarceration penalty, in short, is a moral judgment that yields partially to instrumental evidence without being reducible to it. Decomposing the moral-cultural account further, we find its mechanisms do not behave alike: program design proves inert on every measure we test, community belonging raises support broadly without narrowing the incarceration penalty itself, and, strikingly, meritocratic desert cues do not shrink the incarceration penalty but widen it. A hard-working incarcerated applicant is compared against a hard-working person near the poverty line, and that comparison is exactly the one less eligibility renders least forgiving.

This study contributes to the literature on public opinion and incarceration in several ways. It provides a direct experimental test of the principle of less eligibility as applied to public opinion—whether support for public programs is conditional on the social position of the recipient (incarcerated, on welfare, or below the poverty line). Decomposing the moral-cultural account into distinct mechanisms shows which reform levers are structurally capable of narrowing the incarceration penalty and which are not, a distinction the existing literature has not drawn and one with direct implications for which reform strategies are likely to face headwinds. And the penalty’s relationship to punitive attitudes and their sources is revealing in its own right: it scales with punitiveness, anti-rehabilitative orientations, and conservative ideology, but is untouched by direct exposure to incarceration or crime victimization—evidence that less eligibility functions as a widely shared cultural logic rather than an expression of personal animus.

2 Theory and Hypotheses

The principle of less eligibility has a long pedigree in both the philosophy and politics of punishment. In its original utilitarian formulation, the principle holds that no person subject to punishment should be better situated than the poorest free laborer: if prison conditions equaled or exceeded life outside for the struggling poor, incarceration would lose its deterrent force (Bentham, 1843; Bentham, 1791).³

Less eligibility is a comparative standard. It asks not how well the punished fare on their own terms, but how they fare *relative to a free counterpart of similar standing*. That comparison carries two important implications. First, it means the principle can hold at very different absolute levels of generosity—a society can treat its prisoners relatively well and still satisfy less eligibility, so long as free people outside prison are treated even better. Testing the principle therefore requires measuring a gap, not a level: the difference between support for an incarcerated person and support for an otherwise comparable free person, a quantity we call the *incarceration penalty*. Second, that gap depends on which free person serves as the comparison. That comparator was, for Bentham, the poorest free laborer, and the low-wage worker remains the more demanding benchmark: on this reading, punishment is meant to do two things at once—deter crime and encourage work—so that if prison conditions rivaled those of the working poor, people on the margins of the labor market would have less reason to keep working and more reason to turn to crime instead. A softer version substitutes the welfare recipient for the low-wage worker: it asks only that punishment deter crime, not that it also encourage work, so conditions need only make crime less attractive than welfare receipt, not less attractive than a paycheck (Bonnet, 2025).

³Bentham’s reasoning was transmitted into policy through Edwin Chadwick, his secretary, who co-drafted the 1834 Royal Commission Report on the Poor Laws; the Report codified “less eligibility” as its governing administrative principle, which the Poor Law Amendment Act (1834) then implemented. See Piven and Cloward (1993).

Subsequent empirical work has tested this deterrence logic and has, for the most part, rejected it.⁴ Harsher prison conditions not only fail to increase deterrence; higher-service prisons actually reduce recidivism (Chen and J. M. Shapiro, 2007; Drago et al., 2011; Mastrobuoni and Terlizzese, 2022). Rusche and Kirchheimer, in a work first published in 1939, revisited the principle in a Marxian register, arguing that punishment severity tracks the economic value of free labor.⁵ When labor is scarce and free workers are doing relatively well, prison conditions can improve while still remaining below the free-labor floor. When unemployment is high and the reserve army of labor is desperate, the floor falls—and prison must fall further to remain beneath it, producing harsher conditions and higher rates of incarceration as a mechanism for disciplining the surplus population. The best systematic test of this prediction is Michalowski and Carlson (1999); they find no single stable relationship between unemployment and imprisonment rates in U.S. data, concluding that the principle holds only contingently, mediated by welfare-state institutions that vary across periods.

But less eligibility has a third interpretation, most prominently associated with David Garland (1990), that treats the principle not as a behavioral or political-economic claim but as a moral-cultural belief that structures public sympathy toward prisoners. On this account, less eligibility functions as a folk principle of moral accounting that compares what prisoners receive to what law-abiding citizens go without. It recurs in debates about everything from prison food to educational programming and functions as a default objection to reform (Sieh, 1989; Johnston and Wozniak, 2021a; Johnston and Wozniak, 2021b). The principle, in this reading, marks a moral boundary between the deserving poor and the criminal poor. Subsequent treatments have given this logic a harder edge, treating any expenditure on incarcerated individuals beyond bare subsistence as presumptively impermissible on fairness grounds. The ‘no frills’ prison movement of the 1990s is probably the clearest example, with advocates pushing to strip cable TV, weightlifting equipment, and various forms of instruction from prisoners, extending the less eligibility logic even to calories consumed. In the words of Rep. Dick Zimmer, the author of the “no-frills” amendment, “Some criminals have come to view jail as an almost acceptable life style because amenities are better for them on the inside than on the outside.” The measure, passed by a voice vote in the House, would have required Federal prison officials to “provide prisoners the least amount of amenities and personal comforts consistent with constitutional requirements and good order and discipline” (Peterson, 1995; Labotka, 2021).⁶ When invoked, less eligibility tends to function as an iron constraint on possibility.

Of the three interpretations of less eligibility, this moral-cultural account speaks most directly to contemporary debates over rehabilitative programming, yet it remains the least empirically developed. The broader public opinion literature on criminal justice documents a consistent punitive orientation among Americans. Enns (2016), for example, constructs a latent punitiveness measure from items spanning the death penalty, court permissiveness, and mandatory minimums, and shows that it broadly tracks changes in crime rates and media coverage over time (Enns, 2016; Pickett, 2019). However, punitiveness does not preclude support for rehabilitation. As mentioned earlier, Miller (2016) finds that most Americans favor what she terms “punishment plus”—strict sanctions *and* effective programming—suggesting that the operative question is not whether the public supports rehabilitation in the abstract but whether that support is conditional on the identity of the recipient. Less eligibility implies that it is: the same program may be evaluated differently depending on whether the applicant is incarcerated or merely impoverished, and it is that comparative judgment that no prior study has directly tested.

Our core hypothesis is:

⁴Jail administrators will often reference seasonal fluctuations in homeless jail populations (the desire for “three hots and a cot”), for example, but there is no clear evidence that marginal populations cycle through incarceration in response to material deprivation.

⁵See Rusche et al. (2003). Foucault’s *Discipline and Punish* and Melossi and Pavarini’s *The Prison and the Factory* take up their approach, but largely set aside Rusche and Kirchheimer’s claims about less eligibility. For a review of this tradition, see De Giorgi (2006) and Harcourt (2011).

⁶For more on the no-frills movement in the 1990s and the wider punitive turn of which it was part, see generally Gottschalk (2006) and Garland (2001).

H1 (Less Eligibility; incarceration penalty): Incarcerated applicants will receive lower support for the same government program than otherwise comparable individuals near the poverty line or on welfare.

We refer to the gap in support between incarcerated and free-poor applicants—our primary point of comparison throughout the paper—as the “incarceration penalty.” Incarcerated applicants also face a smaller, distinct gap relative to welfare recipients, which we refer to as the “welfare-referenced incarceration penalty” wherever the distinction matters. Welfare serves as a second free-world baseline, not merely poverty, because it isolates what is specific to incarceration: welfare applicants are needy and already stigmatized as state dependents, while poverty-line applicants are needy but not stigmatized (Gilens, 1999). Comparing incarcerated applicants against both lets us net out generic dependency stigma, attributing what remains to incarceration itself.

The moral-cultural version of less eligibility, we suggest, is a bundle of potential mechanisms masquerading as one. Here, we distinguish three. The first, *community belonging*, admits two readings that generate different predictions for our design. In Garland’s cultural sociology, criminal conviction itself places individuals outside the boundaries of the political or moral community: legal status is the line of exclusion, and no independent signal of local attachment should be able to move someone out of a category that conviction has already assigned them to (Garland, 1990). A second reading treats community membership as a graded social fact rather than a categorical one—closer to the “identity” criterion in the deservingness-heuristics literature, where perceived closeness to “us” shapes how deserving a recipient appears, independent of formal status (Oorschot, 2000). On this reading, an applicant with deep local ties—family nearby, a shared history, personal connections in the respondent’s community—should be read as more of an insider than one without them, whatever their legal status, and community-belonging cues should narrow the incarceration penalty rather than merely raise support in general.

Our design manipulates whether an applicant has roots in the respondent’s community independently of legal status, which lets us adjudicate between these two readings directly. If belonging is categorical and conviction-bound, roots should raise support broadly without closing the gap between incarcerated and free-poor applicants: the cue never touches the thing doing the excluding, which is legal status itself. If belonging is a graded social fact, roots should narrow that gap instead. Because incarceration is the strongest outsider signal in the design, belonging cues have the most ground to make up there, so the same evidence of local ties should raise support more for the incarcerated applicant than for the free-poor applicant, closing part of the distance between them. We test the same logic against the parallel, non-carceral penalty faced by welfare applicants, for whom no comparable legal boundary exists—giving us a case where the graded-membership account should apply cleanly, as a point of comparison for the incarcerated case.

H2A (Effect of community belonging cues): Positive information about an applicant’s ties to the community will increase support for the program: an applicant described as belonging to the respondent’s community will receive more support than one described as not belonging.

H2B (Community belonging moderates incarceration penalty): The incarceration penalty will be moderated by community belonging: the support gap between incarcerated and free-poor applicants will differ depending on whether the applicant is described as belonging to the respondent’s community or not.

A second mechanism, grounded in *meritocratic* ideology, treats incarcerated individuals as “line-cutters”—recipients of state resources who have not earned their place through the deferred gratification and rule-following meritocratic norms prescribe. Hochschild (2016)’s account of the “deep

story” animating resentment of government programs captures this logic: the sense of standing patiently in line only to watch others cut ahead with state assistance. When meritocratic desert drives the comparison, applicants described as hard-working should recover moral standing and the penalty should narrow. Experimental work on punitive framing is broadly consistent with this expectation: sympathetic cues attenuate punitive responses, though effects are typically modest (Peffley and Hurwitz, 2007; Ramirez, 2013). We test this account through the applicant’s described *effort*, varying whether the applicant is described as hard-working and reliable or as unreliable.

A competing prediction, however, follows from attribution research. The fundamental attribution error documents a systematic tendency to overweight personal agency as a cause of behavior (Ross, 1977): describing an incarcerated applicant as hard-working does not simply raise their moral standing, it also heightens the inference that they had the capacity to succeed through legitimate means and chose not to. Effort evidence complicates the comparison in a second way as well, by reshaping the benchmark itself. A hard-working person near the poverty line is the paradigmatic deserving poor, someone who deferred gratification and played by the rules; against that benchmark, a hard-working incarcerated applicant is precisely the figure Hochschild’s deep story describes—someone who could have stood in line and chose to cut instead. Both effects point the same way, toward a widening rather than a narrowing of the gap.

H3A (Effect of effort cues): Positive signals about applicant effort will increase support for the program: an applicant described as hard-working and reliable will receive more support than one described as unreliable.

H3B (Effort cues & incarceration penalty): The incarceration penalty will be partially offset by effort cues: the incarceration penalty will be smaller when an applicant is described as hard-working and reliable than when described as unreliable.

A third mechanism centers on *visible suffering*: the objection is not to giving resources to prisoners *per se*, but to resources that soften the appearance of punishment and thereby undermine its expressive function. The United States has a long tradition of harsh and degrading punishment (Whitman, 2003). Penal programs that violate this tradition risk drawing opprobrium and being condemned as too soft. A strict, intensive vocational program may be evaluated quite differently from a tablet distribution program, not because of cost or effectiveness but because of what each signals about the punitive character of incarceration. This logic implies an interaction, not just a main effect: program design should matter most for incarcerated applicants, since only their case carries a punitive signal to undermine. But a demanding, structured design could also raise support on its own, for any applicant, by functioning as a conditionality cue akin to a work requirement for welfare programs—evidence the recipient must earn the benefit rather than simply receive it. We test both possibilities through *program design*, contrasting a demanding, structured program against a flexible, supportive one.

H4A (Effect of program design): Support will be higher for programs described as demanding and structured compared to programs described as flexible and supportive.

H4B (Program design moderates incarceration penalty): The incarceration penalty will be moderated by program design: describing the program as demanding and structured rather than flexible and supportive will raise support more for incarcerated applicants than for poverty-line applicants, narrowing the incarceration penalty, with welfare applicants intermediate between the two.

Common to all three mechanism accounts is a prediction about the role of instrumental evidence. If the incarceration penalty reflects a moral boundary judgment rather than a cost-benefit calculation,

information documenting program effectiveness should narrow the gap without closing it. We test this logic directly:

H5A (Effect of cost-benefit information): Cost-benefit information about the program indicating reduced recidivism, improvement in employment outcomes, and taxpayer savings will increase program support.

H5B (Moral dominance): The incarceration penalty will survive favorable cost-benefit evidence: information indicating that the incarcerated applicant benefits from the program more than an otherwise similar free applicant will narrow the incarceration penalty without eliminating it.

The moral-cultural account implies that less eligibility operates as a broadly shared cultural norm. Nonetheless, its intensity should vary with respondent characteristics that prime the relevant moral logic. We examine four. Punitive and anti-rehabilitative orientations provide a construct validity benchmark: if the experimental penalty tracks what it should, more punitive respondents should apply a larger incarceration penalty. We expect this pattern to hold, and treat it as a check on the measure rather than a substantive finding. Political ideology is the more theoretically informative predictor: conservative orientations are associated with a sharper distinction between deserving and undeserving recipients of public support (Feldman and Zaller, 1992; Skitka and Tetlock, 1993; cf. Petersen, 2012), and we expect both the incarceration penalty and the offense severity gradient to be larger among conservative respondents. Direct experience with the carceral state should attenuate the incarceration penalty—proximity to incarcerated individuals makes the comparison case less abstract, enabling empathy and blurring the moral boundary that less eligibility implies (Lerman and Weaver, 2014; Soss and Weaver, 2017; Unnever and Cullen, 2009). Crime victimization, by contrast, should amplify the penalty, heightening the salience of the less eligibility comparison and strengthening punitive orientations toward incarcerated applicants.

3 Methods

We test these claims using a conjoint survey experiment. Conjoint analysis—now standard in studies of multidimensional preference (Hainmueller, Hopkins, et al., 2014)—allows us to decompose respondents’ program support evaluations into the contributions of individual applicant characteristics, including the social position attribute (incarcerated, impoverished, welfare beneficiary) that operationalizes the less eligibility comparison. Our design, hypotheses, and analysis strategy were pre-registered prior to data collection; the pre-analysis plan is available at the OSF.⁷

3.1 Experimental Design

To examine whether incarceration status reduces public support for access to government programs, we conducted a vignette-based conjoint survey experiment. In each conjoint task, respondents evaluated a hypothetical applicant to a government program offering one year of free college-level courses and job counseling. Following a single-profile conjoint design, program and applicant attributes vary independently across vignettes, enabling identification of the causal effect of each attribute.⁸ Each respondent completed six conjoint tasks.

⁷[URL blinded for peer review, anonymized copy available in Appendix H], registered June 22, 2026. The order of hypotheses differs between the paper and the pre-analysis plan. Appendix G includes a crosswalk between the two. Nonetheless, all hypotheses in the PAP are tested in the paper and its appendices.

⁸We note that Hainmueller, Hangartner, et al. (2015) find that paired-profile conjoint designs track real-world behavioral benchmarks somewhat more closely than single-profile vignette designs in the context of immigration

Our design includes eight randomly assigned attributes: social position (the primary treatment), offense severity (constrained as described below), effort, community belonging, program design, cost-benefit information, applicant race, and applicant name. Table 1 summarizes the attributes, their levels, and assignment probabilities. Appendix B presents randomization and attribute independence checks; those checks show randomization occurred as designed. Appendix Table B.4 shows balance for the difference attributes across a number of pre-treatment variables.

Table 1: Conjoint experiment attributes, levels, and assignment probabilities

Attribute	Level	p
Social Position	Near the poverty line	.25
	Currently on welfare	.25
	Incarcerated in a state prison	.50
Offense Severity ^a	Shoplifting	.33
	Burglary of an unoccupied home	.33
	Armed robbery	.33
Effort	Hard-working and reliable	.50
	Unreliable	.50
Community Belonging	Belongs to the community	.50
	Does not belong to the community	.50
Program Design	Structured and demanding	.50
	Flexible and supportive	.50
Cost-Benefit Information	Equal benefit	.50
	Unequal benefit	.50
Race	White	.25
	Black	.25
	Latino	.25
	Asian	.25
Name ^b	Albert, Henry, Isaac, Joe, Jordan, Samuel	0.16

^a Constrained to appear only when Social Position = Incarcerated; probabilities are conditional on that level.

^b Drawn without replacement across profiles seen by each respondent.

3.2 Attributes

The *social position* attribute is the primary treatment and captures the applicant’s circumstances relative to the program: near the poverty line, currently on welfare, or incarcerated in a state prison. To maximize statistical power for detecting both the incarceration penalty—the central quantity of interest—and the offense severity effect, the incarcerated level is assigned with probability 0.50, while the poverty and welfare levels are each assigned with probability 0.25. Comparing the incarcerated level against both free-world baselines separately, rather than collapsing poverty and welfare into a single comparison group, is what lets us isolate an incarceration-specific penalty from a more general penalty attached to stigmatized program dependency (see Section 2).

We consider whether *offense severity* alters the incarceration penalty by presenting prisoners with one of three convictions: shoplifting, burglary of an unoccupied home, and armed robbery. We restrict offense severity to acquisitive crimes, as these are the focus of the less eligibility principle

attitudes; we view the fit to our specific decision context and the social-desirability considerations above as the more relevant considerations for this application. See Horiuchi et al. (2022).

(Bonnet, 2025).⁹ These three offenses—shoplifting of \$500 in merchandise, burglary of an unoccupied home, and armed robbery—are typically felonies in most U.S. states.¹⁰ The ordering follows the severity rankings established by Wolfgang et al. (1985).¹¹

The remaining attributes capture dimensions of deservingness and program features, which may serve as distinct components of moral accounting. Applicant *effort* distinguishes hard-working and reliable from unreliable characterizations. *Community belonging* indicates whether the applicant belongs or does not belong to the respondent's broader community. *Program design* varies whether the program is described as structured and demanding versus flexible and supportive. The *cost-benefit information* attribute indicates whether the program's benefits are presented as equal to or greater than comparable investments. Applicant *race* is randomized across four levels (White, Black, Latino, and Asian). Finally, *applicant name* is drawn without replacement from a pool of six names selected for their racial ambiguity, included solely to enhance profile realism.

While our main hypotheses do not rely on the severity attribute, we find it important to include it in our vignettes for two reasons. First, it allows us to evaluate whether desert is influenced by offense severity.¹² Second, being silent about offense severity in the vignettes would not make it disappear from respondents' minds: it would simply invite them to fill the gap with their own priors, and those priors can masquerade as (and confound) the effects we are trying to measure. The same logic applies to race and ethnicity: left unspecified, respondents may infer an applicant's racial or ethnic background or assume a link between certain offenses and certain groups, and either inference would distort our estimated effects. The inclusion of both these attributes allows us to rule out both these possibilities. And by randomly assigning them we can further evaluate supplemental hypotheses involving them.

Appendix A contains an illustrative vignette as well as the full language of each of the attribute levels.

3.3 Outcomes

For each vignette, we measure two outcomes, both of which assess respondents' support for "a new publicly funded program that provides one year of college-level courses and job counseling." The primary outcome is a binary policy choice in which respondents indicate whether their state should offer the program to applicants like the one described (yes/no). The secondary outcome is a five-point scale measuring how deserving the applicant is perceived to be of receiving the program. Per our pre-analysis plan, the binary choice is the primary outcome variable in all main analyses; the deservingness scale is used in robustness checks in Appendix D.¹³

⁹Because offense severity is only substantively meaningful when the applicant is incarcerated, the *offense severity* attribute is constrained to appear only in profiles where social position is set to incarcerated. This design feature produces a structured missingness in the severity attribute that is addressed in the estimation strategy below. When included, severity is drawn with equal probability from three levels: shoplifting, burglary of an unoccupied home, and armed robbery.

¹⁰Burglary of an unoccupied home and armed robbery are felonies in virtually every state regardless of the dollar amount involved, while \$500 in stolen merchandise was selected as a plausible lower bound for felony-level theft—the threshold at or above which shoplifting rises from a misdemeanor to a felony in a meaningful subset of states, even though the exact cutoff varies considerably by jurisdiction.

¹¹Paul Robinson and colleagues have used experimental methods to map lay judgments of criminal liability, finding that individuals rely primarily on assessments of moral blameworthiness rather than on consequentialist criteria. Across diverse populations, there is substantial convergence on a core of blameworthy conduct—physical aggression, property appropriation, deceptive exchange—even amid variation on less canonical offenses (Robinson et al., 2007).

¹²This severity grading is also a hypothesis included in our pre-analysis plan.

¹³The binary choice and scale results are largely consistent with each other. Minor differences are noted where relevant and in Appendix D.

3.4 Sample

We recruited 2,000 respondents via Prolific, an online participant recruitment platform, to complete the survey on Qualtrics.¹⁴ We rely on Prolific’s ‘representative sample’ option, which cross-stratifies participants on five age groups (18–24, 25–34, 35–44, 45–54, and 55+), two sex categories (male and female), and three political affiliation groups (Democrat, Republican, and Independent) matched to population proportions. Participation was restricted to United States residents who speak English. We exclude ten respondents who participated in a prior pilot study we conducted on the same platform.

To protect data quality, we enabled Prolific’s authenticity checks, which include identity verification, participant quality monitoring, automated detection of extremely rapid completions, and bot screening. After excluding respondents with low human scores and those not answering any of the conjoint tasks, we are left with a sample of $N = 1,929$ respondents and a total of 11,547 completed conjoint tasks. Table B.1 in Appendix B presents summary statistics for the variables in our study and details on variable definitions.

Our survey instrument includes three attention checks: an instruction check administered immediately before the conjoint tasks that asks how many vignettes respondents will be asked to complete, and two factual manipulation checks administered after the final task that ask respondents to identify the type of program described and to recall the effort attribute of the final profile. Attention check performance is used to assess robustness in the appendices rather than as an exclusion criterion in primary analyses (see Appendix F), as the exclusion of inattentive respondents may bias the sample as inattentiveness is usually non-random (Alvarez et al., 2019).

3.5 Estimation Strategy

The primary estimand is the Average Marginal Component Effect (AMCE) of each attribute on the probability of supporting the program. We estimate AMCEs using the `cjoint` package in R (Hainmueller, Hopkins, et al., 2014), with standard errors clustered by respondent throughout.

Because the social position attribute has three levels, testing the less eligibility principle requires specifying not one comparison, but two. Poverty is used as the reference category throughout, so the model directly estimates two AMCEs: incarcerated-versus-poverty and welfare-versus-poverty. The incarcerated-versus-poverty AMCE is the direct test of H1 and is what we call the incarceration penalty throughout the paper: a negative estimate means incarcerated applicants receive less support than poor but free applicants. Because welfare applicants already carry the stigma of visible state dependency without incarceration’s legal status, the incarcerated-versus-welfare AMCE—what we call the welfare-referenced incarceration penalty—measures an incarceration penalty, net of that generic dependency stigma. We report the welfare-referenced incarceration penalty alongside the incarceration penalty throughout the results. The model also estimates a third quantity, the welfare-versus-poverty AMCE, which we call the welfare penalty: it captures the stigma attached to welfare receipt independent of incarceration, against which the incarceration penalty’s magnitude can be benchmarked.

We rely on two model specifications. Because the social position attribute is assigned with unequal probabilities, the main specification declares these probabilities using the `makeDesign` option in `cjoint` and excludes the offense severity attribute. This specification is used to estimate the main effects of social position, effort, community belonging, program design, cost-benefit information, and race.

¹⁴This study was approved by [Institution name and IRB protocol number blinded for peer review].

Because offense severity is perfectly collinear with the incarcerated level of the social position attribute—severity is only randomized among incarcerated profiles—hypotheses involving severity require an alternative specification that collapses social position and offense severity into a single five-level attribute: poverty, welfare, incarcerated-robbery, incarcerated-burglary, and incarcerated-shoplifting. The corresponding unconditional assignment probabilities—poverty (1/4), welfare (1/4), incarcerated-robbery (1/6), incarcerated-burglary (1/6), and incarcerated-shoplifting (1/6)—are declared via `makeDesign`.¹⁵

For hypotheses involving interactions between conjoint attributes, we estimate Average Component Interaction Effects (ACIE) by augmenting the relevant specification with the appropriate interaction term. For analyses involving interactions between conjoint attributes and respondent characteristics—including political ideology, incarceration contact, crime victimization, and crime news consumption—we estimate heterogeneous effects using the respondent-level moderator framework in `cjoint`.

Power analyses using the `cjpowR` package (Freitag and Schuessler, 2024) and reported in the PAP, targeting 80% power at a significance level of $\alpha = 0.05$ and assuming an AMCE of 0.05, indicate that our sample provides power approaching 1 for main-effect hypotheses, and 0.89 for the offense severity gradient. For interaction hypotheses involving two binary attributes, expected power is approximately 0.78.

4 Results

Figure 1 presents the Average Marginal Component Effects (AMCE) from the main design (see Appendix E for tables reporting the full models for results reported in this section). Our results show that the logic of less eligibility is clearly in operation, in concordance with H1: support for the program is 7.7 percentage points lower when the hypothetical applicant is incarcerated, compared to when the applicant is below the poverty line ($p < 0.001$)—the incarceration penalty—and 4.9 percentage points lower when compared to applicants currently on welfare ($p < 0.001$), the welfare-referenced incarceration penalty. Our results also reflect the literature on welfare and deservingness (e.g., Rasinski, 1989), showing 2.8 percentage points lower support for the program when the applicant is described as on welfare compared to cases in which the applicant is below the poverty line ($p = 0.004$), the welfare penalty. In Figure C.2 in the Appendix, we show that the incarceration penalty grows monotonically with respondents’ explicit endorsement of less eligibility in a post-experiment question, but is not confined to endorsers: the penalty persists, although more weakly, among respondents who mildly disagree with the principle, and disappears only among those who disagree strongly. This suggests that less eligibility functions as a behavioral norm that operates independently of whether individuals consciously subscribe to it.¹⁶

Among the attributes related to aspects of deservingness, we find strong evidence of both community belonging and effort cues. Consistent with H2A, participants were 9.4 ($p < 0.001$) percentage points more likely to support offering the program when the applicant had roots and belonging in the respondent’s community, compared to when they did not. Consistent with H3A, we find that when the applicant is described as hard-working and reliable, program support is 59.6 ($p < 0.001$) percentage points higher than when the applicant is described as unreliable.

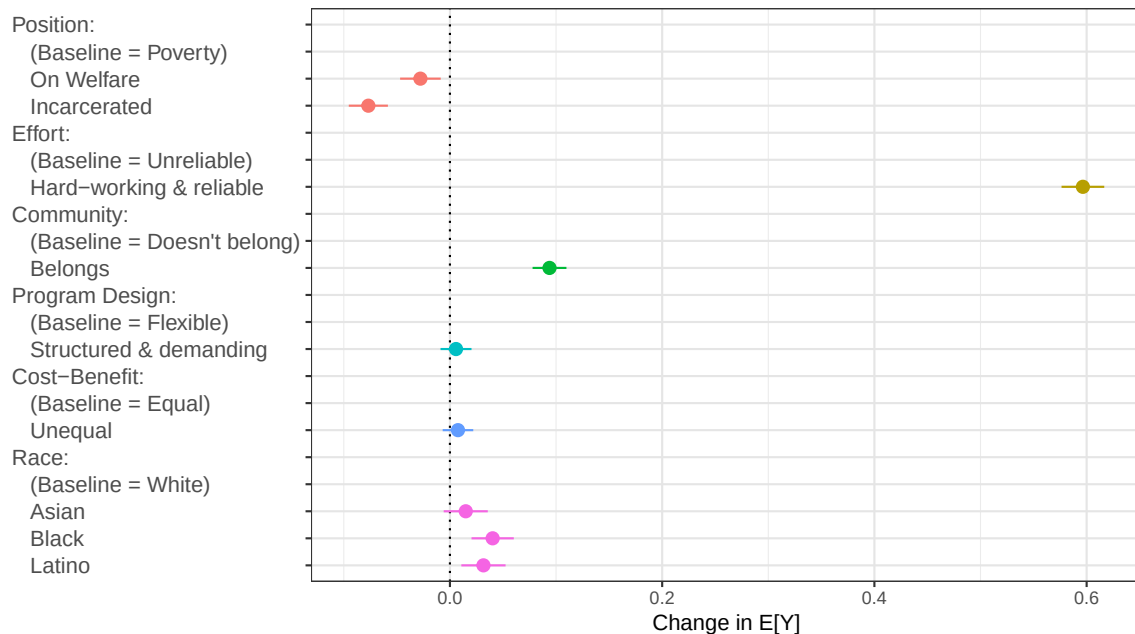
¹⁵The effect of the remaining attributes is the same (up to a small numerical difference) between the two specifications because the attributes were otherwise randomized independently.

¹⁶This parallels a broader pattern in the psychology of punishment and welfare judgment. Carlsmith, Darley, and Robinson, for example, find that lay punishment judgments are driven by just-deserts reasoning even among respondents who state a preference for deterrence-based justifications (Carlsmith et al., 2002). Petersen documents an analogous pattern in welfare attitudes: respondents automatically categorize recipients as lazy or unlucky, and this categorization predicts support independently of the ideological commitments they report holding (Petersen, 2012).

We find no evidence, however, that program design and instrumental cost-benefit information influence support. For hypothesis H4A, we find no difference in support between a program described as structured and demanding versus one described as flexible and supportive ($p = 0.442$). And for hypothesis H5A, we find no difference in program support between a program described as having three-times the benefit for incarcerated individuals versus one that provides equal benefit ($p = 0.303$). It is important to note that these null effects are not the result of low statistical power, as the statistical power to detect even a small effect in these attributes is very high, considering our sample size and number of conjoint tasks.

In terms of race, we find that, relative to White applicants, respondents tended to be 4.0 percentage points more favorable towards applicants described as Black ($p < 0.001$) and 3.2 percentage points more favorable towards applicants described as Latino ($p = 0.003$), but not towards applicants described as Asian ($p = 0.159$). As we discuss below (see Figure 5) and revisit in the discussion, this pattern is concentrated among ideologically Liberal respondents.

Figure 1: Average marginal component effects on choice



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses.

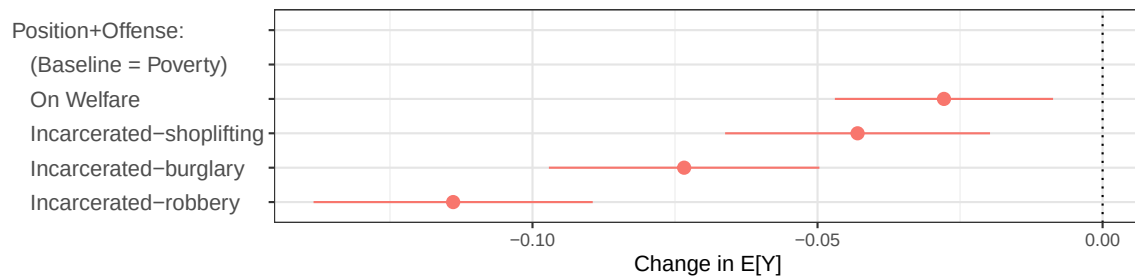
In Figure 2 we present estimates from our alternative specification that includes the offense severity attribute. The results on the attributes of effort, community belonging, program design, cost-benefit information, and race are the same, up to some minor numerical variation (as should be the case given the independent randomization of the attributes), so we focus exclusively on the severity results.¹⁷ Our findings show a clear offense severity ordering: compared to when the applicant is described as on the poverty line, support for the program is 4.3 percentage points lower when the applicant is described as being imprisoned due to shoplifting, 7.3 percentage points lower when described as having committed burglary of an unoccupied home, and 11.4 percentage points lower when described as having committed armed robbery. The differences between all three categories of offense severity are statistically significant.¹⁸ Interestingly, the evaluation of profiles in which

¹⁷For full estimates from this alternative specification, see the Appendix Figure C.1.

¹⁸The difference between the shoplifting level and the burglary level has a one-sided p-value of 0.011; and the difference between the burglary level and the armed-robbery level a one-sided p-value of 0.001.

the applicant is in prison due to shoplifting are similar to those in which the applicant is described as being currently on welfare ($p = 0.203$). This equivalence suggests that the stigma attached to welfare recipients is roughly commensurate with that of the least serious criminal offense. That is, in the public's moral accounting, being on welfare already places an applicant near the bottom of the deservingness hierarchy that incarceration merely deepens. (Or, put differently, respondents consider being on welfare as morally equivalent to being incarcerated for shoplifting.)

Figure 2: Offense severity grading



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Estimates come from a model that includes all attributes. Figure with all attributes in Appendix Figure C.1.

4.1 Experimental Moderators of Less Eligibility

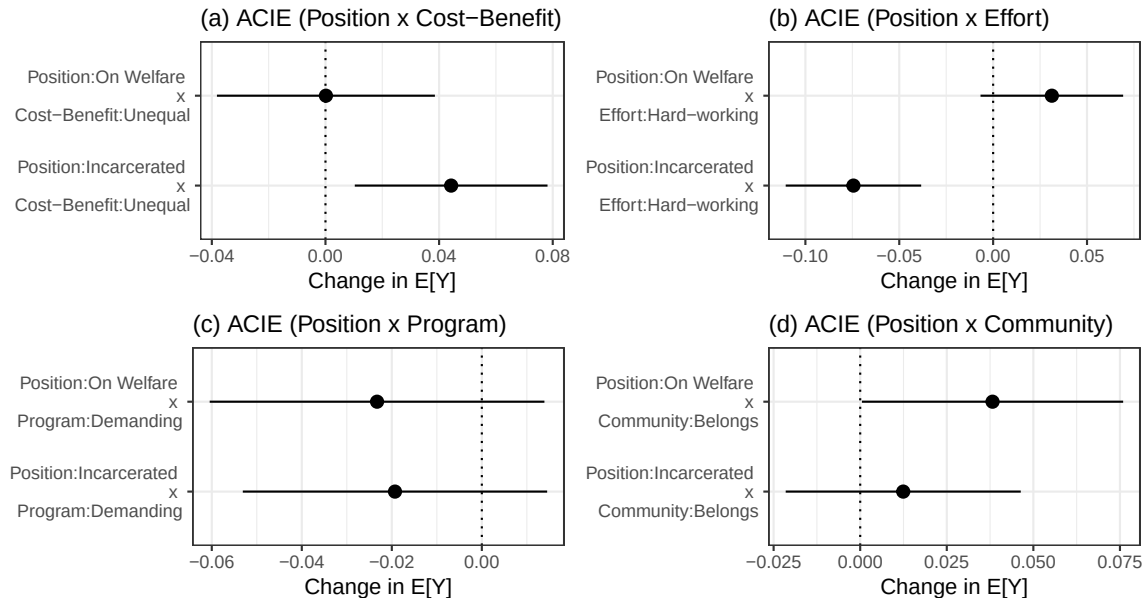
Throughout this section, we report Average Component Interaction Effects (ACIEs) for interactions with the social position attribute, which we use to test hypotheses H2B, H3B, H4B, and H5B. Each ACIE is a difference-in-differences: it compares how much a given attribute changes support for incarcerated applicants against how much the same attribute changes support for free-poor applicants. A positive ACIE narrows the incarceration penalty; a negative ACIE widens it. These estimates describe movement in the gap itself, not the total effect of an attribute on either group alone—those total effects, where informative, are reported separately in footnotes.

Panel (a) in Figure 3 evaluates the moral dominance hypothesis (H5B). While cost-benefit information has no main effect on program support (Figure 1), it reduces the incarceration penalty by 4.4 percentage points ($p = 0.011$)—roughly half—without fully eliminating it, consistent with H5B. The same information does not shift support for welfare applicants. When the incarcerated applicant is shown to benefit three times as much as a comparable low-income applicant—not merely as much—respondents update their support for that applicant. The update, however, is only partial: a residual incarceration penalty remains.

The remaining panels of Figure 3 evaluate the extent to which the less eligibility effect we document is moderated by other key conjoint attributes. Contrary to expectations in H3B, we find that describing an applicant as hard-working and reliable *exacerbates* the incarceration penalty by 7.4 percentage points ($p < 0.001$). This widening is not driven by incarcerated applicants being penalized for appearing reliable; rather, it reflects that reliability simply helps free-poor applicants more than it helps incarcerated ones, so the gap between the two groups grows even as both gain from the trait.¹⁹

¹⁹The predicted probability of supporting the program for a person below the poverty line increases from approximately 29.7 percentage points when the applicant is described as unreliable to approximately 92.4 percentage points when the applicants is described as reliable, an increase of 62.7 percentage points. For incarcerated applicants, support instead increases from 25.7 to 81.0 percentage points; a large but relatively smaller increase. Hence, the incarcerated penalty increases with the effort attribute.

Figure 3: ACIEs: attributes that moderate less eligibility



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Each ACIE comes from a separate model that includes an interaction term between social position and the corresponding attribute (and linear terms for the remaining attributes in the conjoint). See AppendixE for tables with the full models.

Panels (c) and (d) show that the incarceration penalty does not vary with program structure (H4B), nor with applicant's belonging to the respondents' community (H2B). We do find, however, that the welfare penalty is reduced when the applicant belongs to the respondent's community ($p = 0.047$), although the effect is not robust when using the scale outcome (see Appendix D).

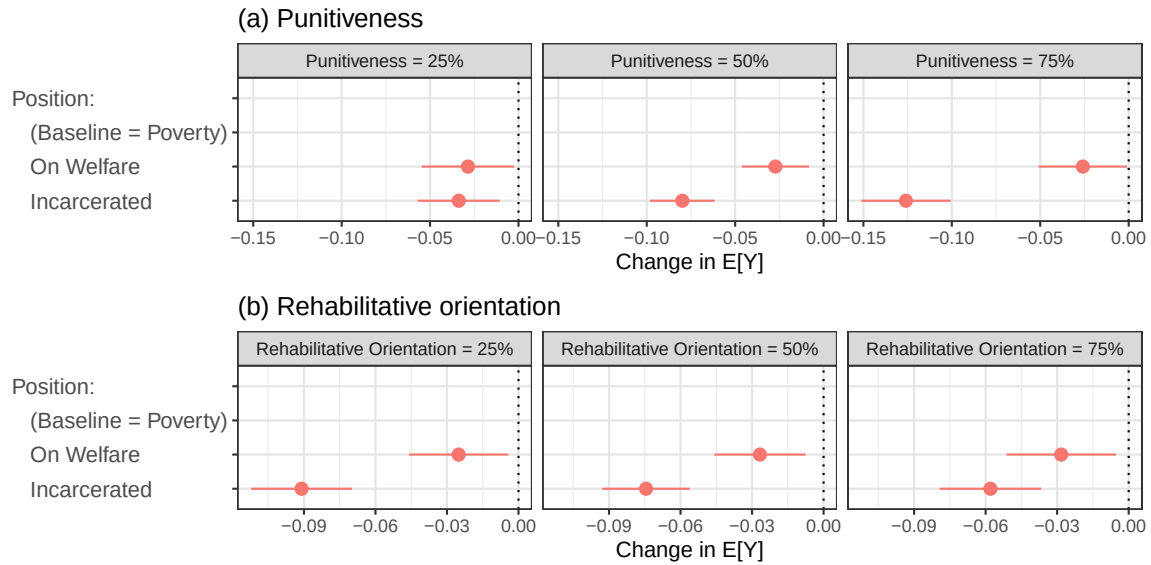
4.2 Respondent-Level Moderators of Less Eligibility

We next examine heterogeneous effects pre-registered in our pre-analysis plan, exploring how the incarceration penalty varies with respondent ideology, incarceration contact, victimization, and punitive attitudes. Appendix C reports additional heterogeneous effects.

We begin with the most predictable heterogeneous effect. As expected, more punitive respondents impose a substantially larger incarceration penalty ($p < 0.001$), while the welfare penalty does not vary with punitiveness—confirming that the effect is incarceration-specific rather than a generalized hostility to program beneficiaries. Rehabilitative orientation follows the mirror image: respondents with weaker rehabilitative attitudes apply a larger incarceration penalty ($p = 0.003$), again with no corresponding shift for welfare applicants (Figure 4). These patterns are reassuring benchmarks rather than substantive findings.

We turn now to moderators that are theoretically more open: ideology, incarceration contact, and victimization. Figure 5 presents heterogeneous results based on respondent's ideology, collapsing our 7-point ideological self-placement scale into Liberal, Moderate, and Conservative. Our results show that there is a larger incarceration penalty and a steeper offense grading among Conservatives compared to Liberals, although virtually no difference between Conservatives and Moderates. On the other hand, we find no ideological heterogeneity in the evaluation of the program design component,

Figure 4: Heterogeneous effects by respondent's punitiveness and rehabilitative orientations



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Each panel comes from a separate model that includes an interaction term between social position and the respondent's punitiveness (panel a) or rehabilitative orientation (panel b), plus linear terms for the remaining attributes in the conjoint. Estimates are shown at the 25th, 50th, and 75th percentile of each respondent-level index. See Appendix E for tables with the full models.

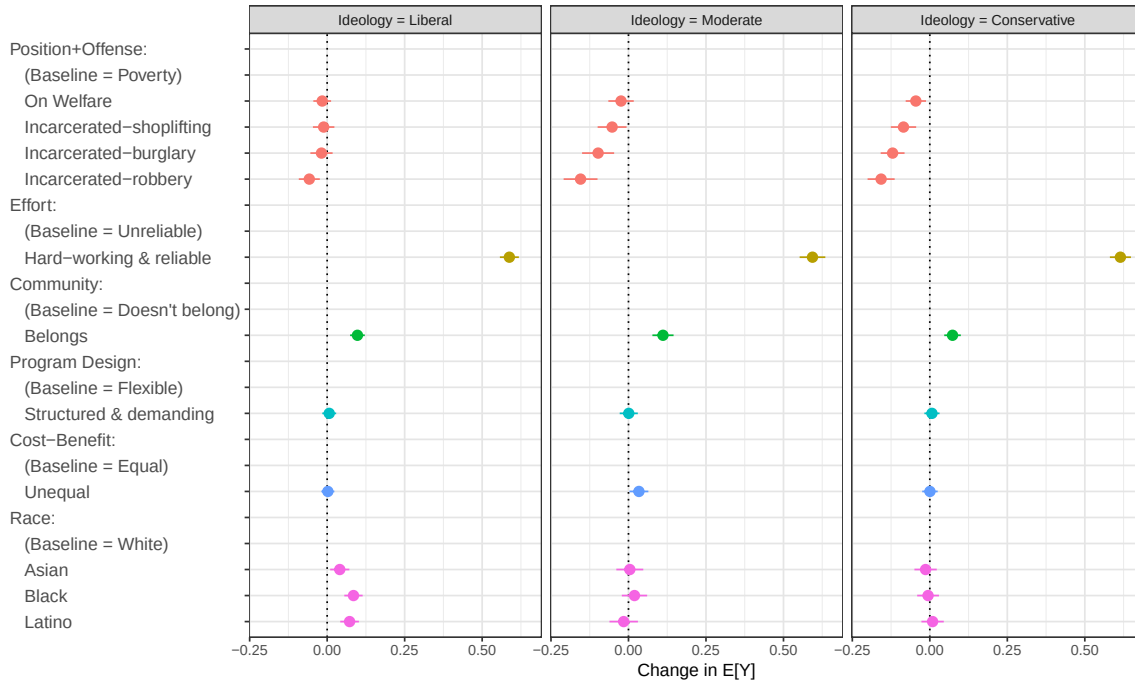
finding no effect among the three ideological groups. Similarly, we find that the null effect of the program cost-benefit attribute is observed among both Liberal and Conservative respondents, although benefit information does increase support for the program among respondents who identify as Moderates ($p = 0.029$).

Applicant race interacts with respondent ideology in a specific way. While there is a gap between Liberals and Conservatives in program support for non-White applicants, particularly Black applicants, this gap is explained by Liberal respondents being more favorable towards Black and Latino applicants than towards White applicants, and not by Conservative respondents being less favorable towards those groups. In fact, for Conservative respondents, the race attribute is not statistically significant.²⁰

The next two moderators operate through a different channel: direct personal experience with the carceral state. Respondents who have been incarcerated, or who are close to someone who has, may bring different priors to these evaluations—as may those who have been victims of crime. As reported in the Appendix (Figure C.7), heterogeneous effects by respondents' own incarceration experience and incarceration of a family member or close friend both tell a similar story. The incarceration penalty persists even among those who have been incarcerated or have had a loved one incarcerated. While the incarceration penalty is slightly smaller among those with personal contact with the carceral state—1.1 percentage points smaller among those formerly incarcerated and 2.2

²⁰When using the 7-point ideology scale, we find a gradient of favorability towards non-White applicants running along the left side of the scale, with respondents identifying as Very Liberal being the most favorable towards Black and Latino applicants, decreasing to no statistically significant favorability among Moderates. Among Liberals, the favorability towards Black and Latino applicants is present among respondents who are White, respondents who are Black, and also respondents who are of other race or ethnicity. We do not find a favorability gradient among Conservatives, with Somewhat Conservative and Very Conservative respondents having similarly null (and non-significant) race attribute effects.

Figure 5: Heterogeneous effects by respondent's ideological orientation



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Estimates come from a single model that interacts respondent ideology with every profile attribute. Panels show the conditional AMCE of each attribute among respondents in each ideology category.

percentage points among those with a loved one who has been incarcerated—this difference is not statistically significant in either case. Victimization (Figure C.6 in the Appendix) shows the same pattern. Respondents who have been victims of crime are 2.6 percentage points less supportive of the program when the applicant is incarcerated, but this difference is not statistically significant ($p = 0.328$), nor are the effects for property crime victims ($p = 0.417$) or violent crime victims ($p = 0.686$) separately.

The pattern across respondent-level moderators is itself a finding. Punitiveness and political ideology moderate the incarceration penalty in expected directions. But neither direct personal exposure to incarceration—as a formerly incarcerated person, as someone close to one, or as a crime victim—meaningfully alters the core result. The penalty does not depend on personal investment in punishment to operate. If less eligibility was simply the expression of personal punitiveness, we would expect it to concentrate among the most personally aggrieved respondents. It does not.

5 Discussion

Two findings support this study's central claim that less eligibility is a real, causally identifiable force in public opinion. The first finding is that the American public applies a meaningful and causally identifiable penalty to incarcerated applicants for the same government program—a penalty that survives across offense types, persists even when the applicant is described as hard-working and community-embedded, and is only partially offset by favorable cost-benefit evidence. The second finding is that the three mechanisms the moral-cultural account of less eligibility bundles together—effort, suffering, and belonging—do not behave alike. The visible suffering account, which predicts

that program design should shape evaluations by signaling how punitive incarceration remains, finds no support: program design is inert both as a main effect and as a moderator of the incarceration penalty. The community belonging and meritocratic desert accounts are both active, but in ways that complicate their policy implications—and, in the case of merit, in a direction that inverts the prediction that favorable moral cues should narrow the incarceration penalty. We address the implications of each finding in turn.

Beginning with less eligibility itself, we find that an incarceration penalty exists. Respondents are resistant to providing a social service program to incarcerated individuals, as seen in evaluations of applicants with various attributes. Their resistance has a direct, monotonic relationship with explicit endorsement of the less eligibility principle: “People in prison should not have access to amenities or programs that are unavailable to law-abiding citizens who are struggling financially.” That relationship does not stop at the boundary of professed belief, however: respondents who mildly disagree with the principle still apply a modest incarceration penalty, even though they reject it when asked directly. Only among respondents who explicitly and strongly disagree does the penalty disappear entirely. This pattern indicates that less eligibility operates less as a considered belief and more as a background norm that shapes judgment for all but its most committed opponents.

But support for less eligibility is not uniform and cultural cues shape how it manifests. When we decompose the various components of the moral-cultural account of less eligibility, we find they are not equally active. Contrary to expectations, program design features that touch on the degree of suffering imposed are inconsequential, both as a main effect and as a moderator of the incarceration penalty. This null result is inconsistent with the visible suffering account, which predicts that program design should shape evaluations by signaling whether incarceration remains sufficiently punitive. In our vignette, the suffering involved adherence to a rigorous schedule and frequent testing, as opposed to more physically painful acts. This design may have been too subtle to signal a meaningful form of suffering or hardship. We therefore interpret this finding cautiously: it may reflect a program-design manipulation too subtle to move opinion, or it may indicate that the visible-suffering mechanism simply does not operate in this context.

In contrast, community belonging raises support for everyone (those incarcerated, on welfare, or in poverty); it does not, however, close the gap between incarcerated and free-poor applicants. An incarcerated person described as community-embedded gets the same belonging boost as anyone else, but the incarceration penalty doesn’t move. The contrast with welfare is instructive: community belonging *does attenuate* the welfare penalty, suggesting that for welfare recipients the deservingness evaluation is open to modification through membership signals. For incarcerated applicants, that does not appear to be the case.

The most counterintuitive result in the study concerns attributions of effort. Describing an incarcerated applicant as hard-working and reliable does not reduce the incarceration penalty—it appears to amplify it. This does not mean effort hurts incarcerated applicants in absolute terms: being described as hard-working raises support for all applicants substantially, including incarcerated ones.²¹ The paradox is relative. The gap between incarcerated and free-poor applicants is larger when both are described as hard-working than when both are described as unreliable. Being described as hard working benefits everyone—it just benefits prisoners significantly less.

Two mechanisms, not mutually exclusive, potentially explain the pattern. The first follows from the fundamental attribution error discussed earlier (Ross, 1977): evidence of effort invites an inference about capacity, and for incarcerated applicants that inference may cut against them. A hard-working prisoner might be seen as someone who had the ability to succeed through legitimate means but chose not to—dampening support. The second mechanism is a property of the ACIE itself. Because the estimate is relative, an effort cue does more than describe the incarcerated applicant. It also speaks to the free-poor applicant’s own claim to desert. Effort is *the* signature trait

²¹This is in line with other research on deservingness heuristics; see Petersen (2012).

of the trope of “the deserving poor” (Hochschild, 2016). Crediting the incarcerated applicant with effort activates that trope for the comparison case as well—and, as a consequence, the free-poor applicant may gain more from it.

If moral desert cues cannot close the gap—and in the case of effort, actively widen it—cost-benefit information stands apart. It is the only lever we test that significantly narrows the incarceration-specific gap (Figure 3, panel a), and the size of that narrowing is worth benchmarking against what the vignette actually tells respondents. The unequal-benefit condition does not merely assert that the program is effective; it states, in the same sentence, that the incarcerated participant benefits three times as much from the identical program as the comparable free-poor applicant. Taken at face value, that is an efficiency case for funding the incarcerated applicant *over* the free-poor one. A purely consequentialist respondent, told that the same investment does three times the good in one recipient’s hands, should shift support toward or past parity. Respondents do not. The gap narrows by roughly half and then holds; the incarcerated applicant still trails a comparison case the vignette has just described as the less efficient use of the same investment. That the update is only partial, given how favorable the instrumental case is, is more informative than the fact that an update occurs at all. Among the desert cues we test, cost-benefit evidence is the only one that moves the gap in the right direction. Even the best lever available, it appears, cannot fully dislodge the boundary it moves against.

Likewise, the incarceration penalty persists even among those with the closest ties to the carceral state. Formerly incarcerated respondents apply a penalty similar in magnitude to respondents who have never been incarcerated—direct personal experience does not detectably soften the core result. Respondents with an incarcerated family member or close friend register a somewhat lower point estimate, a pattern that would be consistent with a humanization account in which social closeness makes the comparison case less abstract. But even that shift is modest and falls short of conventional statistical significance. Similarly, victim identity offers no purchase on the penalty at all. Taken together, these patterns describe a boundary largely indifferent to personal exposure to crime or incarceration.

These findings carry policy implications, but drawing them out requires keeping two different quantities apart: how much an attribute raises support for incarcerated applicants in absolute terms, and how much it narrows the incarceration penalty—the gap between incarcerated and free-poor applicants that is the actual object of the less eligibility comparison. The two do not move together, and the distance between them turns out to be the central policy lesson of the study. Some levers move neither quantity. Program design leaves both untouched: making a program appear more demanding, or more consistent with the visible image of punishment, neither raises support nor narrows the gap. Reform energy spent on this aspect of program design looks to be wasted. Cost-benefit evidence does better, but only by half: it narrows the penalty without closing it, chipping at a moral constraint rather than dissolving it.

Two other levers tell a more encouraging story, at least at first. Community belonging raises support for incarcerated applicants by roughly the same margin it raises support for everyone else (about 9 percentage points overall), and effort framing raises it far more: describing an incarcerated applicant as hard-working and reliable increases support for that applicant by 55.3 percentage points on its own. Public sentiment toward incarcerated people, in other words, is not fixed. It moves, and it moves substantially, in response to ordinary information about who a person is. But less eligibility was never a claim about the absolute level of sympathy the public extends to incarcerated people. In its original formulation the principle is explicitly comparative—a person subject to punishment should not be *better off than* the poorest free laborer. The gap between these groups may persist in radically different societies or historical eras, regardless of the conditions that free laborers encounter (Bonnet, 2025). The relevant test, then, is not whether an attribute can make incarcerated applicants more popular. It is whether that attribute can bring them level with the free-poor comparison case the principle invokes. On that test, both levers come up short, for different reasons. Community belonging raises the floor for everyone at once, so the gap it might have closed simply rises along with

it. Effort framing fails more instructively. Since a hard-working person near the poverty line is the paradigmatic deserving poor (Gilens, 1999), crediting an incarcerated applicant with the same virtue does not blunt the comparison. It actually sharpens it, by making the free-poor *more* sympathetic.

Why should the comparative gap matter for policy purposes, rather than the absolute level of support a program enjoys on its own? Because the objections that actually stop prison programming are almost never phrased in absolute terms. Critics do not argue that a job-training program lacks support in isolation; they argue that it is unfair for an incarcerated person to receive a benefit that a comparable, law-abiding person struggling to get by does not—the same comparison Bentham built into the principle, and the same one that resurfaces in no-frills amendments and the penal populism of local sheriffs a century and a half later.²² That objection remains available however much absolute goodwill a program accumulates, so long as the incarcerated recipient can still be placed level with, or ahead of, the free-poor comparison case. This is why the persistence of the gap, and not the movability of absolute support, is the finding that should concern reformers. Our results point toward the boundary of moral exclusion itself—the prior question of who counts as deserving of public concern at all—as the more precise target.²³ This most clearly arises in disenfranchisement laws: like Florida’s Amendment 4, restoration schemes typically ask whether a person has served long enough, paid enough, or behaved well enough to earn their place back (Morris, 2021).

More generally, defenses of prison programs can be split along this line: some work inside the less-eligibility comparison, others try to step outside it. Appeals to effort and belonging are inside strategies. They, in effect, accept the comparison and try to win it by improving the incarcerated applicant’s side. However, that approach has a ceiling—as we’ve seen, any gain on one side tends to be matched on the other. Outside strategies avoid the comparison altogether. Duty-of-care arguments, for example, do not compare those incarcerated to anyone. They ask who is responsible for a life the state itself controls (Dolovich, 2009). Rights-based claims to basic entitlements work the same way, fixing what is owed independent of any comparison class.²⁴ Instrumental arguments sit in a third position, adjacent to the comparison rather than inside or outside it. Cost-benefit evidence does not contest desert; it attempts to substitute an efficiency calculation for an explicitly moral one, with limited success. This suggests a key avenue for future study: not toward assessing better versions of inside strategies, which our results suggest are structurally limited, but toward testing whether legitimacy for incarcerated applicants can be won on alternative grounds.

Our study has limitations. First, the study examines a single program type—one year of college courses and job counseling. Education and vocational training are arguably the most sympathetic case for incarcerated applicants, programs whose rehabilitative logic is widely accepted and whose costs are relatively modest. If the incarceration penalty persists even here, it almost certainly persists elsewhere; but whether it is larger for less sympathetic program types—cash transfers, housing assistance, early release—is an open question. The estimates reported here are plausibly conservative with respect to the range of policy contexts where less eligibility operates.

Second, our race results resist a simple discrimination reading. We had anticipated, in line with the literature on racial attitudes, that Conservative respondents would penalize non-White applicants; instead, the favorability gap is generated by Liberals rating Black and Latino applicants above White ones, with no significant race effect among Conservatives. Our design cannot tell whether this reflects genuine attitudes about structural disadvantage or a social-desirability response to a more salient race cue than real-world decisions would present, so the pattern should be read as descriptive rather than as evidence for or against the presence of prejudice.²⁵

²²As a case study, consider former Sheriff Joe Arpaio of Maricopa County, AZ. See Finnegan (2009).

²³On the “democratic boundary problem,” see Whelan (1983).

²⁴Other non-comparative principles could ground similar arguments. Judith Resnik’s recent “anti-ruination principle” holds that a government committed to equal citizenship cannot set out to ruin a person’s life chances, regardless of what anyone else is owed (Resnik, 2025). These arguments also split along a further axis, between structural claims that indict mass incarceration as a systemic wrong (Gottschalk, 2006) and individual claims, like the duty-of-care and rights-based arguments above, about what is owed to a particular person.

²⁵A parallel test of racialized offense-severity effects was null but somewhat underpowered.

Third, the effort paradox is the most counterintuitive finding in the study and the one most in need of replication and formal theorization. We proposed two potential mechanisms above, but the present design cannot distinguish between them. Future work should develop experimental contrasts capable of separating those accounts, and should test whether the pattern replicates across program types, applicant characteristics, and samples before it is taken as a stable feature of how merit framing interacts with the less eligibility comparison. One such contrast follows from a distinction welfare-attitudes research already draws between perceived recipient motivation and work-requirement policy design (Petersen, 2012; Shaw and R. Shapiro, 2002). Our effort manipulation tested only the former, a description rather than a verified requirement; future work should test whether verification itself can close the incarceration penalty. This is a question about persuasion, not policy: work requirements carry well-documented costs, including administrative burden and the exclusion of motivated applicants who cannot meet them.

6 Conclusion

The principle of less eligibility has long been invoked to explain public resistance to rehabilitative prison programming, yet it has functioned more as a discursive shorthand than a testable claim. This study provides the first direct experimental evidence that the principle operates as a causal force in public opinion: incarcerated applicants face a substantial and robust penalty in program eligibility evaluations that cannot be accounted for by offense severity, program design, community membership, or cost-benefit evidence. The penalty is moral rather than instrumental in character—anchored in a comparative judgment about desert that cost-benefit evidence can narrow but not dislodge.

Disaggregating the moral-cultural account of less eligibility reveals that not all of its components carry equal weight—and that one carries costs the prison reform literature has not anticipated. Program design effects are null: how demanding or structured a program appears leaves public opinion unchanged. Community belonging raises support for incarcerated applicants but does not close the gap with the free poor. And meritocratic desert framing—the appeal to a hard-working, rehabilitation-oriented applicant that advocates most commonly deploy—does not reduce the incarceration penalty. It actually amplifies it. When an incarcerated applicant is described as hard-working, the comparison case is too, and that comparison is precisely the one less eligibility assessments make most unfavorable.

None of this means public opinion is fixed. Sentiment toward incarcerated applicants moves, and moves substantially, in response to ordinary information—a hard-working incarcerated applicant is far more popular than an unreliable one, and cost-benefit evidence narrows the incarceration penalty by nearly half. What does not appear to move is the comparative judgment less eligibility names: the sense that whatever an incarcerated person has earned, it should still fall short of what a comparable person outside prison walls has earned.

References

- Alvarez, R. Michael et al. (2019). “Paying Attention to Inattentive Survey Respondents”. In: *Political Analysis* 27.2, pp. 145–162. DOI: doi:10.1017/pan.2018.57.
- Bentham, Jeremy (1791). *Panopticon; or, the Inspection-House: Containing the Idea of a New Principle of Construction Applicable to Any Sort of Establishment, in Which Persons of Any Description Are to Be Kept under Inspection*. London: T. Payne.
- (1843). “Principles of Penal Law”. In: *The Works of Jeremy Bentham*. Ed. by John Bowring. Vol. 1. Edinburgh: William Tait, pp. 365–580.
- Bonnet, François (Dec. 2025). “Less Eligibility, Welfare, and Punishment: The Econometric Evidence”. In: *Crime and Justice* 54, pp. 329–413. DOI: 10.1086/735664.
- Carlsmith, Kevin M., John M. Darley, and Paul H. Robinson (2002). “Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment”. In: *Journal of Personality and Social Psychology* 83.2, pp. 284–299.
- Chen, M. Keith and Jesse M. Shapiro (2007). “Do Harsher Prison Conditions Reduce Recidivism? A Discontinuity-Based Approach”. In: *American Law and Economics Review* 9.1, pp. 1–29. DOI: 10.1093/aler/ahm006.
- Cholbi, Michael J (2002). “A felon’s right to vote”. In: *Law and Philosophy* 21.4/5, pp. 543–565.
- De Giorgi, Alessandro (2006). *Re-Thinking the Political Economy of Punishment: Perspectives on Post-Fordism and Penal Politics*. Advances in Criminology. Aldershot, England; Burlington, VT: Ashgate. ISBN: 0754626105 (alk. paper).
- Dilts, Andrew (2014). *Punishment and Inclusion*. New York, NY: Fordham University Press.
- Dolovich, Sharon (2009). “Cruelty, Prison Conditions, and the Eighth Amendment”. In: *New York University Law Review* 84.4.
- Drago, Francesco, Roberto Galbiati, and Pietro Vertova (2011). “Prison Conditions and Recidivism”. In: *American Law and Economics Review* 13.1, pp. 103–130. DOI: 10.1093/aler/ahq024.
- Enns, Peter K (2016). *Incarceration Nation*. Cambridge University Press.
- Feldman, Stanley and John Zaller (Feb. 1992). “The Political Culture of Ambivalence: Ideological Responses to the Welfare State”. In: *American Journal of Political Science* 36.1, pp. 268–307.
- Finnegan, William (2009). “Sheriff Joe”. In: *The New Yorker*.
- Freitag, Markus and Julian Schuessler (2024). *cjpowR: A Priori Power Analyses for Conjoint Experiments*. R Package. URL: <https://github.com/m-freitag/cjpowR>.
- Garland, David (1990). *Punishment and Modern Society: A Study in Social Theory*. Chicago: University of Chicago Press.
- (2001). *The Culture of Control: Crime and social order in contemporary society*. Oxford University Press.
- Gilens, Martin (1999). *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*. University of Chicago Press.
- Gottschalk, Marie (2006). *The Prison and the Gallows: The Politics of Mass Incarceration in America*. Cambridge University Press.
- Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto (2015). “Validating Vignette and Conjoint Survey Experiments against Real-World Behavior”. In: *Proceedings of the National Academy of Sciences* 112.8, pp. 2395–2400. DOI: 10.1073/pnas.1416587112.
- Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto (2014). “Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments”. In: *Political Analysis* 22.1, pp. 1–30. DOI: 10.1093/pan/mpt024.
- Harcourt, Bernard E. (2011). *The Illusion of Free Markets: Punishment and the Myth of Natural Order*. Cambridge, MA: Harvard University Press.
- Hochschild, Arlie Russell (2016). *Strangers in Their Own Land: Anger and Mourning on the American Right*. New York: The New Press.
- Horiuchi, Yusaku, Zachary D. Markovich, and Teppei Yamamoto (2022). “Does Conjoint Analysis Mitigate Social Desirability Bias?” In: *Political Analysis* 30.4, pp. 535–549. DOI: 10.1017/pan.2021.30.

- Jensen, Carsten and Michael Bang Petersen (2017). “The Deservingness Heuristic and the Politics of Health Care”. In: *American Journal of Political Science* 61.1, pp. 68–83.
- Johnston, Travis and Kevin Wozniak (2021a). “Framing Effects and Group Differences in Public Opinion about Prison Pell Grants”. In: *PS: Political Science & Politics* 54.3, pp. 411–415.
- (2021b). “Public Opinion and the Politics of Collateral Consequence Policies”. In: *Punishment & Society* 23.2, pp. 190–217. DOI: 10.1177/1462474520941942.
- Labotka, Lori (2021). “The less eligible eaters: Calorie counts, no-frills, and vending machines in prison”. In: *Signs and Society* 9.1, pp. 61–88.
- Lerman, Amy and Vesla Weaver (2014). *Arresting Citizenship: The Democratic Consequences of American Crime Control*. Chicago: University of Chicago Press.
- Manza, Jeff and Christopher Uggen (2006). *Locked Out: Felon Disenfranchisement and American Democracy*. Oxford University Press.
- Mastrobuoni, Giovanni and Daniele Terlizzese (2022). “Leave the Door Open? Prison Conditions and Recidivism”. In: *American Economic Journal: Applied Economics* 14.4, pp. 200–233. DOI: 10.1257/app.20190038.
- Michalowski, Raymond J. and Susan M. Carlson (1999). “Unemployment, Imprisonment, and Social Structures of Accumulation: Historical Contingency in the Rusche-Kircheimer Hypothesis”. In: *Criminology : an interdisciplinary journal* 37.2, pp. 217–250. DOI: 10.1111/j.1745-9125.1999.tb00485.x.
- Miller, Lisa (2016). *The Myth of Mob Rule: Violent Crime and Democratic Politics*. Oxford: Oxford University Press.
- Morris, Kevin (2021). “Turnout and Amendment Four: Mobilizing Eligible Voters Close to Formerly Incarcerated Floridians”. In: *American Political Science Review* 115.3, pp. 805–820. DOI: 10.1017/S0003055421000253.
- Oorschot, Wim van (2000). “Who Should Get What, and Why? On Deservingness Criteria and the Conditionality of Solidarity among the Public”. In: *Policy & Politics* 28.1, pp. 33–48. DOI: 10.1332/0305573002500811.
- Peffley, Mark and Jon Hurwitz (2007). “Persuasion and Resistance: Race and the Death Penalty in America”. In: *American Journal of Political Science* 51.4, pp. 996–1012. DOI: 10.1111/j.1540-5907.2007.00293.x.
- Petersen, Michael Bang (2012). “Social Welfare as Small-Scale Help: Evolutionary Psychology and the Deservingness Heuristic”. In: *American Journal of Political Science* 56.1, pp. 1–16. DOI: 10.1111/j.1540-5907.2011.00545.x.
- Peterson, Iver (July 10, 1995). “Cutting Down on Amenities To Achieve No-Frills Jails”. In: *The New York Times*. URL: <https://www.nytimes.com/1995/07/10/nyregion/cutting-down-on-amenities-to-achieve-no-frills-jails.html>.
- Pickett, Justin T (2019). “Public Opinion and Criminal Justice Policy: Theory and Research”. In: *Annual Review of Criminology* 2.1, pp. 405–428.
- Piven, Frances Fox and Richard A. Cloward (1993). *Regulating the Poor: The Functions of Public Welfare*. Updated. New York, NY: Vintage Books.
- Ramirez, Mark D. (2013). “Punitive Sentiment”. In: *Criminology : an interdisciplinary journal* 51.2, pp. 329–364. DOI: 10.1111/1745-9125.12007.
- Rasinski, Kenneth A (1989). “The effect of question wording on public support for government spending”. In: *Public Opinion Quarterly* 53.3, pp. 388–394.
- Resnik, Judith (2025). *Impermissible Punishments: How Prison Became a Problem for Democracy*. University of Chicago Press.
- Robey, Jason P., Michael Massoglia, and Michael T. Light (2023). “A Generational Shift: Race and the Declining Lifetime Risk of Imprisonment”. In: *Demography* 60.4, pp. 977–1003. DOI: 10.1215/00703370-10863378.
- Robinson, Paul H., Robert Kurzban, and Owen D. Jones (2007). “The Origins of Shared Intuitions of Justice”. In: *Vanderbilt Law Review* 60.6, pp. 1633–1688.
- Roehrkasse, Alexander F. and Christopher Wildeman (Dec. 2022). “Lifetime Risk of Imprisonment in the United States Remains High and Starkly Unequal”. In: *Science Advances* 8.48, eabo3395. DOI: 10.1126/sciadv.abo3395.

- Ross, Lee (1977). “The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process”. In: *Advances in Experimental Social Psychology*. Ed. by Leonard Berkowitz. Vol. 10. New York: Academic Press, pp. 173–220.
- Rusche, G., O. Kirchheimer, and D. Melossi (2003). *Punishment and Social Structure*. Transaction Pub. ISBN: 0-7658-0921-4.
- Shaw, Greg and Robert Shapiro (2002). “The Polls-Trends: Poverty and Public Assistance”. In: *Public Opinion Quarterly* 66.1, pp. 105–128. DOI: 10.1086/338741.
- Sieh, Edward W. (1989). “Less Eligibility: The Upper Limits of Penal Policy”. In: *Criminal Justice Policy Review* 3.2, pp. 159–183. DOI: 10.1177/088740348900300204.
- Skitka, Linda J. and Philip E. Tetlock (1993). “Providing Public Assistance: Cognitive and Motivational Processes Underlying Liberal and Conservative Policy Preferences”. In: *Journal of Personality and Social Psychology* 65.6, pp. 1205–1223.
- Soss, Joe and Vesla Weaver (2017). “Police Are Our Government: Politics, Political Science, and the Policing of Race–Class Subjugated Communities”. In: *Annual Review of Political Science* 20, pp. 565–591.
- Unnever, James D. and Francis T. Cullen (2009). “Empathetic Identification and Punitiveness”. In: *Theoretical Criminology* 13.3, pp. 283–312. DOI: 10.1177/1362480609336495.
- Whelan, Frederick G (1983). “Prologue: Democratic theory and the boundary problem”. In: *Nomos* 25, pp. 13–47.
- Whitman, James Q (2003). *Harsh Justice: Criminal punishment and the widening divide between America and Europe*. Oxford University Press.
- Wolfgang, Marvin E. et al. (1985). *The National Survey of Crime Severity*. Tech. rep. NCJ-96017. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Zimring, Franklin E. (2020). *The Insidious Momentum of American Mass Incarceration*. New York: Oxford University Press.

Appendices for Incarceration and Deservingness: Experimental Evidence on Less Eligibility

A Conjoint Vignette and Attribute Values

The following is an example of a vignette for an incarcerated participant:

Your state is considering a new publicly funded program that provides one year of free college-level courses and job counseling.

The program is flexible and supportive. Participants set their own pace, learning in a comfortable classroom environment with tablets and lounge-style seating, and receive individualized guidance without strict deadlines. A nonpartisan state audit found that this program increases employment, reduces crime, and saves taxpayer money ---with equal benefit for incarcerated participants and similar low-income residents who are not incarcerated and completed the same program.

Demand for this program far exceeds available funding. Some applicants will be accepted but many others will be rejected. The state must decide who to prioritize.

Consider the following applicant: Joe is a 28-year-old Asian man. He has a reputation for being unreliable and unmotivated: he avoids tasks when he can, often fails to follow through on basic responsibilities, and rarely puts in consistent effort. He belongs to your broader community. He has roots in your area, shares the values and experiences of people in your area, and has personal ties to people who live there. He is currently incarcerated in a state prison with a prior conviction for armed robbery of a store.

The following list contains the full text of each attribute level:

- Social Position

[Poverty] He is currently living near the poverty line.

[On Welfare] He is currently on welfare.

[Incarcerated] He is currently incarcerated in a state prison.

- Offense Severity

[shoplifting] with a prior conviction for shoplifting merchandise worth \$500.

[burglary] with a prior conviction for burglary of an unoccupied home.

[robbery] with a prior conviction for armed robbery of a store.

- Effort

[Hard-working & reliable] He has a reputation for being dependable and hard-working: he takes on tasks without complaint, follows through on responsibilities, and is known for putting in steady effort day to day.

[Unreliable] He has a reputation for being unreliable and unmotivated: he avoids tasks when he can, often fails to follow through on basic responsibilities, and rarely puts in consistent effort.

- Community Belonging

[Belongs] He belongs to your broader community. He has roots in your area, shares the values and experiences of people in your area, and has personal ties to people who live there.

[Doesn't belong] He does not belong to your broader community. He has no roots in your area, does not share the values and experiences of people in your area, and has no personal ties to people who live there.

- Program Design

[Structured & demanding] The program is highly structured and demanding. Participants follow a strict daily schedule, complete long hours of coursework, and are evaluated through frequent tests and mandatory study periods.

[Flexible & supportive] The program is flexible and supportive. Participants set their own pace, learn in a comfortable classroom environment with tablets and lounge-style seating, and receive individualized guidance without strict deadlines.

- Cost-Benefit

[Unequal] with incarcerated participants benefiting three times as much as similar low-income residents who are not incarcerated and completed the same program.

[Equal] with equal benefit for incarcerated participants and similar low-income residents who are not incarcerated and completed the same program.

- Race: Asian, Black, Latino, White.

- Name: Albert, Henry, Isaac, Joe, Jordan, Samuel.

B Summary Statistics, Randomization, and Covariate Balance Checks

B.1 Summary Statistics & Variable Description

Table B.1: Summary statistics

Variable	Mean/N	SD/%	Min	Max
Choice (share Yes)	0.6	0.3	0.0	1.0
Deservingness Scale	1.8	0.8	0.0	4.0
Age	45.8	15.9	18.0	84.0
Punitiveness	-0.1	2.3	-4.0	4.0
Rehabilitative Orientation	1.9	1.8	-4.0	4.0
News Consumption	2.3	0.9	0.0	4.0
LE Direct 1	-0.2	1.2	-2.0	2.0
LE Direct 2	-0.3	1.2	-2.0	2.0
Survey Duration (sec)	450.7	246.7	112.0	3386.0
White	0.7	0.5	0.0	1.0
Foreign Born	0.1	0.3	0.0	1.0
Student	0.1	0.3	0.0	1.0
Sex (male)	0.5	0.5	0.0	1.0
Formerly Incarcerated	0.0	0.1	0.0	1.0
Family/Friend Incarcerated	0.5	0.5	0.0	1.0
Any Victimization	0.2	0.4	0.0	1.0
Property Victimization	0.1	0.3	0.0	1.0
Violent Victimization	0.1	0.3	0.0	1.0
Passed IMC	1.0	0.1	0.0	1.0
Passed FMC1	1.0	0.2	0.0	1.0
Passed FMC2	0.5	0.5	0.0	1.0
Incarceration Contact Count	0.6	0.7	0.0	2.0
Victimization Count	1.2	0.6	1.0	5.0
Total Checks Passed	2.5	0.5	0.0	3.0
<i>Employment</i>				
Full-Time	847	43.9		
Other	514	26.6		
Party-Time	338	17.5		
Unemployed	230	11.9		
<i>Political Affiliation</i>				
Democrat	574	29.8		
Independent	823	42.7		
Republican	532	27.6		
<i>Ideology (7-pt)</i>				
Very liberal	255	13.2		
Liberal	375	19.5		
Somewhat liberal	210	10.9		
Moderate	447	23.2		
Somewhat conservative	235	12.2		
Conservative	290	15.0		
Very conservative	116	6.0		

Respondent-level summary statistics. Choice and Deservingness scale (conjoint outcomes) are average across each respondent's six conjoint tasks. Binary variables are coded 0/1 and reported as shares (mean). Categorical variables with more than 2 categories report N and percent per category.

Table B.1 presents the summary statistics for all our non-experimental variables (see next subsection for randomization checks involving them). Most of these variables come from questions included

in our survey, whereas a few others come from the Prolific participant information. The following list indicates how each variable is measured (and provides any additional details).

[Choice] is the primary outcome and is asked immediately after each vignettes: “If you had to choose, should your state offer this program to people like [applicant name]?” Answers “Yes” are coded as 1, and “No” are coded as zero.

[Deservingness Scale] is the secondary outcome used in Appendix D. It is a numerical rating (1-5) of candidate deservingness in response to the question “On a scale from 1 to 5, how deserving is [applicant name] of receiving this program?”

[Age] is respondent’s age from the Prolific profile.

[Punitiveness] measures respondents’ punitiveness and is derived from agreement with two statements: the first is “Counts in this country are not harsh enough in dealing with criminals”, the second is “Prison sentences should be made much tougher to send a message to would-be criminals.” Respondents indicated their degree of agreement on a 5-point scale (Strongly Agree to Strongly Disagree). Each question was first recoded into a numerical scale from -2 to +2, and the punitiveness index was built as the sum of the two numerical scales. Higher values indicate higher punitiveness.

[Rehabilitative Orientation] measures respondents’ agreement that prison and the criminal justice system’s role is to rehabilitate offenders. Similar to punitiveness, it is derived from agreement with two statements: the first is “One of the main goals of prison should be to rehabilitate offenders so they can become productive members of society,” and the second is “With the right programs and support, most people who commit crimes can turn their lives around.” Respondents indicated their degree of agreement on a 5-point scale (Strongly Agree to Strongly Disagree). Each question was first recoded into a numerical scale from -2 to +2, and the rehabilitative orientation index was built as the sum of the two numerical scales. Higher values indicate a stronger rehabilitative orientation.

[News Consumption] is based on a question asking respondents “How often do you follow news stories about crime.” Response options were: Never, Rarely, Sometimes, Often, and Very Often. For simplicity of analysis, this variable was transformed into a numeric scale, ranging from 0 to 4.

[LE Direct 1] is a direct endorsement of the Less Eligibility principle, measured as agreement with the following statement: “People in prison should not have access to amenities or programs that are unavailable to law-abiding citizens.” Respondents indicated their degree of agreement on a 5-point scale (Strongly Agree to Strongly Disagree), which was recoded into a numerical scale from -2 to +2 for analysis.

[LE Direct 2] is a direct endorsement of the Less Eligibility principle, measured as agreement with the following statement: “As long as someone is in prison, their well-being should be a lower priority than the well-being of law-abiding citizens.” Respondents indicated their degree of agreement on a 5-point scale (Strongly Agree to Strongly Disagree), which was recoded into a numerical scale from -2 to +2 for analysis.

[Survey Duration] is the time, in seconds, taken to complete the survey, based on Qualtrics start and end times.

[White] is an indicator for respondents who identify as White in our survey.

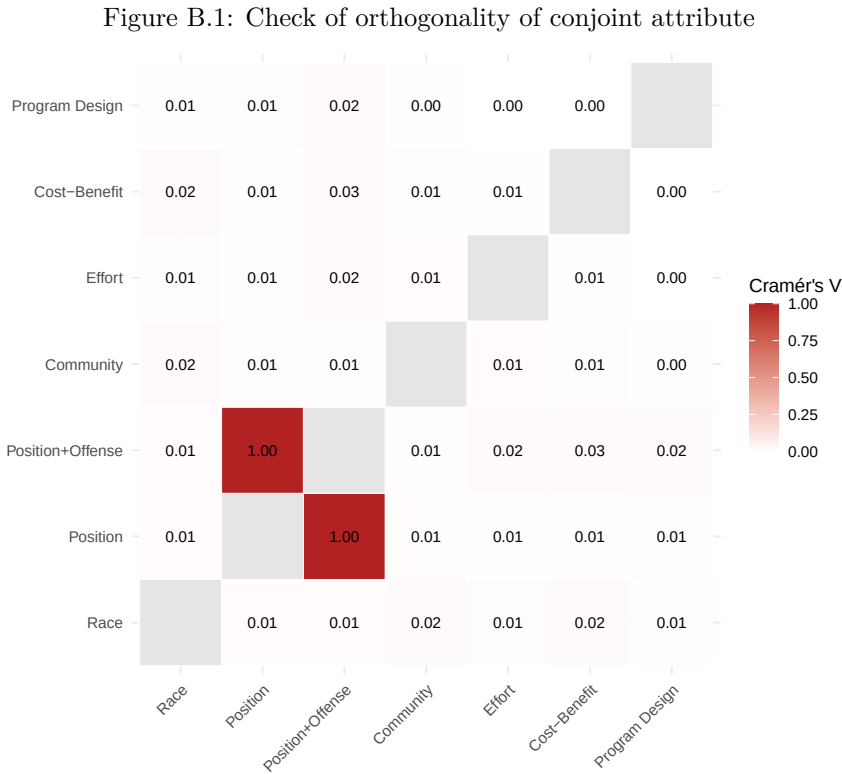
[Foreign Born] is an indicator for respondents whose Prolific profile indicates birth in a country other than the United States.

[Student] is an indicator for respondents whose Prolific profile indicates they are currently a student.

- [Sex] is an indicator for respondents whose Prolific profile indicates they are men.
- [Formerly Incarcerated] comes from responses to the question “Have you ever been incarcerated in jail or in prison, even briefly.” The question admits the following answers: Yes, No, Prefer not to Say. For some of the analyses, Prefer not to Say is considered a separate answer. For the incarceration contact count (see below), it is considered missing. Table B.1 shows the proportion answering “Yes.”
- [Family/Friend Incarcerated] comes from responses to the question “To the best of your knowledge, has a family member or a close friend ever been incarcerated in a jail or prison,” which admits answers Yes and No.
- [Incarceration Contact Count] counts whether the respondents themselves have a personal incarceration contact and whether their family/friends do.
- [Any Victimization] indicates whether the respondent has been a victim of a crime in the previous 12 months based on responses to the question “In the last 12 months, have you personally experienced any of the following?” The options include “Someone broke into or attempted to break into my home,” “Something was stolen from me (including a car, bike, wallet, phone, etc),” “I was physically attacked or assaulted,” “I was threatened with violence,” and “None of the above.”
- [Property Victimization] is based on the same question as Any Victimization, but indicates whether the respondent selected at least one of the following options: “Someone broke into or attempted to break into my home” or “Something was stolen from me (including a car, bike, wallet, phone, etc).”
- [Violent Victimization] is based on the same question as Any Victimization, but indicates whether the respondent selected at least one of the following options: “I was physically attacked or assaulted” or “I was threatened with violence.”
- [Victimization Count] counts the number of options selected (other than “None of the Above”) in the victimization question.
- [Passed IMC] indicates whether respondents passed an instruction check presented after instructions and before the first conjoint vignette and asking “How many applicant profiles will you review in this survey?”
- [Passed FMC1] indicates whether respondents passed a factual check presented immediately after the final conjoint vignette asking respondents to “Please select the answer that matches what you read in the previous profiles. What did the program provide?” with options: “One year of mental health care,” “One year of college and job counseling,” “One year of anger management courses,” and “I did not read the profiles carefully.”
- [Passed FMC2] indicates whether respondents passed a factual manipulation check presented immediately after FMC1. It asked respondents “Thinking about the last applicant you read about, which of the following best described their reputation?” Question options are: “Unreliable and unmotivated,” “Dependable and hard-working,” “Not mentioned,” and “I did not read the profiles carefully.”
- [Total Checks Passed] counts the number of checks passed by each respondent.
- [Employment] comes from respondents’ Prolific profiles and indicates their employment status: Full-time, Part-time, Unemployed, and Other (including out of the work force).
- [Political Affiliation] comes from respondents’ Prolific profiles and indicates their party identification: Democrat, Independent, and Republican.
- [Ideology (7-pt)] comes from a question asking “In politics, people sometimes talk about liberal and conservative. In general, would you describe yourself as:” with options: Very Liberal, Liberal, Somewhat Liberal, Moderate, Somewhat Conservative, Conservative, and Very Conservative.

B.2 Randomization and Covariate Balance Checks

Figure B.1 reports pairwise independence across attributes, using Cramer’s V. With the exception of Position and Position+Offense, which are mechanically linked by design, all pairwise associations are essentially zero, as expected.



Heatmap with pairwise Cramer’s V across all profile attributes. Values near 0 indicate the two attributes are independently randomized. Offense severity is only included in the collapsed attribute of Position+Offense.

Table B.2 presents a randomization check for the attributes in the study, across all conjoint tasks. The offense severity attribute is not presented separately because it is conditionally randomized only when social position is incarcerated. Instead, we present it collapsed into the Position+Offense attribute. In all cases, the deviation from the target is below 1 percentage point, the largest being for the Cost-Benefit attribute, with a difference of 0.7 percentage points from target. Based on the χ^2 test, all attributes pass the randomization check.

Table B.2: Randomization checks

Attribute	Level	Target %	Observed %	Diff. (pp)	χ^2	p -value
Race	White	25.0	24.8	-0.2	1.04	0.791
	Asian	25.0	25.2	0.2		
	Black	25.0	25.3	0.3		
	Latino	25.0	24.7	-0.3		
Position	Poverty	25.0	24.6	-0.4	1.30	0.522
	Incarcerated	50.0	50.4	0.4		
	On Welfare	25.0	25.0	0.0		
Position+Offense	Poverty	25.0	24.6	-0.4	1.65	0.800
	Incarcerated-burglary	16.7	16.9	0.3		
	Incarcerated-robbery	16.7	16.8	0.2		
	Incarcerated-shoplifting	16.7	16.6	-0.0		
	On Welfare	25.0	25.0	0.0		
Community	Doesn't belong	50.0	49.4	-0.6	1.49	0.223
	Belongs	50.0	50.6	0.6		
Effort	Unreliable	50.0	50.3	0.3	0.37	0.545
	Hard-working & reliable	50.0	49.7	-0.3		
Cost-Benefit	Equal	50.0	50.7	0.7	2.53	0.112
	Unequal	50.0	49.3	-0.7		
Program Design	Flexible	50.0	50.4	0.4	0.69	0.408
	Structured & demanding	50.0	49.6	-0.4		

Randomization balance for each profile attribute. Target % is the intended level; Observed % is the empirical frequency across all profiles; Diff. (pp) is the difference between target and observed; χ^2 and p -value compare the observed distribution and target levels. Since offense severity is only randomized for those incarcerated, it is presented only in the collapsed attribute Position+Offense.

Table B.3 tests whether attribute levels were balanced across conjoint tasks. All attributes show no significant association with task order, indicating full randomization across profiles.

Table B.3: Randomization balance across conjoint tasks

Attribute	χ^2	p -value
Race	11.45	0.720
Position	6.10	0.807
Position+Offense	20.61	0.421
Community	5.93	0.313
Effort	6.62	0.251
Cost-Benefit	0.79	0.978
Program Design	7.56	0.182

Table reports χ^2 tests of association for the attribute levels shown across conjoint tasks. Null results indicate independent randomization across conjoint tasks.

Table B.4 presents a check of covariate balance. We test eight respondent-level covariates for which data is recorded prior to the conjoint questions and are therefore pre-treatment. Six of these covariates are categorical: respondent race (White vs. non-White), employment status (Full-Time, Part-Time, Unemployed, or Other),²⁶ foreign born status, student status, sex, and political affiliation

²⁶Other includes retired and those doing unpaid work.

(Democrat, Republican, Independent). The other two covariates are numerical: respondent age, and Prolific experience (number of prior study approvals).

For each attribute-covariate pair, we test the null hypothesis that the covariate is unrelated to the attribute level shown to the respondent. Because covariates are fixed at the respondent level while attributes are randomized independently for each task (see above), we pool all task-level observations and estimate a linear model with standard errors clustered by respondent. For binary covariates and the two numerical covariates we provide a single cluster-robust Wald test of whether the attribute's level coefficients are jointly indistinguishable from zero. For the covariates with more than two categories, we instead conduct the same test for each non-reference level as a separate outcome.

Overall, the results show that covariate balance is achieved. Across the 77 tests, the vast majority of them have p-values larger than 10%, with only a handful with p-values falling between 5% and 10% (and none falling below 5%).

Table B.4: Balance tests

Covariate	Race	Position	Position+ Offense	Community	Effort	Cost-Benefit	Program Framing
Race: White	4.48 (.214)	2.24 (.327)	3.05 (.550)	0.16 (.691)	0.01 (.919)	0.58 (.447)	3.27 (.071)
Employment: Other	1.82 (.611)	1.17 (.557)	2.06 (.725)	3.42 (.064)	0.84 (.359)	0.81 (.368)	0.33 (.567)
Employment: Party-Time	5.12 (.163)	0.93 (.627)	1.86 (.762)	0.55 (.458)	1.36 (.244)	1.92 (.166)	0.01 (.911)
Employment: Unemployed	6.09 (.107)	0.42 (.811)	1.20 (.879)	0.68 (.411)	1.02 (.313)	2.87 (.090)	0.51 (.474)
Foreign born: Yes	4.35 (.226)	4.93 (.085)	7.68 (.104)	2.21 (.137)	2.85 (.091)	0.22 (.636)	1.74 (.187)
Student: Yes	2.45 (.484)	2.25 (.324)	2.26 (.688)	0.14 (.704)	0.11 (.739)	0.56 (.453)	0.36 (.546)
Sex: Male	2.54 (.468)	1.37 (.504)	2.64 (.621)	0.00 (.959)	0.48 (.489)	1.03 (.310)	1.80 (.180)
Pol. Affiliation: Independent	0.69 (.876)	0.81 (.668)	1.37 (.849)	0.02 (.899)	0.03 (.868)	0.25 (.616)	0.00 (.955)
Pol. Affiliation: Republican	3.57 (.312)	2.18 (.336)	5.65 (.227)	0.08 (.780)	0.33 (.568)	0.00 (.979)	0.06 (.811)
Age	2.11 (.549)	0.65 (.723)	1.19 (.880)	1.03 (.309)	2.67 (.103)	2.91 (.088)	2.48 (.115)
Experience	1.54 (.674)	3.11 (.211)	3.82 (.431)	1.06 (.304)	0.46 (.499)	0.15 (.701)	0.91 (.339)

Each cell reports the Wald χ^2 statistics and associated p-value in parenthesis from a test of whether the covariate is balanced across levels of the column attribute. For covariates with more than 2 categories, all but the reference category are included. Standard errors clustered by respondent.

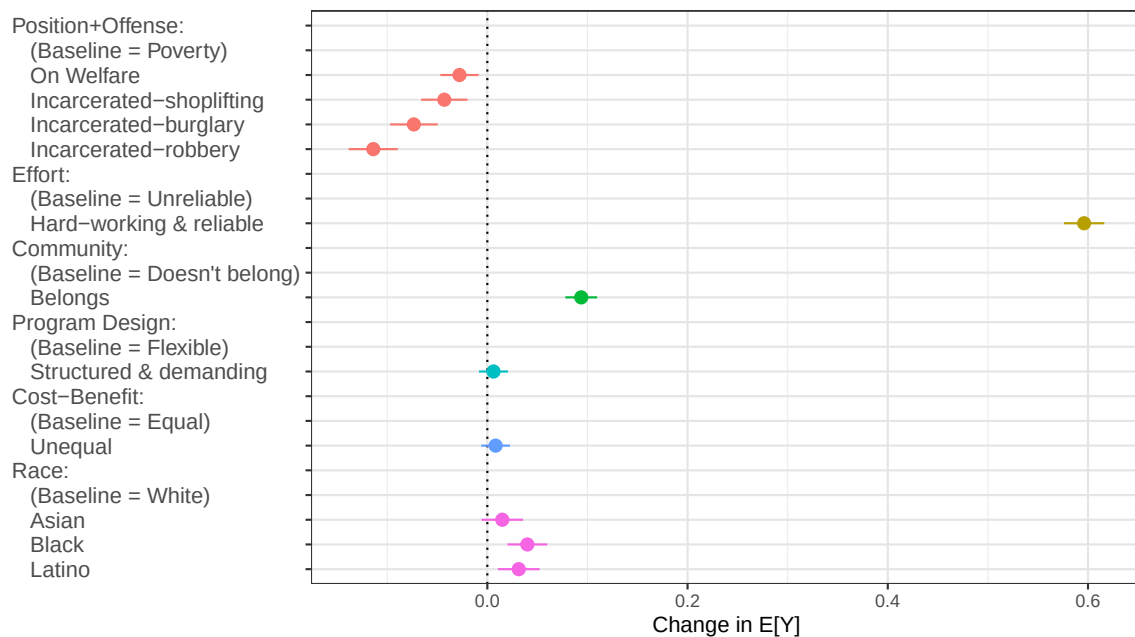
C Additional Results

This section presents additional results that rely on Choice as the outcome of the conjoint experiment. In some instances, the results presented here expand on those discussed or presented in the manuscript. Other results are designed to test additional hypotheses included in our Pre-Analysis Plan but not fully discussed in the manuscript. Appendix E contains tables that report the full models associated with all figures reported in this appendix and in the manuscript.

C.1 Main Effects with Offense Severity, Full Results

Figure C.1 presents the Average Marginal Component Effects (AMCE) for all attributes in the model that includes offense severity. The coefficients for the Position+Offense attribute are reported in the paper in Figure 2. Given that attributes (other than the connection between social position and offense severity) are independently assigned, the effects for effort, community belonging, program design, cost-benefit information, and race are equivalent to those presented in Figure 1, up to small numerical discrepancies.

Figure C.1: Average marginal component effects on choice, including offense severity



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses.

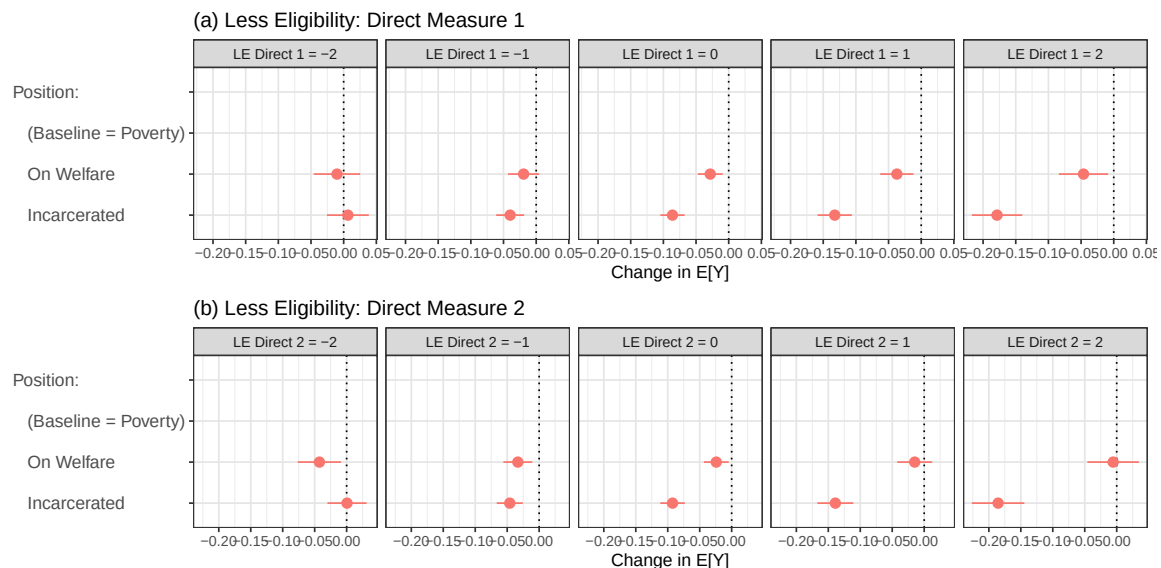
C.2 Incarceration Penalty and Direct Endorsement of Less Eligibility Statement

Here we explore whether our experimental measurement of less eligibility is consistent with respondents' levels of agreement with two statements that directly measure less eligibility. The first statement asks respondents to indicate the extent to which they agree that "People in prison should not have access to amenities or programs that are unavailable to law-abiding citizens." The second

asks about “As long as someone is in prison, their well-being should be a lower priority than the well-being of law-abiding citizens.”

The findings are in Figure C.2. As the results show, for respondents that strongly disagree with each of the two measurements (-2), we find no experimental less eligibility: the Incarcerated level of the position attribute is approximately zero and not statistically significant. For respondents who disagree (-1) with the direct statements of less eligibility, there is a mild incarceration penalty. This incarceration penalty increases with the level of agreement with the direct endorsement measures. Thus, our results show that our experimental less eligibility findings are consistent with individuals’ direct endorsement of the principle, but that the behavioral norm is more widespread than the conscious and explicit subscription to it.

Figure C.2: Heterogeneous effects by direct LE questions

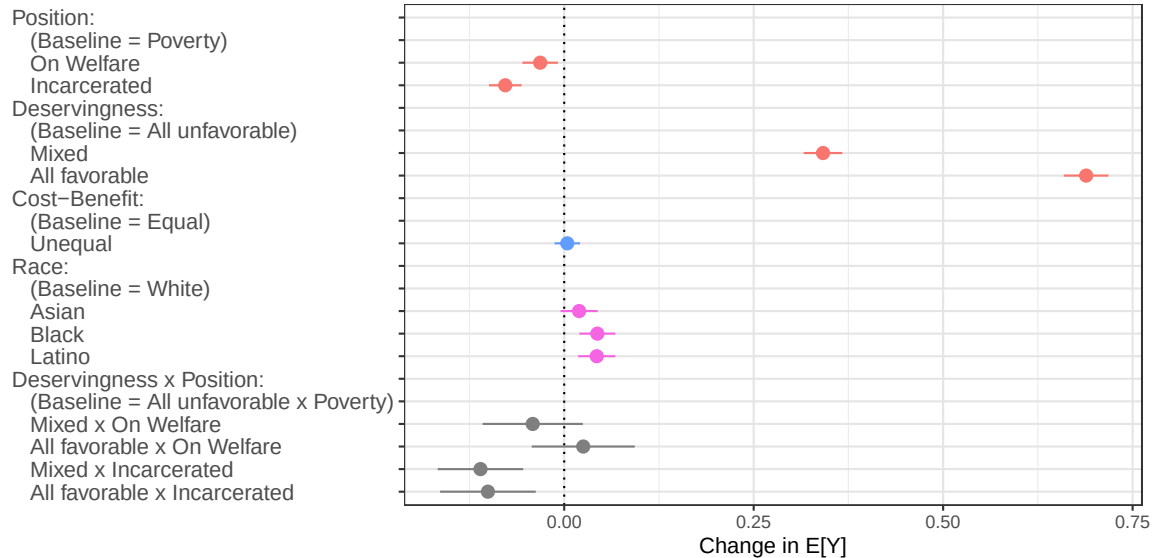


Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Each panel comes from a separate model that includes an interaction term between social position and a direct measure of Less Eligibility, plus linear terms for the remaining attributes in the conjoint. Both items are measured on a 5-point scale from -2 (Strongly Disagree) to 2 (Strongly Agree), and estimates are shown at each point of the scale.

C.3 Less Eligibility as a Bundle

We also constructed a contrast between respondents receiving all three favorable desert signals simultaneously (hard-working, community member, and demanding program design) and those receiving all three unfavorable signals. The findings using this bundle are presented in Figure C.3. Profiles in which applicants receive all favorable desert attributes are 70.8 percentage points more likely to lead to program support. Using the desert bundle as a moderator for the incarceration penalty, we find that highly deserving applicants receive a higher incarceration penalty than completely undeserving ones (mirroring the findings for effort); we find no statistically significant moderating effects in the case of applicants on welfare.

Figure C.3: Desert bundle

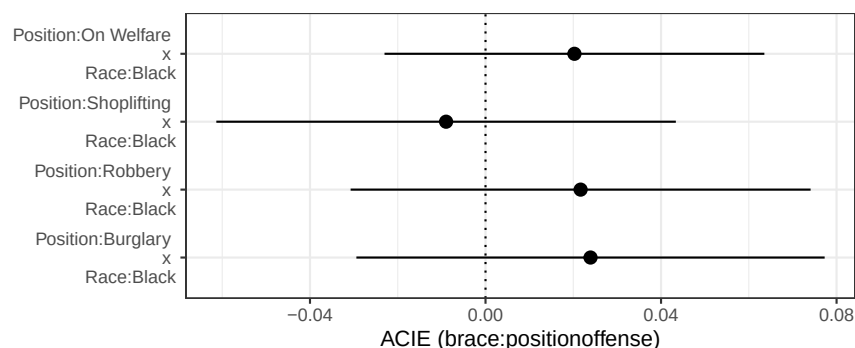


Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Estimates come from a single model that interacts social position with a deservingness bundle (combining effort, community belonging, and program design into a single three-level indicator: all unfavorable, all mixed, or all favorable), plus linear terms for the cost-benefit and race attributes. The final rows show the ACIE for the position x deservingness bundle interaction.

C.4 Race and Offense Severity

We also tested whether the gradient of the offense severity scale is larger for Black applicants than for non-Black applicants, consistent with the racialization of crime in American public perception. We find no significant interaction effects between the applicant race and offense severity attributes.

Figure C.4: ACIE: race & offense severity interaction

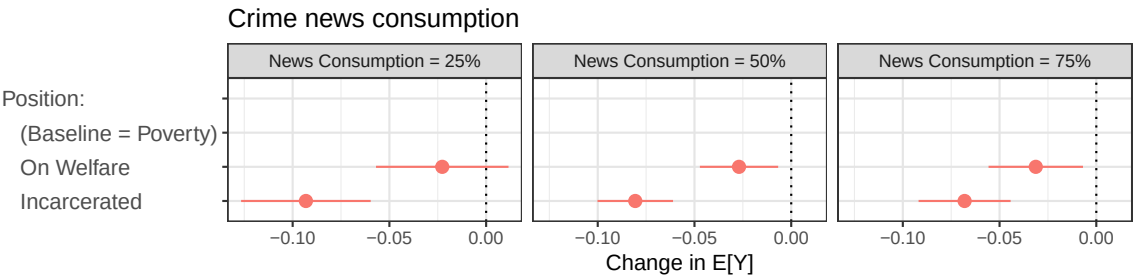


Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Estimates come from a single model that interacts a binary race attribute (Black vs. Non-Black) with the severity specification of social position (poverty, welfare, or incarceration for shoplifting, burglary, or robbery), plus linear terms for effort, community belonging, program design, and cost-benefit information. Only ACIEs shown.

C.5 Crime News

In Figure C.5 we test whether the less eligibility effect varies by respondents’ consumption of crime news, finding no heterogeneity with respect to this variable ($p = 0.253$).

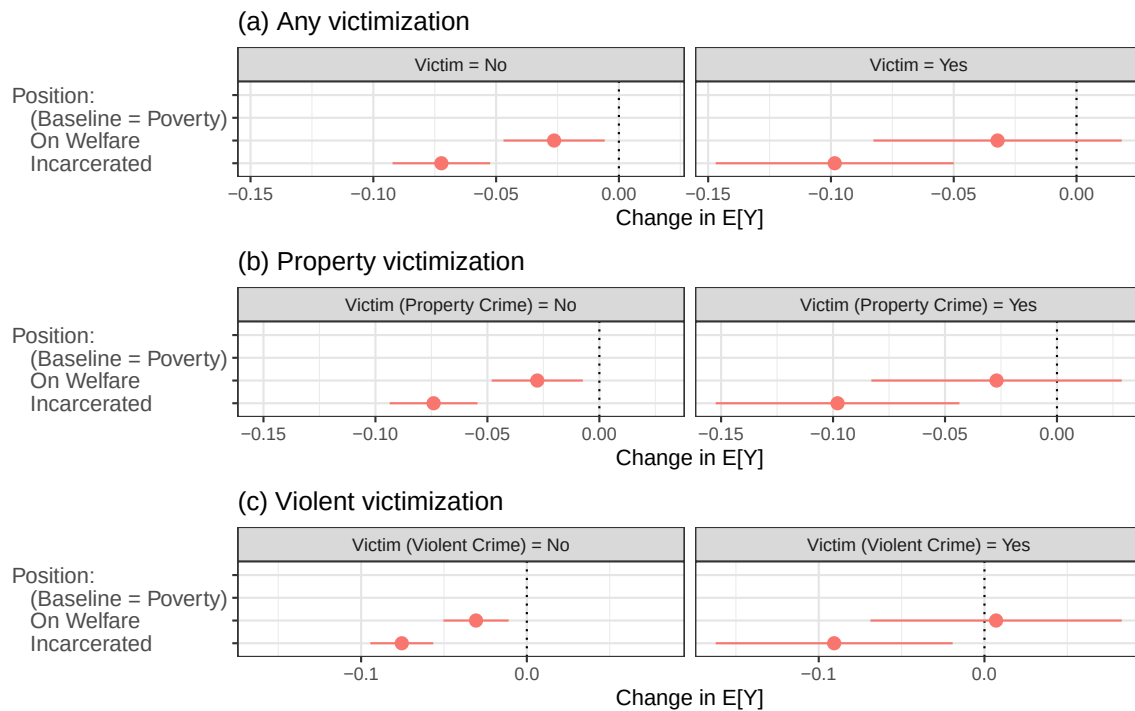
Figure C.5: Heterogeneous effects by respondent’s crime news consumption frequency



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Estimates come from a model that includes an interaction term between social position and respondents’ frequency of crime news consumption (measured on a 5-point scale), plus linear terms for the remaining attributes in the conjoint. Estimates are shown at three representative values of the news consumption measure.

C.6 Victimization Results

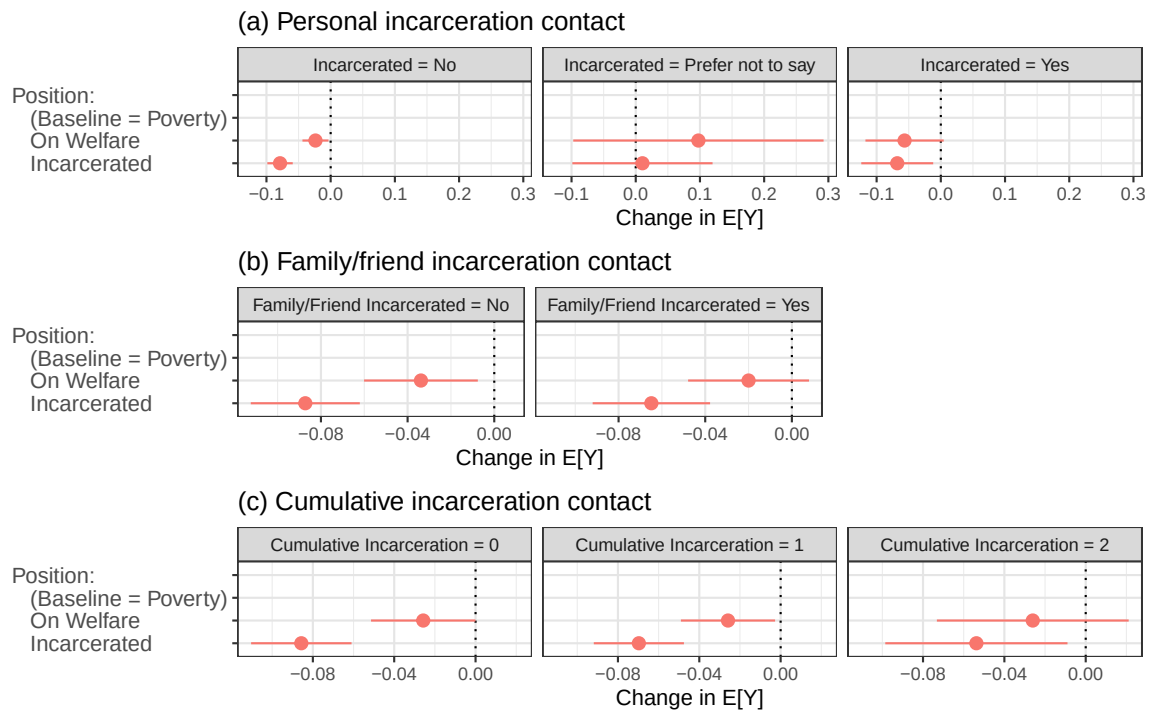
Figure C.6: Heterogeneous effects by respondent's crime victimization



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels of each attribute shown in parentheses. Each panel comes from a separate model that includes an interaction term between social position and a binary indicator of respondent victimization history, plus linear terms for the remaining attributes in the conjoint. Panel (a) indicates any prior victimization; panel (b) indicates property victimization (having something stolen or a break-in); panel (c) indicates violent victimization (having been physically attacked or threatened with violence).

C.7 Incarceration Contact

Figure C.7: Heterogeneous effects by respondent's contact with incarceration



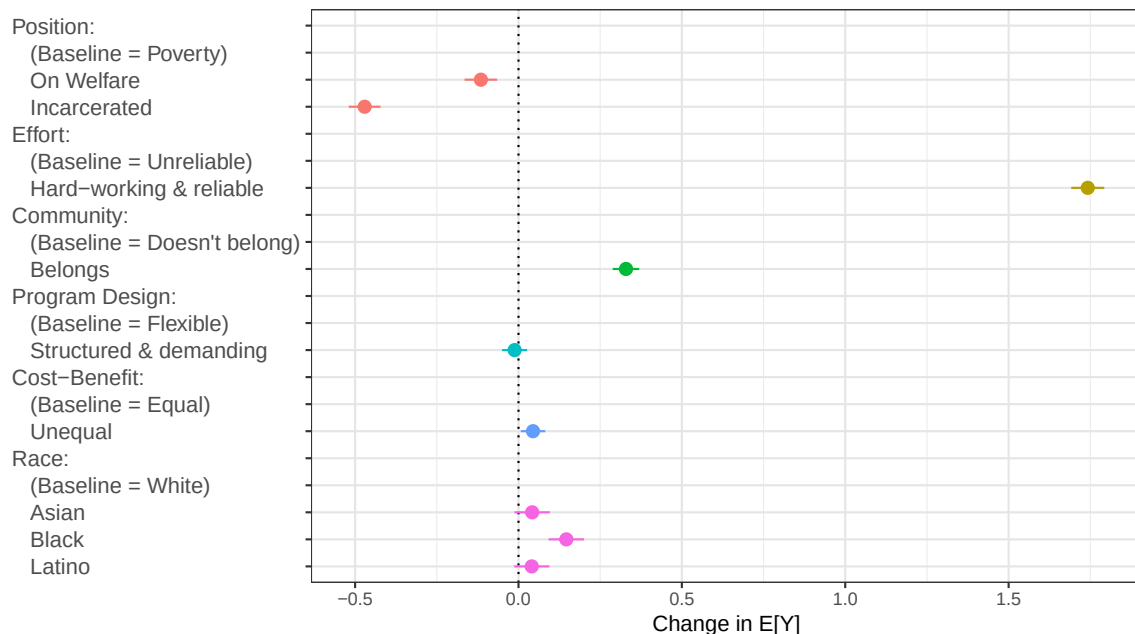
Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Each panel comes from a separate model that includes an interaction term between social position and a measure of respondent incarceration contact, plus linear terms for the remaining attributes in the conjoint. Panel (a) indicates the respondents' own history of incarceration; panel (b) indicates incarceration of a close family member or friend; and panel (c) is a cumulative count summing both forms of incarceration contact (0, 1, or 2).

D Robustness Checks: Scale Outcome

In this section we present the same estimates as in the main text, but utilizing our numerical 5-point deservingness scale as the outcome variable, instead of the binary choice outcome. The analysis is identical in all cases, changing only the outcome variable. We reproduce all the estimates presented in the paper, as well as those in Appendix C, using this alternative outcome.

Figure D.1 presents similar results to those in Figure 1. There are only some minor differences. First, only Black applicants—not Latino applicants—are favored relative to White applicants when using the scale outcome. Second, the contrast between incarceration penalty and the welfare penalty is stronger. And third, the effect of program cost-benefit is positive and statistically significant ($p = 0.020$) although the substantive size of the effect is very small: an increase in support of 0.044 on a scale from 1 to 5.

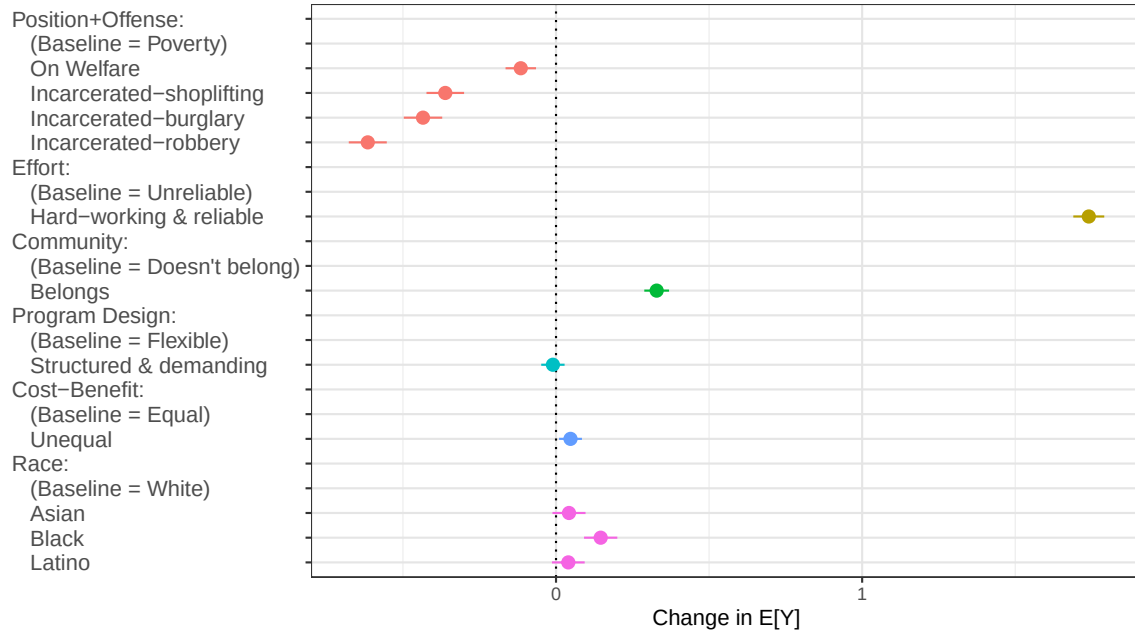
Figure D.1: Average marginal component effects on scale



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute are shown in parentheses. Estimates are average marginal component effects (AMCEs) on the numerical 5-point deservingness scale outcome. Replicates the main choice-outcome specification (Figure 1) using the continuous scale measure instead of the binary one.

Figure D.2 reproduces Figure C.1 (and thus Figure 2). The results on offense severity are substantively equivalent to those presented in the main text. The only minor difference is that the evaluation penalty for shoplifting is statistically distinguishable from that of being on welfare.

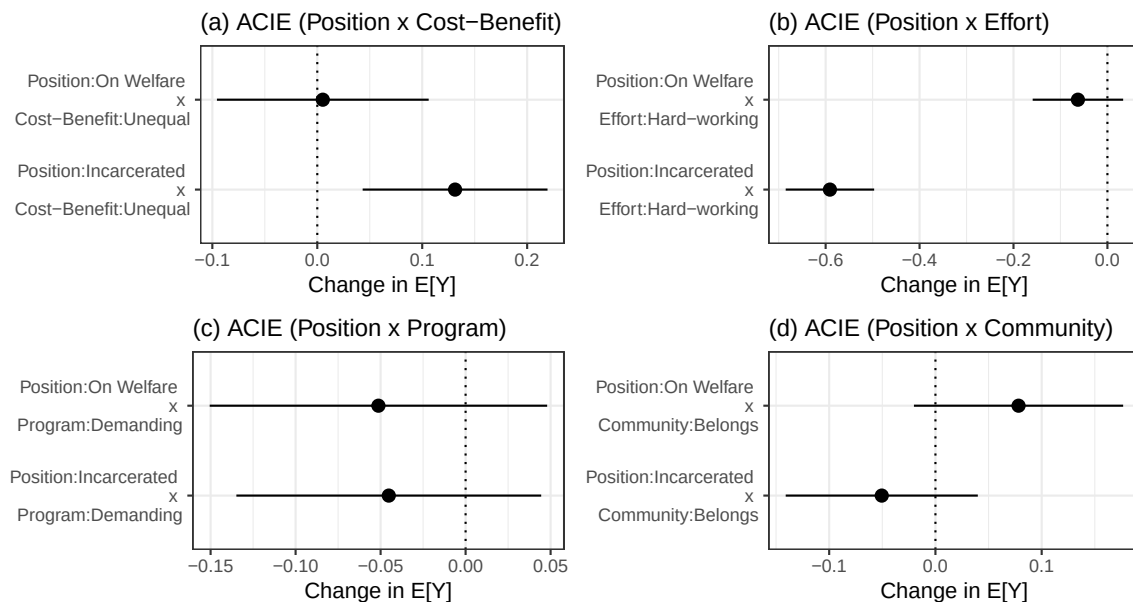
Figure D.2: Offense severity grading using scale outcome



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute are shown in parentheses. Estimates are average marginal component effects (AMCEs) on the numerical 5-point deservingness scale outcome, for the model that uses the 5-level combined attribute of social position and offense severity. Replicates the main choice-outcome specification (Figures C.1 and 2) using the continuous scale measure instead of the binary one.

Figure D.3 reproduces Figure 3. The results are largely equivalent, with the only difference being the disappearance of the positive and statistically significant interaction between the welfare position and community belonging attribute (which was only marginally significant in the manuscript, when using the choice outcome).

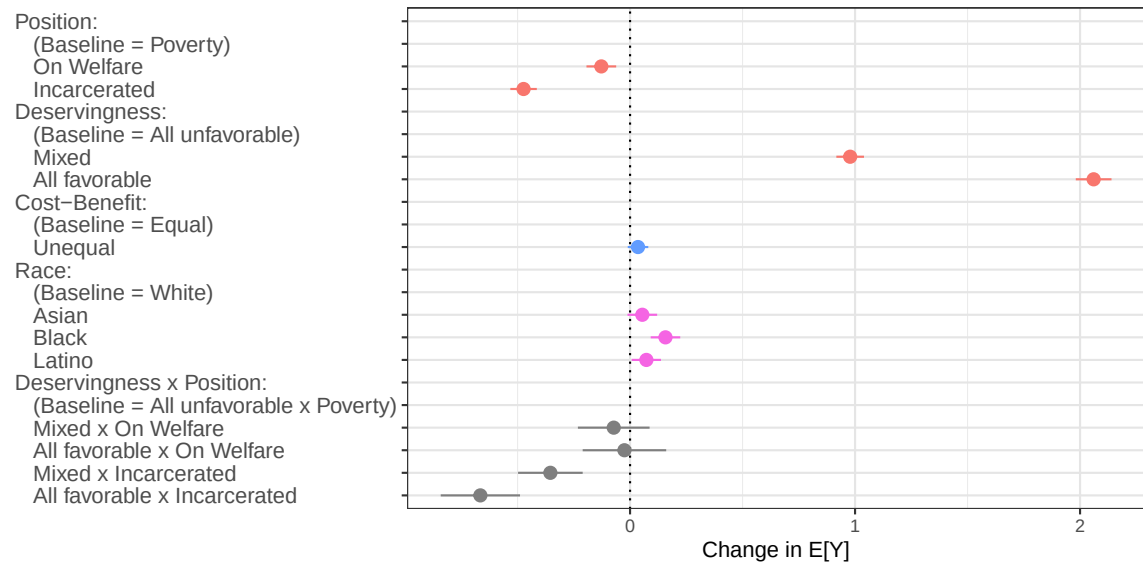
Figure D.3: ACIEs: attributes that moderate less eligibility, using scale outcome



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Each panel comes from a separate model using the numerical 5-point deservingness scale outcome that includes an interaction term between social position and the corresponding attribute (and linear terms for the remaining attributes in the conjoint). Panel (a) shows the ACIE for position $\times$ cost-benefit; panel (b) for position $\times$ effort; panel (c) for position $\times$ program design; panel (d) for position $\times$ community belonging. This replicates the choice-outcome moderator figure (Figure 3) using the continuous scale measure instead of the binary choice one.

Figure D.4 replicates C.3. These results are slightly different than those presented in the main text when considering the interaction between our desert bundle (combined) attribute and less eligibility. In particular, we find a statistically significant desert grading moderating less eligibility: incarcerated applicants in cases in which all desert attributes are positive are more harshly penalized compared to instances of mixed desert attributes, and instances of all negative desert attributes. It should be noted, however, that the desert bundle variable overcomes this harsher penalty so that overall support for the applicant increases with increasing desert (but at a flatter rate for incarcerated ones).

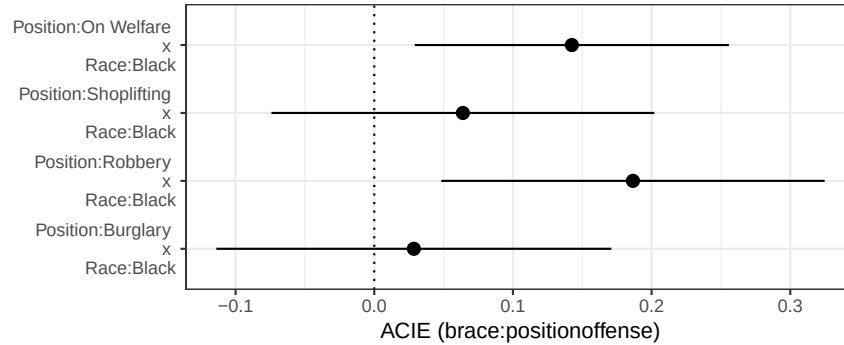
Figure D.4: Desert bundle using scale outcome



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Estimates come from a single model on the numerical 5-point deservingness scale outcome that interacts social position with a deservingness bundle (combining effort, community belonging, and program design into a single three-level indicator: all unfavorable, mixed, or all favorable), plus linear terms for cost-benefit and race. The final rows show the ACIE for the position $\times$ deservingness bundle interaction. This replicates the choice-outcome deservingness bundle figure (Figure C.3) using the continuous scale measure instead of the binary choice one.

Figure D.5 replicates Figure C.4. The results are somewhat different than those presented in the main text. Our findings show that respondents tend to favor, in relative terms, Black applicants on welfare relative to non-Black applicants on welfare. Similarly, respondents tend to favor Black applicants incarcerated for robbery relative to non-Black applicants incarcerated for the same reason. The interactions with applicants incarcerated for shoplifting and burglary remain statistically non-significant.

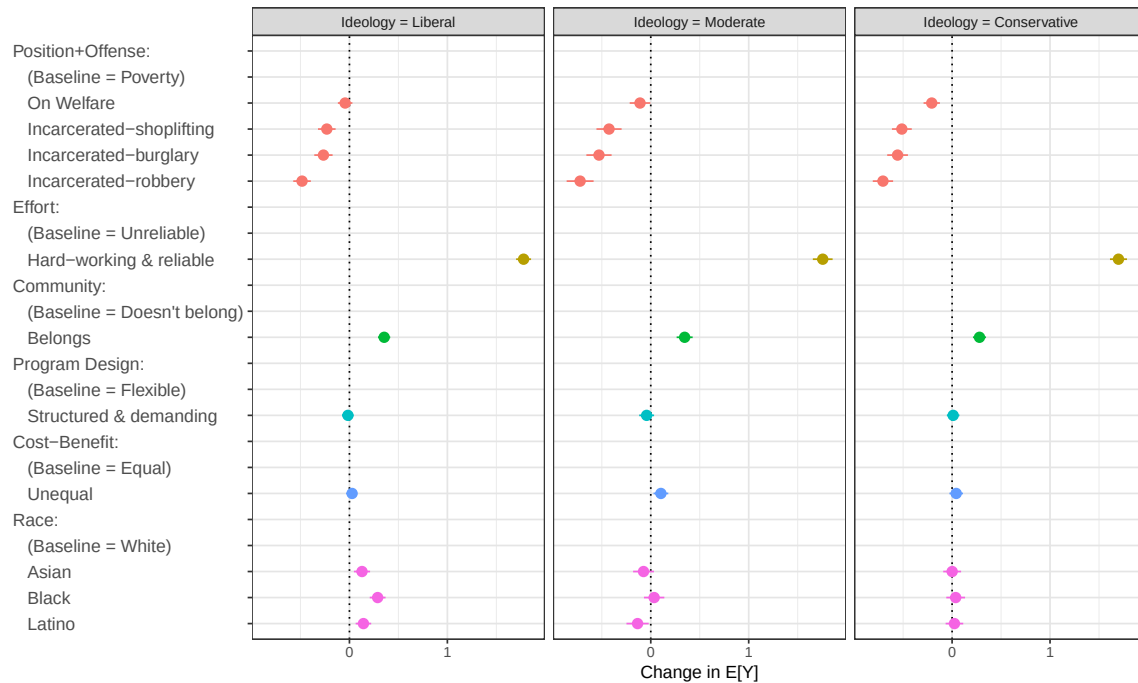
Figure D.5: ACIE: race & offense severity interaction using scale outcome



Confidence intervals at the 95% level. Standard errors clustered by respondent. Estimates are the average conditional interaction effect (ACIE) for race (binarized to Black vs. non-Black) interacted with offense severity, on the numerical 5-point deservingness scale outcome, from a model that also includes linear terms for effort, community belonging, program design, and cost-benefit. This replicates the choice-outcome race $\times$ severity ACIE figure (Figure C.4) using the continuous scale measure instead of the binary choice one.

Figure D.6 replicates Figure 5. The findings are equivalent to those described in the main text.

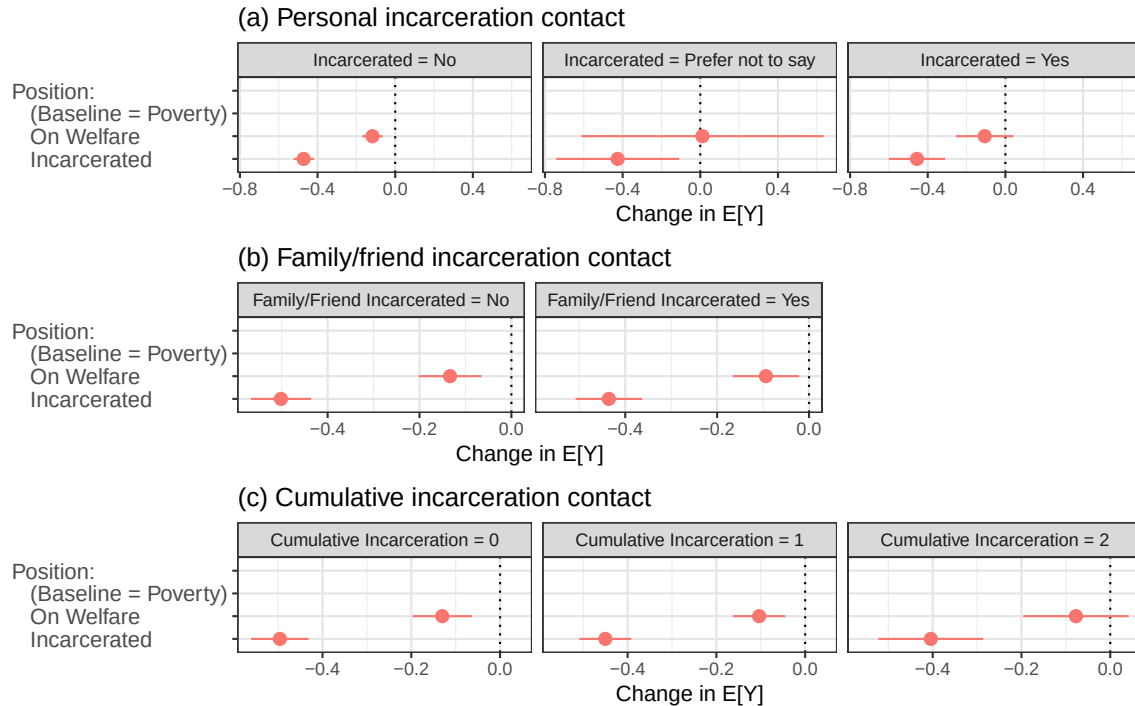
Figure D.6: Heterogeneous effects by respondents' ideological orientation, using scale outcome



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Estimates come from a single model on the numerical 5-point deservingness scale outcome that interacts respondent ideology with every profile attribute, using the severity specification of social position (poverty, welfare, or incarceration for shoplifting, burglary, or robbery). Each facet shows the conditional AMCE of each attribute among respondents in each ideology category. This replicates the choice-outcome ideology moderation figure (Figure 5) using the continuous scale measure instead of the binary choice one.

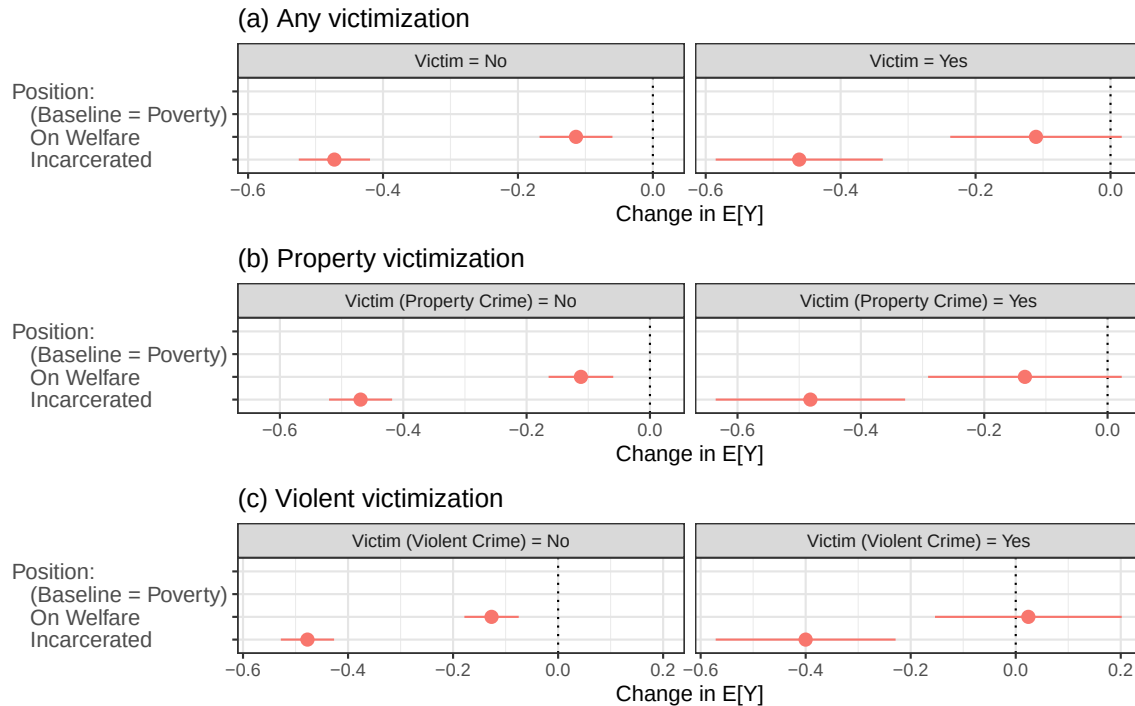
The remaining figures replicate the remaining heterogeneous effects of less eligibility on incarceration contact, crime victimization, punitiveness and rehabilitative orientations, and crime news consumption. The results are substantively the same in all of these cases.

Figure D.7: Heterogeneous effects by respondents' contact with incarceration, using scale outcome



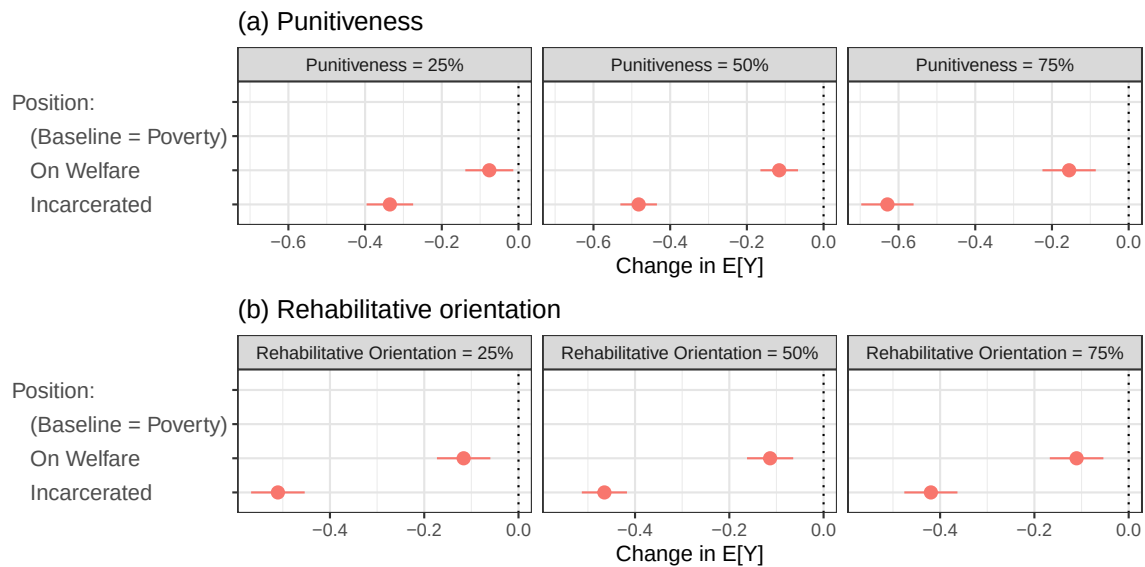
Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Each panel comes from a separate model on the numerical 5-point deservingness scale outcome that includes an interaction term between social position and a measure of respondent incarceration contact, plus linear terms for the remaining attributes in the conjoint. Panel (a) indicates the respondent's own history of incarceration; panel (b) indicates incarceration of a close family member or friend; panel (c) is a cumulative count summing both forms of contact (0, 1, or 2). This replicates the choice-outcome incarceration contact figure (Figure C.7) using the continuous scale measure instead of the binary choice one.

Figure D.8: Heterogeneous effects by respondents' crime victimization, using scale outcome



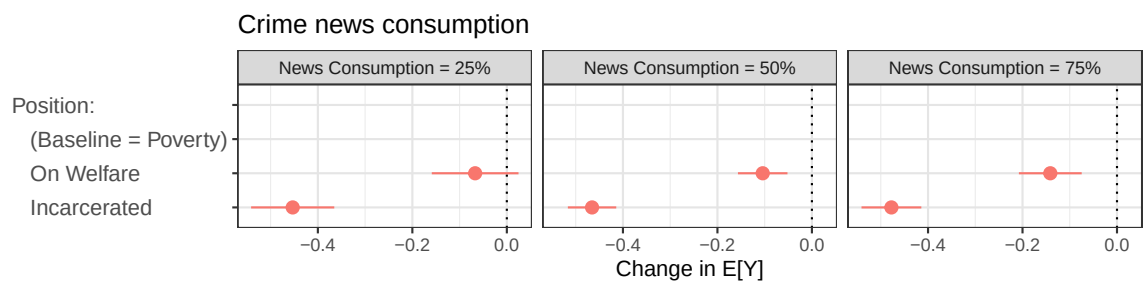
Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Each panel comes from a separate model on the numerical 5-point deservingness scale outcome that includes an interaction term between social position and a measure of respondent victimization history, plus linear terms for the remaining attributes in the conjoint. Panel (a) indicates any prior victimization; panel (b) indicates property victimization (having something stolen or a break-in); panel (c) indicates violent victimization (having been physically attacked or threatened with violence). This replicates the choice-outcome victimization figure (Figure C.6) using the continuous scale measure instead of the binary choice one.

Figure D.9: Heterogeneous effects by respondents' punitiveness and rehabilitative orientations, using scale outcome



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Each panel comes from a separate model on the numerical 5-point deservingness scale outcome that includes an interaction term between social position and a respondent-level attitudinal measure, plus linear terms for the remaining attributes in the conjoint. Panel (a) uses respondents' punitiveness (agreement with punitive crime-policy statements); panel (b) uses respondents' rehabilitative orientation (agreement with rehabilitation-oriented statements). Estimates are shown at the 25th, 50th, and 75th percentile of each measure. This replicates the choice-outcome punitiveness/rehabilitation figure (Figure 4) using the continuous scale measure instead of the binary choice one.

Figure D.10: Heterogeneous effects by respondents' crime news consumption frequency, using scale outcome



Confidence intervals at the 95% level. Standard errors clustered by respondent. Baseline levels for each attribute shown in parentheses. Estimates come from a model on the numerical 5-point deservingness scale outcome that includes an interaction term between social position and respondents' frequency of crime news consumption, plus linear terms for the remaining attributes in the conjoint. Estimates are shown at three representative values of the news consumption measure, labeled 25%, 50%, and 75%. This replicates the choice-outcome news consumption figure (Figure C.5) using the continuous scale measure instead of the binary choice one.

E Full Model Tables for Choice Outcome

This appendix contains the tables with the full models corresponding to the figures presented in the main text, as well as the figures in Appendix C.

Table E.1: Table with full models from Figures 1, 2, and 3

	Main	Severity	LE x Benefit	LE x Effort	LE x Program	LE x Communi- ty
Cost-Benefit	0.008	0.008	0.007	0.008	0.008	0.008
Unequal	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Effort	0.596***	0.596***	0.596***	0.597***	0.596***	0.596***
Hardworking	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Community	0.094***	0.094***	0.094***	0.094***	0.094***	0.094***
Belongs	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Position	-0.028**		-0.028**	-0.027**	-0.028**	-0.028**
Welfare	(0.010)		(0.010)	(0.010)	(0.010)	(0.010)
Position	-0.077***		-0.077***	-0.077***	-0.077***	-0.077***
Incarcerated	(0.009)		(0.009)	(0.009)	(0.009)	(0.009)
Race	0.015	0.015	0.014	0.014	0.015	0.015
Asian	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Race	0.040***	0.040***	0.040***	0.040***	0.040***	0.040***
Black	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Race	0.032**	0.031**	0.031**	0.031**	0.031**	0.032**
Latino	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Program	0.006	0.006	0.005	0.006	0.006	0.006
Structured	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Position+Offense		-0.028**				
Welfare		(0.010)				
Position+Offense		-0.043***				
Incarcerated-shoplifting		(0.012)				
Position+Offense		-0.073***				
Incarcerated-burglary		(0.012)				
Position+Offense		-0.114***				
Incarcerated-robbery		(0.012)				
Position x Cost-Benefit			0.000			
Unequal x Welfare			(0.020)			
Position x Cost-Benefit			0.044*			
Unequal x Incarcerated			(0.017)			
Position x Effort				0.031		
Hardworking x Welfare				(0.019)		
Position x Effort				-0.074***		
Hardworking x Incarcerated				(0.018)		
Position x Program					-0.023	
Welfare x Structured					(0.019)	
Position x Program					-0.019	
Incarcerated x Structured					(0.017)	
Position x Community						0.038*
Belongs x Welfare						(0.019)
Position x Community						0.012
Belongs x Incarcerated						(0.017)
N	11547	11547	11547	11547	11547	11547

Table reports AMCEs and ACIEs across six model specifications. Column “Main” corresponds to Figure 1; column “Severity” to Figure 2; the remaining columns to the models in Figure 3. Standard errors, clustered by respondent, shown in parentheses. * $p < .05$; ** $p < .01$; *** $p < .001$

Table E.2: Table with full models from Figures C.4 and C.3

	Race x Offense	Deservingness Bundle
Cost-Benefit	0.009	0.004
Unequal	(0.007)	(0.009)
Race	0.025**	
Black	(0.008)	
Effort	0.596***	
Hardworking	(0.010)	
Community	0.094***	
Belongs	(0.008)	
Position+Offense	-0.028**	
Welfare	(0.010)	
Position+Offense	-0.043***	
Incarcerated-shoplifting	(0.012)	
Position+Offense	-0.073***	
Incarcerated-burglary	(0.012)	
Position+Offense	-0.114***	
Incarcerated-robbery	(0.012)	
Program	0.006	
Structured	(0.007)	
Race x Position+Offense	0.020	
Black x Welfare	(0.022)	
Race x Position+Offense	-0.009	
Black x Incarcerated-shoplifting	(0.027)	
Race x Position+Offense	0.024	
Black x Incarcerated-burglary	(0.027)	
Race x Position+Offense	0.022	
Black x Incarcerated-robbery	(0.027)	
Deservingness		0.342***
Mixed		(0.013)
Deservingness		0.689***
All favorable		(0.015)
Position		-0.032**
Welfare		(0.012)
Position		-0.078***
Incarcerated		(0.011)
Race		0.020
Asian		(0.013)
Race		0.044***
Black		(0.012)
Race		0.043***
Latino		(0.012)
Deservingness x Position		-0.042
Mixed x Welfare		(0.034)
Deservingness x Position		0.025
All favorable x Welfare		(0.035)
Deservingness x Position		-0.110***
Mixed x Incarcerated		(0.029)
Deservingness x Position		-0.101**
All favorable x Incarcerated		(0.032)
N	11547	11547

Table reports AMCEs and ACIEs across two model specifications. Column “Race x Offense” corresponds to Figure C.4; column “Deservingness Bundle” to Figure C.3. Standard errors, clustered by respondent, shown in parentheses.

* $p < .05$; ** $p < .01$; *** $p < .001$

Table E.3: Table with full models from Figure C.7

	Formerly Incarcerated	Close Family/Friend Incarcerated	Cumulative Incarceration Contact
Cost-Benefit	0.008	0.008	0.008
Unequal	(0.007)	(0.007)	(0.007)
Effort	0.597***	0.597***	0.598***
Hardworking	(0.010)	(0.010)	(0.010)
Community	0.095***	0.093***	0.094***
Belongs	(0.008)	(0.008)	(0.008)
Race	0.014	0.015	0.014
Asian	(0.011)	(0.011)	(0.011)
Race	0.040***	0.040***	0.040***
Black	(0.010)	(0.010)	(0.010)
Race	0.032**	0.031**	0.032**
Latino	(0.011)	(0.011)	(0.011)
Program	0.006	0.006	0.006
Structured	(0.007)	(0.007)	(0.007)
Position	-0.024*	-0.034*	-0.026*
Welfare	(0.010)	(0.013)	(0.013)
Position	-0.079***	-0.087***	-0.086***
Incarcerated	(0.010)	(0.013)	(0.013)
Position x Formerly Incarcerated	0.121		
Welfare x Prefer not to say	(0.100)		
Position x Formerly Incarcerated	0.089		
Incarcerated x Prefer not to say	(0.057)		
Position x Formerly Incarcerated	-0.033		
Welfare x Yes	(0.033)		
Position x Formerly Incarcerated	0.011		
Incarcerated x Yes	(0.030)		
Position x Close Family/Friend Incarcerated		0.014	
Welfare x Yes		(0.020)	
Position x Close Family/Friend Incarcerated		0.022	
Incarcerated x Yes		(0.019)	
Position x Cumulative Incarceration Contact			-0.000
Welfare			(0.015)
Position x Cumulative Incarceration Contact			0.016
Incarcerated			(0.015)
N	11487	11541	11481

Table reports AMCEs and heterogeneous effects across three models corresponding to those in Figure C.7. Standard errors, clustered by respondent, shown in parentheses. * $p < .05$; ** $p < .01$; *** $p < .001$

Table E.4: Table with full models from Figure C.6

	Crime Victim (Any)	Crime Victim (Property)	Crime Victim (Violent)
Cost-Benefit	0.008	0.008	0.008
Unequal	(0.007)	(0.007)	(0.007)
Effort	0.596***	0.596***	0.596***
Hardworking	(0.010)	(0.010)	(0.010)
Community	0.094***	0.094***	0.094***
Belongs	(0.008)	(0.008)	(0.008)
Race	0.015	0.015	0.015
Asian	(0.011)	(0.011)	(0.011)
Race	0.040***	0.040***	0.040***
Black	(0.010)	(0.010)	(0.010)
Race	0.032**	0.032**	0.032**
Latino	(0.011)	(0.011)	(0.011)
Program	0.006	0.006	0.006
Structured	(0.007)	(0.007)	(0.007)
Position	-0.026*	-0.028**	-0.031**
Welfare	(0.011)	(0.010)	(0.010)
Position	-0.072***	-0.074***	-0.075***
Incarcerated	(0.010)	(0.010)	(0.010)
Position x Victim	-0.006		
Welfare x Yes	(0.028)		
Position x Victim	-0.026		
Incarcerated x Yes	(0.027)		
Position x Victim (Property Crime)		0.001	
Welfare x Yes		(0.030)	
Position x Victim (Property Crime)		-0.024	
Incarcerated x Yes		(0.030)	
Position x Victim (Violent Crime)			0.038
Welfare x Yes			(0.040)
Position x Victim (Violent Crime)			-0.015
Incarcerated x Yes			(0.038)
N	11547	11547	11547

Table reports AMCEs and heterogeneous effects across three models corresponding to those in Figure C.6. Standard errors, clustered by respondent, shown in parentheses. * $p < .05$; ** $p < .01$; *** $p < .001$

Table E.5: Table with full models from Figure 4, Figure C.2, and Figure C.5

	Punitiveness	Rehabilitative Orientation	Direct LE 1	Direct LE 2	Crime News Consumption
Cost-Benefit	0.008	0.008	0.008	0.008	0.008
Unequal	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Effort	0.598***	0.597***	0.598***	0.597***	0.597***
Hardworking	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Community	0.093***	0.094***	0.094***	0.094***	0.094***
Belongs	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Position	-0.027**	-0.027**	-0.027**	-0.027**	-0.028**
Welfare	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Position	-0.076***	-0.076***	-0.076***	-0.076***	-0.077***
Incarcerated	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Race	0.016	0.016	0.016	0.016	0.015
Asian	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Race	0.040***	0.041***	0.041***	0.041***	0.040***
Black	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Race	0.033**	0.032**	0.032**	0.032**	0.031**
Latino	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Program	0.006	0.006	0.006	0.006	0.005
Structured	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Position	-0.027**	-0.023	-0.028**	-0.024*	-0.018
Welfare	(0.010)	(0.014)	(0.010)	(0.010)	(0.027)
Position	-0.080***	-0.107***	-0.086***	-0.092***	-0.106***
Incarcerated	(0.009)	(0.014)	(0.009)	(0.010)	(0.027)
Position x Punitiveness	0.001				
Welfare	(0.004)				
Position x Punitiveness	-0.023***				
Incarcerated	(0.004)				
Position x Rehabilitative Orientation		-0.002			
Welfare		(0.005)			
Position x Rehabilitative Orientation		0.016**			
Incarcerated		(0.005)			
Position x LE Direct 1			-0.009		
Welfare			(0.008)		
Position x LE Direct 1			-0.046***		
Incarcerated			(0.008)		
Position x LE Direct 2				0.009	
Welfare				(0.008)	
Position x LE Direct 2				-0.047***	
Incarcerated				(0.008)	
Position x News Consumption					-0.004
Welfare					(0.011)
Position x News Consumption					0.013
Incarcerated					(0.011)
N	11488	11511	11517	11523	11541

Table reports AMCEs and heterogeneous effects across five models corresponding to those in Figure C.2, Figure 4, and Figure C.5. Standard errors, clustered by respondent, shown in parentheses. * $p < .05$; ** $p < .01$; *** $p < .001$

Table E.6: Table with full models from Figure 5

	Ideology
Cost-Benefit	0.002
Unequal	(0.011)
Effort	0.588***
Hardworking	(0.016)
Community	0.098***
Belongs	(0.012)
Position+Offense	-0.016
Welfare	(0.015)
Position+Offense	-0.011
Incarcerated-shoplifting	(0.018)
Position+Offense	-0.018
Incarcerated-burglary	(0.018)
Position+Offense	-0.058***
Incarcerated-robbery	(0.017)
Race	0.041**
Asian	(0.016)
Race	0.085***
Black	(0.015)
Race	0.072***
Latino	(0.015)
Program	0.007
Structured	(0.011)
Position+Offense x Ideology	-0.008
Moderate x Welfare	(0.026)
Position+Offense x Ideology	-0.029
Conservative x Welfare	(0.022)
Position+Offense x Ideology	-0.042
Moderate x Incarcerated-shoplifting	(0.030)
Position+Offense x Ideology	-0.074**
Conservative x Incarcerated-shoplifting	(0.027)
Position+Offense x Ideology	-0.080*
Moderate x Incarcerated-burglary	(0.032)
Position+Offense x Ideology	-0.102***
Conservative x Incarcerated-burglary	(0.027)
Position+Offense x Ideology	-0.097**
Moderate x Incarcerated-robbery	(0.033)
Position+Offense x Ideology	-0.099***
Conservative x Incarcerated-robbery	(0.028)
Effort x Ideology	0.006
Hardworking x Moderate	(0.026)
Effort x Ideology	0.027
Hardworking x Conservative	(0.023)
Community x Ideology	0.014
Moderate x Belongs	(0.021)
Community x Ideology	-0.024
Conservative x Belongs	(0.018)
Program x Ideology	-0.006
Moderate x Structured	(0.019)
Program x Ideology	0.000
Conservative x Structured	(0.017)
Cost-Benefit x Ideology	0.031
Unequal x Moderate	(0.019)
Cost-Benefit x Ideology	-0.002
Unequal x Conservative	(0.017)
Race x Ideology	-0.036
Moderate x Asian	(0.027)
Race x Ideology	-0.054*
Conservative x Asian	(0.024)
Race x Ideology	-0.066*
Moderate x Black	(0.026)
Race x Ideology	-0.091***
Conservative x Black	(0.023)
Race x Ideology	-0.088**
Moderate x Latino	(0.028)
Race x Ideology	-0.063**
Conservative x Latino	(0.024)
N	11541

xxix

Table reports AMCEs and heterogeneous effects for the model corresponding to those in Figure 5. Standard errors, clustered by respondent, shown in parentheses. * $p < .05$; ** $p < .01$; *** $p < .001$

F Robustness: Attention, Speed, and Fatigue

Table F.1 presents estimates corresponding to Figure 1 but separately for each conjoint task to test for task-order effects or fatigue; the table includes the main results for comparison (All Tasks column). The results are consistent across conjoint tasks. Other than expected changes in statistical significance (due to smaller sample size for individual tasks), the results are largely consistent across tasks.

Table F.1: AMCEs of main specification, by task

	All Tasks	Task1	Task2	Task3	Task4	Task5	Task6
Cost-Benefit	0.008	-0.001	0.019	0.009	-0.007	0.009	0.015
Unequal	(0.007)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Effort	0.596***	0.517***	0.604***	0.615***	0.620***	0.611***	0.610***
Hardworking	(0.010)	(0.019)	(0.018)	(0.018)	(0.017)	(0.018)	(0.017)
Community	0.094***	0.108***	0.090***	0.095***	0.097***	0.066***	0.108***
Belongs	(0.008)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Position	-0.028**	-0.058*	-0.013	-0.000	-0.022	-0.036	-0.037
Welfare	(0.010)	(0.026)	(0.024)	(0.022)	(0.024)	(0.024)	(0.024)
Position	-0.077***	-0.101***	-0.064**	-0.050*	-0.069**	-0.105***	-0.071***
Incarcerated	(0.009)	(0.023)	(0.022)	(0.021)	(0.021)	(0.021)	(0.021)
Race	0.015	-0.016	0.015	0.006	0.018	0.025	0.043
Asian	(0.011)	(0.027)	(0.026)	(0.025)	(0.026)	(0.024)	(0.025)
Race	0.040***	0.059*	0.042	0.017	0.037	0.056*	0.030
Black	(0.010)	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Race	0.032**	-0.005	0.020	0.057*	0.019	0.054*	0.041
Latino	(0.011)	(0.028)	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)
Program	0.006	0.008	-0.007	-0.032	0.029	0.032	0.002
Structured	(0.007)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
N	11547	1925	1924	1925	1926	1923	1924

Table reports AMCEs from separate models for each of the six conjoint tasks as a robustness for task-order or fatigue effects. Baseline categories are omitted from the table. Standard errors shown in parentheses. Standard errors for the “All Tasks” model are clustered by respondent. * $p < .05$; ** $p < .01$; *** $p < .001$

Table F.2 presents estimates of the same model as in Figure 1 for subsets of respondents. First, we show estimates among respondents who passed all three manipulation checks, among respondents who passed two of the manipulation checks, and respondents who only passed one of the manipulation checks (there are too few respondents who failed all checks to estimate the model in that subsample). The other 4 columns estimate the same model on subsamples that split respondents by quartiles of survey completion time.

While there are some differences in effect sizes (and significance levels), the results are generally consistent. The only sample that has clearer deviations is the one composed of respondents who passed exactly 1 manipulation check. However, that sample is very small, composed of only 197 respondent-tasks.

Table F.2: AMCEs of main specification, by manipulation checks and completion speed

	Full Sample	Passed All 3	Passed 2	Passed 1	Dura- tionQ1	Dura- tionQ2	Dura- tionQ3	Dura- tionQ4
Cost-Benefit	0.008 (0.007)	0.004 (0.010)	0.006 (0.011)	0.097 (0.067)	0.022 (0.015)	0.019 (0.014)	0.002 (0.014)	-0.011 (0.015)
Unequal Effort	0.596*** (0.010)	0.603*** (0.015)	0.599*** (0.016)	0.346*** (0.084)	0.568*** (0.021)	0.614*** (0.020)	0.619*** (0.020)	0.587*** (0.021)
Hardworking Community	0.094*** (0.008)	0.103*** (0.011)	0.081*** (0.012)	0.082 (0.077)	0.081*** (0.017)	0.093*** (0.016)	0.102*** (0.016)	0.101*** (0.016)
Belongs Position	-0.028** (0.010)	-0.023 (0.013)	-0.030* (0.015)	-0.099 (0.079)	-0.034 (0.020)	0.001 (0.019)	-0.024 (0.018)	-0.053** (0.020)
Welfare Position	-0.077*** (0.009)	-0.060*** (0.012)	-0.088*** (0.015)	-0.368*** (0.091)	-0.081*** (0.019)	-0.059** (0.019)	-0.077*** (0.018)	-0.087*** (0.018)
Incarcerated Race	0.015 (0.011)	0.008 (0.014)	0.026 (0.016)	-0.026 (0.104)	0.037 (0.023)	-0.005 (0.020)	-0.001 (0.020)	0.027 (0.022)
Asian Race	0.040*** (0.010)	0.037** (0.013)	0.046** (0.016)	0.062 (0.093)	0.044* (0.021)	0.031 (0.019)	0.040* (0.020)	0.046* (0.021)
Black Race	0.032** (0.011)	0.031* (0.014)	0.033* (0.016)	0.036 (0.074)	0.025 (0.022)	0.008 (0.020)	0.062** (0.020)	0.033 (0.023)
Latino Program	0.006 (0.007)	-0.002 (0.010)	0.017 (0.011)	-0.018 (0.065)	-0.016 (0.016)	-0.014 (0.015)	0.032* (0.014)	0.019 (0.015)
Structured N	11547	6443	4896	197	2887	2887	2887	2886

Table reports AMCEs from robustness checks on the choice outcome. The first column corresponds to the original model reported in Figure 1. The next three columns focus on respondents who passed only 1, exactly 2, or all 3 of the manipulation checks. The final four columns split the sample by quartiles of completion speed. Baseline categories are omitted from the table. Standard errors, clustered by respondent, are shown in parentheses. * $p < .05$; ** $p < .01$; *** $p < .001$

G Hypotheses Crosswalk

All hypotheses presented in the Pre-Analysis Plan (see Appendix H) are analyzed in the paper, either in the main text or in the appendices. Some of the hypotheses from the PAP are presented as the main central hypotheses in the manuscript (less eligibility, main effect of conjoint attributes, conjoint attributes moderation of less eligibility), whereas others are presented in a more informal way. Most of the PAP hypotheses are presented in figures in the main text or discussed there. Only a handful are discussed in the Appendix only.

Table G.1: Hypothesis crosswalk

PAP Hypothesis	Paper Hypothesis	Figure/Table (Choice Outcome)
H1	H1	Figure 1; Table E.1
H2		Figure 2; Table E.1
H3	H3A	Figure 1; Table E.1
H4	H2A	Figure 1; Table E.1
H5	H4A	Figure 1; Table E.1
H6	H5A	Figure 1; Table E.1
H7	H5B	Figure 3; Table E.1
H8		Figure 5; Table E.6
H9		Figure 1; Table E.1
H10		Figure C.4; Table E.2
H11		Figure 5; Table E.6
H12	H3B	Figure 3; Table E.1
H13	H4B	Figure 3; Table E.1
H14	H2B	Figure 3; Table E.1
H15		Figure C.3; Table E.2
H16		Figure C.7; Table E.5
H17		Figure C.6; Table E.4
H18		Figure 4; Table E.5
H19		Figure C.5; Table E.5

Correspondence of hypotheses from the Pre-Analysis Plan (see Appendix H) and the paper. An empty cell in the Paper Hypothesis column indicates that the hypothesis was not assigned a name in the manuscript (but it is still evaluated). The Figure/Table column shows the main figure used to evaluate the hypothesis, and the Appendix table that contains that model's results.

H Pre-Analysis Plan

Pre-Analysis Plan for Incarceration and Deservingness: Experimental Evidence on Less Eligibility

Posted to OSF: June 22, 2026

1 Overview and Theoretical Framework

This study examines whether incarceration status reduces public support for access to government programs, consistent with a principle of *less eligibility*. The core theoretical claim is that incarcerated individuals are normatively evaluated as less deserving than similarly situated low-income individuals in the community, even when programs produce substantial social benefits.

The design permits separation of various moral-status effects both from each other and from instrumental cost-benefit reasoning.

2 Design

Respondents complete a vignette-based conjoint experiment in which they evaluate an applicant to a government program that would provide selected applicants with one year of free college-level courses and job counseling. Program and applicant attributes vary across vignettes. All applicant attributes are independently randomized subject to constraints described below. Each respondent completes this task a total of 6 times.

Our single vignette design includes a total of 8 randomly assigned attributes: social position (primary treatment), offense severity (constrained), effort, community belonging, program structure, cost-benefit information, applicant race, and applicant name.

The social position attribute is the primary treatment and refers to the status of the applicant to the program. The attribute's levels are: near the poverty line, currently on welfare, and incarcerated in a state prison. The poverty and welfare levels are given a 0.25 probability of occurring in the vignette, whereas the incarcerated level is given a 0.5 probability.

If the applicant is incarcerated, we randomly vary the severity of the offense committed by the applicant: shoplifting, burglary of an unoccupied home, and armed robbery. All severity levels are assigned with equal probability. Profiles are constrained in the sense that offense severity is only included for vignettes in which social position is incarcerated, but not for vignettes in which social position is poverty or welfare.

The next suite of attributes is designed to capture different aspects of deservingness. Applicants are distinguished by their general effort (hard-working vs. unreliable), and the extent to which they belong to the social community (belongs vs. does not belong). The program itself is distinguished by its structure (structured and demanding vs. flexible and supportive), and by its cost-benefit analysis (equal benefit vs. unequal benefit).

The final two attributes we randomize relate to the applicant to the program: race (with levels White, Black, Latino, and Asian), and applicant name. Applicant names are only used to make the profiles more realistic, and are drawn without replacement from a list of 6 names selected due to their sufficient racial ambiguity (Albert, Henry, Isaac, Joe, Jordan, and Samuel).

3 Outcomes

For each vignette, we measure two outcomes. The first is a binary policy choice (yes/no), where we ask the respondents whether their state should offer the program to people like the applicant. The second is a 1 to 5 scale measuring opposition/support for the applicant receiving the program. The binary choice is the primary outcome that will be used in the analysis, with the support scale used as a robustness and polarization measure.

4 Hypotheses

We have a variety of hypotheses that we intend to test with this framework, some of which are primary hypotheses and others are more exploratory in nature.

- H1: Incarceration Penalty (Less Eligibility).** Applicants who are incarcerated will receive lower support than applicants who are either near the poverty line or on welfare.
- H2: Offense Severity Ordering.** Among incarcerated applicants, support will be higher among those who committed shoplifting, than among those who committed burglary, than among those who committed armed robbery.
- H3: Effort Effect.** Applicants described as hard-working and reliable will receive greater support than applicants described as unreliable.
- H4: Community Belonging Effect.** Applicants described as belonging to the respondents' broader community will receive greater support than those described as not belonging.
- H5: Program Structure Effect.** Highly structured and demanding programs will receive greater support than flexible and supportive programs.
- H6: Instrumental Benefit Effect.** Cost-benefit information about the program indicating reduced recidivism, improvement in employment outcomes, and taxpayer savings will increase support.
- H7: Moral Dominance Condition.** Cost-benefit information will reduce but not completely eliminate the incarceration penalty relative to comparable non-incarcerated applicants.

H8: Ideological Moderation. On a standard conservative-to-liberal self-placement scale, we expect:

- A larger incarceration penalty among conservatives.
- A steeper offense gradient among conservatives.
- Greater responsiveness to harshness framing among conservatives.
- Greater responsiveness to cost-benefit information among liberals.

H9: Race Main Effect. Applicants described as Black will receive lower support than applicants described as non-Black, net of all other attributes. We expect this effect to hold across social position conditions, reflecting generalized racial bias in deservingness judgments, independent of incarceration status.

H10: Race $\times$ Offense Severity Interaction. The penalty associated with moving up the offense severity gradient will be larger for Black applicants than for non-Black applicants. That is, the gap in support between low- and high-severity offenses will be steeper for Black incarcerated applicants, consistent with the racialization of crime in American public perception. Formally, we expect the offense gradient to be moderated by applicant race, with the largest penalties accruing to Black applicants at the high end of the severity scale.

H11: Race $\times$ Ideology Interaction. The gap in support between Black and non-Black applicants will be larger among conservatives than among liberals, consistent with the racial attitudes literature.

H12: Effort $\times$ Social Position Interaction. The effect of effort cues on support will vary across social position conditions, following the logic of deservingness baselines. Applicants near the poverty line start near a ceiling of perceived deservingness, leaving limited room for positive effort information to increase support further. Incarcerated applicants start near a floor of perceived deservingness, leaving substantial room for positive effort information to move support upward. We therefore expect:

- The effort effect to be largest for incarcerated applicants, for whom a hard-working characterization provides a redemptive signal that actively counters the low-deservingness prior.
- The effort effect to be smallest for poverty-line applicants, for whom hard-working is closer to the assumed default and adds little additional information.
- Welfare applicants to fall between these two extremes.

This asymmetry implies that the effort attribute does more work – and is more diagnostic – as social position descends the deservingness hierarchy.

H13: Program Structure $\times$ Social Position Interaction. The effect of program structure on support will be concentrated among welfare and incarcerated applicants and will be negligible for poverty-line applicants. A demanding, structured program may signal to respondents that recipients are being appropriately required to earn their benefits, which is a more salient consideration when the applicant’s deservingness is uncertain or low. For poverty-line applicants, whose deservingness is presumed, the structure framing carries less normative weight. We therefore expect the program structure effect to be largest for incarcerated applicants and smallest for poverty-line applicants, with welfare applicants intermediate.

H14: Community Belonging $\times$ Social Position Interaction. The effect of community belonging information on support will differ across social position conditions, though the direction of this interaction is theoretically ambiguous and is therefore treated as an exploratory hypothesis.

One mechanism predicts a smaller community belonging effect for incarcerated applicants: if respondents read incarceration itself as an implicit signal of community non-membership – consistent with the sociological concept of civic death – then the explicit community belonging manipulation may be redundant for incarcerated participants. The “does not belong” condition confirms what incarceration already implied, and the “belongs” condition may be discounted as inconsistent with the incarceration cue. Under this mechanism, the community belonging manipulation does less independent work for incarcerated applicants than for non-incarcerated ones.

An alternative mechanism predicts a larger community belonging effect for incarcerated applicants. If respondents are genuinely uncertain about whether an incarcerated person belongs to their community, the explicit manipulation resolves that uncertainty and should be highly diagnostic, producing a larger effect than for poverty-line applicants whose community membership is not in question.

H15: Deservingness Bundle and Social Position. We construct a contrast between respondents receiving all three favorable deservingness signals simultaneously (hard-working, community member, demanding program) and those receiving all three unfavorable signals (unreliable, does not belong, flexible). We predict that this high-versus-low deservingness contrast will be larger in magnitude for incarcerated applicants than for non-incarcerated applicants, consistent with the floor-to-ceiling logic underlying H12 and H13.

As an exploratory extension, we will examine whether the three dimensions combine additively or exhibit diminishing returns as positive signals accumulate.

H16: Incarceration Contact and the Less Eligibility Effect. Respondents with direct or vicarious experience of incarceration will exhibit a smaller incarceration penalty than respondents with no such contact. We distinguish two types of contact in decreasing order of intimacy and expected effect size:

- Personal incarceration experience (self-reported history of ever being incarcerated): Strongest attenuation of the incarceration penalty, reflecting both greater personal identification with incarcerated applicants and direct experiential knowledge of prison conditions.
- Family member or close friend incarcerated: Intermediate attenuation, reflecting vicarious humanization through close ties.

We will construct a cumulative contact index (0-2) as a summary measure alongside separate indicators for each type. We will additionally examine whether incarceration contact predicts higher rehabilitative orientation scores (see H18), consistent with a pathway from contact to attitudinal liberalization to reduced less eligibility penalty.

H17: Victimization and the Less Eligibility Effect. Respondents who have experienced recent victimization (within the past 12 months) will exhibit a larger incarceration penalty than non-victims. We distinguish two types of victimization:

- Violent victimization (physical assault or threat of violence): Expected to produce a larger penalty than property victimization, reflecting a stronger retributive response to personal physical threat.

- Property victimization (burglary/attempted break-in or theft): Expected to increase the penalty, but less strongly than violent victimization.

We will examine both an additive index across victimization types and each category separately. We additionally expect victimization to predict higher punitiveness scores (see H18), consistent with a mechanism in which experienced harm shifts respondents towards retributive attitudes that amplify the incarceration penalty.

H18: Punitiveness, Rehabilitative Orientation, and the Less Eligibility Effect.

We measure three attitudinal constructs using Likert items from the post-experiment survey:

- Less eligibility endorsement (2-item scale): Agreement that incarcerated individuals should not access programs available to struggling law-abiding citizens; and that prisoner well-being should be deprioritized relative to non-prisoners.
- Punitiveness (2-item scale): Agreement that courts are insufficiently hard; and that sentences should be toughened.
- Rehabilitative orientation (2-item scale): Agreement that rehabilitation is a primary goal of prison and that most people who commit crimes can turn their lives around with appropriate programs and support.

From these, we consider:

H18a Moderation of experimental effect: Respondents scoring higher on punitiveness will exhibit a larger incarceration penalty. Respondents scoring higher on rehabilitative orientation will exhibit a smaller incarceration penalty and greater responsiveness to the cost-benefit information treatment.

H18b Punitiveness and rehabilitation as outcomes - Observational: Punitiveness and rehabilitative orientation scores serve as outcomes in observational analyses that do not rely on experimental variation. We expect:

- * Incarceration contact (self, family, friend) to predict lower punitiveness and higher rehabilitative orientation.
- * Recent victimization (especially violent) to predict higher punitiveness and lower rehabilitative orientation.
- * More frequent crime news consumption to predict higher punitiveness.
- * Conservative ideology to predict higher punitiveness and lower rehabilitative orientation.

These observational patterns provide non-experimental corroboration for the attitudinal mechanisms underlying the experimental moderation hypotheses (H16, H17, H19, and H8, respectively).

H19: Crime News Consumption and the Less Eligibility Effect. Respondents who follow crime news more frequently will exhibit a larger incarceration penalty, consistent with theory: repeated exposure to crime coverage increases the perceived prevalence and severity of crime and reinforces punitive attitudes toward offenders.

5 Sample

Our target sample is $N = 2,000$ respondents recruited via Prolific, an online participant recruitment platform to complete the survey on Qualtrics. Prolific operates a pre-screened

pool in which all participants have completed identity verification and are subject to monitoring and quality tracking. We will rely on Prolific’s built-in ‘representative samples’ which cross-stratifies on 5 age groups (18-24, 25-34, 35-44, 45-54, and 55+), 2 sex groups (male and female), and 3 political affiliation groups (Democrat, Republican, and Independent) matched to population data. Survey slots are filled to maintain the proportions of those groups. The sample is limited to participants residing in the United States and who speak English. We exclude from the participant pool a small set of respondents (10) who were recruited for a small pilot on Prolific.

To protect data quality, Prolific’s Authenticity Checks were enabled during data collection, including identity verification, quality monitoring of the participant pools, automatic detection for extremely fast completion, and a bot authenticity check to screen out non-human participants. Three attention checks are included within the survey instrument. The first is an instruction check that appears immediately before the conjoint vignettes and asks respondents how many vignettes they will be exposed to. The second and third are factual manipulation checks that appear immediately after the final conjoint vignette and ask about the type of program that appeared in the vignettes (multiple choice) and the value of the effort attribute in the final vignette (multiple choice as well). The attention checks will be used to explore the robustness of the results.

6 Analysis

The main outcome of interest is the choice variable (whether to offer the program), while the scale (support for applicant receiving the program) variable will be used in robustness checks.

The analysis will rely on the `cjoint` package on R to estimate Average Marginal Component Effects (AMCE), Average Component Interaction Effects (ACIE), and heterogeneous effects that depend on respondent-varying interactions (e.g., ideology). In all cases, standard errors will be clustered by respondent.

For analyses that do not require a focus on the severity attribute, the main specification will include all the attributes with the exception of the severity attribute. To account for the fact that the social position attribute does not have equal probabilities, we will declare a design using the `makeDesign` option in the `cjoint` package. This main specification will be used to test hypotheses H1 (Less Eligibility), H3 (Effort Effect), H4 (Community Belonging Effect), H5 (Program Structure Effect), H6 (Instrumental Benefit Effect), and H9 (Race Main Effect).

For analyses that require a focus on the severity attribute, we will use an alternative specification. This alternative specification combines the social position and offense severity attributes into a single attribute with the following levels: poverty, welfare, incarcerated-robbery, incarcerated-burglary, and incarcerated-shoplift. The reason for combining these two attributes into a single one is that they are otherwise perfectly correlated (severity is only randomized for those who are incarcerated). To account for the fact that the levels of the combined attribute do not have equal probabilities, we rely on the `makeDesign` option in `cjoint` to declare these unequal probabilities: poverty (1/4), welfare (1/4), incarcerated-robbery (1/6), incarcerated-burglary (1/6), and incarcerated-shoplift (1/6). This approach will be used to test hypotheses that require consideration of offense severity: H2 (Offense Severity Ordering) and H7 (Moral Dominance Condition).

For hypotheses that involve the moderation/interaction of one conjoint attribute by another conjoint attribute, we will rely on the estimation of Average Component Interaction Effects (ACIE). Each ACIE will be evaluated separately. In cases in which severity is not relevant to the analysis, we will rely on a version of the main specification that additionally includes an interaction term between the relevant attributes. This is the case for the hypotheses that interact social position with effort (H12), program structure (H13), and community belonging (H14). For hypothesis H10, which considers the interaction between the race and offense severity attributes, we will rely on the specification that uses the combined attribute with the addition of the relevant interaction between the attributes.

For H15, which considers the deservingness bundle, we will define a new design in which the three attributes considered in the bundle are combined into a single one with three levels: all three favorable (0.125), three unfavorable (0.125), and mixed (0.75). A model with all remaining attributes and an interaction between the bundle and the social position attributes will be estimated to test this hypothesis.

For the hypotheses that involve the moderation/interaction of a conjoint attribute with a respondent characteristic, we will estimate heterogeneous effects within the `cjoint` package. For hypotheses H8 (Ideological Moderation) and H11 (Race $\times$ Ideology Moderation), we will interact the conjoint attributes with a three-point ideology scale (Conservative, Moderate, Liberal). For hypothesis H16 (Incarceration Contact and Less Eligibility Effect), we will interact the incarceration contact variables (personal, family/friend, and cumulative index) with the corresponding conjoint attribute. Similarly, for H17 (Victimization and Less Eligibility Effect), we will interact victimization (violent, property, any) with the corresponding conjoint attribute. In the same manner, for H19 (Crime News Consumption and Less Eligibility Effect), we will interact the crime consumption variable with the corresponding conjoint attribute.

For hypothesis H18a, which considers the moderation of the less eligibility effect by respondents' punitiveness and rehabilitative orientation scores, we will interact each of these scores with the corresponding conjoint attributes.

Finally, for hypothesis H18b (Observational), we will estimate linear regressions in which the outcomes are the punitiveness and rehabilitative orientation scores, and the independent variables are: respondent gender, race, age, employment status, ideological orientation, incarceration contact dummies, victimization dummies, and crime news consumption.

7 Power Analysis

The relatively large sample size of $N = 2,000$ was chosen to gain reasonably sufficient statistical power to test several of the hypotheses that involve interaction effects and heterogeneous effects. We rely on Freitag and Schuessler's `cjpowR` for power calculations.¹ In all cases, we calculate required minimum sample sizes with a target of 80% power, assuming an AMCE of 0.05, and relying on a standard significance level of $\alpha = 0.05$.

For the estimation of Average Marginal Component Effects (AMCE), the relevant char-

¹Freitag, Markus and Julian Schuessler (2024). "cjpowR - A Priori Power Analyses for Conjoint Experiments" R Package.

acteristic is the number of levels of an attribute. For the attributes that have two levels (program structure, community belonging, effort, and cost-benefit attributes), we find a minimum required sample size of 522. With a target sample size of 2,000 respondents, the statistical power for detecting effects in these attributes approaches 1.

For the key attribute of social position, we expect a similar power for detecting a difference between the incarcerated level and the other two levels, due to the fact that the incarcerated level occurs 50% of the time (see Design section). However, for distinguishing the levels of poverty and welfare from each other, each of which occurs 25% of the time, we expect a minimum required sample size of 1,044 (the equivalent of a 4-level attribute) and thus an expected power of 0.97 with 2,000 respondents.

For the attribute of race, which has 4 levels, the minimum required sample size is 1,044 respondents, and the expected power with 2,000 respondents is 0.97.

The offense severity gradient hypothesis (H2) is equivalent to an attribute with 6 levels (because each offense level occurs with 1/6 probability). The minimum required sample size for this hypothesis is thus 1,566 respondents, and the expected power with a sample size of 2,000 respondents is 0.89.

Thus, based on this power analysis and assuming effect sizes of 0.05, the target sample size of 2,000 respondents should provide us with sufficient statistical power to test hypotheses H1, H2 (more weakly), H3, H4, H5, H6, and H9.

The remaining hypotheses rely on either interactions between conjoint attributes or with respondent characteristics. For the estimation of Average Component Interaction Effects (ACIE) that involve two attributes with the functional equivalent of two levels, the minimum required sample size is 2,080, and the expected power with a sample size of 2,000 respondents is 0.78. This means that the statistical power will possibly be sufficiently high for testing hypotheses H7, H12, H13, and H14, but likely insufficient (unless effect sizes are larger) for hypotheses H10 and H15.

For the hypotheses involving heterogeneous effects, estimating power a priori is difficult because we do not know the distribution of the respondent characteristics being used. However, for the two hypotheses involving ideology (H8 and H11), we expect that power will be sufficiently high given the relatively large differences in views between Democrats and Republicans on many issues, which will likely lead to large heterogeneous effects.