Essays in Public and Behavioral Economics

by

Charlie Rafkin

A.B. Mathematics, Dartmouth College, 2016

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ECONOMICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2024

Authored by: Charlie Rafkin
Department of Economics
May 15, 2024

Certified by: Amy Finkelstein
John & Jennie S. MacDonald Professor of Economics, Thesis Supervisor

Certified by: Frank Schilbach
Associate Professor of Economics, Thesis Supervisor

Certified by: James Poterba
Mitsui Professor of Economics, Thesis Supervisor

Accepted by: Isaiah Andrews
Professor of Economics
Chairman, Departmental Committee on Graduate Theses

Essays in Public and Behavioral Economics

by

Charlie Rafkin

Submitted to the Department of Economics
on May 15, 2024 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ECONOMICS

ABSTRACT

This thesis examines how psychological forces and non-standard preferences affect poverty and the design of social welfare programs for low-income households.

The first chapter, "Eviction as Bargaining Failure: Hostility and Misperceptions in the Rental Housing Market" (co-authored with Evan Soltas), studies the causes of evictions from rental housing and the welfare impact of policy interventions to address them. Court evictions from rental housing are common but could be avoided if landlords and tenants bargained instead. Such evictions are inefficient if they are costlier than bargaining. We test for two potential causes of inefficient eviction — hostile social preferences and misperceptions — by conducting lab-in-the-field experiments in Memphis, Tennessee with 1,808 tenants at risk of eviction and 371 landlords of at-risk tenants. We detect heterogeneous social preferences: 24% of tenants and 15% of landlords exhibit hostility, giving up money to hurt the other in real-stakes Dictator Games, yet more than 50% of both are highly altruistic. Both parties misperceive court or bargaining payoffs in ways that undermine bargaining. Motivated by the possibility of inefficient eviction, we evaluate the Emergency Rental Assistance Program, a prominent policy intervention, and find small impacts on eviction in an event-study design. To quantify the share of evictions that are inefficient, we estimate a bargaining model using the lab-in-the-field and event-study evidence. Due to hostile social preferences and misperceptions, one in four evictions results from inefficient bargaining failure. More than half would be inefficient without altruism. Social preferences weaken policy: participation in emergency rental assistance is selected on social preferences, which attenuates the program's impacts despite the presence of inefficiency.

The second chapter, "The Welfare Effects of Eligibility Expansions: Theory and Evidence from SNAP" (co-authored with Jenna Anders), studies the U.S. rollout of eligibility

expansions in the Supplemental Nutrition Assistance Program. Using administrative data from the U.S. Department of Agriculture, we show that expanding eligibility raises enrollment among the inframarginal (always-eligible) population. Using an online experiment and an administrative survey, we find evidence that information frictions, rather than stigma, drive the new take-up. To interpret our findings, we develop a general model of the optimal eligibility threshold for welfare programs with incomplete take-up. Given our empirical results and certain modeling assumptions, the SNAP eligibility threshold is lower than optimal.

The third chapter, "Preferences for Rights" (co-authored with Aviv Caspi and Julia Gilman), observes that public discourse about in-kind transfers often appeals to "preferences for rights" — for instance, the "right to health care" or "right to counsel" for indigent legal defense. Preferences for rights are "non-welfarist" if the person values the right per se, holding fixed how the right instrumentally affects others' utilities. We test for non-welfarist preferences for rights, and their relationship to redistributive choices, with incentivized online experiments (N = 1,800). Participants face choices about allocating rights goods (lawyers, health care) and benchmark goods (bus passes, YMCA memberships) to tenants facing eviction. We implement a share of choices. In two of three experiments, more than half of participants allocate rights goods in ways that are consistent with preferences for rights and dominated if preferences were entirely welfarist. Dominated behaviors are more common with rights goods than benchmarks. In a fourth experiment, those with preferences for rights also exhibit "anti-targeting," where they redistribute lawyers and health care more universally than benchmark goods to recipients whose incomes differ. At least 26% of participants are non-welfarist, while at most 31% are welfarist.

Thesis supervisor: Amy Finkelstein
Title: John & Jennie S. MacDonald Professor of Economics

Thesis supervisor: Frank Schilbach
Title: Associate Professor of Economics

Thesis supervisor: James Poterba
Title: Mitsui Professor of Economics

Acknowledgments

I am enormously grateful to my advisors: Amy Finkelstein, Frank Schilbach, and Jim Poterba. So many times, Amy has helped me salvage an insight buried underneath pages of sloppy thinking. To flesh out an idea, examine a hypothesis, or clean up some unclear prose, I start by asking how Amy would proceed. Frank encouraged my creativity, and taught me how to do research on non-standard topics without sacrificing rigor. He showed tremendous patience — no one should have to suffer through as many iterations of my job-market talk as he did — and care for my personal well-being during the toughest parts of grad school. Jim's way of cutting to the core economics made a huge impression on me. All three of my advisors set a high bar for my work, but never failed to exhibit warmth and good humor. I will miss our joint advising meetings, and I hope to pay forward their generosity.

I have also benefited from remarkable mentoring. Paul Novosad started coaching me when I was a raw undergrad with no research skills, and has been a collaborator and sounding board for nearly a decade. During my predoc, John Beshears, James Choi, Brigitte Madrian, and David Laibson taught me new skills and gave me freedom to explore my interests. Hunt Allcott helped me become a more rigorous thinker, and showed me how to execute a research project (and dinner party). Abhijit Banerjee's well-timed suggestions to pursue other tacks that only he could see yielded big improvements to my work.

My other co-authors Jenna Anders, Sam Asher, Aviv Caspi, Julia Gilman, Advik Shreekumar, Adam Solomon, Evan Soltas, and Pierre-Luc Vautrey have likewise been huge influences. They have taught me so much economics, humored me when I proposed dead ends or nitpicked minutia, and made doing research worthwhile. I look forward to continuing to learn from them in future work. Evan deserves a particular thank you for co-authoring my job-market paper; his knowledge and friendship were key to the project, and to my grad-school trajectory.

I have found wonderful friends during grad school. Partick Schwarz schooled me for years in everything we did — from problem sets, to rock climbing, to chess — and his humor and mature perspective made the PhD far more enjoyable. Lively debates and constant jokes with Jon Cohen and Anna Russo powered me through grad school's ups and downs. Basil Halperin and Joel Flynn provided levity and a different point of view on research and life. It has been a blast to learn from, and goof around with, Pedro Bessone, Roberto Corrao, Lisa Ho, Ed Kong, Omeed Maghzian, Shakked Noy, Abby Ostriker, Lucy Page, Hannah Ruebeck, Nagisa Tadjfar, and Sean Wang, among many others. I also thank my non-econ friends, who send me funny memes, and support me in many other ways besides that.

My work in Memphis was made possible via collaborations with The Works, Inc. and the Memphis/Shelby County Emergency Rental and Utilities Assistance Program. I thank Mairi Albertson, Kayla Billingsley, Ashley Cash, Karen Gause, Dorcas Young Griffin, and especially Roshun Austin, Steve Barlow, Brian Rees, and Jenna Richardson, for their commitment to research and extensive help with implementation.

Without my parents' and sisters' love, I would not be a fraction of the person or economist. I am profoundly shaped by my parents' broad interests, grit, concern for others, and sense of humor. Anything good in my work bears the marks of their intellectual influence and social conscience.

I owe my greatest gratitude to my partner, Margaret Haltom. The most important lessons of this period of dissertation writing have come from spending time with someone with such clarity of purpose, brilliant insight, and genuine creativity. Margaret introduced me to many of the topics in my dissertation, and sacrificed much for our relationship during grad school. The combination of silliness and solace that I have found in her partnership are precious to me.

Contents

List of Figures

List of Tables

Chapter 1

Eviction as Bargaining Failure: Hostility and Misperceptions in the Rental Housing Market

This chapter is coauthored with Evan Soltas.[1]

[1]Disclosure: Funding and in-kind assistance for this project was provided by The Works, Inc. (TWI), a nonprofit organization that is supported in part by a grant from the Memphis/Shelby County Emergency Rental Assistance Program (ERAP). Rafkin received financial support from TWI's ERAP grant to cover graduate school tuition and release time from teaching. Rafkin's partner was TWI's Director of Emergency Rent Assistance and Housing Policy for part of the time when this project was being carried out. The authors' data use agreements with TWI, the City of Memphis/Shelby County, and the Legal Services Corporation provide the authors with full editorial control with regard to the reporting of research findings. We thank our collaborators at the City of Memphis and Shelby County, Tennessee and TWI, including Mairi Albertson, Roshun Austin, Steve Barlow, Kayla Billingsley, Webb Brewer, Ashley Cash, Karen Gause, Margaret Haltom, Nicholas Thompson, Dorcas Young Griffin, and Paul Young. We thank Abhijit Banerjee, Amy Finkelstein, Jim Poterba, and Frank Schilbach for their mentorship, guidance, and support. For helpful feedback, we thank Hunt Allcott, Jenna Anders, David Autor, Sam Asher, Leonardo Bursztyn, Theo Caputi, Eric Chyn, Jon Cohen, Rob Collinson, Stefano DellaVigna, Esther Duflo, Sarah Eichmeyer, Joel Flynn, Peter Ganong, Laura Gee, Jon Gruber, Nathan Hendren, Peter Hepburn, Simon Jäger, Ray Kluender, David Laibson, Jing Li, Jeffrey Liebman, Sara Lowes, Elisa Macchi, Sendhil Mullainathan, Paul Novosad, Shakked Noy, Jett Pettus, Chris Roth, Anna Russo, Tobias Salz, Andrei Shleifer, Chris Snyder, Doug Staiger, Dmitry Taubinsky, Muhamet Yildiz, Daniel Waldinger, Winnie van Dijk, and Jonathan Zinman, as well as seminar participants at Dartmouth and MIT. We thank Jenna Richardson for project management. We thank the Legal Services Corporation for sharing data on eviction court and Daniel Bernstein for his guidance on using the data. In addition to funding disclosed above, we acknowledge funding from the Lynde and Harry Bradley Foundation; the Hausman Dissertation Fellowship; The Institute of Consumer Money Management (ICMM) Pre-doctoral Fellowship on Consumer Financial Management, awarded through the National Bureau of Economic Research; the Lincoln Institute of Land Policy; the MIT Shultz Fund; and the National Science Foundation Graduate Research Fellowship under Grant No. 1122374. This study was approved by MIT's Committee on the Use of Humans as Experimental Subjects under protocol #2102000316 and the experiments were pre-registered at the AEA RCT Registry under AEARCTR-0008053 (landlord experiment), AEARCTR-0008436 (general population samples), and AEARCTR-0008975 (tenant experiment).

1.1 Introduction

Evictions are costly and common. Tenants are often traumatized and lose possessions when evicted (Desmond, 2016), and landlords face costs from vacancies and property damage. Formal evictions — i.e., those involving a court order — increase homelessness, induce financial distress, and impose court costs on both parties (Collinson et al., 2024). Despite these costs, courts in the United States give eviction orders to around 2% of renters each year (Gromis et al., 2022; Graetz et al., 2023).[2] During the pandemic, evictions' costs motivated the $40-billion Emergency Rental Assistance Program (ERAP), which nearly doubled the $50 billion spent annually on low-income rental assistance. Advocates have proposed a permanent version of ERAP, describing the program as "the most important eviction prevention policy in American history."[3]

In a Coasean world with frictionless landlord–tenant bargaining, all formal evictions would be privately efficient. Then, since landlords and tenants could otherwise bargain to avoid court, formal evictions would only occur when their benefits to the parties exceed their costs. According to this view, evictions may be costly but still better than bargained alternatives. Eviction's private costs have attracted substantial scholarly and policy attention. But if evictions are efficient, the case for policy intervention to stop evictions hinges on other arguments like externalities or redistribution, rather than eviction's private costs.[4]

Still, two forces suggest that landlord–tenant bargaining is not always frictionless, so some evictions could be undesirable because of private costs alone. First, hostility (i.e., anti-altruistic social preferences) in landlord–tenant relationships could impede efficient bargaining. In that case, formal evictions might even be "repugnant" — that is, normatively undesirable if the social planner does not respect hostile preferences. Second, misperceptions about eviction or bargaining costs, perhaps arising from eviction's complicated legal environment, may cause inefficient evictions. Whether formal evictions are driven by bargaining costs, hostility, or misperceptions affects the desirability and efficacy of anti-eviction policy versus other ways of assisting the poor.

This paper seeks to answer three main questions. How prevalent are hostility and

[2]Surveys suggest formal evictions constitute perhaps 15% (Gromis and Desmond, 2021) to 33% (Desmond and Shollenberger, 2015) of all forced moves.

[3]For example, the Eviction Crisis Act introduced in the Senate would allocate $3 billion to permanent emergency rental assistance for renters facing eviction. The quotation is from "Fact Sheet: White House Summit on Building Lasting Eviction Prevention Reform," The White House, August 2022.

[4]For instance, Collinson et al. (2024) find fiscal externalities via, e.g., stays in homeless shelters.

misperceptions among tenants facing eviction and their landlords? Do hostility and misperceptions cause a meaningful share of formal evictions? And what are the consequences of emergency rental assistance for formal evictions — efficient, inefficient, and repugnant — and welfare?

We study these questions in four parts. First, we outline a simple model that shows how formal evictions may emerge from bargaining costs, hostility, and misperceptions. Second, we test for the presence and quantitative importance of hostile social preferences and misperceptions by conducting lab-in-the-field experiments with landlords and tenants facing eviction, linking their behaviors in the experiment to realized formal evictions. Third, we study ERAP's causal effects with event studies, and use data from the experiment to explain its effectiveness. Fourth, we combine the experimental and observational moments with the model to quantitatively estimate the importance of each force for evictions, conduct welfare analysis, and analyze counterfactuals.

Section 1.2 presents a simple bargaining model that illustrates when evictions occur, whether eviction is efficient or inefficient, and how policy affects eviction and welfare. We define eviction as formal eviction and view informal evictions as implicitly bargained outcomes that avoid court costs.[5] Absent misperceptions and social preferences, this framework yields a classical benchmark: eviction occurs if and only if efficient, that is, if and only if net bargaining costs exceed net court costs. By contrast, misperceptions (e.g., if landlords overestimate court payoffs) and hostile social preferences can cause inefficient or repugnant evictions. We then introduce a government program modeled after ERAP in which landlords receive a payment for the tenant's back rents but pay a private take-up cost. The program's effect on evictions depends on court costs, misperceptions, and social preferences among enrollees. For instance, those who enroll could be altruists who would rarely evict, attenuating ERAP's effect. Moreover, if ERAP does stop evictions, these forces also govern whether ERAP stops efficient, inefficient, or repugnant ones. The nature of evictions and the impact of anti-eviction policy are thus empirical questions and the focus of our analysis.

Our setting of Memphis, Tennessee is well-suited for studying evictions (Section 1.3). Memphis is a large city with high rates of poverty and housing insecurity. Bargaining

[5]Although causal evidence is scarce, formal evictions may be costlier than informal evictions. For tenants, formal evictions are publicly observable by credit agencies and future landlords. For landlords, formal evictions cost money to file and time to manage. In Appendix A.5, we present evidence from tenant experiment participants that court eviction is particularly costly. Finally, the negative causal effects of eviction in Collinson et al. (2024) identify the effect of formal evictions in a sample where the not-formally-evicted comparison group is likely to experience informal eviction.

to avoid formal eviction is common in Memphis. For instance, 57% of surveyed tenants have formed repayment plans to repay back rents over time to landlords, and about 40% of court processes do not conclude in evictions because of bargaining. We partner with the housing agencies that administer Memphis's ERAP to recruit 1,808 tenant applicants and 371 landlords of tenant applicants for lab-in-the-field experiments. ERAP applicants and experiment participants have rental debts exceeding $3,000, three–four times their monthly household incomes, and are at high risk for eviction.

We examine social preferences by conducting more than 4,000 real-stakes Dictator Games (DGs) with landlords and tenants (Section 1.4). We modify the standard DG to allow players to exhibit hostility as well as altruism. We call players "hostile" if they choose a dominated bundle in which they forgo money to lower another player's payment. Players are "altruistic" if they forgo money to raise the other's payment.

Both hostility and altruism are prevalent. About 15% of landlords and 24% of tenants are hostile toward their own tenants and own landlords respectively. Suggestive evidence on matching suggests that about one in three landlord–tenant pairs has at least one hostile party. Yet most landlord–tenant relationships are altruistic: 69% of landlords halve their payment to ensure their own tenant receives an equal payment. 53% of tenants do the same for their own landlord.[6]

Next, we test for misperceptions that affect the perceived payoffs from eviction or bargaining and therefore suggest the potential for inefficiency (Section 1.5). With landlords, we elicit beliefs about back rents that they can attempt to recoup in court. With tenants, we elicit beliefs about landlords' hostility and bargaining rates after initiating the court process. We test for misperceptions by: (i) comparing prior beliefs to known ground truth; (ii) testing whether information treatments induce revisions of related beliefs; and (iii) testing whether information treatments affect revealed-preference bargaining behaviors, like whether tenants propose repayment plans for back rents to landlords.

Rates of misperceptions are high. Our most conservative tests imply that 25% of tenants and 17% of landlords have misperceptions exceeding 20 pp. Other tests give misperception rates of 41–79%. As further evidence that beliefs affect behaviors, and of

[6]To rule out elicitation error, we randomize if participants play against their own or a random landlord or tenant. We find, for instance, that landlords are significantly more hostile to their own tenant than random tenants. We find higher rates of hostility and altruism in the landlord and tenant samples than in nationally representative samples we recruit to play the same game against anonymous landlords and tenants. The rates of altruism exceed the 30% of subjects who give half or more in a classical DG (Engel, 2011). As a comparison to hostility, less than 10% typically reject even offers in ultimatum games (Camerer, 2003). We also randomize stakes among tenants and find that altruism and hostility persist in DGs with $1,000 stakes.

independent interest, we find that correcting misperceptions affects some bargaining and ERAP take-up decisions, with larger impacts for landlords than tenants.

We then ask whether these experimental elicitations can help us understand ERAP's impacts and eviction behavior beyond the experiment (Section 1.6). Given high rates of misperceptions and hostility, many evictions may be inefficient or repugnant. A natural question is then whether policy intervention stops evictions. To study the Memphis ERAP's causal effects, we use administrative data on around 4,700 enrollees in a cash-only arm of the program. Conducting event studies around payment, we find statistically significant, but sometimes short-lived effects on eviction filings (i.e., the orders that start the court eviction process), and null or small effects of ERAP on eviction judgments (i.e., ultimate formal evictions). These estimates suggest that ERAP was not cost effective. The 95% confidence intervals in our best-case specification suggest that it cost more than $70,000, distributed across tenants, to avert one judgment for 6 months. These results cast doubt on advocates' claims that ERAP meaningfully reduced eviction.

To shed light on ERAP's small effects on eviction, we link the experiment data to administrative records on ERAP take-up. The model indicates that perverse selection on altruism could reduce ERAP's treatment effects: program enrollees who find ERAP desirable, despite its take-up costs, could be altruists with low counterfactual eviction propensities. Indeed, we find that altruists are more likely to be paid and are paid faster. Social preferences may thus influence policy effectiveness in the Memphis ERAP.

We additionally confirm that hostility, misperceptions, and court costs correlate with evictions in the direction that the model predicts. We link the experiment to court eviction records. An index which combines experimental proxies for the three forces predicts eviction filings. Of the three forces, an indicator for hostility is the strongest predictor, itself correlating with a 10 pp increase in filings among tenants and 15 pp among landlords (off base rates of 33% and 23%). Hostility also correlates with tenant bargaining offers and reported housing-code violations.

Using our estimates and the model, we conduct welfare analysis of eviction and ERAP (Section 1.7). An advantage of the lab-in-the-field approach is that we directly measure economic primitives — social preferences and beliefs — which other research must calibrate or estimate. However, we still do not observe costs and a few other primitives. We estimate the remaining primitives by matching model-implied behavior among experiment participants to moments from the experiment and ERAP program evaluation, using the Method of Simulated Moments.

Our main result from the model is that about one in four evictions is inefficient, and two-thirds of the inefficient evictions are repugnant. Different model assumptions yield that between 15–41% of evictions are inefficient. Given most evictions are efficient, we find that stopping the average eviction would reduce private surplus by at least $870 in our primary specification. The frictions we document are quantitatively meaningful, since if all evictions were efficient, stopping the average eviction would reduce surplus by at least $2,000. Even so, eviction's private costs alone cannot rationalize eviction policy. Naturally, implications could differ for policies that stop particularly inefficient evictions, or if there are other rationales for eviction policy like externalities or high social marginal welfare weights on tenants facing eviction (Saez and Stantcheva, 2016).[7]

The interaction of nonclassical forces partly explains why inefficient evictions are rare despite high rates of misperceptions and hostility: altruism offsets misperceptions. Absent altruism, the rate of inefficient evictions would rise to 56%. Intuitively, many landlords misperceive eviction payoffs but are altruistic and so do not pursue inefficient eviction. Since the joint distribution of altruism and misperceptions proves key, this mechanism highlights the value of collecting both misperceptions and social preferences in one experiment.

We draw several other conclusions. The model confirms that perverse selection on altruism depresses ERAP's Treatment Effect on the Treated (TOT): absent landlord altruism, ERAP would stop about half of evictions. We also find that a quarter of the evictions ERAP does stop are inefficient, meaning the program is no better-targeted than if it stopped evictions at random. Policy counterfactuals of (i) instating fines if landlords take ERAP and then evict, or (ii) targeting the program based on demographics can raise its TOT substantially without sacrificing targeting.

Related Literature. First, we add to research on housing insecurity and evictions. In examining the causes and efficiency of evictions, we study different questions than research that considers the effects of housing insecurity.[8] Although the sociology literature emphasizes the impact of landlord–tenant relationships on eviction (see Section 1.2), economists have largely neglected this mechanism. We contribute an economic framework

[7] The last argument may have particular weight, as we estimate that tenants' eviction costs exceed $4,000. In this case, advocates could view the objective of anti-eviction policy as purely to redistribute rather than to stop evictions.

[8] One antecedent to our modeling approach is Hoy and Jimenez (1991), who consider squatter evictions and bargaining. Recent empirical contributions to the literature on housing insecurity include Palmer et al. (2019), Cohen (2020), Fetzer et al. (2020), Abramson (2021), Geddes and Holz (2022), and Collinson et al. (2024).

that captures the role of social preferences, as well as data confirming their quantitative relevance among the housing-insecure. Along with Collinson et al. (2023), we conduct among the first empirical evaluations of ERAP. Prior analyses report summary statistics, rather than estimate ERAP's causal effects (e.g., Aiken et al., 2022).

Second, we contribute to the literature in behavioral public finance (Bernheim and Taubinsky, 2018a) by studying how social preferences and information frictions affect poverty and welfare policy. Although a mature literature quantifies social preferences in various lab settings (e.g., Levitt and List, 2007a), empirical evidence on how social preferences affect policy design is much rarer.[9] Our finding that social preferences can affect program take-up and targeting adds to the literature on benefit program design (Currie, 2004; Bhargava and Manoli, 2015; Finkelstein and Notowidigdo, 2019), which considers other nonclassical forces, like information frictions, rather than social preferences. The importance of altruism for program design could apply elsewhere if, for example, program participant–social worker relationships affect take-up. Meanwhile, misperceptions that cause inefficient bargaining failure yield a different rationale for policy intervention than in previous public-finance research on corrective taxation or nudges to address misperceptions (e.g., Allcott et al., 2022; List et al., 2023). To study misperceptions, we use laboratory techniques (Fuster and Zafar, 2022; Haaland et al., 2023) in a high-stakes field setting.[10,11]

As a third contribution, we use experiments motivated by bargaining theory to conduct empirical tests of the bargaining literature. Bargaining theory emphasizes that biased beliefs can generate bargaining inefficiencies (Yildiz, 2011; Vasserman and Yildiz, 2019). How social preferences affect bargaining has attracted less theoretical attention (see Pollak, 2022, for a recent exception). Our finding that altruism offsets information frictions is consonant with Friedberg and Stern (2014), who find the same in the context of divorce. Empirical research in other high-stakes settings finds that social preferences affect behaviors (Hjort, 2014; Ashraf and Bandiera, 2018; Lowes, 2021; Blouin, 2022; Ramos-Toro,

[9]Public economics has long considered how altruism affects charitable donations (Andreoni, 1990, 1993; Andreoni and Miller, 2002; DellaVigna et al., 2012), bequests (Becker, 1974), and contributions to public goods (reviewed in Ledyard, 1995). The social preferences we consider are distinct from signaling motives or social norms (e.g., Allcott, 2011; Bursztyn and Jensen, 2017), and echo research on altruism and tax morale (Luttmer and Singhal, 2014).

[10]Our welfare analysis of eviction and ERAP applies a "structural behavioral economics" approach (DellaVigna, 2018), similar to recent work in housing (e.g., Andersen et al., 2022) and beyond.

[11]A recent literature has examined misperceptions in housing markets in particular, usually focusing on spending or investment choices (Armona et al., 2019; Bottan and Perez-Truglia, 2021; Chopra et al., 2023; Fairweather et al., 2023).

2022), but does not link these forces to bargaining.[12]

1.2 A Model of Bargaining in the Shadow of Eviction

This section keeps the model as simple as possible. Extensions and proofs are in Appendix C.3.

1.2.1 Setup

Environment. Our model features two types of agents, Landlords and Tenants. Landlords and tenants have complete information about the other's preferences and beliefs. We index landlord–tenant relationships by i.

The tenant makes a take-it-or-leave-it bargaining offer $o_i \in \mathbb{R}$ to her landlord to repay exogenous rental debt d_i. The offer o_i can be positive or negative; for instance, the landlord may pay the tenant in a "cash for keys" arrangement ($o_i < 0$). This section considers eviction to be formal eviction. We view agreements to stay or leave but which avoid court costs as a form of bargaining. If bargaining occurs, the tenant pays o_i and cannot abscond.[13]

Formal eviction occurs when at least one party perceives it as profitable relative to bargaining. A fraction of tenants does not pay an eviction judgment, either because they abscond or win the case. If the landlord evicts, she recoups $p_i d_i$ in back rents from the tenant, where p_i represents the probability of receiving a full repayment from the tenant.[14]

We normalize bargaining costs to zero. Landlords and tenants respectively face net court costs of $k_{Li}, k_{Ti} \in \mathbb{R}$, where $k_{ji} > 0$ denotes that eviction is costlier to party j than bargaining. This model captures credit constraints by changing the relative costs between

[12]Empirically, Freyberger and Larsen (2021) and Larsen (2021) find departures from efficient bargaining, and that these inefficiencies mostly do not reflect incomplete information as in Myerson and Satterthwaite (1983). Bargaining behavior consistent with fairness norms or heuristics is common (Backus et al., 2020; Keniston et al., 2022), and recent work finds that information interventions can improve bargaining outcomes in court (Sadka et al., 2020). Other work estimates relationship quality or misperceptions to rationalize observed bargaining (e.g., Merlo and Tang, 2019).

[13]Equivalently, one can permit the tenant to abscond with some probability and view o_i as the expected payment net of the tenant's absconding risk.

[14]While we informally describe p_i as a probability, one can recast p_i as any scalar that affects the expected value of the transfer the tenant makes to the landlord. For instance, p_i need not be limited to the interval $[0, 1]$. The real value of the court transfer may be negative, if the tenant does not pay rent during the court process.

court and bargaining. If tenants can only bargain at high cost, because borrowing to obtain cash and transfer to landlords is expensive, that reduces k_{Ti} or makes it negative.

Payoffs. Utility is linear in money for both landlords and tenants. Payoffs V_{Li} and V_{Ti} satisfy:

$$V_{Li} = \begin{cases} p_i d_i - k_{Li}, \text{ if evicts} \\ o_i, \text{ if bargains} \end{cases} \quad \text{and} \quad V_{Ti} = \begin{cases} -p_i d_i - k_{Ti}, \text{ if evicted} \\ -o_i, \text{ if bargains.} \end{cases} \tag{1.1}$$

1.2.2 Classical Benchmark

When do landlords and tenants bargain to resolve debts, versus evict?

Proposition 1.1. *Eviction occurs if and only if*

$$k_{Li} + k_{Ti} < 0. \tag{1.2}$$

This result affirms that, in our basic environment stripped of nonclassical features, a Coase theorem holds. Consider offer $o_i \in [p_i d_i - k_{Li}, p_i d_i + k_{Ti}]$. Offers within this interval are accepted, as both parties are weakly better off than going to court. The interval is well-defined as long as joint bargaining surplus is positive: $k_{Li} + k_{Ti} \geq 0$. Equivalently, eviction occurs if and only if it is efficient, where "efficient" means that joint surplus from eviction is positive.

The benchmark model does not imply that bargaining eliminates all evictions. To the contrary, evictions are possible as long as landlords or tenants have strong reasons to prefer going to court, such that joint surplus is negative. For example, landlords' costs of appearing "soft" to other tenants may exceed their legal fees: $k_{Li} < 0$. Alternatively, if landlords only accept positive cash offers, but tenants face high borrowing costs to obtain liquidity and bargain, then going to court may be cheaper than bargaining for tenants: $k_{Ti} < 0$.

1.2.3 Misperceptions

Motivation for Incorporating Misperceptions. Misperceptions are natural to consider. First, economic theory emphasizes beliefs as a cause of bargaining breakdown (Yildiz, 2011). Second, the sociology and law literatures emphasize how tenants lack information about the complicated eviction process (e.g., Bezdek, 1992; Chisholm et al., 2020). Relatedly, landlords could lack information about tenants' ability to pay.

Setup. Landlords and tenants may now disagree about the probability that the tenant pays an eviction judgment. The terms p_{Li} and p_{Ti} denote the landlord's and the tenant's beliefs about the probability that the tenant pays if evicted. We say there are "misperceptions" if beliefs do not coincide, $p_{Li} \neq p_{Ti}$. We write the difference in beliefs, or "misperception wedge," as $\Delta p_i := p_{Li} - p_{Ti}$.[15] We often consider $\Delta p_i > 0$, as this condition implies that landlords see formal eviction as more favorable for the landlord's payoffs than tenants do.

Eviction with Misperceptions. Misperceptions change the eviction condition as follows:

Proposition 1.2. *With misperceptions, eviction occurs if and only if*

$$k_{Li} + k_{Ti} < \Delta p_i d_i. \tag{1.3}$$

We illustrate Proposition 1.2 in Diagram 1.1. Without misperceptions, evictions occur only if joint surplus $k_{Li} + k_{Ti}$ is less than zero (green shaded region). If $\Delta p_i > 0$, then misperceptions generate more evictions (blue shaded region), as eviction now occurs for $k_{Ti} + k_{Li} \in (0, \Delta p_i d_i)$ and would not occur with coincident beliefs. The additional evictions that misperceptions cause are inefficient, as bargaining would yield joint surplus.

Symmetrically, misperceptions reduce scope for eviction if $\Delta p_i d_i < 0$ — that is, if landlords are more pessimistic than tenants. Then misperceptions generate inefficient bargaining. We do not stress this case because our empirical evidence suggests a positive misperception wedge.

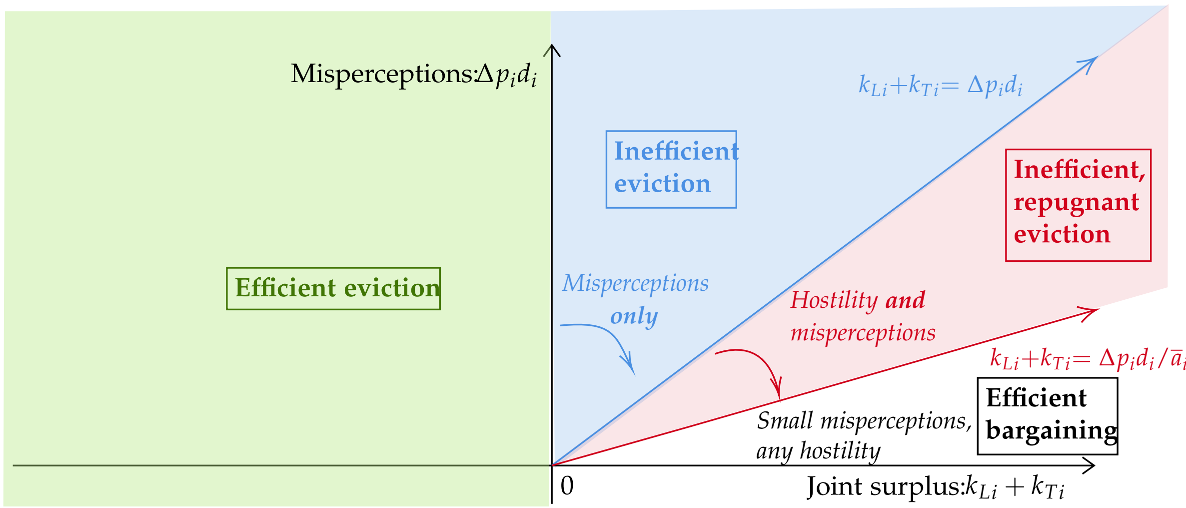
1.2.4 Social Preferences

Motivation for Incorporating Social Preferences. Social preferences between landlords and tenants are a second natural way to extend the model. The popular media regularly highlights examples of extremely deteriorated landlord–tenant relationships.[16] The vast literature on social preferences, including in other high-stakes settings (e.g., Hjort, 2014),

[15] We use the term "misperceptions" as shorthand for "non-coincident beliefs." With equal misperceptions, the misperception wedge is zero. We model a complete-information game with non-coincident beliefs. But inefficiency can also obtain in an incomplete-information game with uncertainty about the opponent's valuation (Myerson and Satterthwaite, 1983; Ausubel et al., 2002).

[16] For instance, *The New York Times* describes a landlord whose tenants "curse and spit at her and owe more than $23,000 in rent" (Haag, 2021).

Diagram 1.1: Comparative Statics



suggests they might matter here. Yet economists have not stressed social preferences among landlords and tenants.[17]

Meanwhile, law and sociology literatures emphasize the importance of landlord–tenant relationships (e.g., Desmond, 2016; Garboden and Rosen, 2019; Balzarini and Boyd, 2021). In a law review, Bell (1984) describes evictions due to landlords' "coercive, punitive, or malicious reasons" as warranting special tenant protections, as they have "no social value" and "only satisf[y] the landlord's vindictiveness" (p. 537). In sociology, Vaughn (1964), Akers and Seymour (2018), and Chisholm et al. (2020) describe eviction as emerging from coercive power dynamics between landlords and tenants. Conversely, Gilderbloom (1985) observes that inexperienced landlords sacrifice profits out of "concerns for tenants' welfare" (p. 159). Indeed, interviews suggest that altruism motivates landlord participation in housing assistance programs (Aubry et al., 2015) and Moving to Opportunity (Cossyleon et al., 2020). A recent sociology review summarizes that "social relationships," "resource constraints," and "supply-side actors and policy" all contribute to housing insecurity (DeLuca and Rosen, 2022, p. 344).

Set-up. Becker (1974)-type altruism parameters $a_{Li}, a_{Ti} \in (-1, 1)$ now enter parties' utilities. The value $a_{ji} = 0$ implies that party j's utility is unaffected by the other party's utility; $a_{ji} > 0$ implies altruism; and $a_{ji} < 0$ implies hostility (anti-altruism or "spite" in

[17] An exception is the literature on racial discrimination in housing markets (e.g., Ewens et al., 2014).

Levine, 1998). Utility remains linear. Payoffs are:

$$V_{Li} = \begin{cases} p_{Li}(1-a_{Li})d_i - k_{Li} - a_{Li}k_{Ti}, \text{ if evicts} \\ (1-a_{Li})o_i, \text{ if bargains} \end{cases}$$
$$V_{Ti} = \begin{cases} -p_{Ti}(1-a_{Ti})d_i - k_{Ti} - a_{Ti}k_{Li}, \text{ if evicted} \\ -(1-a_{Ti})o_i, \text{ if bargains.} \end{cases} \tag{1.4}$$

The altruism parameter scales the other's non-altruistic utility. For instance, if the landlord gets 20 utils of non-altruistic utility and the tenant has $a_{Ti} = 0.5$, then the tenant gets 10 utils plus her own payoff. Alternatively, $a_{Ti} = -0.5$ implies that the tenant gets –10 utils plus her own payoff: she is made worse off when the landlord gets positive utility.[18]

Hostility is distinct from signaling. Signaling implies that a party derives an *instrumental* benefit from appearing harsh and affects the net court cost k_{ji}. Hostility captures *non-instrumental* reasons why one party may get utility from harming the other.

Eviction with Social Preferences. The next proposition provides a condition for eviction with social preferences:

Proposition 1.3. *With social preferences and misperceptions, eviction occurs if and only if*

$$k_{Li} + k_{Ti} < \frac{\Delta p_i d_i}{\overline{a}_i}, \tag{1.5}$$

for "compound altruism" $\overline{a}_i := (1 - a_{Li}a_{Ti})/((1-a_{Li})(1-a_{Ti}))$.

To interpret Equation (1.5), note that compound altruism $\overline{a}_i$ is strictly positive and strictly increasing in either party j's altruism: $\frac{\partial \overline{a}_i}{\partial a_{ji}} > 0$. We collect comparative statics across propositions:

Corollary 1.1 (Comparative Statics)**.** Let $k_c^* := 0$, $k_m^* := \Delta p_i d_i$, and $k_a^* := \Delta p_i / \overline{a}_i$ be the "eviction thresholds" in the setups with <u>c</u>lassical preferences, <u>m</u>isperceptions, and <u>a</u>ltruism and misperceptions. Then a positive misperception wedge raises the eviction

[18]Individuals do not internalize the others' social payoffs. That is, the tenant ignores that the landlord may get altruistic utility from giving the tenant money, and vice-versa. Our simple way of modeling social preferences is agnostic about whether they emerge from impure altruism (Andreoni, 1990), moral obligations (Rabin, 1995), inequity aversion (Fehr and Schmidt, 1999), or reciprocity (Charness and Rabin, 2002), among other models. A material share of the population having genuinely hostile preferences could rationalize how posturing as hostile (Abreu and Gul, 2000) might persist.

threshold relative to the classical setup, and hostility raises the eviction threshold relative to misperceptions alone: (i) $\frac{\partial(k_m^* - k_c^*)}{\partial(\Delta p_i)} > 0$ and (ii) if $\Delta p_i > 0$, $\frac{\partial(k_a^* - k_{\mathrm{m}}^*)}{\partial(-\bar{a}_i)} > 0$.

Equation (1.5) gives three insights. First, hostility can cause evictions. More hostility ($\bar{a}_i \downarrow$) flattens the slope of Diagram 1.1's red ray and increases inefficient evictions (red shaded region), assuming positive joint surplus ($k_{Li} + k_{Ti} > 0$) and misperception wedge ($\Delta p_i d_i > 0$). We call evictions that occur only because of hostile preferences "repugnant" — that is, "repugnant" evictions are those which would not occur if a_{ji} were 0 for each j where $a_{ji} < 0$. Such repugnant evictions will always be inefficient.[19]

Second, altruism also affects eviction and efficiency. The symmetric comparative static is that if eviction is inefficient, more altruism ($\bar{a}_i \uparrow$) raises the chance of efficient bargaining because it steepens the slope of Diagram 1.1's red ray. Altruism can therefore sustain bargaining that would otherwise fail due to misperceptions. However, with opposite-signed misperceptions, altruism can cause efficient eviction, in the case where $\Delta p_i d_i < k_{Li} + k_{Ti} < \Delta p_i d_i / \bar{a}_i < 0$.

Third, with zero misperception wedge, social preferences play no role. In that case, Equation (1.5) reduces to the benchmark of Equation (1.2).[20] However, hostility can cause eviction from even modest misperceptions, since enough hostility can make $\bar{a}_i$ arbitrarily small.

1.2.5 The Role of Emergency Rental Assistance

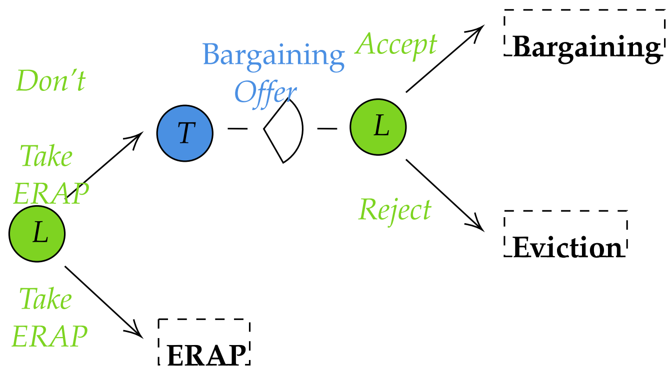
Setup. We now consider an emergency rental program, intended as a stylized version of ERAP. Diagram 1.2 shows the full game with the policy intervention. The landlord decides whether to take up rental assistance, which pays the full back rent d_i but requires paying a take-up cost k_{Li}^P. If the landlord takes up the program, then eviction is impossible. If the landlord declines the program, then the tenant can make a bargaining offer as in the prior subsection.[21]

[19]Efficiency concepts with altruism are philosophically challenging (Friedman, 1988). If $a_T \neq a_L$, then utility is not transferrable and the natural efficiency concept has the more altruistic party transfer infinite wealth to the other. To sidestep this issue, we say that the efficient outcome is the natural definition for *equal* altruists ($a_T = a_L$): the outcome is efficient when eviction occurs if and only if $(1 + a_T)k_L + (1 + a_L)k_T < 0$. That is, eviction occurs if and only if joint surplus, adjusted for equal altruism, is negative. This definition happens to coincide with the definition of an efficient outcome when parties are both classical.

[20]In our model, with coincident beliefs, parties with social preferences reach a bargained solution if there is joint surplus from bargaining. For any altruism parameters, a higher altruism-adjusted surplus can always be achieved outside court. Social preferences only change how surplus is divided.

[21]We ignore the tenant's application choice. One way of thinking about this game is that it is played among the sample of tenants who have already applied for the program, and who need the landlord to

Diagram 1.2: Timing and Game Tree



Note: The arc for the tenant's decision indicates a continuous choice problem.

The payoffs are identical to the prior subsection if landlord does not take up. The landlord gets $d_i - k^P_{Li}$ if she takes up. The tenant gets 0 if the landlord takes up.

Take-Up Decision. This environment yields the following threshold condition for program take-up by the landlord:

Proposition 1.4. *Emergency rental assistance take-up occurs if and only if*

$$k_{Li} + k_{Ti} + w_i \geq \frac{\Delta p_i d_i}{\bar{a}_i}, \tag{1.6}$$

where the "program wedge" w_i is defined as:

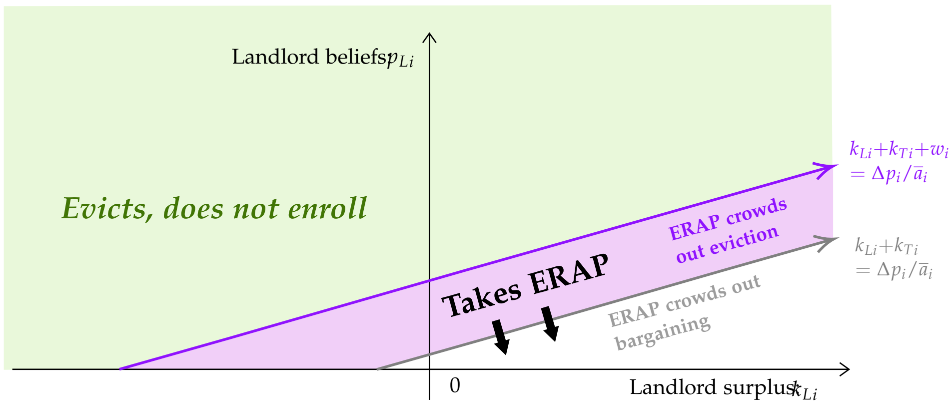
$$w_i = \frac{1}{\bar{a}_i}\left(\underbrace{\frac{d_i - k^P_{Li}}{1 - a_{Li}}}_{\substack{\text{Altruism-adjusted} \\ \text{net ERAP payment}}} - \underbrace{\left(p_{Ti} d_i + \frac{k_{Ti} + a_{Ti} k_{Li}}{1 - a_{Ti}}\right)}_{\text{Outside option adjustment}}\right). \tag{1.7}$$

Proposition 1.4 indicates that the program can "crowd out" (i.e., cause not to occur) bargaining, or efficient eviction, or inefficient eviction. Notice that Equation (1.7) reverses the sign of the inequality of Equation (1.5), but is augmented with the program wedge w_i. If $w_i < 0$ and take-up occurs, then an eviction would not have occurred without the program: ERAP entirely crowds out bargaining. If $w_i > 0$ and take-up occurs, then it may also crowd out some eviction.

To develop intuition for the program wedge, observe that the first term within

accept it.

Diagram 1.3: Effects of ERAP



Note: The diagram considers a case with $a_{Ti} = 0$, such that the wedge is constant for all k_{Li}.

parentheses is an altruism-adjusted ERAP net payment (benefit d_i less take-up cost k^P_{Li}). As this term rises, ERAP becomes more valuable. The second term accounts for the fact that ERAP forecloses bargaining. As the landlord's outside option of bargaining improves (e.g., the tenant's court cost rises), the second term rises, which shrinks the wedge overall.

Diagram 1.3 shows take-up by eviction versus bargaining for a fixed $w_i > 0$, now in p_{Li}/k_{Li} space (i.e., fixing other parameters, so translating the comparative statics in Diagram 1.1). For this configuration of parameters, landlords enroll for all values to the right of the gray line. ERAP crowds out all bargaining (white region) as well as some evictions (purple region). If $w_i < 0$, which corresponds to shifting the purple line to the right of the gray line, then ERAP exclusively crowds out bargaining.

Even absent nonclassical forces, ERAP is not perfectly targeted. The landlord's net surplus from enrolling in ERAP is:

$$S_{Li} := \underbrace{d_i - k^P_{Li}}_{\text{ERAP payoff}} - \underbrace{\left(p_{Li}(1-a_{Li})d_i - k_{Li} - a_{Li}k_{Ti}\right)}_{\text{Outside option}}. \tag{1.8}$$

Enrollment occurs if and only if landlord net surplus S_{Li} is positive, as the condition $S_{Li} \geq 0$ is equivalent to (1.6). Consider the classical benchmark with zero altruism. The landlord's net surplus S_{Li} is increasing in the landlord's court costs k_{Li}. Simultaneously, higher court costs also raise the value of bargaining. Thus, even in the benchmark, ERAP risks enrollment among pairs who otherwise bargain.

Still, altruism amplifies the risk of inframarginality. Observe that surplus S_{Li} is increasing in a_{Li}. Intuitively, enrolling in ERAP costs landlords a fixed k^P_{Li}. Landlords' outside option of evicting tenants or bargaining with them is decreasing in altruism. When enrolling, altruists gain more utility per dollar of take-up cost. Perversely, altruism also makes bargaining more likely (Diagram 1.1). We call this force, in which the pairs who enroll are altruists who seldom evict in the first place, "perverse selection on altruism."

If ERAP does stop evictions, the evictions that it prevents could be efficient, inefficient or repugnant. The values of costs, misperceptions, and social preferences govern whether households at the margin of inefficient eviction enroll.

Appendix C.3 extends the framework to include a two-period dynamic setting with endogenous rental debts, concave tenant utility, and Nash Bargaining. Sufficiently low poverty or enough impatience can recover the same forces as in the main text.

1.3 Institutional Details and Experiment Sample

1.3.1 Background

Setting. Our setting is the City of Memphis, Tennessee and its surrounding county of Shelby County. Memphis has a population of 620,000, is 65% Black, and has a poverty rate of 24%, according to U.S. Census Bureau estimates for 2022. Shelby County has a population of 920,000. In 2016, according to estimates from Eviction Lab (Gromis and Desmond, 2021), there were 48 eviction judgments per 1,000 renter households in Shelby County. Shelby County was in the top quartile of eviction filings in 2018 among counties with at least 100,000 renter households.

Eviction Process. Most evictions in Shelby County are caused by nonpayment of rent, though property damage or other violations can also prompt eviction. Once a tenant fails to pay rent for a single month, landlords typically gain the legal right to evict. To formally evict a tenant, landlords must first serve an eviction notice that gives the tenant 14 days to pay. If the tenant fails to pay, the landlord files for an eviction warrant (a *filing*), which begins the court process and includes a hearing date. If the judge rules in favor of the landlord at the hearing (a *judgment*), the tenant must vacate the property and, if they do not do so within ten days, the landlord may obtain a writ of possession from the county sheriff. A writ authorizes them to remove the tenant and their belongings against their will from the property.

Landlords can either seek judgments for possession only or judgments for money (owed back rents or damages) as well as possession. Of the roughly 26,600 evictions filings in 2019, 71% yielded judgments, of which 54% included money. Money judgments require the tenant pay the landlord overdue back rents, and give the landlord the legal right to garnish tenants' wages. But they require that landlords give "personal service" of the eviction notice, which can be costly as tenants can intentionally evade landlords.

Figure A.1 shows a time series of the share of filings that result in judgments within 150 days of filing. Pandemic-related court delays and the eviction moratorium reduced judgment rates below 20% until early 2021. After the eviction moratorium expired in Shelby County in spring 2021, only 40% of filings yielded judgments. The lower level of judgments persists throughout our study period.

ERAP. Under the CARES Act and Consolidated Appropriations Act of 2021 (CARES II), state and local governments received funding to form Emergency Rental and Utilities Assistance Programs (ERAPs). The programs paid overdue debts of rent and utilities that accrued during the coronavirus pandemic. We partner with the Memphis and Shelby County ERAP, which operated as an integrated program. Through December 2022, the Shelby County ERAP distributed at least $100 million in assistance to around 20,000 households.[22] By comparison, Section 8 supports about 35,000 households in all of Tennessee.

Memphis ERAP shared application and payment data with the landlord and tenant contact information that we use to solicit participation in survey experiments. See additional ERAP program details in Section 1.6.2.

1.3.2 Experiment Overview and Recruitment

We conduct separate experiments with landlords and tenants. Figure 1.1 presents a visualization of the survey flow for each. There are two main sections to each experiment: a part that uses Dictator Games to elicit social preferences (Section 1.4) and a part that elicits misperceptions (Section 1.5). Additional details are in Appendix C.4.

Experiment Sample Recruitment. We recruited experiment participants from Memphis ERAP application records. We used completed and partial applications, so not all participants ultimately received ERAP payment.

[22] More precise estimates are difficult to obtain because we only have complete data on some parts of the program. Section 1.6 studies an arm of the program that paid more than $25 million to about 5,000 unique households.

We conducted all surveys online. Recruitment materials informed participants that they would receive a $20 gift card for completing the survey and could earn other rewards. Participants were informed about the purpose of the study and provided consent before beginning. Responses are identified and linked to ERAP records.

We conducted the landlord survey experiment from August 29 to October 28, 2021, sending several reminders over this time period by email and text message. We contacted 3,966 unique email addresses associated to landlords and property managers listed on tenant ERAP applications. We received 371 valid responses, for a response rate of about 9% (12% conditioning on a valid landlord email address).

We contacted 16,861 unique tenants via email over 14 survey waves from February 14 to May 30, 2022. We sent one email reminder, along with text-message invitations and one reminder to tenants with valid phone numbers. We obtained 1,808 valid responses in total, for a response rate of 11% (12% conditioning on a valid email address).

Ethics. Given the vulnerability of our study population, we took care to ensure our experiment created minimal risk of harm. We designed the surveys in partnership with ERAP personnel and eviction defense attorneys in Memphis. We vetted all information provided, experimental elicitations, and outcome modules to ensure they would help both parties. We only provided information that we believed would reduce risk of eviction. Elicitations that had any potential for adverse outcomes (e.g., choices in Dictator Games) were strictly anonymous. Landlords and tenants benefited from survey participation, as they were offered payment, information about eviction and ERAP, and opportunities to participate in ERAP or negotiate to avoid eviction.

1.3.3 Demographics

Tenant Characteristics. Study participants are low-income and housing-insecure (Table 1.1, Column 1). The experimental sample is 81% female and 88% Black (Panel A). They have incomes of less than $1,000 per month and owed a median of around $3,500 when they applied for ERAP. About a third have ever been evicted, and almost 90% have had overdue rents (Panel B).[23]

Landlord Characteristics. 58% are Black, and 62% of landlord participants are female (Table 1.2, Column 1). 62% are landlords, and 29% identify as property managers. We

[23] We survey tenants at different times in the ERAP process. Back rents at the time of the survey are lower than at application because some tenants are evicted, move, or are paid before the survey.

refer to both groups as "landlords." 48% report owning or managing 10 or more units, with the remainder about evenly divided between managing 1 or 2 units and 3 to 10 units. Because large landlords may have multiple property managers or employees, we have many observations from the larger Memphis landlords and property management companies.

Selection. We use administrative data on tenant applications to examine whether experiment participants differ from the overall ERAP pool (Table 1.1, Columns 2–4). Overall differences are relatively small, with the exception of the share female (9 pp higher among participants) and monthly incomes ($160 higher among participants). We do not have demographic data on non-participating landlords, but we see information about their tenants. Participating landlords' tenants owe more in back rents and less in utilities. The tenants of participating landlords are also slightly more likely to be white.

Balance and Attrition. We test balance for our survey experiments by examining characteristics of people who complete the study (Appendix C.4 shows attrition funnels and balance tables). One of five treatments across experiments, the landlord information experiment, is unbalanced on observables ($p = 0.01$). Results are unchanged when we include controls or use post-double-selection Lasso to select controls (Belloni et al., 2014).

Conditional on completing a short demographic questionnaire that starts the survey, 77% of tenants and 65% of landlords complete the surveys. Our sample consists of participants who complete the whole survey.[24]

1.4 Altruism and Hostility: Measurement and Results

1.4.1 Measurement

Background on Dictator Games. We adapt the laboratory experiment called the "Dictator Game" (DG). In a standard DG, the Dictator is given an endowment. The experimenter elicits how much of the endowment the Dictator wants to give to another player, the "Opponent." Classical economic models predict that the Dictator gives $0. A meta-analysis finds that 65% of Dictators give a positive value, and the conditional average given is around 40% (Engel, 2011).

[24]Attrition rates are constant across treatments (joint p-value > 0.2 for both surveys, Table A.7). 96% of tenants and 91% of landlords complete the DGs. Appendix C.4 shows results for the Dictator Games among this larger sample, including those who attrit later.

Modified Dictator Game. We modify the DG so that it can measure both altruism and hostility. Consider a Dictator who gives \$0 in the standard DG. It is ambiguous whether she has classical preferences and wishes to maximize her own payment, or if she has hostile preferences and derives utility from reducing her opponent's payment.

In our modification, the Dictator chooses between two bundles: (\$$s$ to Dictator, \$0 to Opponent) and (\$$x$ to Dictator, \$$x$ to Opponent). In the first bundle, the Dictator receives \$$s$ and the Opponent gets nothing. In the second bundle, both players receive \$$x$. We vary s within-subject until we identify (bounds on) the participant's "indifference point" $S(x)$ — i.e., the value that renders her indifferent between $(s, 0)$ and (x, x).

With classical preferences, $S(x) = x$: the Dictator would prefer whichever bundle gives a larger private payoff. With altruism, $S(x) > x$: the Dictator would forgo some private payoff so that the Opponent receives a payoff. With hostility, $S(x) < x$: the Dictator would forgo some private payoff so that the Opponent receives \$0.

The game exactly delivers the Becker (1974) altruism parameter. For individual i, granting linear preferences, their altruism parameter $a_i \equiv (S_i(x) - x)/x$, as $S_i(x)$ must satisfy

$$\underbrace{S_i(x) + a_i \cdot 0}_{\text{Payoff from } (S(x),0) \text{ bundle}} = \underbrace{x + a_i \cdot x}_{\text{Payoff from } (x,x) \text{ bundle}} . \tag{1.9}$$

Relative to the large literature employing the standard DG, this modification is relatively novel and has the benefit of eliciting both altruism and hostility at once.[25]

Elicitation and Opponents. Bundles are Amazon gift cards. We elicit bounds on indifference points using a multiple price list: we repeatedly ask whether a participant prefers $(s, 0)$ or (x, x), varying s until the participant switches her response. We implement a small share of choices to ensure incentive-compatibility (see Appendix C.4). We provide plain-English instructions (see Figure A.2A for an example elicitation). We inform participants that it is in their best interest to respond truthfully (Danz et al., 2022).

Landlords play the game twice against three potential opponents (Figure 1.1): their own, named tenant (with probability 2/3); a random, unnamed tenant (with probability 1/3); and a random, unnamed landlord (all landlord participants). Tenants play the game twice against a symmetric group of three potential opponents: their own, named landlord (with probability 2/3); a random, unnamed landlord (with probability 1/3); and

[25] The modification takes inspiration from other DG adjustments (e.g., Charness and Rabin, 2002; Kranton et al., 2016) and the joy-of-destruction game (Abbink and Sadrieh, 2009; Abbink and Herrmann, 2011). Contemporaneous work on religious conflict in Nigeria implements a similar DG modification (Ortiz, 2023).

a random, unnamed tenant (all tenant participants). Landlords' DG behaviors against own versus random tenants indicate whether preferences reflect the specific relationship versus attitudes toward the entire class of tenants. Similar logic holds for interpreting tenants' behaviors toward own versus random landlords.

Instrumental reasons for helping or harming the opponent, such as signaling motives, are unlikely to explain DG behavior. We inform all participants when they provide consent that their responses will be anonymous. We also remind tenants in the DG that "the gift card will not be associated with your name and won't count as rent" and include an extra confirmation check. Even if someone receives a gift card, they cannot deduce that it came from their own landlord or tenant, as they could receive a gift card from a random landlord or tenant.[26] Some free responses for both parties do mention instrumental reasons for their choice, so we cannot rule out the concern entirely.

Stakes. Landlords play the game where the alternative $\$x$ is a \$10 gift-card given to each party and $\$s$ can vary between \$1 and \$20. Among tenants, we randomly vary $x \in \{10, 100, 1000\}$ across participants and correspondingly vary s between $x/10$ and $2x$. For instance, we ask some tenants whether they prefer (\$900 self, \$0 opponent) or (\$1,000 self, \$1,000 opponent), and others whether they prefer (\$9, \$0) or (\$10, \$10). This subexperiment lets us test whether tenant altruism and hostility are purely low-stakes phenomena.

Benefits of DGs. DGs have several strengths. First, there should be a high burden of proof to argue that participants derive utility from reducing each other's utility. Observing participants actually burn money to prevent the opponent from getting a payoff arguably provides stronger evidence than free-response reflections or similar. Second, the game delivers the Becker (1974) measure of social preferences, facilitating quantitative analysis. Third, experimental manipulation lets us compare affect across opponents. To raise confidence in the DG results, we show that qualitative measures and simpler quantitative elicitations correlate with DG behavior.

[26] 88% of tenants correctly answer a comprehension check about anonymity. With landlords, we discuss anonymity only at the point of consent. Related real-stakes measures of hostility with no room for instrumental motives have similar levels and are highly correlated for both landlords and tenants (Appendix C.4).

1.4.2 Results

Landlords and tenants exhibit substantial hostility and altruism (Figure 1.2).[27] 15% of landlords are hostile toward their own tenant (Panel A), meaning they forgo money so that their own tenant receives nothing. 7% of landlords are hostile toward a random tenant, implying a 9-pp difference in hostility between own versus random tenants (p-value of difference = 0.006). Meanwhile, landlords are similarly hostile toward own tenants as random landlords. Taken together, these differences suggest landlords' acrimony toward their own tenants is relationship-specific and not distaste for renters as a class.

Tenants are more hostile than landlords (Panel B). 24% of tenants are hostile toward their own landlord, and 22% are hostile toward a random landlord, both of which are significantly higher than rates of landlord hostility toward own tenants ($p < 0.01$) and random tenants ($p < 0.001$). Tenants are 10 pp more hostile toward their own landlords versus random tenants ($p < 0.001$), while the difference in hostility between their own landlord and random landlords is statistically insignificant ($p = 0.279$). The fact that tenants are hostile toward random landlords suggests that tenants' acrimony is generalized and not only relationship-specific.

As further evidence that hostility is intense, and that minor elicitation noise does not drive the results, more than half of landlords and tenants who express any hostility are "highly hostile," meaning they prefer the bundle $(x/10, 0)$ to (x, x). Differences in high hostility are also significant (Table A.16). For instance, 9.0% of landlords are highly hostile to their own tenants, versus 3.6% who are highly hostile to random tenants (p-value of difference: 0.031).[28]

Altogether, 24–39% of landlords and tenants have at least one party exhibiting hostility toward their own tenant/landlord, depending on matching between hostile landlords and tenants. Suggestive evidence about assortative matching between landlords and tenants suggests that about one third of landlord–tenant pairs have at least one hostile party

[27] Table A.16 summarizes results and presents statistical tests. Figure A.4 shows histograms of indifference points.

[28] The relatively high rates of hostility toward random tenants and landlords are partially driven by anchoring. Recall that each participant does the DG twice, in random order. When we consider behavior in only the *first* DG each participant plays, 3% of landlords and 8% of tenants are hostile to random tenants (Figure A.6 and Table A.17), similar to rejection rates of even offers in ultimatum games with generic opponents (Camerer, 2003). Reassuringly, in this non-anchored sample, the levels of hostility toward own tenants and own landlords are similar to our primary estimates. They are statistically different from the benchmarks despite cutting the sample in half. Residual hostility among tenants toward other tenants could reflect negative affect toward other low-income households, which sociology research documents is prevalent among the poor (Shildrick and MacDonald, 2013).

(Appendix A.5). If we conservatively posit that only own-tenant/own-landlord minus random-tenant differences give true hostility, rather than elicitation errors, we still obtain estimates of 9–19%. Even if just 9% of parties were hostile, not all will receive evictions (we match parties to evictions in Section 1.6). Thus, these rates of hostility could cause a meaningful share of ultimate evictions.

Nevertheless, more than two in three landlords are "highly" altruistic toward their own tenant, meaning they forgo doubling their own private payoff to ensure an even split. Landlords are 6.5 pp more likely to be highly altruistic to their own tenant than a random landlord ($p = 0.03$). Tenants are more likely to be highly altruistic toward another tenant rather than own or random landlords. Yet we still see substantial tenant altruism toward landlords, which is notable given our sample's precarious housing situations.

These social preferences are stronger than those in a benchmark. We test whether hostility and altruism are more common in the landlord and tenant samples than among random participants recruited from online survey panels. We conduct the same DG with random participants in Memphis ($N = 275$) and a nationally representative sample ($N = 623$) (see Appendix A.5 for recruitment details). Landlords are 14 pp more likely to be highly altruistic toward their own tenant than the pooled random sample is to a random tenant (Figure A.10, $p < 0.001$). Tenants' hostility toward own landlords exceeds the pooled random sample's hostility toward random tenants by 11 pp ($p < 0.001$) and toward random landlords by 4.4 pp ($p = 0.014$).[29]

Turning to heterogeneity by demographics, we find that large landlords and landlords with eviction experience are significantly more hostile; tenants with high levels of back rents are significantly less hostile; and female and Black tenants are significantly more hostile (Figure A.5).

Free-Response Reflections. Free-response reflections confirm extreme social preferences. One tenant wrote:

> "He also harassed me for rent even though I had applied to ERA. Even sent his property manager to my door with a gun."

This tenant had an indifference point of $S(1{,}000) \in [300, 400]$: she was willing to give up at least \$600 so that her landlord would receive \$0. On the other hand, another tenant

[29] The hostility exceeds the shares who reject even offers in ultimatum games (Camerer, 2003). However, the rates of hostility are similar to those collected by Kranton et al. (2016) among college students when allocating to members of a different political party (20%), or the shares who choose to destroy in joy-of-destruction games where subjects can pay to destroy others' bundles (26% in the hidden treatment in Abbink and Herrmann 2011).

wrote that her landlord "is a very good person and is willing to help the very best she can." Some participants also appealed to religion. One landlord wrote, "I am a Christian and I try to live out my faith."

Validation and Experimenter Demand. One might worry that the high levels of altruism and hostility reflect inattention, confusion, or other artifacts of the lab setting. In Appendix C.4, we describe our attention checks and show that dropping inattentive participants does not affect results. Appendix C.4 also shows that behavior in the DG is highly correlated with: (i) additional measures of landlord–tenant relationships that we included in the surveys, (ii) tenant sentiment about landlords extracted from simple Likert scales and natural-language processing analysis of free-response questions, and (iii) externally validated measures of preferences that extend beyond the laboratory from Falk et al. (2018). Another benefit of these exercises is that several of them are especially unlikely to have instrumental value, suggesting that behaviors in the DG indeed reflect social preferences and not other motives.

In a particularly informative check, we ask participants whether they want to enroll their own landlord or own tenant in a lottery to win a gift card (Figure A.14). Enrolling the opponent is costless (see details in Appendix C.4). The shares of participants who reject enrolling the opponent are similar to the shares who are hostile in the DG. Moreover, behavior in this exercise and the DG is highly correlated.

Differences by opponent further allay elicitation concerns. For instance, tenants are unlikely to be more attentive to DGs played against a random tenant than against their own landlord.

Experimenter demand is unlikely to explain this pattern of results. First, demand effects would push participants to appear more generous than they might behave in private. But in fact, the rates of hostility we find are high relative to similar DGs. Second, some of the above validations are especially insulated from demand concerns. For instance, we conduct tenants' free-response reflections before lab games. Tenants seemed to share honest (often negative) perspectives.

A related concern is that landlords give tenants money because they expect to recoup it as rent. This motive would imply that we overestimate altruism and underestimate hostility. First, observe that this concern cannot explain behavior of landlords who are highly altruistic (nearly two-thirds). Absent altruism, these landlords are weakly better off just taking $20 directly. Second, we use landlords' elicited beliefs about the amount of money they can recoup from tenants in an eviction (Section 1.5) to scale down their implied altruism. Naturally, landlords become less altruistic and more hostile. But preferences remain extreme (Figure A.15).

Selection Tests. An important concern is that people who complete the study if invited are more likely to be altruistic than people who do not. In general, we expect this concern to attenuate our estimates of hostility. In support of this hypothesis, participants who complete the DGs but attrit mid-study have slightly higher levels of hostility and, among tenants, lower rates of high altruism (Table A.16 vs. Table A.19), suggesting that our main results on hostility are conservative. Next, we use entropy balancing (Hainmueller, 2012) to reweight participants on observables to match non-participants. Our estimates of social preferences are essentially unchanged (Table A.18).

Robustness to Stakes and Controls. Using experimental variation from our stakes sub-experiment, we find no evidence that stakes affect behavior (Figure A.9).[30] Landlords' differences between own tenants and random tenants are robust to controls or limiting the sample to attentive participants. This robustness also applies to tenants' differences between own landlords and random tenants (Tables A.11–A.12).

1.5 Misperceptions: Measurement and Results

1.5.1 Measurement

Structure. For both tenants and landlords, we start the misperceptions module by eliciting prior beliefs (i.e., beliefs before being possibly treated with information), after conducting the DGs (Figure 1.1). We then randomly expose half to information, before measuring posterior beliefs.

At a high level, we structure both misperceptions elicitations as follows. We elicit participants' prior beliefs about an observed statistic that we can benchmark to ground truth. We also elicit participants' beliefs about a closely related statistic that we cannot observe. Finally, we elicit posterior beliefs about the unobserved statistic among the treated group only.

Landlords: Recoupment Rates. In the landlord survey, we focus on beliefs about whether tenants repay eviction judgment balances after a court eviction. We use public court

[30]This result contributes to a conflicted literature on the importance of stakes in social-preference experiments. For instance, stakes affect behaviors in ultimatum games in India (Andersen et al., 2011).

records to link eviction balances to evictions. Less than 10% of money judgments in Shelby County eviction court result in any repayment to the landlord within a year. Informing landlords that money judgments are rarely recouped could reduce their perceived return to formal eviction.

Why did we think judgment balances are important to landlords' cost–benefit analysis of formal eviction? First, in part due to the pandemic context, the sample contains many small landlords who lack prior eviction experience. Second, as Section 1.3 notes, roughly half of landlords obtain evictions for *money* and not just *possession*, even though money judgments are costlier to obtain. Money judgments' value relative to judgments for possession is determined by the rental debts recouped from money judgments.

Among landlords, we elicit prior beliefs about the recoupment rate of judgment balances for the *average* tenant in Shelby County court (the observed statistic above). We ask landlords the share of tenants who fully repaid balances between January 2020 and August 2021. We also elicit landlords' subjective probability of recouping judgment balances if they filed an eviction against their *own*, named tenant (the unobserved statistic above). When eliciting beliefs, we included several confirmation questions and visual aids (Figure A.2B shows an example).

Tenants: Two Beliefs. Tenant misperceptions about landlords' social preferences could affect tenant bargaining propensity. To examine this hypothesis, we elicit tenants' beliefs about the share in the landlord experiment who preferred to split a $20 gift card evenly with their own tenant, rather take the gift card for themselves (i.e., their indifference point was more than $20). 69% of landlords preferred the bundle ($10, $10) to the bundle ($20, $0) when their own tenant was the opponent.[31] We also elicit the subjective probability that the tenant's *own* landlord would evenly split the gift card.

Misperceptions about landlords' behavior in court could also affect tenant bargaining propensity. Tenants could believe that eviction filings are empty threats unlikely to cause eviction, or, conversely, that bargaining after court filings is impossible. We elicit tenants' beliefs about the share of landlords who bargain during the formal eviction process to avoid eviction judgments. 31% of Shelby County landlords that initiated the court eviction process (i.e., filed) in 2019 had explicitly withdrawn or settled (i.e., not obtained judgments) by August 2021. We elicit tenant beliefs about this statistic on average, and also their subjective probability that their *own* landlord would drop an eviction if filed.

Treatment and Posteriors. We randomly treat half of participants with information

[31]We elicit this fact soon after tenants play the DG, so participants are familiar with the game.

intended to shift their beliefs. The purpose of shifting beliefs is to study whether correcting beliefs can change behaviors, which has direct policy implications and also sheds light on the relationship between misperceptions and evictions.

For each beliefs module, the information that we provide precisely corresponds to the observed statistic in each elicitation. For instance, we inform landlords that 6 out of 100 cases on average fully repaid judgment balances. The tenant experiment cross-randomizes the two treatments.

After the information, we collect posterior beliefs for treated participants. These posteriors update the prior beliefs about one's own tenant or landlord (the unobserved statistic). For instance, we ask landlords to report posterior beliefs about whether their own tenant would repay debts in a money judgment. The treatment is informative for posterior beliefs if the unobserved statistic is correlated with the observed statistic.

To ensure that participants intend to update, we first ask participants whether their prior belief is "too high," "too low," or "still correct." Then we ask them to report a posterior consistent with the direction of the reported belief update. We impose that the control group's posteriors are equal to their priors since they did not receive information.[32]

Outcomes. We consider three groups of belief outcomes. First, we test whether participants' prior beliefs about the observed statistic are accurate. Second, we test whether participants revise beliefs about the unobserved statistic after receiving information. If prior beliefs are unbiased, they should not respond on average to truthful information. Third, we test whether providing information leads participants to change revealed behaviors.

Our main revealed landlord outcome is whether they choose to receive informational materials about applying for the ERAP. This is a real choice: the program sent materials to landlords who requested them.[33] We see this measure as revealing landlord interest in ERAP participation.

Our main revealed tenant outcome concerns how tenants treat a real opportunity to propose a payment plan to their landlords. A payment plan is an agreement in which the tenant agrees to repay none, some, or all of their rental debts over time. Tenants could always negotiate outside the experiment or decline to propose a plan. However, forming

[32]This procedure assumes that eliciting beliefs does not itself cause beliefs to move in one direction on average. To test this assumption, we collected a pre-registered placebo belief for landlords. Landlords do not update on average about the placebo after receiving information (Appendix A.5).

[33]Because of concerns about power, we pre-registered this and other revealed belief outcomes as secondary. We show other outcomes in Section 1.5.2 and Appendix C.4 and conduct tests for multiple hypotheses.

a payment plan in the experiment may reduce transaction costs, confer legitimacy, or provide useful structure to tenant proposals. We see this measure as revealing tenant interest in bargaining.

In the survey, we ask tenants with back rents if they are interested in a payment plan. If they are, then we ask how much they propose to repay and the payment period. We confirm with tenants several times that we will actually email their payment plan, and we do so after the survey concludes. Payment-plan proposals in the experiment are nonbinding offers. We restrict proposals to involve tenant repayments between zero and the total back rents owed.[34]

Connection to the Model. Landlord's beliefs about tenant repayment exactly correspond to p_{Li}. Tenants' beliefs about landlord behavior in the DG map to beliefs about a_{Li}. Beliefs about landlord propensity to drop eviction filings map to beliefs about k_{Li}. If tenants make bargaining offers based on their beliefs about whether Equation (1.5) is satisfied, then believing either of a_{Li} or k_{Li} is low directly reduces bargaining propensity.

Unlike in the model, the experiments compare landlords and tenants to ground truth, rather than each other's beliefs. We view this as a conservative test about the importance of noncoincident beliefs, although it is possible that both parties' misperceptions exactly cancel.[35]

1.5.2 Results

Landlords are highly optimistic about the probabilities of recouping back rents (Figure 1.3A). Their incentivized beliefs about average tenants are 18 percentage points higher than the true value of 6% (s.e.: 1.4), with substantial mass above 50%.[36]

Landlords who have ever evicted their tenant have 7 pp more accurate beliefs on average (Figure A.16B). Nevertheless, more than 25% of landlords who have ever evicted

[34]We do not allow negative or large payments to avoid the possibility that such proposals could cause retaliation from landlords or otherwise harm tenants.

[35]We did not elicit beliefs about the court eviction process symmetrically because doing so could introduce ethical concerns. For instance, we did not want to provide tenants with information that tenants rarely repay eviction judgments, as that could harm landlords.

[36]Landlords' beliefs about their own tenants are even more optimistic: landlords believe they have a 43-percent chance of recouping back rents from their own tenant (s.e.: 1.8). Landlords could in principle be unbiased about their own tenant if this sample of landlords would indeed have a 43-percent chance of recouping back rents from their own tenants. That seems unlikely, since it would render these landlords extraordinarily effective relative to the average landlord's recoupment rate of 6%. Beliefs about own and average tenants are highly correlated (Figure A.16A), which further suggests that landlords are biased about their own tenant.

believe there is at least a 35-percent chance of recouping back rents from the average tenant.

Tenants, meanwhile, exhibit two beliefs that may suppress bargaining. They underestimate landlord altruism (Figure 1.4A). The average tenant reports that 47% (s.e.: 0.84) of landlords would choose (\$10 self, \$10 own tenant) to (\$20, \$0), when in fact 69% do (vertical red line). Similarly, we find that tenant beliefs about landlords' propensity to drop eviction filings in the future are overly optimistic (Panel B).

Presenting information about the average tenant causes 43% of landlords to update beliefs about their own tenant (s.e.: 3.8 pp). The average unconditional belief update is -12 pp (s.e.: 1.6) (Figure A.8A). Thus, the difference between landlords who have and have not evicted — a measure of eviction experience — is associated with belief updates that are 60% as large as receiving information directly.

Information corrects both tenant beliefs, although less than in the landlord experiment. If given information, 32% and 37% of tenants update beliefs about altruism and dropping filings, respectively. Information about landlord altruism increases beliefs on average by 5.9% (Figure A.19). There is heterogeneity in the belief update, since the information is central in the beliefs distribution. Among people whose beliefs about their own landlord lie above 69%, beliefs shift down by 8 percentage points. For bargaining, the treatment reduces beliefs on average by 5.9 percentage points (Panel D).[37]

We obtain lower bounds on the share with misperceptions by calculating the percent of participants whose prior beliefs are at least 20 pp from ground truth (Figure A.17).[38] 33% of landlords and 79% of tenants have misperceptions by this measure. 17% of landlords and 41% of tenants both have incorrect beliefs and choose to update when told the truth.[39]

Revealed-Preference Outcomes: Landlords. The treatment increases the probability that landlords request informational materials about the ERAP by 11 pp, or 17% (Figure 1.3B, $p = 0.02$). These are large effects for a light-touch information intervention.

These intent-to-treatment estimates are stable or rise if we change how we choose

[37]Tenants are more uncertain than landlords: they are more likely to report 50%, which may suggest cognitive uncertainty (Enke and Graeber, 2019). Dropping tenants reporting 50% has little effect on average misperceptions. Using a Likert scale, we ask tenants to report how certain they are about their prior beliefs. Uncertain tenants report lower beliefs about whether their landlord would split the gift card and whether their landlord would bargain.

[38]We choose 20 pp so that tenants who report 50 are conservatively treated as having correct beliefs.

[39]The only way to obtain small shares of misperceptions using this approach is to consider the share of tenants who have incorrect priors about *both* beliefs, and choose to update. 5% of tenants have misperceptions according to this measure. However, as having one mistaken belief suffices impede bargaining, this test may be overly conservative.

controls or restrict the sample to attentive landlords. As this treatment does exhibit imbalance (Section 2.3), including controls is an important check.

Treatment effects are concentrated among landlords whose prior beliefs were especially incorrect and thus who, on average, updated more in response to the information treatment. We estimate an instrumental-variables specification for the effect of landlord beliefs about recoupment on requests for ERAP materials, using the prior beliefs, in Appendix A.5. We also conduct placebo tests which examine whether unrelated beliefs move due to the information treatment, using additional data we collected on landlord beliefs about time to process an eviction.

We test the effects of information on several other revealed landlord outcomes and find no significant effects of the treatment among any outcome except requesting ERAP materials (Table A.13; see table notes for outcomes). Muted or wrong-signed effects on other outcomes (e.g., tenant referrals) suggests that information may be insufficient to change other behaviors.

A related concern is multiple-hypothesis testing. A stacked test rejects the null that the treatment does not affect any measured outcome ($p = 0.01$). Meanwhile, multiple-hypothesis corrected p-values (Romano and Wolf, 2005) attenuate the effect for requesting materials ($p = 0.10$). Altogether, we recommend caution when interpreting this evidence.

Revealed-Preference Outcomes: Tenants. 74% of tenant survey participants are eligible for payment plans because they owe their landlord money for rent at the time of our survey. Neither information treatment had a significant intent-to-treat effect on whether the tenant requested a payment plan. The absence of these overall effects is unsurprising because the true information is more central in the prior-beliefs distribution. Some tenants update up and others update down, which attenuates the average effect.

As a result, we split tenants by whether their prior beliefs about their own landlord lie above or below the information. We find a moderate effect for the altruism correction (Figure 1.4C) and no detectable effect for the bargaining correction (Panel D). In particular, tenants with optimistic beliefs about their landlord's altruism become 9 pp less likely to request a payment plan ($p = 0.08$). This result suggests that correcting tenant misperceptions can affect real bargaining behaviors, but we acknowledge the moderate effect size and multiple hypotheses.

Survey responses from landlords about payment plans suggest the presence of inefficient bargaining failure. We surveyed landlords to whom we sent payment plans from tenants. Out of 691 payment plans sent (including emails that bounced), we received

96 responses indicating whether the landlord would accept the payment plan (response rate: 14%). 11% of respondents said they would accept the payment plan outright, and 45% said they would accept or discuss the payment plan with their tenant. Even if every non-respondent would reject the proposed payment plan, these results imply that a minimum of 6% of landlords would accept or discuss the payment plan with their tenant. The payment plans were nearly costless to propose to landlords: we simply emailed tenants' survey responses to landlords.

1.6 Connecting the Model and Experiment to Evictions and Policy

1.6.1 The Model and Experiment Predict Evictions

1.6.1.1 Adjusted Surplus Index

We now study whether the forces in the experiment predict evictions and other real-world outcomes, as the model suggests. The model implies a notion of "adjusted surplus" in the landlord–tenant relationship. For each individual, we wish to construct:

$$\theta_i := k_{Li} + k_{Ti} - \frac{\Delta p_i d_i}{\bar{a}_i}, \tag{1.10}$$

for adjusted surplus θ, which comes from Equation (1.5). As adjusted surplus increases, the model predicts eviction to be less likely.

We test whether a proxy for θ_i negatively correlates with eviction risk. The advantage of testing θ_i's correlation with outcomes is that θ_i combines the three forces in a model-consistent way. Moreover, as it is similar to an index formed from the three forces, it can add power. But since it implies strong parametric restrictions, we also show the effects from each force separately.

It is infeasible to construct θ_i directly. First, we do not observe k_{Ti} or k_{Li}. Second, as the landlord and tenant experiments were conducted separately, for a given tenant, we do not know their landlord's a_{Li}, p_{Li}, or k_{Li}, and vice-versa.

We form proxies of landlord and tenant costs as follows (see Appendix B.1 for details). We form proxy $\tilde{k}_{Ti}$ using questions where we ask tenants about their private cost of having an eviction. We form proxy $\tilde{k}_{Li}$ using questions where we ask whether the landlord would accept less than full rent for the given tenant. Appendix B.1 also details how we prepare the misperceptions and altruism data from the experiment to be suitable for this test.

We then form adjusted-surplus proxy $\tilde{\theta}_i$ by plugging in the individual's own misperception, own altruism parameter, and own cost proxy into Equation (1.10). We set their opponent's misperceptions, altruism, and cost at 0.

We link the data from the experiment to eviction filings (Appendix B.1). We consider any eviction filing among the same individual at the same address after January 2018, so the filing itself could affect the index. We restrict to tenants who have back rents at the time of the tenant survey. For power, we pool the landlord and tenant experiments.

Specification. We estimate:

$$Y_i = \beta_0 + \beta X_i + \gamma \text{Controls}_i + \varepsilon_i \tag{1.11}$$

where Y_i is an indicator for having an eviction filing and X_i is a covariate. Controls_i contains an indicator for whether the observation comes from the tenant survey versus the landlord survey, and an indicator for randomized variation in how the landlord's costs were elicited. When covariate X_i is an indicator for having positive adjusted surplus $1(\tilde{\theta}_i > 0)$, the parameter of interest, β, shows the relationship between surplus and eviction filing propensity. We also test whether $\tilde{\theta}_i$'s constituent elements (real costs, hostility, misperceptions) are correlated with evictions.

1.6.1.2 Results

Eviction Filings. Adjusted surplus negatively predicts eviction filings (Figure 1.5, Panel A). Pooling both samples, having positive adjusted surplus is associated with a 10 pp reduction in eviction filings. These effects are economically large relative to the mean of 37%. They are also statistically significant and consistent across landlords and tenants.

Court costs are predictive among landlords, but not tenants. On the other hand, we find that hostility among both parties has a large quantitative association with evictions. Among tenants, hostility toward own landlords is associated with a 10 pp additional propensity to have an eviction filing (off a base rate of 33%). Among landlords, hostility toward own tenants is associated with a 15 pp additional propensity (65% of their base rate of 23%). Finally, misperceptions predict evictions in the pooled sample, with results driven by tenants.

These patterns survive controls (Figure A.20A). Moreover, hostility and costs predict ultimate eviction *judgments* (Figure A.20B), which are rarer and thus worse-powered. The parametric combination given in the adjusted surplus index is suggestive. The patterns also persist if X_i is an indicator for being in the bottom-quartile of surplus, rather than an indicator for being positive (Figures A.20C and D).[40]

These correlations let us conduct back-of-the-envelope exercises to quantify the shares inefficient and repugnant. First, 48% of the parties with eviction filings have above-median misperceptions and 26% are hostile. These figures are upper bounds: if all these parties' filings were caused by misperceptions or hostility, that implies that just under half of evictions are inefficient and 26% are repugnant.[41]

As a second exercise, we consider the marginal predictive effect of above-median misperceptions or hostility from Equation (1.11). As above-median misperceptions are not themselves correlated with judgments, we cannot reject zero inefficiency from this back-of-the-envelope alone. Hostility is associated with a 7.1 pp increase in judgments, off a base rate of 21.0%. As 20.9% in this sample are hostility, that implies that 7% of evictions are repugnant ($\approx 0.071 \times 0.210 \div 0.209$).

Misperceptions and Hostility Are Complements. We find support for the model prediction that hostility and misperceptions are complements (Figure 1.5). Among participants with below-median misperceptions, hostility is not associated with any increase in eviction filing risk. However, hostility has substantial bite among parties with above-median misperceptions, in which case hostile participants have 13 pp larger filing risk than non-hostile participants.

Code Violations. As additional evidence that hostility correlates with behavior outside the lab, we show that highly hostile tenants live in units that have significantly higher rates of 311 calls to Memphis's code enforcement department (Appendix A.5). On the other hand, tenant bargaining offers for the payment plan are not correlated with adjusted surplus (Figure A.21).

1.6.2 Policy Evaluation

Given the magnitudes of hostility and misperceptions, we cannot rule out the possibility of inefficient evictions. A natural next question is whether rental assistance stops evictions, and whether it stops inefficient ones. We present graphical evidence here. In

[40] We observe a discernible increase in hostility in tenants who do the DG in the 100 days after a filing, though this increase also manifests for hostility toward random tenants (Figure A.11).

[41] The 48% is not mechanical, as above-median misperceptions could have been negatively correlated with judgments.

Appendix A.5, we provide a formal event study, as well as details about data construction, identification, balance and robustness checks.

Data and Sample. We use administrative data on ERAP program receipt to form panels of households before and after ERAP payment. Our main sample consists of about 4,700 households who applied between September 1, 2021 and December 31, 2022, and is comparable to those in the experiment and the full sample of ERAP payees (Table 1.1). ERAP payment involved full repayment of rental debts and utility bills, as well as payment of between 1–3 months of future rent payments. Payment was made directly to landlords in most cases.

Graphical Evidence. Figure 1.6 shows, among tenants who receive an ERAP payment, the cumulative shares with filings (blue line) and judgments (orange line) relative to the payment week. These shares rise linearly in the weeks before payment. After payment, the filing share stops rising for about eight weeks and then returns to trend. Judgments *increase* after payment. Extrapolating linear trends fit from the period between two and 16 weeks before payment (dashed lines) suggests filings fell for about eight weeks versus pre-payment trend, while judgments were above trend.

Event Studies. We conduct formal event studies that leverage two sources of variation. First, we use variation in payment timing among households who apply to ERAP around the same time and are paid. Second, we employ comparisons between paid and non-paid households who apply to ERAP, in a differences-in-differences type design.

The event studies confirm the graphical evidence and yield modest results on judgments (Appendix A.5). In the best-case specification for ERAP (Table A.24), we reject that ERAP reduced judgments by more than 7.6 pp, and 95% confidence intervals include 0 ($p = 0.150$). As households receive more than $5,000 on average, it is perhaps surprising that the policy has a small effect. Assuming homogeneity and dividing the average payment amount by the average treatment effect, the best-case fiscal cost to stop a judgment for six months exceeds $70,000. Other specifications suggest that the cost exceeds $100,000.

Meanwhile, we detect effects on eviction filings at 1–2 months, but mixed evidence on filings at 6 months. The best-case fiscal cost to stop a filing for six months exceeds $35,000.

We highlight several limitations. First, this high fiscal cost is partially driven by the fact that judgments are rare in this sample and period (Figure 1.6), which both reduces power and the best-case impact of ERAP. In the best-case specification, despite still implying a large fiscal cost to stop an eviction, the point estimate is around 38% of the maximal effect if ERAP stopped all judgments among payees. Still, other specifications give weaker effects, and the result still implies that the remaining 62% simply take ERAP and evict within six months anyway.[42] The second limitation is that Memphis's ERAP had a distinct arm that provided legal assistance to the most at-risk tenants. We focus on the cash-only arm for external validity, as most ERAPs did not have such a legal services arm, and because the legal arm may complicate identification (see discussion in Appendix A.5). However, this sample restriction reduces power and could understate the overall effects of the Memphis ERAP on eviction, as the legal sample may have had different effects.

1.6.3 Explaining ERAP's Effects

The model and experiment data can potentially explain why so many landlords are inframarginal to ERAP. The model raises the possibility of perverse selection on altruism, wherein ERAP is more desirable for altruists than hostile parties. Relatedly, ultimate ERAP receipt involves cooperation between landlords and tenants. Applicants for ERAP must upload a lease and overdue rent ledger, documents that typically require landlord cooperation to obtain. After application, their landlord also must approve it, although if the landlord declines, the tenant may be able to obtain a check for the overdue rent directly.

Among landlords, we focus on whether hostility measured in the experiment predicts demand for ERAP materials. We also link some tenant experiment participants to program receipt using administrative ERAP data (Appendix B.1). We cannot do the corresponding exercise for landlords because we have a small number of landlord survey participants linked to payment status.

Results. Model forces predict landlord interest in ERAP (Figure A.21). Altruism-adjusted surplus is associated with more than a 10 pp increase in landlord interest. These results are driven by hostility, which is associated with more than a 20 pp reduction.

Tenants who are hostile to their own landlord are less likely to receive funds quickly (Figure 1.7A, black series). We focus on the survival function of open applications as measured by days from initial submission to payment, comparing hostile and non-hostile tenants. In particular, hostile tenants are 15 pp less likely to receive ERAP funds in

[42] On the other hand, these cost estimates may be lower bounds because they only include the cost of the first payment made to a household at a given address.

less than 50 days. Adding controls for prior beliefs, tenant demographics, and tenants' economic conditions changes little (blue series). We also regress ERAP receipt in the administrative data on hostility. Hostile tenants are 15.6 pp less likely to receive payment at all ($p = 0.056$). Kaplan-Meier failure curves, which aggregate these tests, show similar effects (Panel B, Wilcoxon p-value < 0.05).

1.7 Quantitative Analysis of Eviction Model

1.7.1 Empirical Model

We augment the model in Section 1.2 to be suitable for empirical estimation. Appendix C.3 contains model and estimation details.

Payoffs. We propose landlord payoffs from four separate end states. We model the landlord's decision to take ERAP payment, beginning from once the tenant has applied. If the landlord declines to take ERAP, she can either get an eviction judgment ("eviction") or bargain not to get a judgment. If the landlord takes ERAP, she receives a payment of the back rents less an eviction cost. She can either get a judgment or bargain, but the judgment and bargaining payoff are less valuable as the tenant has no back rents.

Building on the environment in Section 1.2, the landlord's payoffs V_{Li}^{k} are:

$$\begin{aligned} V_{Li}^{\text{NoERAP,Evict}} &= (1 - a_{Li}) p_{Li} d_i - a_{Li} k_T - k_L + \varepsilon_{i1} && (1.12) \\ V_{Li}^{\text{NoERAP,Bargain}} &= (1 - a_{Li}) n_i^*(d_i) + \varepsilon_{i2} && (1.13) \\ V_{Li}^{\text{ERAP,Evict}} &= d_i - k_L^P - a_{Li} k_T - k_L + \varepsilon_{i3} && (1.14) \\ V_{Li}^{\text{ERAP,Bargain}} &= d_i - k_L^P + (1 - a_{Li}) n_i^*(0) + \varepsilon_{i4}. && (1.15) \end{aligned}$$

Terms without subscripts i are constant across households and are parameters to be estimated or calibrated, whereas terms with subscripts i are data or unobserved idiosyncratic shocks.

The terms $\varepsilon_{i1}, \ldots, \varepsilon_{i4}$ represent idiosyncratic payoff-specific shocks from each state. For this reason, we can interpret $-(\varepsilon_{i1} - \varepsilon_{i2})$ as an extra idiosyncratic eviction cost relative to bargaining, if ERAP were not available.

The landlord's payoffs when she does not take ERAP (Equations 1.12 and 1.13) are the same as Section 1.2, except $n_i^*(\cdot)$ represents the Nash Bargaining payoff (described below). Whenever the landlord takes ERAP, she receives the mechanical ERAP payoff of back rents owed d_i, less the additive take-up cost k_L^P. In addition to the mechanical ERAP payoff, after taking ERAP, the landlord can choose to evict and get only the court costs (Equation 1.14), as there is no rental debt that she can recover. Alternatively, she can bargain with the tenant to avoid eviction and additionally get the Nash Bargaining payoff (Equation 1.15).

Bargaining. We posit asymmetric Nash Bargaining over surplus from avoiding court, which yields standard closed-form solutions for $n_i^*(d_i)$ and $n_i^*(0)$ (see expressions in Appendix C.3). These Nash payoffs naturally depend on tenant bargaining power $\beta \in [0, 1]$. When $\beta = 1$, payoffs correspond to a take-it-or-leave-it offer by the tenant, as in Section 1.2.

Nash Bargaining can occur in both the take-up and non-take-up case. In the non-take-up case, landlords and tenants bargain over tenant rental debts and court costs, as in Section 1.2. In the take-up case, they bargain over court costs only, as rental debts are fully paid by ERAP and not recoverable in court. Even absent the possibility of recouping rental debts, landlords still wish to evict the tenant if the sum of court costs and the shock is negative. For instance, court can be net profitable for landlords if formal eviction sends a message to other tenants or allows them to turn over the unit faster.

Nash Bargaining occurs if and only if bargaining is weakly profitable for both parties. In the case where the landlord does not take ERAP, eviction occurs if and only if:

$$k_L + k_T + \varepsilon_{i2} - \varepsilon_{i1} \leq \frac{(p_{Li} - p_{Ti})d_i}{\bar{a}_i}, \tag{1.16}$$

whereas when the landlord does take ERAP, eviction occurs if and only if:

$$k_L + k_T + \varepsilon_{i4} - \varepsilon_{i3} \leq 0. \tag{1.17}$$

Observe that these equations are lightly augmented versions of Equation (1.5). The expression in Equation (1.17) comes from substituting $d_i = 0$ into the eviction condition in Equation (1.16) and exchanging the shocks.

Timing and Landlord Maximization Problem. Landlords draw shocks ε_{ik}. These values and landlord payoffs are publicly observed by the tenant. These shocks define whether either of Equations (1.16) and (1.17) are satisfied. There are four distinct combinations of whether eviction or bargaining is possible when the landlord does and does not take up ERAP, yielding four possible maximization problems depending on the realization of the shocks. The landlord then chooses the actions ERAP/No ERAP and Evict/Bargain which maximize her payoffs, depending on which maximization problem she faces (see full problem in Appendix C.3).

Discussion. Despite its simplicity, the model captures many relevant economic forces.

Valuations of the physical property. There may be a portion of the landlord payoff (owing to possession of the physical property) that she automatically recoups with an eviction judgment, regardless of tenant behavior. We can think of this component as entering the landlord payoff from an eviction judgment via k_{Li}, as well as the landlord–tenant transfer.

Credit constraints. As in Section 1.2, tenant credit constraints enter the model through the bargaining costs and subjective beliefs that the tenant will repay if a court eviction occurs.

Money judgments versus judgments for possession. To reduce the number of landlord actions, the model assumes all judgments are for money. In reality, about half are exclusively for possession of the property (Section 1.3). Note that, in principle, one can think of $p_i d_i$ as representing the total expected value of the landlord–tenant transfer in court, which could include possession of the property only. However, in our empirical implementation, we treat p_i as beliefs about recoupment, and d_i as back rents. We explore the consequences of this simplification in robustness.

1.7.2 Estimation

We let ε_{ik} be type-I extreme value with scale parameter σ.[43] We seek to estimate the vector of parameters $\Theta := \{k_L, k_T, \beta, \sigma\}$. The values $X_i := \{a_{Li}, a_{Ti}, p_{Li}, p_{Ti}, d_i, k_L^P\}$ are either observed in the data from the landlord and tenant experiments or calibrated using that data.

The random shocks ε_{ik}, together with latent heterogeneity in the data, induce model-implied shares of landlords in each of the end states. We use the experiment to obtain the distribution of beliefs, altruism, and other data among the landlords and tenants (X_i). Given this distribution of inputs into model-implied landlord choices, we solve for the parameters that match a set of moments using the Method of Simulated Moments (MSM). Simulation-based estimation is useful because the shocks change the choice-set restrictions (Equations 1.16 and 1.17) and then give non-standard take-up probabilities.

We estimate four parameters from 24 moment conditions. We target the following

[43]The type-I extreme value shocks impose an Independence of Irrelevant Alternatives (IIA) assumption, which forecloses richer substitution patterns between judgments and bargaining in our model.

moments: ERAP's treatment effects on judgments; the mean judgment rates among landlords who take up ERAP; the take-up rate among landlords; the bargaining offer made by tenants in the payment plans; and interactions between experimental variables and predicted or observed eviction judgments, landlord take-up, and the tenant bargaining offer. Intuitively, the estimator uses the functional form above to match the event study's "well-identified moment" and the correlations in the experiment. We form standard errors by block bootstrapping both the treatment effect moment and the experiment data.[44]

Data and Calibration. A key advantage is that most of X_i can be read directly from the experiment data. For instance, we plug in a_{Li}, a_{Ti}, and p_{Li} from the experiment. Thus, we directly observe key primitives that are typically estimated or calibrated in other research. Still, we do not observe take-up costs. We calibrate $k_L^P = \$500$ and show sensitivity to this assumption.[45]

We make several additional decisions to facilitate estimating the model (Appendix C.3). First, we use the tenant proposals about payment plans that we collected in the tenant survey to proxy for the Nash solution $n_i^*(d_i)$. The data we collected on tenant proposals constitute an unusual and rich trove of information about informal landlord–tenant negotiations. Second, we simulate assortative matching on altruism at the midpoint of our suggestive estimates (Appendix A.5). Standard errors take into account the random matching process. Third, as we do not exactly elicit tenant beliefs about repayment probabilities, we assume that their beliefs that we do elicit scale one-for-one with beliefs about repayment and present how different choices affect results.

What role does selection into survey participation play? Our experiment sampling frame was all tenants, or landlords of tenants, who applied or began applying. The model begins once a tenant applies. If the experiment data are representative of the sampling frame, then they need no adjustment. Our diagnostics in Section 1.4 do not suggest important reasons to be concerned.

We permit correlated altruism and misperceptions within landlords/tenants by us-

[44]Our estimator may be biased because the shocks enter both the maximand and the choice restrictions. In particular, we do not account for potential bias induced by minimizing the distance between the data and a nonlinear function of the shocks. This problem is a cousin to the familiar bias in Maximum Simulated Likelihood (Train, 2009).

[45]This parameter includes all costs associated with interacting with ERAP and payment delays. Our calibration may exceed a strict accounting. For instance, suppose: landlords discount future payments at 8% per year; it takes two months to receive payment; it takes three hours to interact with ERAP and produce relevant materials; and landlords value their time at \$50 per hour. Under these assumptions, ERAP costs are lower than \$500. However, landlords also expressed significant frustration at ERAP which seemed to exceed the strict accounting costs. We can explain this frustration if interacting with ERAP and managing applications impose additional hassle costs.

ing the empirical distributions captured in the experiment. However, an important assumption, explored in robustness checks, is that landlord and tenant eviction costs are orthogonal to altruism and misperceptions. We also assume that tenants have homogenous costs, whereas landlords have heterogeneity via ε_{ki}.

1.7.3 Results

Parameter Estimates and Model Fit. The model delivers reasonable parameter estimates (Table 1.3). Total unconditional eviction costs (i.e., the sum of k_{Li} and k_T) are positive (\$3,400 on average), indicating that court is costlier than bargaining. However, landlords' mean eviction cost is negative, at –\$641, while tenants' is highly positive (exceeding \$4,000). The standard deviation of landlord eviction costs (given by σ) is about \$3,500, suggesting that eviction could be efficient for many parties. Tenants have high bargaining power ($\beta > 0.8$). To assess model fit, we compare several model-simulated moments with their targets and find they are broadly similar (Panel B and Table A.26).[46]

Efficient and Inefficient Evictions. We use the model to quantify the shares of evictions and bargaining that are efficient versus inefficient. Figure 1.8A shows an empirical version of the model comparative statics in Diagram 1.1 (see Table A.25 for levels). We consider a state of the world without ERAP, shutting down those payoffs and simulating landlord behavior if the two non-ERAP payoffs are the only ones available. Altogether, 19.1 pp of the sample obtains evictions. This low percentage reflects that eviction judgments are extreme events, even among our sample.

Of those who get evictions, 23.9% are inefficient (s.e.: 4.7%). We obtain this share by counting the number who obtain evictions despite positive surplus $k_L + k_T + \varepsilon_{i2} - \varepsilon_{i1}$. The remaining three quarters are efficient.[47] Hostility causes about two thirds of inefficient evictions, meaning about one in six evictions is repugnant. We attribute an eviction to hostility if it would not occur if hostile parties all had classical preferences (i.e., if we set $a_{Li} = a_{Ti} = 0$ when less than 0).

These results lie between the back-of-the-envelope figures for inefficient/repugnant evictions that we computed in Section 1.6. After all, we target the correlations between misperceptions/social preferences and judgments in the model directly. The other

[46]Standard errors are large for decomposing landlord versus tenant costs. Nevertheless, the results below show that the main model conclusions about inefficient evictions are estimated precisely.

[47]Note that even if the means of k_{Li} plus k_T are larger than 0, the large scale parameter for landlord eviction costs means a substantial share can have a negative sum.

moments that we target do not greatly alter the conclusions from the simple back-of-the-envelope.

On the other hand, bargaining is far more common than eviction, and nearly all bargaining is efficient (Table A.25B).

Given most evictions are efficient, policy that stops evictions destroys joint surplus on average (Panel B). In principle, as one in four evictions is inefficient, the average eviction cost could be positive if inefficient evictions are four times as costly as efficient evictions are beneficial. However, stopping the average eviction costs \$871 of private surplus if we do not normatively respect social preferences (i.e., the average $k_{Li} + k_T = -871$) and \$2,033 if we do normatively respect social preferences (i.e., the average $(1 + a_{Ti})k_{Li} + (1 + a_{Li})k_T = -2\,033$).[48]

Even though the average eviction has positive private surplus, misperceptions and social preferences are still key to the economics of eviction. First, Panel B shows that the average efficient eviction yields more than \$2,000 in surplus without respecting social preferences. Thus, non-classical forces reduce the average surplus from eviction by half.

Second, a partial explanation for why few evictions are inefficient is that altruism offsets misperceptions. We show this point by conducting an exercise in which we shut down altruism — replacing social preferences as being classical if altruistic — and resimulate behavior assuming the same value of other primitives (Panel C). In this exercise, 56% of evictions would be inefficient, and the average eviction has a positive court cost. Thus, high rates of altruism stop many pairs from inefficiently evicting. The underlying feature of the data that is responsible for this result is the positive correlation in the experiment between landlord altruism and the belief that their own tenant would repay in an eviction (Figure A.24). Consequently, this result highlights the value of collecting the joint distribution of preferences and beliefs.

Our finding that the average eviction has positive private surplus rejects the view that evictions' private costs exceed their benefits on average. A given anti-eviction policy can still be justified if, among other arguments: (i) the policy targets inefficient evictions better than average, so that the average eviction averted has positive costs; (ii) evictions' externalities exceed their net private benefits; or (iii) the policy induces self-targeting. The potential presence of bargaining failure is a less compelling policy rationale in isolation. If the objective is to assist low-income households undergoing a costly and traumatic

[48]Since we shut down the ERAP payoffs in this exercise, $k_{Li} := k_L - (\varepsilon_{i1} - \varepsilon_{i2})$. Intuitively, as tenants are altruistic on average and many landlords who evict have negative court costs, social preferences only amplify the efficiency costs of stopping the average eviction.

event, direct transfers to tenants may be more efficient than stopping the eviction per se. That said, we do find sufficiently prevalent bargaining failure to meaningfully reduce the magnitude of externalities that would justify eviction policy.

Perverse Selection on Altruism. Table 1.4 varies nonclassical forces and court costs (Panel A) and considers policy counterfactuals (Panel B).

First, we conduct a positive analysis of ERAP's low treatment effects, confirming that altruism reduces ERAP's effectiveness. In our primary simulation (Row 1), ERAP has a small TOT on evictions (Column 3). A natural way of scaling the TOT is to divide Column 3 by Column 5, the baseline rate of counterfactual judgments. When we shut off both social preferences and misperceptions entirely, the TOT becomes *positive* (Row 2). But underscoring the results in Section 1.6, shutting off social preferences (Row 3) or particularly landlord altruism (Row 4) would cause ERAP to have a –15.7 pp effect on judgments. The rate of evictions also rises, as eviction is more desirable without altruism, but the TOT as a percent of the counterfactual judgment rate is almost 60%.[49] Adjusting misperceptions (Rows 7–8) makes a smaller difference for ERAP's TOT. Doubling costs for both parties would lead to a large ERAP TOT (–3.5 pp out of 6.3 pp), but counterfactual evictions then become very rare (Row 10).

As foreshadowed in Section 1.2, even absent non-classical forces, ERAP is not perfectly targeted. A classical force pushes toward enrollment among inframarginal "never-evictors," since ERAP is more desirable for landlords with high court costs. Altruism amplifies this force, raising the value of ERAP and reducing the value of eviction.

However, we augment the model in Section 1.2 to allow pairs to evict even with ERAP. Such "always-evictors" take ERAP and still pursue eviction (Equation 1.14). The model then captures a countervailing mechanism, as court costs and altruism both *reduce* inframarginality from always-eviction. They both make eviction less desirable, including when in ERAP. Still, the never-evict force is more quantitatively important than the always-evict force. Eliminating altruism amplifies ERAP's TOT (Rows 4–5), whereas eliminating hostility attenuates ERAP's TOT (Row 6).

ERAP's Targeting and Welfare Impacts. Now, we conduct a normative analysis of ERAP. Even if most evictions are efficient, a given eviction intervention could still improve welfare if well-targeted. However, in the primary estimate, 23% of the evictions that

[49]The difference between Rows 4 and 5 is explained by tenant altruism, which generates favorable Nash bargaining offers for landlords. Shutting down tenant altruism makes ERAP less effective because tenant altruism has a larger proportionate effect on the Nash bargaining outcome when landlords take up (Equation A.5 vs. A.6).

ERAP stops are inefficient (Row 1, Column 4), meaning it is about as well-targeted as stopping the average eviction.

We next present a suggestive analysis of ERAP's welfare impacts, when the planner does and does not normatively respect altruism (Columns 6 and 7). We compute the average gain from ERAP as the landlord's net idiosyncratic payoff from enrollment, plus the effect of ERAP on eviction times the net eviction cost, less the landlord's take-up cost (see Appendix C.3 for details). ERAP's other impacts run through the mechanical ERAP transfer or intrahousehold bargaining payments, both of which we assume have zero social value or cost. In this sense, we assume that the government can finance ERAP without distortionary taxation and employ a marginal cost of public funds of 1. We ignore other administrative costs or fiscal externalities. Given these limitations, we treat the exercise cautiously, noting that we probably produce an upper bound on welfare generated from ERAP.

ERAP increases welfare in the baseline specification. Intuitively, ERAP yields gains to inframarginals, as landlords enroll when they receive a large draw of the idiosyncratic shock. ERAP has a small effect on stopping evictions. If the effect were larger, that would reduce surplus on average. Echoing our finding of perverse selection on altruism, landlord altruism attenuates ERAP's welfare impacts (Row 4 vs. Row 1).

ERAP gives \$211 (Row 1, Column 6) or \$83 (Column 7) in net welfare gains, if society does not or does respect altruism. We scale payoffs by landlord and tenant altruism in Column 7 but always count transfers as having zero social value. These gains involve transferring at least \$2,300 on average (the average back rents at the time of the tenant survey) and administrative costs (which we ignore). The transfer is not waste, especially as landlords in our study are often lower income themselves. Still, the increase in welfare from ERAP only comes from aspects of the program that are unrelated to stopping evictions.[50]

Counterfactuals. Finally, the model lets us study how counterfactual policies affect ERAP's TOT and targeting (Rows 11–13). As one example, we simulate fining landlords \$2,000 if they take ERAP and then evict a tenant (Row 11). This intervention raises ERAP's TOT to about one third of total counterfactual evictions. We also simulate restricting

[50]Relatedly, ERAP's Marginal Value of Public Funds (Hendren and Sprung-Keyser, 2020), respecting altruism, is $\frac{T+83}{T+\text{FE}}$, for ERAP transfer $T \approx \$2{,}300$ and fiscal externality FE. FE includes administration costs and fiscal externalities that run through ERAP's effects on evictions. This calculation places equal weight on landlord and tenant WTP, so intrahousehold transfers do not enter the MVPF numerator. Given administration costs of around 10%, ERAP has a similar MVPF to providing a non-distortionary cash transfer.

eligibility among those in the top 10% most predicted to have a large treatment effect from an OLS regression.[51] Doing so would generate a TOT that is about half of the counterfactual judgment rate. It would also raise the share of evictions that are inefficient to about 40% of evictions prevented, though it would reduce ERAP's welfare impacts. While it may be infeasible to completely restrict eligibility to those households, the program could certainly expand outreach.

Robustness. Changing model moments or assumptions does not greatly affect the main conclusions above: that a material but minority share of evictions are inefficient, that hostility drives the inefficiency, and that eliminating landlord altruism increases ERAP's TOT (Table A.27). Across different checks or assumptions, 15–41% of evictions are inefficient.

We discuss several especially important tests. The main conclusions are relatively robust to our assumption about landlords' take-up cost k_L^P. If we double k_L^P to $1,000 (Table A.27, Row 8), we obtain that about a third of evictions are inefficient. While we see that as an implausibly high take-up cost, future research that quantifies take-up costs would be valuable.

We also explore the importance of the assumption that landlords' eviction cost k_{Li} is uncorrelated with p_{Li} and $\overline{a}_i$. We replace landlords with above-median misperceptions as having 0.2 s.d. higher eviction costs. This value would approximately generate the observed correlation between misperceptions and cost proxies in the experiment. Allowing such correlation has modest effects (Table A.27, Row 10), raising the share inefficient to about 30%.

While measurement error could affect the magnitude of our results, measurement error would need to be severe to overturn our conclusions. We focus on non-classical measurement error that would cause systematic bias in our elicitations due to, e.g., lack of numeracy in the study population. First, we simulate the share of efficient/inefficient evictions that would obtain if we rescale social preferences as $\hat{a}_{ji} := sa_{ji}$ for different $s < 1$ (Figure A.25A) and posit the same value of estimated and calibrated parameters. We do not study $s > 1$ because the model requires $a_{ji} < 1$. If social preferences are overstated by half ($s = 0.5$), then about one in three evictions would be inefficient. The reason

[51] The demographics exclude variables like beliefs or altruism that are not available on the application. The landlord variables are: age, gender, race, education, landlord report of tenant's tenure in landlord experiment, whether the participant reports being a landlord (versus property manager or other), rent in the unit, number of units, years of experience. The tenant variables are: race, gender, age, education, whether they have formed a payment plan, whether they have overdue rent, back rent, monthly rent, monthly income, and an employment dummy. We also include interactions between all demographics.

measurement error would generate more inefficiency is because we find more altruism overall, so reducing social preferences on net generates more inefficient evictions from misperceptions. Second, we scale misperceptions for various $s \in \mathbb{R}^+$ (Panel B) and find our conclusions are also reasonably insensitive. One reason that these results are not highly sensitive to measurement error is that we find a large scale factor for eviction costs, which generates many efficient evictions among those who obtain a high cost draw.[52]

Our results are most sensitive to our assumptions about assortative matching (Figure A.25C). Our primary specification simulates assortative matching at the midpoint of suggestive estimates. That is, we assume that 27.5% of tenants and landlords are assortatively matched on altruism or hostility, and the remainder are matched at random (Appendix A.5.2). If all pairs are matched at random, then the share of evictions that are inefficient falls to 15%, holding other parameters constant. If all pairs are matched on social preferences, then the share inefficient rises to 41%. More matching on social preferences generates more inefficiency because compound altruism $\overline{a}_i$ becomes large as a_{ji} goes to 1 for either party j, regardless of the other party $-j$'s hostility. Sufficient altruism thus dampens the impact of even intense hostility.

1.8 Conclusion

Formal evictions represent the culmination of an unsuccessful bargaining process. If misperceptions or hostility cause bargaining failure, some evictions could be inefficient or repugnant. Belief elicitations and Dictator Games with landlords and tenants facing eviction in Memphis, Tennessee suggest substantial misperceptions and strong social preferences. These nonclassical forces predict eviction, render material shares of evictions inefficient and repugnant, and sustain bargaining that would otherwise not occur.

Even so, just one in four evictions is efficient, and stopping the average eviction destroys private surplus. Thus, concerns that evictions are privately inefficient cannot rationalize eviction policy, unless the policy targets inefficient evictions in particular. Other arguments, such as externalities or welfare weights on the evicted, can rationalize eviction intervention. Credible empirical research suggests that evictions indeed cause fiscal externalities via, for instance, stays in homeless shelters and hospitalizations (Collinson

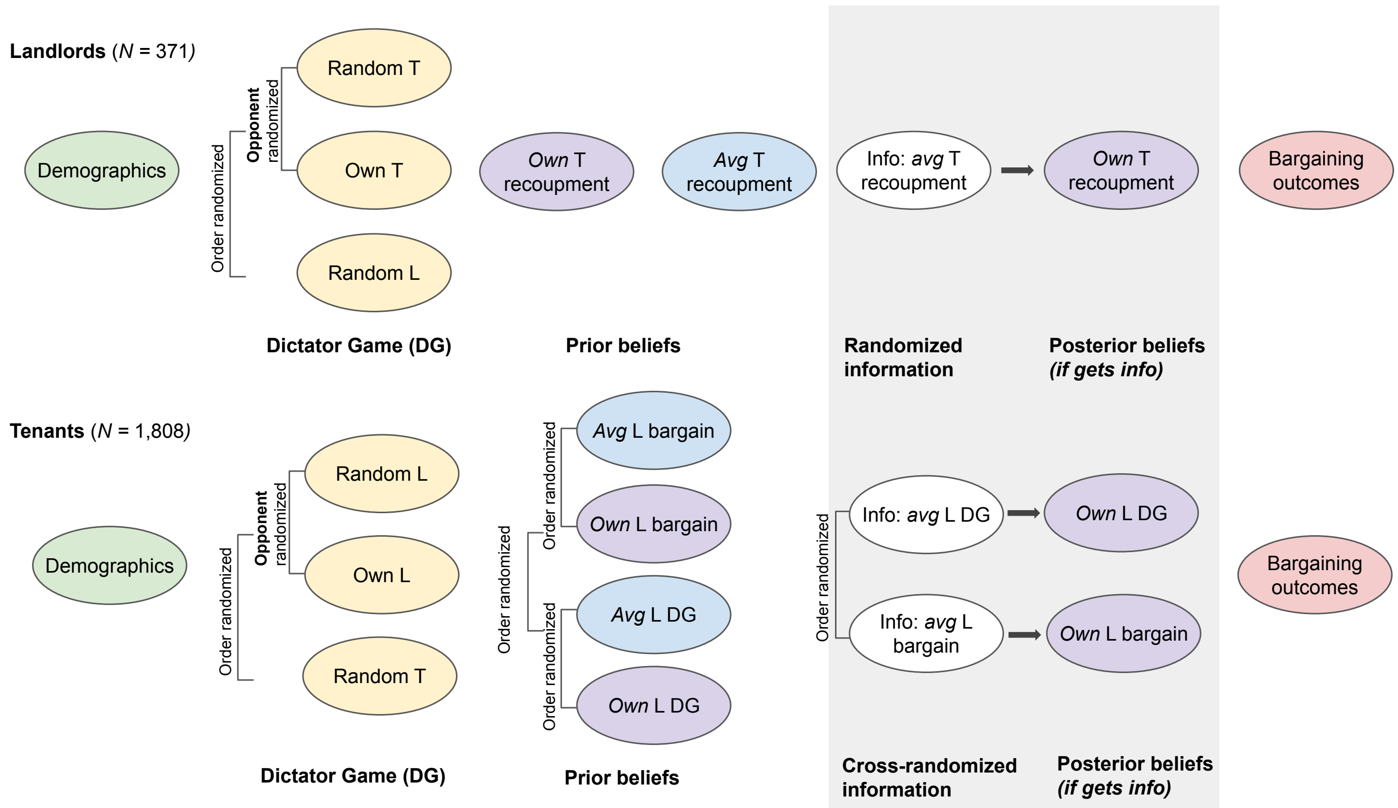
[52]A related concern is that we assume all judgments are for money, when half of evictions are for possession. We simulate Δp_i from the landlord data as being $0.5(p_{Li} - 0.06)$, holding all parameters fixed. This exercise moves inefficiency to about 21%. Intuitively, with any misperception wedge, the wedge is greatly amplified by hostility.

et al., 2024). We find enough inefficiency to meaningfully reduce the magnitude of externalities that make eviction intervention socially desirable. Future work that builds on Collinson et al. (2024)'s estimates would be valuable.

The presence of misperceptions and hostility has important implications for eviction policy and beyond. An illuminating literature suggests that misperceptions and mistakes are key to the economics of poverty and welfare programs (Mullainathan and Shafir, 2013; Bhargava and Manoli, 2015; Finkelstein and Notowidigdo, 2019). Intense social preferences have received less attention. Our evaluation of Memphis's Emergency Rental and Utilities Assistance Program suggests these forces reduce the policy's cost-effectiveness and affect its targeting. Yet social preferences are likely important for housing policies other than ERAP, as Section 8 vouchers and other low-income housing policies also rely on landlord–tenant cooperation. Indeed, misperceptions and hostility may affect efficient bargaining in settings outside housing — including, for instance, marriage/divorce or labor strikes. When these forces are present, they have the potential to influence the desirability and effectiveness of policy intervention.

1.9 Figures

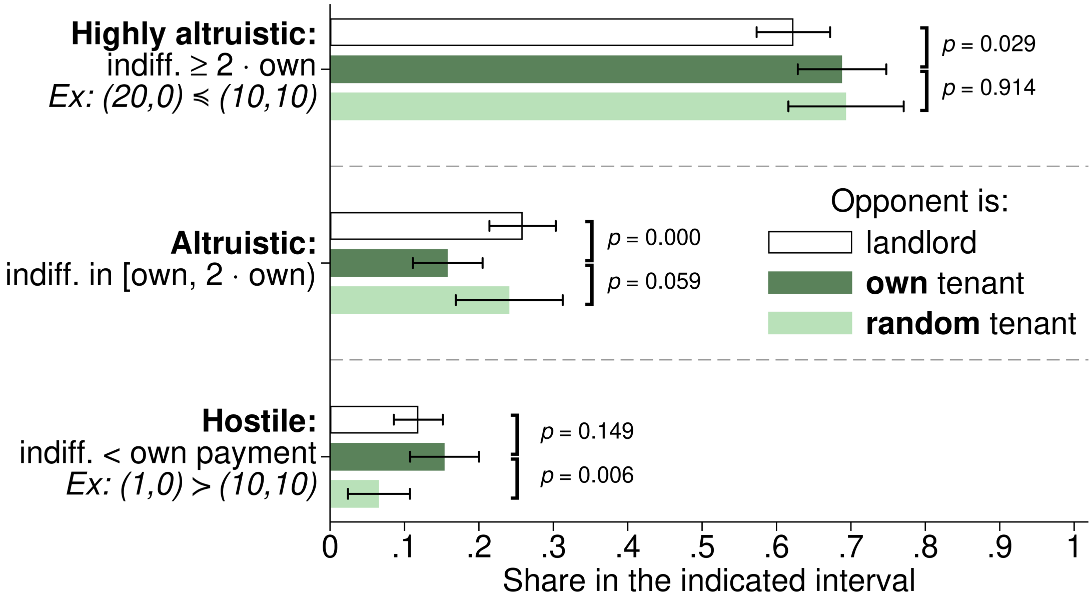
Figure 1.1: Survey Flow



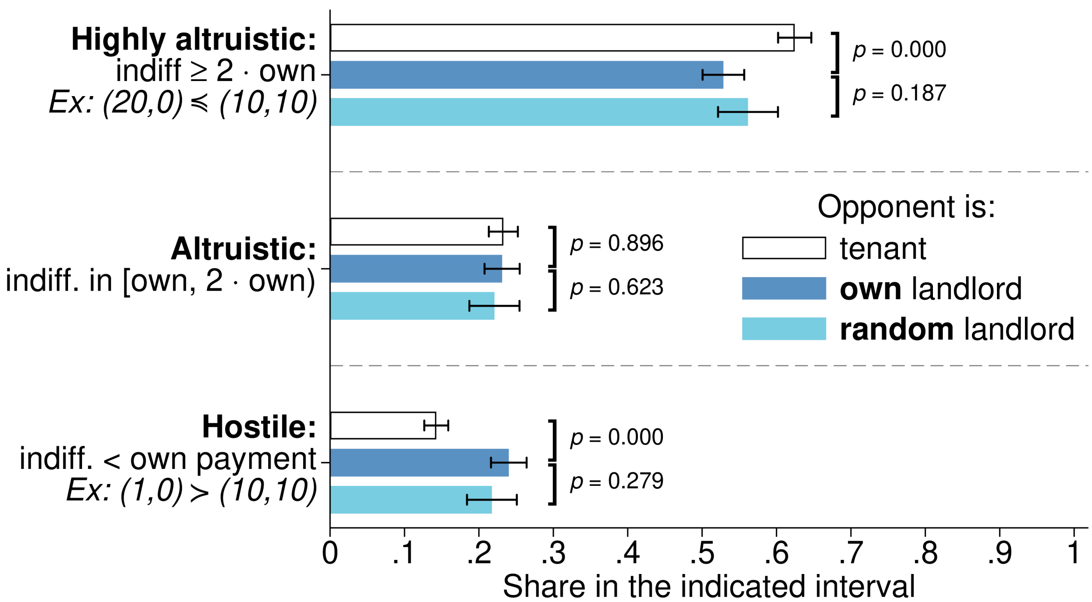
Note: This figure depicts the survey flow for the landlord and tenant experiments. The yellow ovals show the opponent for the dictator game; each participant played the game twice. We elicited all depicted prior beliefs for each participant. We randomize the order of prior beliefs for tenants only. We provided information only to randomly selected participants, with $p = 0.5$ for each treatment. We elicited posteriors only for participants to whom we provided information. We provided information about the average landlord or tenant, but we elicited posterior information only about the participant's own landlord or tenant. Several secondary elicitations are omitted for legibility.

Figure 1.2: Behavior in Modified Dictator Game

(a) Landlord participants ($N = 371$)



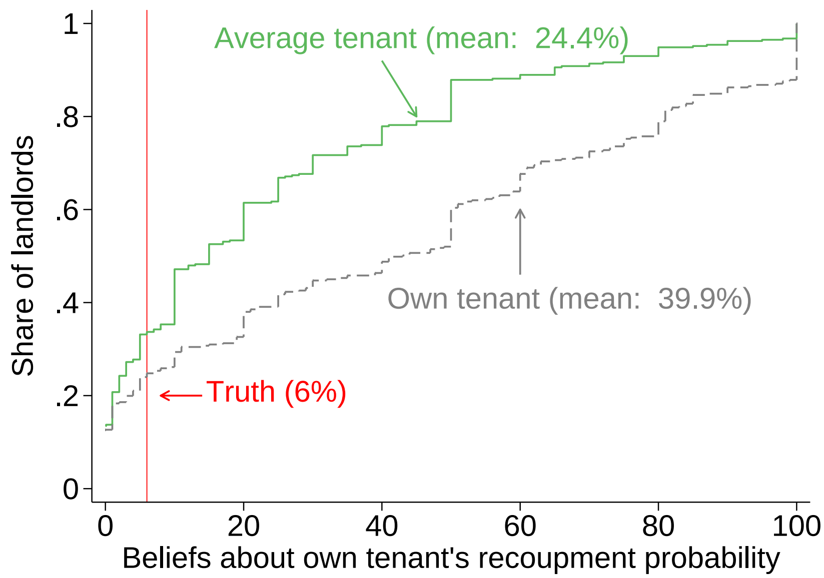
(b) Tenant participants ($N = 1{,}808$)



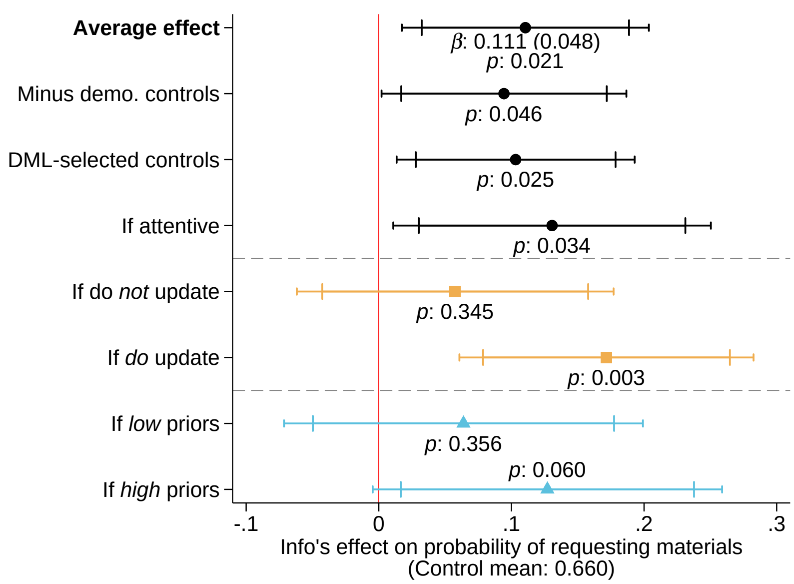
Note: This figure summarizes the results from the Dictator Games among landlord participants (Panel A) and tenant participants (Panel B). Landlords were randomized to play the game against their own tenants (named) or against random tenants (unnamed). All landlords additionally played the game against random landlords (unnamed). Tenants were randomized to play the game against their own landlord or a random landlord. All tenants played the game against random tenants. We elicit bounds on the the point $S(x)$ at which a participant is indifferent between the bundle (\$$s$ self, \$0 other) and (\$$x$ self, \$$x$ other). If $S(x)/x < 1$, then the player is hostile; if $S(x)/x > 1$, then the player is altruistic. We show the share of people who are "highly altruistic" ($S(x)/x > 2$), "altruistic" ($S(x)/x \in (1, 2)$), and "hostile" ($S(x)/x < 1$). We elicit bounds on $S(x)$ using a multiple price list. Our elicitation gives bounds on the indifference point, explaining why no one is exactly classical.

Figure 1.3: Landlord Misperceptions

(a) Priors about recoupment



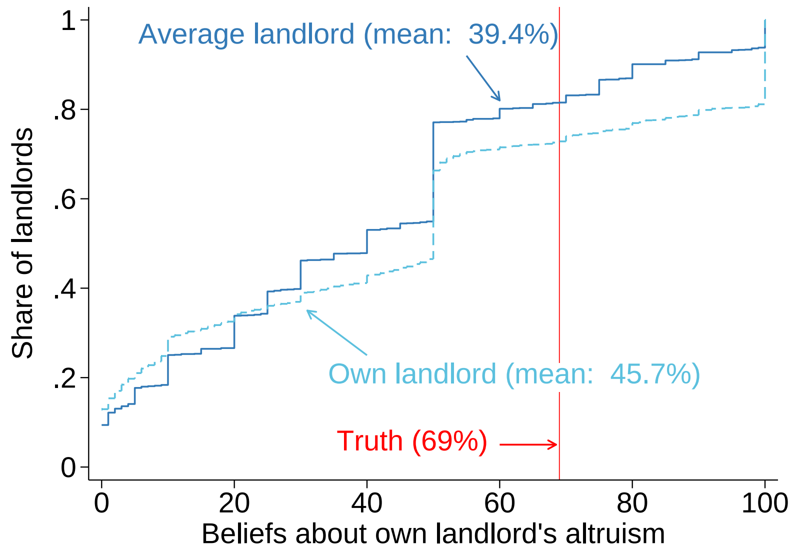
(b) ITT: requests materials



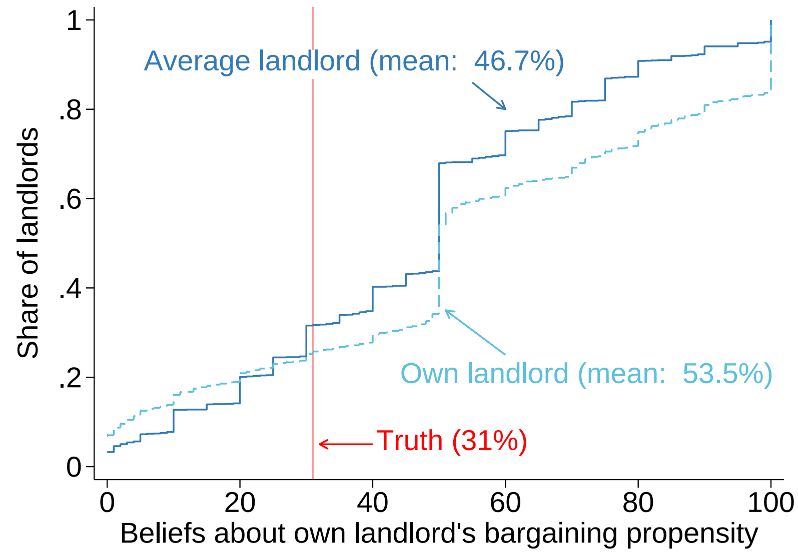
Note: This figure shows data from the landlord sample ($N = 371$). Panel A shows the cumulative distribution function of landlord priors about their own tenant and the average tenant. Landlords recouped 6 percent of back rents during the time period that we asked about (red line). Panel B shows the ITT of providing information on whether the landlord requests materials about ERAP.

Figure 1.4: Tenant Misperceptions (Altruism)

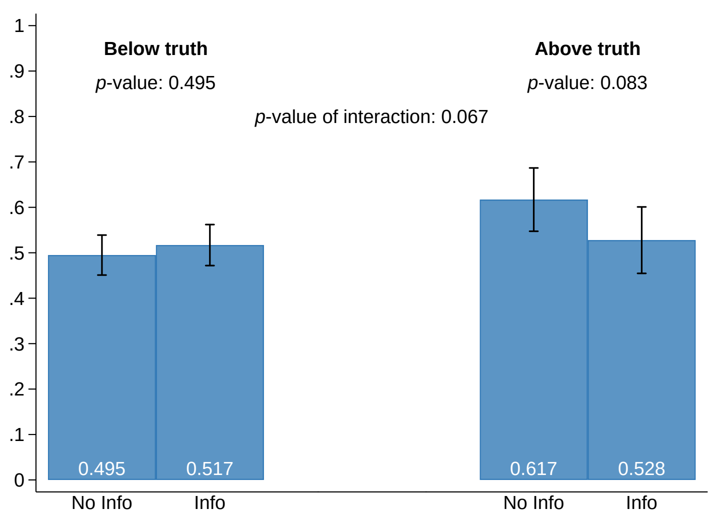
(a) Priors about landlord altruism



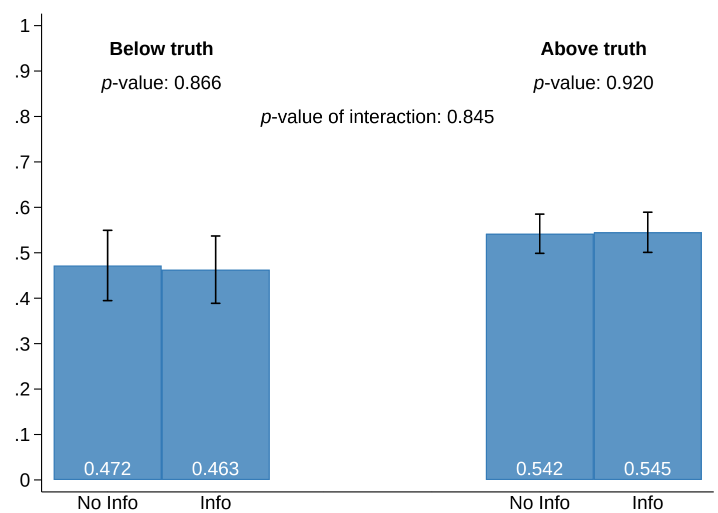
(b) Priors about landlord bargaining



(c) ITT: effect on requesting a payment plan (altruism)



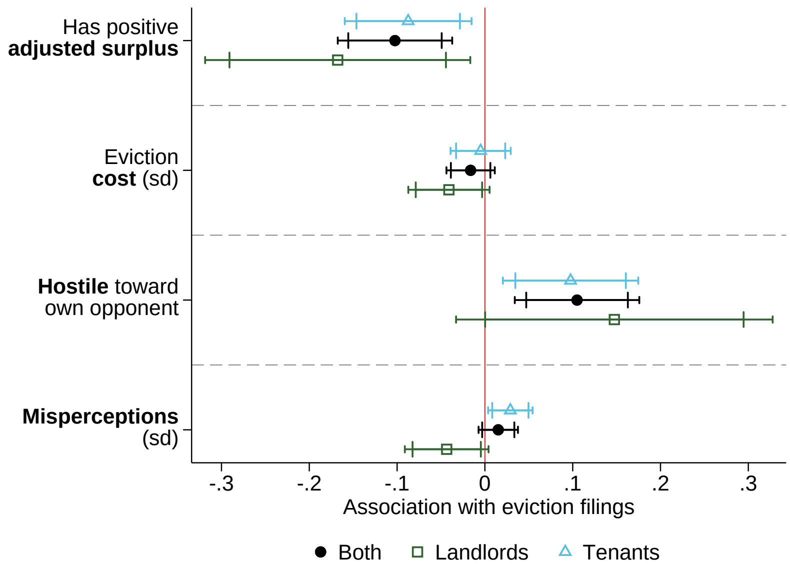
(d) ITT: effect on requesting a payment plan (bargaining)



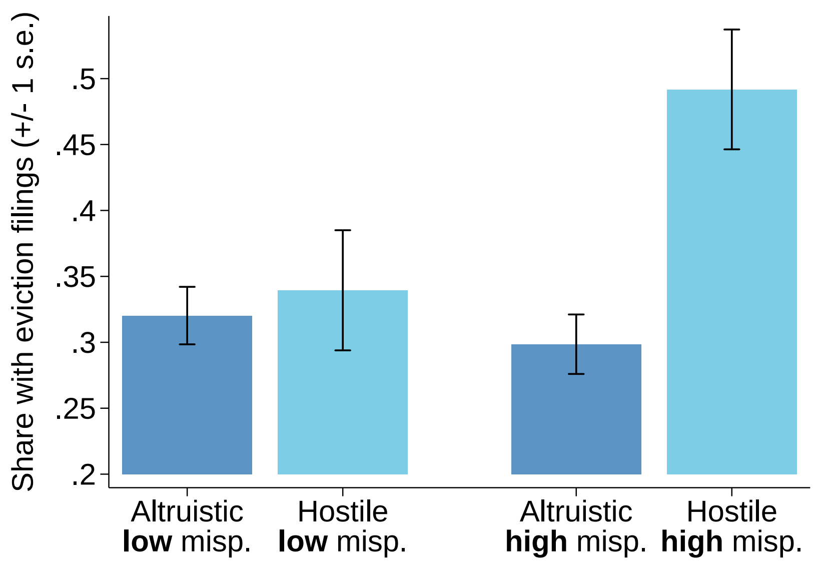
Note: Panels A and B show cumulative distribution functions of tenant priors for beliefs about landlord altruism (i.e., the share of landlords who prefer (x self, x own tenant) to ($2x$ self, 0 own tenant)) and bargaining in court conditional on filing an eviction. The truth is indicated in a vertical red line. We cross-randomized information treatments. Panels C and D shows intent-to-treat effects of providing information on whether the participant requested a payment plan, splitting the sample by whether they are *above* or *below* the information provided. The sample is the tenants who were eligible to request a payment plan.

Figure 1.5: Model Validations

(a) Adjusted Surplus Predicts Eviction Filings

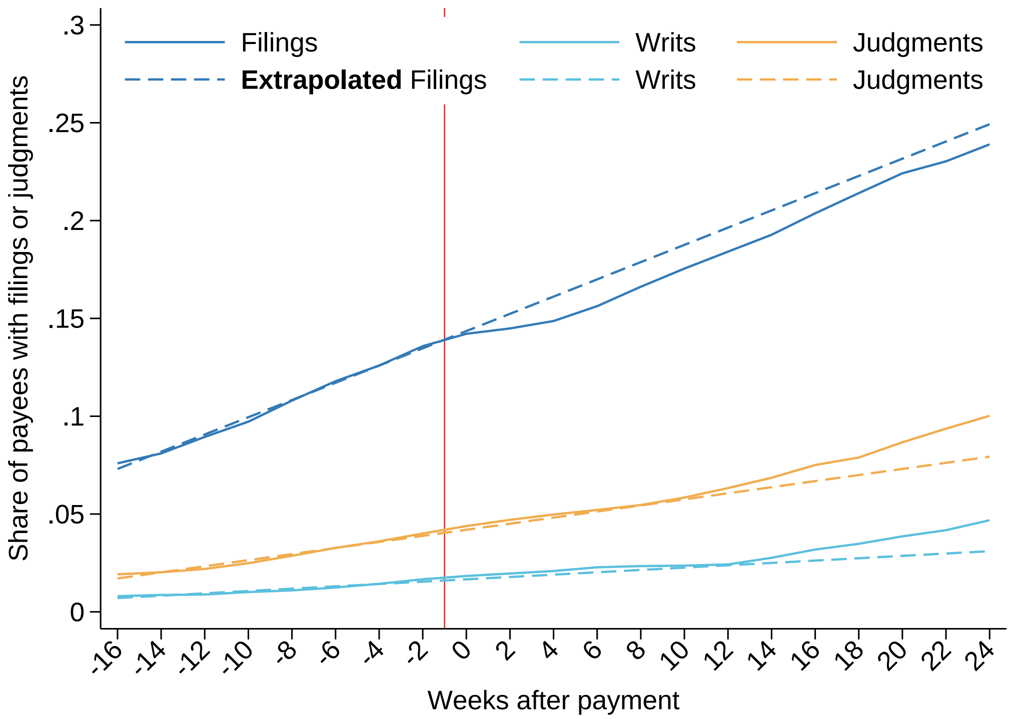


(b) Social Preferences Predict Evictions Only With Misperceptions



Note: Panel A presents estimates of $\hat{\beta}$ from Equation (1.10), where the covariate X_i is either positive adjusted surplus, a proxy of eviction cost, hostility, or misperceptions. Appendix A.5 gives details on how these are formed. Panel B splits the data by having above- or below-median misperceptions and hostility. It presents mean eviction filings. The outcome in both panels is whether the participant (either tenant of landlord in landlord survey, or tenant in tenant survey) gets an eviction filing at the address in the experiment between January 2019 and June 2023. In the surplus and hostility regressions, we keep only the landlords or tenants who played the DG against their own tenant. The tenant sample contains only tenants with positive back rents at the time of the survey. The specifications that involve landlord costs control for experimental variation in how the landlord costs were elicited. All pooled specifications include an experiment fixed effect.

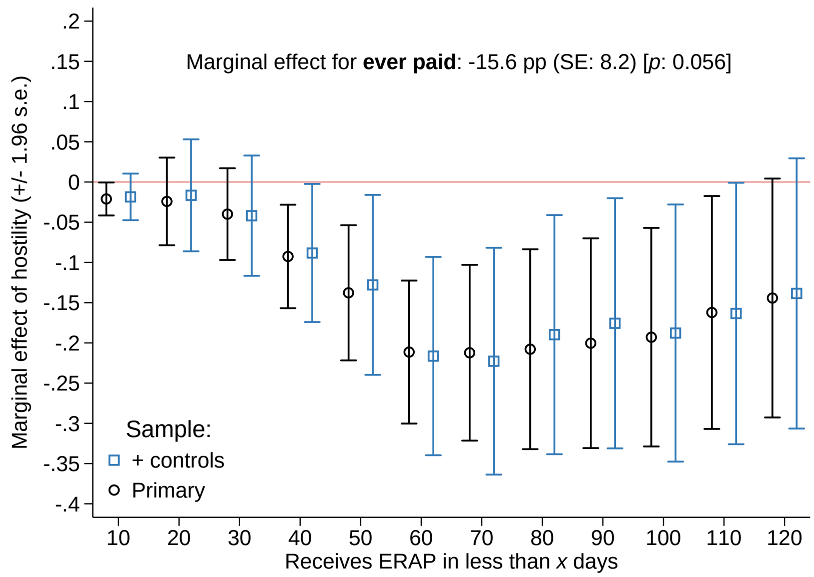
Figure 1.6: Rental Assistance Filings and Judgments by ERAP Payment



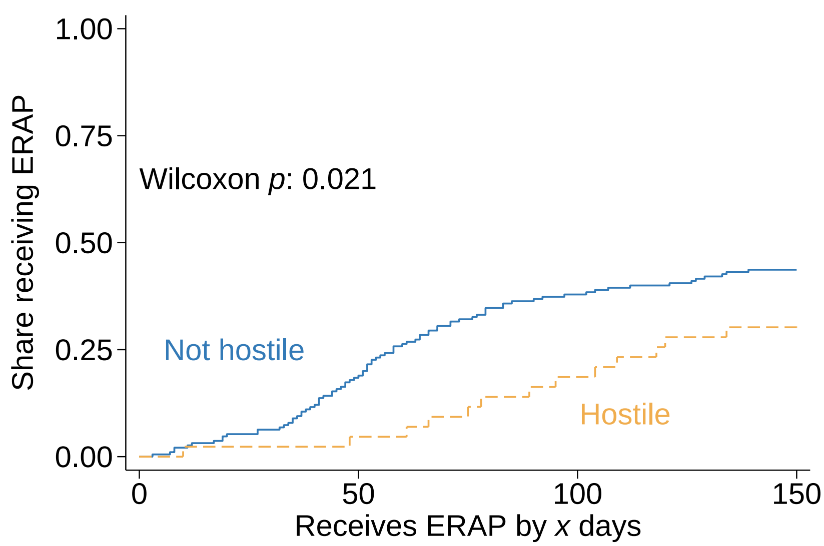
Note: This figure shows filing and judgment rates for tenants who receive payments from the Memphis-Shelby County Emergency Rental and Utilities Assistance Program. The red line indicates the date of payment for that tenant. The dashed lines show linear extrapolations from 16 to 2 weeks before payment.

Figure 1.7: Tenant Hostility and Speed of ERAP Receipt

(a) Receives ERAP Quickly



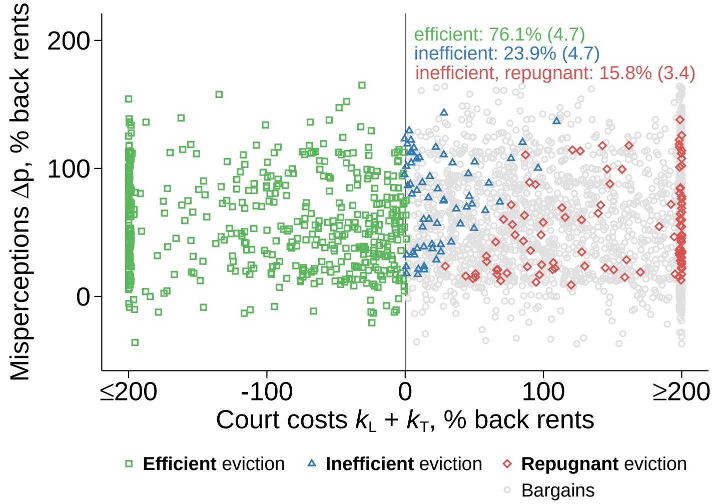
(b) Kaplan-Meier Failure Curves



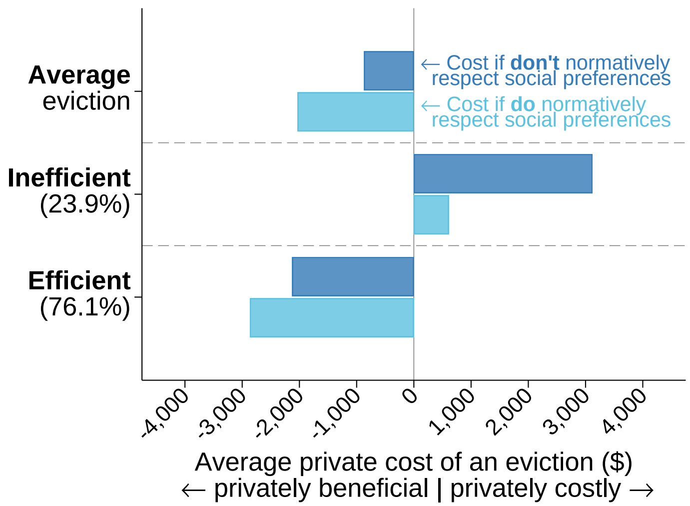
Note: This figure presents tests of whether hostile tenants (defined as in Section 1.4) receive ERAP funds within the number of days listed on the horizontal axis. Panel A shows regressions of receiving ERAP withn the indicated number of days on tenant hostility. Controls for beliefs and demographics are: controls for prior beliefs about own/average landlord bargaining behavior or behavior in the DG; controls for demographic variables: race, gender, age, and education; controls for economic variables: log back rent owed, log monthly rent, log monthly income, and employment status. Panel B presents Kaplan-Meier curves and a Wilcoxon test for differences. This figure shows the sample of tenants who: play the DG against their own landlord; had not moved between applying for ERAP and participating in our study; did not enter the separate ERAP eviction representation process; and applied after September 1, 2021, since that is when data on ERAP reciept is available.

Figure 1.8: Decomposing Eviction: Efficient and Inefficient Evictions

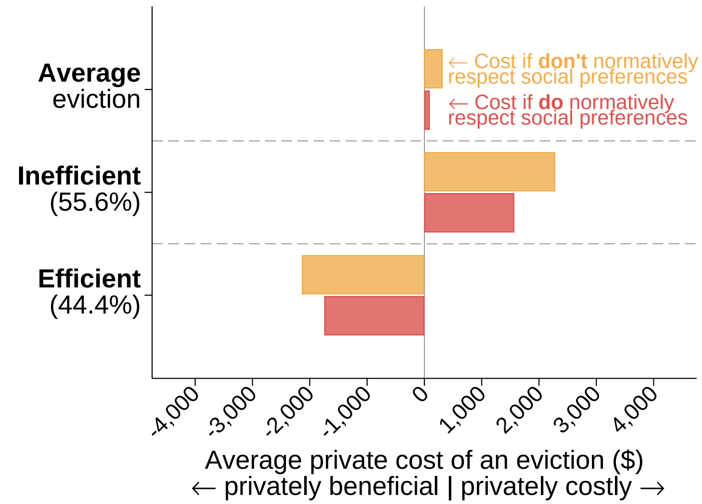
(a) Empirical Analog of Diagram 1.1



(b) Magnitudes of Eviction Costs



(c) Magnitudes of Eviction Costs Absent Altruism



Note: Estimates from 200 bootstraps. Panel A presents the simulated estimates of the share of evictions in each region of Diagram 1.1. We shut off ERAP payoffs and simulate using the procedures in Section 1.7. Parentheses in Panel A display standard errors. Panel A plots a 2% sample but percentages are from the full sample. Panel B shows the average cost of an eviction. The bars that do not normatively respect social preferences report the average $k_{Li} + k_T$ among those who evict, where $k_{Li} := k_L - (\varepsilon_{i1} - \varepsilon_{i2})$. The bars that do normatively respect social preferences report the average $(1 + a_{Ti})k_{Li} + (1 + a_{Li})k_{Ti}$. Panel C replaces altruists as having classical social preferences and resimulates their behavior and costs, holding all other parameters and primitives constant.

1.10 Tables

Table 1.1: Tenant Demographics

Sample:	Experiment Participant (1)	Experiment Non-Participant (2)	Difference (1 – 2) (3)	SE Difference (4)	Memphis ERAP (Paid) (5)	Memphis ERAP (Paid & Nonlegal) (6)	Shelby County (7)
A. Observed for Experiment Participants and Non-Participants							
Age	34	35	-1.1	0.27	37	39	45
Female	0.81	0.72	0.09	0.010	0.77	0.76	0.53
Black	0.88	0.89	-0.01	0.008	0.92	0.91	0.53
Disabled	0.14	0.14	-0.00	0.009	0.10	0.11	0.14
Household size	2.1	1.9	0.2	0.04	2.3	2.3	2.2
Employed	0.53	0.51	0.01	0.012	0.43	0.40	0.64
Household monthly income	891	733	157	29	1,185	1,172	5,408†
Monthly rent	872	900	-28	12	949	978	834
Back rent owed at application†	3,585	3,900	-315	173	4,000	4,155	
B. Observed for Experiment Participants Only							
Some college +	0.56				0.55		
Never married	0.84						
Ever evicted	0.33						
Ever overdue rents	0.86						
Ever formed a payment plan	0.57						
Back rent owed at survey†	1,014						
Paid by ERAP at survey	0.55						
N	1,808	15,062			10,111	4,742	5,586

Note: This table shows demographic means for tenants in our survey experiments (Columns 1 and 2), ERAP samples (Columns 5 and 6), and in the 2019 ACS for Shelby County (Column 7). The observation count includes people who appear in any row but each row excludes missing values. Columns 3 and 4 show the difference and standard error on the difference between experiment participants and non-participants. Panel B shows additional demographic data collected in the tenant experiment. Monthly income is conditional on having non-zero income. †: displays median.

Table 1.2: Landlord Demographics

Variable	(1) Participant	(2) Participant's Tenant	(3) Non-Participant's Tenant	(4) Difference (2 – 3)	(5) SE Difference
A. Observed for Landlords and Tenants					
Age	49	39	37	2.1	0.68
Female	.62	0.66	0.69	-0.03	0.026
Black	.58	0.77	0.84	-0.07	0.023
White	.32	0.09	0.06	0.03	0.015
B. Observed for Tenants Only					
Veteran		0.04	0.02	0.01	0.010
Back rent owed†		2,700	2,400	300	210
Utilities owed		387	535	-148	81
C. Observed for Landlords Only					
Years of experience	13				
Some college +	.85				
Landlord	.62				
Property manager	.29				
Small landlord (1–2 units)	.26				
Medium landlord (3–10 units)	.24				
Large landlord (10+ units)	.48				
Ever evicted	.71				
N	371	371	3,595		

Note: This table presents demographic characteristics of the landlord sample. Cells present means. Column 1 shows demographic characteristics collected in the survey experiment. Columns 2 and 3 use administrative data from the Emergency Rental and Utilities Assistance Program to obtain demographic information about the tenants of the landlord sample participants, as well as the landlords who were invited to participate but did not. In the landlord survey, landlords are asked about a randomly selected reference tenant (the person against whom they play the Dictator Games below). We compare the reference tenant of the landlords who were invited to the study but did not participate to the tenants of landlords who did participate. Columns 4 and 5 show difference in means between Columns 2 and 3 and standard errors on the difference. Some demographic information was collected only in the survey and is presented in Panel C. †: displays median. Back rents are reported owed at application.

Table 1.3: Empirical Bargaining Model: Parameter Estimates and Model Fit

A. Parameter Estimates

Explanation	**Parameter**	**Estimate**
Total unconditional eviction cost ($)	—	3,401 (1,087) {3,507}
Landlord cost ($)	k_L	-641 (1,806)
Tenant cost ($)	k_T	4,042 (1,654)
Tenant bargaining power	β	0.84 (0.12)

B. Model Fit

Selected Moments	**Estimated Value**	**Targeted Value**
1. ERAP mean judgment rate (unconditional)	0.158	0.091
2. Treatment effect on judgments (unconditional)	-0.016	-0.009
3. Landlord take-up rate	0.668	0.648

Note: Estimates from 200 bootstraps. Panel A shows the estimated parameters from the empirical model presented in Section 1.7. The number in braces shows the standard deviation of unconditional eviction costs, which is the standard deviation of estimated $-(\varepsilon_{i1} - \varepsilon_{i2})$. Panel B shows the model fit to several targeted moments. For the full set of moments, see Table A.26.

Table 1.4: Empirical Bargaining Model: Mechanisms and Counterfactuals

	Descriptives		Targeting			ERAP Welfare Impact	
	Judgment % (absent ERAP)	Takeup (%)	TOT: judgments (pp)	Inefficient judgment (%) (if prevented)	Counterfactual judgment (%) (enrollees)	Raw ($)	Altruism-adjusted ($)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Mechanisms							
1. Primary	19.1	66.8	-1.6	22.7	17.5	211	83
2. No social prefs., misperceptions	14.6	68.3	5.7	0.0	11.9	-124	-124
3. No social preferences	28.1	61.0	-5.4	55.4	23.7	206	206
4. No landlord altruism	23.0	56.9	-15.7	31.3	26.7	802	834
5. No altruism	32.6	60.1	-9.1	62.5	27.7	372	245
6. No hostility	16.1	67.7	0.8	10.4	15.0	96	110
7. No misperceptions	14.6	68.4	2.2	0.0	13.6	54	31
8. High misperceptions	20.3	65.0	-2.0	24.4	17.8	236	121
9. Mean T and L court costs are 0	53.7	66.6	8.8	8.4	48.1	326	398
10. Double T and L court costs	6.7	66.8	-3.5	50.8	6.3	330	290
B. Counterfactual Policies							
11. ↑ eviction penalty	19.1	63.2	-6.3	22.7	17.2	356	327
12. ↓ take-up costs	19.1	72.4	-2.2	22.8	17.9	135	-31
13. Targeted on demographics	26.2	85.7	-10.6	41.2	25.2	119	-307

Note: Estimates from 200 bootstraps. We simulate behavior using data from the experiment and ERAP, using the model in Section 1.7. Row 1 is our primary specification, using parameters estimated via the Method of Simulated Moments procedure displayed in Table 1.3. Row 2–10 change social preferences, misperceptions, or costs. Rows 11 and 12 raise ERAP's eviction judgment cost (adding a cost of $2,000 to Equation 1.14) or eliminate take-up costs ($k_L^P \to 0$). Row 13 conducts a targeting counterfactual by restricting ERAP eligibility to households whom an OLS procedure determines are in the top 10% most likely to bargain without ERAP. Column 1 shows the share of households that would pursue eviction judgments if ERAP did not exist. Column 2 shows the share of landlord–tenant pairs who enroll. Columns 3 and 4 show the treatment effect on judgments, and the share of judgments averted that are inefficient. Even if the estimate in Column 3 is positive, the estimate in Column 4 is non-missing because we consider any individual whose eviction is stopped. Column 5 shows the total percentage rate of judgments. Thus, Column 3 divided by Column 5 ($\times$100) shows a treatment effect in percentage terms. Columns 6 and 7 show the per-household welfare impact of ERAP among enrollees, depending on if the planner does not versus does normatively respect altruism. Column 7 does not scale governmental or intrahousehold transfers by altruism.

Chapter 2

The Welfare Effects of Eligibility Expansions: Theory and Evidence from SNAP

This chapter is coauthored with Jenna Anders.[1]

2.1 Introduction

Social programs in the United States are characterized by incomplete take-up, and there is substantial heterogeneity in take-up across programs. Meanwhile, there is also heterogeneity in eligibility criteria across programs. In fact, in some social programs, such as the Supplemental Nutrition Assistance Program (SNAP, also known as food stamps), the eligibility threshold even varies across states. There is a suggestive positive correlation between U.S. welfare programs' eligibility thresholds and take-up: programs with less

[1]We thank Hunt Allcott, Abhijit Banerjee, Judi Bartfeld, Leo Bursztyn, Clément de Chaisemartin, Raj Chetty, Jon Cohen, John Conlon, Valerie Chuang, Zoë Cullen, Esther Duflo, Amy Finkelstein, Peter Ganong, Benny Goldman, Jon Gruber, Craig Gundersen, Basil Halperin, Emma Harrington, Nathan Hendren, Lisa Ho, Jeffrey Liebman, Stephen Morris, Whitney Newey, Ben Olken, Emily Oster, Amanda Pallais, Dev Patel, Jim Poterba, Indira Puri, Frank Schilbach, Amy Ellen Schwartz, Jesse Shapiro, Johannes Spinnewijn, Evan Soltas, Dmitry Taubinsky, and participants at workshops at Harvard and MIT for helpful discussions. We thank Rian Flynn and Amber Zheng for excellent research assistance. We thank Timothy Harris for sharing data on SNAP work requirement waivers. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374 and Grant No. 1745303; by Harvard's Foundations of Human Behavior Initiative; and by the Harvard GSAS Professional Development Fund for PhD Students. The online experiment was pre-registered at the AEA RCT Registry under AEARCTR-0005566, and survey instruments are available at Rafkin's website. The experiment received exempt status from MIT's Committee on the Use of Humans as Experimental Subjects (#E1962) and Harvard's Institutional Review Board (#IRB20-0326).

stringent income eligibility thresholds have higher take-up rates (Figure 2.1).

Regardless of the across-program correlation, the within-program relationship between take-up rates and eligibility is consequential. In the simplest model of how to set eligibility thresholds, policymakers trade off giving larger benefits to only the poorest people or spreading the benefit more thinly to a larger number of people. But if eligibility thresholds affect take-up within the eligible population, the policymaker no longer faces this basic trade-off alone. Targeting benefits only to the poorest households could decrease take-up for these groups. As a result, it is important to determine whether there is a causal relationship between the eligibility threshold and take-up of the already-eligible population.

Does the eligibility threshold affect take-up of social programs? If so, how does this phenomenon affect programs' optimal eligibility? In this paper, we provide novel evidence that the eligibility threshold affects take-up among low-income individuals who are always eligible for SNAP, regardless of the threshold. We explore the mechanisms underlying this take-up response using an online experiment and analysis of a government-commissioned survey on incomplete SNAP take-up. To interpret our findings, we propose a general model of welfare program participation that allows us to study optimal policy when the eligibility threshold endogenously affects take-up. The model makes precise how mechanisms — namely, stigma and incomplete information — affect welfare considerations, and we estimate the model empirically.

We focus on SNAP for several reasons. First, it is a large program (with an annual budget of about $70 billion) that forms an important part of the U.S. public assistance system. Second, SNAP eligibility rules are at the center of an ongoing public discussion.[2] Third, SNAP publishes anonymized public-use administrative data (the Department of Agriculture's Quality Control files), which we use to form our main outcome of log enrollment counts. The administrative data alleviate concerns that the results could reflect the mismeasurement of individuals' eligibility status or program participation reporting biases (Kreider et al., 2012; Meyer et al., 2015).

We begin by providing evidence that eligibility expansions in SNAP raised enrollment among the lowest-income individuals who are always SNAP-eligible. States can choose to expand SNAP eligibility standards beyond the federal minimum of 130% of the Federal Poverty Level (FPL). We focus on individuals at 50–115% of the FPL, a group eligible for SNAP in every state because they are poorer than the federal minimum eligibility.

[2]The Trump administration proposed eliminating state discretion in eligibility thresholds (Federal Register, 2019).

Leveraging an event-study design (using variation across states and years), we find that raising the eligibility threshold by 10 percentage points (pp) of the FPL (e.g., from 130% to 140%) boosts enrollment by over 1 percent among the inframarginal group that was always eligible for SNAP. Our setting also yields a clean placebo test: the policy change that permits states to change their SNAP eligibility threshold also gave other bureaucratic benefits to states, and we show that states which adopted the policy *without* expanding SNAP eligibility saw no increases in SNAP enrollment. As another way of benchmarking the magnitude, we find that for every person who joins SNAP because she becomes newly eligible, 0.9 inframarginal people join the program. We conduct a model-free cost-effectiveness exercise and find that the mechanical cost of raising the means test enough to increase inframarginal enrollment by 1 pp is $2.2 billion per year — about the same as the mechanical cost of increasing the SNAP benefit enough to achieve the same goal.

This take-up response among the inframarginal population is consistent with a small literature documenting a similar phenomenon for public health insurance programs, where it is called a "welcome-mat effect" or a "woodwork effect" (because already-eligible individuals appear "out of the woodwork" to take up the health program). We use the term "inframarginal effects" to avoid negative or positive connotations.

To further connect our findings to the literature on incomplete social program take-up (Currie, 2004; Bhargava and Manoli, 2015; Finkelstein and Notowidigdo, 2019), we next turn to uncovering the mechanisms underlying inframarginal effects. One hypothesis, motivated by models of social signaling (e.g., Bursztyn and Jensen, 2017), is that raising the income threshold could reduce SNAP stigma: with less stringent eligibility rules, taking up SNAP no longer conveys as much information about one's type. To test this hypothesis, we conduct an online experiment with a nationally representative sample of more than 2,000 participants. We provide truthful information about the eligibility threshold in *one* state to shock beliefs about the mean eligibility threshold *across* states. The experimental variation increases participants' beliefs about the share of individuals who are eligible for SNAP in the entire U.S. by 9 percentage points on average (standard error: 0.8 pp), and decreases an index of stigma by -0.050 standard deviations (SE: 0.027, $p = 0.061$). Effects on stigma are larger among people who are SNAP-eligible.

A second hypothesis is that relaxing the eligibility restrictions increases information about SNAP. To test this hypothesis, we analyze microdata from the Food Stamp Program Access Study (FSPAS), a nationally representative survey on SNAP awareness and stigma among both SNAP enrollees and non-enrollees conducted by the USDA (Bartlett et al., 2004). To our knowledge, this is the first academic analysis of this rich dataset on SNAP take-up mechanisms. Using the FSPAS, we identify demographic groups that are likely subject to SNAP awareness and stigma. We find that the demographic groups with the largest inframarginal effects are those with low levels of SNAP awareness and *not* the groups whose stigma is most sensitive to eligibility thresholds from the experiment. Thus, combining the FSPAS and the online experiment, we find that relaxing the eligibility threshold does reduce SNAP stigma, but information appears to play a larger role in the decisions of people who newly take-up.

To assess the quantitative importance of each mechanism, and determine the implications of inframarginal effects for social welfare, we propose a general economic framework for analyzing optimal eligibility in the presence of inframarginal effects. Individuals who are eligible for a welfare program take up the program benefit as long the benefit exceeds a private take-up cost (e.g., stigma) and they are aware of the program.[3] Both the cost and information (awareness) can depend on the eligibility threshold. The model emphasizes that the planner trades off (i) a standard redistributive motive in which she values giving a bigger benefit to people with higher welfare weights against (ii) a new motive, inframarginal effects, in which relaxing eligibility thresholds raises take-up.[4] We derive an optimality condition for the eligibility threshold in which our key empirical fact, the inframarginal effect, enters as an observable elasticity (Chetty, 2009; Kleven, 2021).

The model also explains the role of the two candidate mechanisms, stigma and information frictions. Similar to in a Baily (1978)-Chetty (2006) framework, the optimality condition features a fiscal externality of the inframarginal effects and recipients' willingness to pay (WTP) for a higher eligibility threshold. Recipients' WTP depends on why inframarginal effects exist. First, suppose that inframarginal effects are mostly driven by behavioral responses to changing costs (e.g., through stigma). Then, the new enrollees driving the inframarginal effects are just indifferent between taking up and not, so they do not value the eligibility expansion — a standard envelope condition. However, those who would have enrolled regardless now pay lower stigma costs to take up. The optimal eligibility threshold trades off the reduced stigma among inframarginals with the fiscal externality of new take-up. In this way, the model cleanly isolates two countervailing

[3] While we focus on stigma costs, our framework permits any cost that depends on the eligibility threshold. Another possible cost embedded within our framework is uncertainty about eligibility, as in e.g. Kleven and Kopczuk (2011).

[4] In our benchmark model, we hold labor supply constant, but we show in the Appendix that similar intuitions apply in a more elaborate environment.

forces that govern the welfare effects of reducing stigma. On the other hand, suppose instead that inframarginal effects reflect improved awareness of the program. Then, the new take-up now confers first-order utility gains, since people who lacked awareness were not previously optimizing.

Our framework lets us conduct normative analysis about whether the planner should raise the eligibility threshold. We study when the "naïve" planner who ignores inframarginal effects but otherwise behaves optimally will set the eligibility threshold too low — or equivalently, the benefit size too high — relative to a "sophisticated" planner who is aware of inframarginal effects. We characterize a simple sufficient condition: the naïve planner will always set the threshold too low if information agents' take-up is weakly more elastic to a change in the threshold than stigma agents' take-up. Thus, the model yields a direct empirical test for whether the existence of inframarginal effects implies that the eligibility threshold should rise.

We proceed to implement this test using a model-based decomposition of the mechanisms. On the one hand, the experiment suggests that eligibility changes reduce stigma. On the other hand, the FSPAS analysis suggests that stigma agents are not those who newly take-up in response to eligibility changes. We propose a decomposition that lets us empirically estimate the contributions of stigma and information in a regression framework. We conclude that that the types of people who are marginal to the eligibility increase are those who are misinformed, not those who are subject to stigma. Put another way, we find that the eligibility increase reduces take-up costs among people who always take up. Those who newly take-up do so because they were previously uninformed, so they capture the full utility gain of the program. When we implement our test, we reject that stigma agents are more elastic than information agents. The upshot of this test is that the eligibility threshold is set too low, if current policy naïvely ignores inframarginal effects.

Finally, we combine the model with our empirical estimates to conduct analysis of the optimal eligibility threshold. As noted, our propositions developed in the model deliver that the social planner will set the eligibility threshold too low if she ignores inframarginal effects and information is more important than stigma in driving inframarginal effects. But how large will the planner's mistake be? Traditionally, local analysis in the spirit of Baily (1978)-Chetty (2006) does not inform the analyst about whether the planner's mistake is large or small when the optimality condition does not hold exactly. In our first exercise, we propose a new method of estimating the magnitude of the planner's mistake, using only the local optimality condition. The core idea is to solve for how much the planner must misperceive the population's risk aversion in order that the optimality condition holds in the context of the naïve model. We find that the naïve planner would overestimate risk aversion by 30%, which corresponds to overvaluing the marginal utility of inframarginal types who always take up and hence over-transferring to them. In a second exercise, we impose more parametric structure to solve for the globally optimal eligibility threshold implied by our model and empirical estimates. We find that the optimal threshold will be 13% too low if the planner ignores inframarginal effects.

Contributions and related literature. Every social program makes some determination about program eligibility (even if the program is universal). Yet much of the vast literature on program design focuses on other policy instruments besides the eligibility threshold. To quantify this, we collected all 278 papers published in the *American Economic Review* between 2010–2018 and the *Quarterly Journal of Economics* between 2010–2019 that met one of 33 search terms about social welfare programs (see Figure 2.2 and Appendix B.1 for details). Seventy-six of them were primarily about effects or design of social welfare programs, 49 of which involved the study of a specific policy instrument. Yet only 7 (14% of the 49) examined eligibility criteria as a policy instrument that the planner could manipulate to improve welfare. On the other hand, 25 of the 76 papers about welfare programs consider eligibility thresholds as a source of variation for estimating the program's treatment effect. In sum, while economists regularly exploit eligibility thresholds for causal inference, they are often neglected as an aspect of optimal program design. Our paper is among the first to combine empirical estimates of endogenous take-up from eligibility thresholds with a theoretical model that permits welfare analyses of current program rules.

Our work advances several literatures. First, we add to the large body of research in public economics that deals with the optimal design of social programs. Much of this work considers the optimal *benefit level* when take-up is distorted by moral hazard (Baily, 1978; Gruber, 1997; Chetty, 2006, 2008; Hendren et al., Forthcoming). Kroft (2008) introduced to this literature a new fiscal externality which is closer to ours — social spillovers, which are one potential microfoundation for inframarginal effects — and explored how this phenomenon affects optimal benefit size. Relative to Kroft (2008), we emphasize how the mechanism underlying peer effects drives different welfare effects, and we consider the implications for choosing the optimal eligibility threshold. Altogether, analyses of optimal eligibility are rare in this literature.[5,6]

As an additional contribution to the program-design literature, we propose a new strategy to help researchers assess the magnitude of the social planner's mistake when a given social optimality condition does not hold precisely, as is common when empirically testing Baily (1978)-Chetty (2006) conditions. Our approach uses only the local optimality condition and does not require extrapolation with additional parametric assumptions. This contribution therefore relates to other methods of conducting welfare analysis like the Marginal Value of Public Funds (Hendren and Sprung-Keyser, 2020).

Second, we contribute to the public-finance literature on barriers to social program take-up (Moffitt, 1983; Aizer, 2003; Currie, 2004; Heckman and Smith, 2004; Bhargava and Manoli, 2015; Friedrichsen et al., 2018)[7] and the role of social spillovers in program take-up (Bertrand et al., 2000; Dahl et al., 2014). Similar to Finkelstein and Notowidigdo (2019), we consider the welfare implications of these barriers to take-up, but doing so in the context of eligibility thresholds allows us to consider new trade-offs in the model. Our experiment provides clean evidence that aspects of program design may affect stigma costs. We emphasize that reducing stigma introduces two forces — a fiscal externality and a first-order gain to people who always enroll — and provide methods to analyze them empirically. Our discussion of restricted eligibility departs from the most common prior motivation for restricting eligibility, described in Nichols and Zeckhauser (1982), who suggest that limiting program participation can induce self-targeting.[8]

Third, we link research on optimal program design to the growing literature in behavioral public economics (Bernheim and Taubinsky, 2018a). Our analysis suggests that individuals' utility depends on social norms, and government policy plays an important role in shaping these norms, like in Lindbeck et al. (1999). Economists have only begun

[5]Fetter and Lockwood (2018) is a recent example that studies optimal eligibility for old-age insurance. Other papers, e.g. Diamond and Sheshenski (1995), Low and Pistaferri (2015), and Golosov and Tsyvinski (2005) study optimal eligibility in the context of disability insurance.

[6]In many social programs, the eligibility threshold is defined by the benefit size and the slope of the benefit schedule. Thus studies of benefit levels may also contribute to our understanding of eligibility thresholds. Our empirical setting allows us to isolate the effect of eligibility thresholds alone, since the benefit remains constant. Our framework proposes a setting where the policymaker can set eligibility separately from the benefit.

[7]This literature includes several papers studying stigma surrounding SNAP take-up and the rollout of the Electronic Benefits Transfer (Daponte et al., 1999; Currie and Grogger, 2001; Atasoy, 2009; Klerman and Danielson, 2011; Manchester and Mumford, 2012; Eck, 2018).

[8]Two related papers, Kleven and Kopczuk (2011) and Hanna and Olken (2018), model the means test as an instrument for optimal program targeting in the presence of exclusion (Type I) and inclusion (Type II) errors. Our model differs from these papers by emphasizing how the eligibility threshold might directly affect stigma and information.

to explore how policy may influence psychological forces like shame or guilt, which may in turn may have important consequences for social welfare. For instance, we provide empirical support for the claim, promulgated in sociology and historical discussion of welfare programs, that programs like Social Security are not stigmatized precisely because they are not means tested (e.g., Katz, 1986). Our model-based decomposition of information shows a novel strategy for isolating information from stigma, which has proven to be difficult in many contexts (Chandrasekhar et al., 2019).

Fourth, we contribute to the study of the Supplemental Nutrition Assistance Program, the subject of a wide-ranging literature.[9] We draw on the data used in Ganong and Liebman (2018), who study how changes in the economic environment, coupled with changes in SNAP program design, affected SNAP enrollment through 2012. Relative to prior work, we highlight a previously unappreciated phenomenon in SNAP (inframarginal effects), show how inframarginal effects affect SNAP stigma and information, and consider their implications for optimal eligibility. As an auxiliary contribution, we also present an academic analysis of the USDA's FSPAS data.

Finally, we contribute to the small literature on inframarginal effects. These effects have received little attention in public economics. The health literature on Medicaid expansions finds evidence of inframarginal effects (Aizer and Grogger, 2003; Sommers and Epstein, 2011; Frean et al., 2017; Sacarny et al., 2022), but it has not considered their implications for optimal program design.[10,11]

[9]Currie (2003) provides a review of the U.S. food assistance programs and Bartfeld et al., eds (2016) gives extensive coverage to additional research on SNAP. Recent research studies how SNAP receipt affects household members' nutrition, health or other outcomes (Almond et al., 2011; Hoynes et al., 2016; Bronchetti et al., 2019; Bailey et al., 2020; Hastings et al., Forthcoming); whether the marginal propensity to consume food out of SNAP benefits differs from that out of cash (Hoynes and Schanzenbach, 2009; Hastings and Shapiro, 2018); and how SNAP affects recipients' labor supply (Hoynes and Schanzenbach, 2012; East, 2018; Harris, 2021). Ratcliffe et al. (2008) study the effect of *categorical eligibility* on SNAP take-up but do not examine the effect of eligibility thresholds. Homonoff and Somerville (2021) study the screening properties of the SNAP recertification process.

[10]There is little evidence that inframarginal effects would generalize outside the Medicaid setting. Much of the inframarginal effects documented in the health literature pertain to *within-household* take-up of the already-eligible population — for instance, new Medicaid take-up among children who are already eligible because children face less stringent Medicaid requirements than adults, as in Sacarny et al. (2022). By contrast, we show that entire households that were already eligible may sign up when eligibility requirements are relaxed.

[11]Outside of the health literature, Leos-Urbel et al. (2013) and Marcus and Yewell (2021) find that eligibility expansions boost take-up among inframarginal recipients of free school breakfast or lunch programs. These authors study reforms that granted universal eligibility; programs with universal eligibility may be very different than programs like SNAP where eligibility remains restricted. Moreover, program take-up among children may be subject to very different social dynamics and information frictions than among adults.

Roadmap. Section 2.2 establishes evidence of an inframarginal effect in SNAP. Section 2.3 discusses mechanisms underlying the effect, and Section 2.4 proposes the model of optimal eligibility thresholds. Section 2.5 presents welfare analysis. Section 2.6 concludes.

2.2 Inframarginal Effects in SNAP

This section documents the empirical relationship that motivates this paper. States that have less stringent eligibility standards tend to have higher take-up in SNAP *among inframarginal people* — people whose incomes are low enough that they are eligible everywhere, regardless of the state's eligibility threshold. Appendix B.1 provides more information about the dataset construction and policy variation.

2.2.1 SNAP Data

We obtain the total number of people who participate in SNAP from the SNAP Quality Control (QC) files, which are administrative data from the U.S. Department of Agriculture on a random sample of SNAP participants (United States Department of Agriculture Food and Nutrition Service, 2019). The data record granular information about household characteristics, benefit size, and incomes of SNAP participants. The data are a repeated cross-section, so we cannot study households over time. Using these files, we construct the total counts of program participants and those below a given income threshold, for each state and year from 1996 until 2016, the last year for which systematic policy data are available.[12] The Quality Control files are administrative data, so they record people's incomes and household size accurately, thereby addressing concerns about measurement error from Meyer et al. (2015) and others. On the other hand, the dataset is relatively small at the state level. There are about 100,000 observations across 51 states and DC in each year from 2001 to 2016.[13]

Sample and outcomes. We begin with a sample of individuals with household income between 0%–130% of the FPL. In this section, we also focus on a sample including only individuals in households with income between 50%–115% of the FPL. We exclude

[12]We construct our dataset by modifying the publicly available replication code for Ganong and Liebman (2018).

[13]Relative to comparable datasets, the QC data are best-suited for our analysis. The Survey of Income and Program Participation is not intended to be representative at the state level. The Panel Study of Income Dynamics and Current Population Survey both may be subject to measurement error about SNAP participation.

individuals between 115%–130% of the FPL to address concerns about measurement error: we might consider an individual near the threshold as "inframarginal" when in fact she would be ineligible under a different eligibility regime because of additional restrictions such as asset tests. We focus on individuals above 50% of the FPL because take-up is very high among individuals below 50% of the FPL, regardless of the state's eligibility threshold. Thus there is little scope for increased take-up among this group.

Using this sample, our main outcome is a measure of take-up *counts*. In particular, we use log total enrollment within specific income groups — in our main regression, among people earning 50–115% of the FPL. Almost all individuals in this range are eligible for SNAP in every state. This allows us to study inframarginal recipients; we are not counting increased enrollment among people who are newly eligible. Compared to take-up rates, this outcome has the advantage of not involving imperfect measures of the share of people who are *eligible* for the program as the outcome variable. Instead, we rely on the assumption that the number of people who are eligible regardless of the eligibility threshold (e.g., the number of people in households earning 50–115% FPL) is not correlated with the eligibility threshold beyond the controls we include; we provide support for this assumption below.

We form take-up *rates* as a secondary outcome. Following Ganong and Liebman (2018), we divide the number enrolled (from the QC data) by the number of people within a given band of the income distribution in the state from the Current Population Survey's Annual Social and Economic Supplement (CPS ASEC) (Ruggles et al., 2020). For instance, in our main specification, the denominator is the number of people in the CPS who are between 50–115% of the FPL. Crucially, the denominator does not exclude people who are otherwise ineligible for SNAP due to work requirements or asset histories. Thus the take-up rates are likely underestimates. There also may be measurement error in reported incomes in the CPS. We show that measurement error in the CPS data cannot explain our results in Section 2.2.6.

Just as in other work estimating SNAP take-up, a possible limitation to the analysis is that we do not always observe household assets or work histories, which can affect SNAP eligibility. First, federal rules restrict households with sufficient assets from participating in SNAP. In practice, only a small fraction of households are ineligible for SNAP under these asset histories. Second, under the Personal Responsibility and Work Opportunity Act (PRWORA), single households must meet certain work requirements to participate. However, the changes in these requirements do not coincide with changes in the eligibility threshold, and we also show that the results are similar among households with dependents. We also control for the requirements in robustness checks.

Policy changes. Federal rules require that households below 130% of the FPL are eligible to participate in SNAP. Beginning in 2001, states had the option to expand eligibility to additional households up to 200% of the FPL under Broad-Based Categorical Eligibility (BBCE). The SNAP benefit schedule, which is set nationally, does not depend on a state's eligibility rules.

The eligibility thresholds relaxed under the BBCE correspond to gross income tests. Households must also pass a net income test: net of allowable deductions (e.g., an earnings deduction amounting to 20% of their earned income), their income must be below 100% of the Federal Poverty Level. This is true regardless of the gross income test set by the state. Moreover, because the SNAP benefit size falls in net income and is not changed by the BBCE, many people who become newly eligible from the BBCE receive a small SNAP benefit. Nevertheless, this section documents that raising the gross eligibility threshold led to persistent and large increases in inframarginal take-up.

Not every state that adopted the BBCE took the option to expand the eligibility threshold. In Section 2.2.6, we note that adopting the BBCE did entail additional changes to state welfare programs, but we reject that these changes can explain the inframarginal effects we document here. Ultimately, 30 states expanded SNAP eligibility through the BBCE through 2016, four of which adjusted eligibility twice during this period (Figure B.1a). Expansions occurred throughout the period, but they were especially likely to occur in 2001–2002 and 2010–2011. The states that do roll out an eligibility expansion are generally distributed across the country, although there are no states in the Great Plains region that implement an expansion (Figure B.1b).

2.2.2 Econometric Strategy

We estimate an event-study regression that leverages the variation in eligibility provided through the BBCE. We index each event by event-time τ, where $\tau = 0$ represents the first fully treated year. We set $\tau = -1$ in all years for untreated states. We define the "event eligibility rate" in each state s as the eligibility rate as a percent of the FPL after the BBCE expansion in treated states and the federal minimum (130%) in untreated states. We use a balanced panel: we limit the sample to the five years before and after treatment for treated states, and include all years in control states.[14] We normalize our coefficients relative to the year before the event and estimate:

$$y_{s,t,\tau} = \sum_{r \in \mathcal{R}} \eta^r \left(\mathbb{1}(\tau = r)_{s,t,\tau} \times \text{event eligibility rate}_s \right) + \delta_s + \gamma_t + X'_{s,t,\tau}\phi + \varepsilon_{s,t,\tau} \tag{2.1}$$

where $\mathcal{R}$ is the set of event periods, s indexes states, t indexes years, event eligibility rate$_s$ measures the eligibility rate as a ratio of the FPL, δ_s is state fixed effects, and γ_t is year fixed effects.[15] We include X, a vector of additional linear controls for the state unemployment rate, the log of the number of people in a given income group in the state (measured in the CPS), SNAP outreach spending per person earning under 130% FPL in the states (transformed with $\sinh^{-1}$), and an index of other SNAP policies implemented around the same time (as in Ganong and Liebman (2018), henceforth the "Ganong-Liebman index").[16] In our primary estimates, we use $\ln(\text{enrollment}_{s,t,\tau})$ as the dependent variable ($y_{s,t,\tau}$). The coefficient of interest η^r represents the marginal effect of 1 pp increase in the eligibility rate (expressed in terms of the FPL) on enrollment in event-time r. This specification encodes a standard pre-trends test for whether $\eta^r = 0$ when $r < 0$. Our primary specifications are unweighted. We present standard errors clustered at the state level in this and all subsequent analyses that use state-year variation.

We also pool the data in this sample to estimate:

$$y_{s,t} = \eta \text{ eligibility rate}_{s,t} + \delta_s + \gamma_t + X'_{s,t}\phi + \varepsilon_{s,t}. \tag{2.2}$$

The variable eligibility rate$_{s,t}$ represents the eligibility as a percent of the FPL in a given state-year, so η is the average effect on inframarginal people after an eligibility expansion.

Discussion of controls. Given that our state and year fixed effects remove fixed differences in outcomes across states and across years, the identifying assumption is that there are no time-varying within-state trends in enrollment (not absorbed by our time-varying state controls). One concern is that states that impose the eligibility increase have faster population *growth* in the inframarginal sample. To address this concern, we control

[14]We drop the four states with two events in the event-study analysis, as well as the two states that have events too recently to have sufficient post-period data. This leaves 45 states (including the District of Columbia).

[15]For instance, event eligibility rate$_s$ = 1.3 represents that the state has the minimum threshold of 130% of the FPL.

[16]The Ganong-Liebman index is the average of several indicators for the presence of different policies that may influence SNAP take-up, such as whether households can apply to SNAP online. See Appendix B.1 for details on the variables that enter the index.

for the log count of the people within the inframarginal income group from the CPS. The economic environment and the policy environment are also relevant for SNAP take-up (Mabli et al., 2014; Ganong and Liebman, 2018). We control for the state unemployment rate to address the concern that states with eligibility increases may simply have more financial distress, and we include the Ganong-Liebman index to address the concern that states that expand eligibility may also impose other policies relevant to enrollment. We present robustness to additional threats to identification later in this section. Altogether, these controls do not have a dramatic effect on our results. The most important control is for the count of people who are eligible, which we show eliminates a modest (and insignificant) pre-trend in our event study.

2.2.3 Results

Descriptive evidence. Before presenting the formal estimates, we begin by visualizing inframarginal effects in the raw data. In Figure 2.3a, we present total SNAP enrollment per 1,000 people (population-wide) in state-years with eligibility thresholds equal to 130% FPL versus above 130% FPL. We normalize the enrollment by the total population in all states with the relevant income rule to aggregate enrollment counts across states.

First, without the eligibility expansion, very few individuals with household income above 130% FPL take up the program, while with the eligibility expansion, mass appears above 130% FPL where individuals are newly eligible. This confirms that the QC data give sensible estimates of the enrollment counts, and that the eligibility changes relax a binding constraint for some individuals. Second, individuals *below* the threshold also enroll at higher rates with looser eligibility restrictions. These inframarginal effects — the increased enrollment below the threshold — are the subject of our attention.[17]

Figure 2.3b presents a binscatter of the cross-sectional relationship between SNAP take-up among these inframarginal individuals (i.e., earning 0-130% FPL) and the state's eligibility threshold at the state-year level. We observe five different eligibility thresholds chosen by states between 1996–2016. Mean take-up is roughly 10 pp lower in states with eligibility at 130% of the FPL, the most stringent eligibility standard permitted under federal law.

Event-study specifications. For confidence that the raw data reflect inframarginal effects and are not driven by confounds, we turn to our event study (Equation (2.1)). We

[17]We note a slight excess mass around 75% of the FPL, which may be an artifact of the QC data; however, inframarginal effects appear throughout the income distribution.

plot log enrollment among our inframarginal sample by event period, relative to event period -1 (Figure 2.4a). We find no evidence of pre-trends leading up to the policy change. After the policy change, enrollment increases steadily. Figure 2.4b shows that the effect is concentrated among people in households earning over 50% FPL. We also exclude households over 115% to alleviate concerns (described above) about measurement error or unobserved assets. Our benchmark estimates suggest that increasing the eligibility level by 10 pp of the FPL boosts the number enrolled by 1-2 percent in the five years following the policy change.

To show the effect of controls on our empirical estimates, we present in Figure B.2 the event study with state and year fixed effects only (Panel A) and then add the control for the log of the total number of people between 50 to 115% of the FPL (Panel B). Overall the results are similar without controls. With no controls at all (Panel A), we see some visual evidence of a pre-trend prior to treatment, although the trend is small in magnitude and vanishes three years before treatment. Once we control for the log of the CPS population totals (Panel B), any pre-trend vanishes, and the results in Panel B are very close to those in Figure 2.4. Note that we are running log take-up regressions: the moderate importance of controlling for CPS population simply confirms that the denominator of a take-up regression matters and is not on-face concerning.

The event-study figure suggests that results grow over time. The effects are larger in years 4–5 than years 1–3, suggesting that inframarginal effects persist or grow in the medium-term.[18] Such effects might grow even several years later if, for instance, information takes time to spread or cascades once others become eligible. Alternatively, stigma might respond only slowly to changes in the threshold.

Placebo. We conduct a placebo test that offers a useful validation of the above results. We observe nine states implement the BBCE *without* expanding eligibility beyond 130% of the FPL.[19] Most of these states adopted the BBCE around the same time as the states in the main event study (2009–2011). Thus, we study the effect of the BBCE in the states that did *not* expand eligibility but *did* implement the BBCE. To implement the placebo test, we show an event study (as in Equation (2.1)), where treatment represents states that implemented the BBCE but did not expand eligibility (Figure 2.4c).[20] We use log

[18]Tests of the null hypothesis that $\eta^4 = \eta^3$ and $\eta^5 = \eta^3$ both reject with $p < 0.01$. A joint test for both hypotheses also rejects the null with $p < 0.01$.

[19]States can implement the BBCE for bureaucratic reasons, as the policy can simplify program administration, or to relax the SNAP assets test. See Appendix B.1.

[20]We exclude states that did increase eligibility from this test, so the regression includes 19 states. A handful of states which adopted BBCE without changing their eligibility thresholds at that point did

enrollment among the 0–130% of FPL sample as the dependent variable. This event study gives no effect; we find no evidence to support that the short- or long-term effects in placebo states are the same as the 5-year effect in states with an eligibility expansion.

The placebo test suggests that eligibility expansions, and not ancillary features of the BBCE, drive the results. We cannot completely rule out that the BBCE caused unobserved changes in outreach (not captured by our outreach control variable) or transaction costs (not captured by the vector of SNAP policy controls). But such forces would also be inconsistent with the placebo test, unless they only occurred in BBCE states that also raised the eligibility threshold.

Combined estimates. The event-study specification and placebo test confirm the existence of inframarginal effects. To obtain the pooled effect over all periods, and parsimoniously present robustness to different specifications, Table 2.1 estimates Equation (2.2). Our preferred specification (Column 1) uses the sample used in the event study and includes state and year fixed effects, and controls for the state unemployment rate, outreach spending, and the Ganong-Liebman controls.[21] The independent variable is the eligibility threshold as a ratio of the Federal Poverty Level, so that increasing it by 1 corresponds to increasing the threshold by 100% of the FPL. We find that $\eta = 0.107$ and reject $\eta = 0$ at $p < 0.05$. These estimates suggest that raising the eligibility rate by 10 pp of the FPL (e.g., from 130% to 140%) boosts take up by 1.07 percent. The modal eligibility increase in our sample is from 130% to 200% of the FPL, which delivers a 7.5 percent increase in take-up among this sample ($0.7 \times 0.107 \approx 0.75$). The results in Column 1 are consistent with the event study plot.

The rest of Table 2.1 shows that our estimate of inframarginal effects is robust to the particular choice of the specification. Column 2 separates the Ganong-Liebman index into separate indicators for each component variable. Column 3 reverts to the index form of these controls but adds new controls for lagged unemployment and the prevalence of waivers relaxing the SNAP work requirements for able-bodied adults without dependents (ABAWDs), beginning in 2010.[22,23] Column 4 excludes the years 2008–2011 (the Great Recession). Column 5 weights by state-year population. Column 6 computes the treatment

expand eligibility at a later date. Here, we exclude these states, but the results are similar when they are included and we add a control for the eligibility threshold.

[21]We control for the inverse hyperbolic sine of outreach spending to address state-years with zero outreach spending (Burbidge et al., 1988).

[22]We use data on ABAWD waivers from data generously shared by Harris (2021).

[23]Figure B.3b also shows the event study where the sample includes only SNAP recipients in households with children.

effect as the difference between the average of the event study coefficients in the post period and the average of the coefficients in the pre period, weighting all post periods equally. Finally, Column 7 uses all years of data we have (a balanced panel of 50 states and D.C. from 1996–2016), instead of only the event study sample of a 5-year window around the eligibility increase. It also includes states that change eligibility several times or reduce eligibility. Throughout the table, the results are stable: estimates of η range from 0.10 to 0.12.

We also repeat the exercise for two different samples in Table B.1 and find similar results. Panel A shows enrollment responses in the 0-130% FPL sample. The estimates are consistent with the main results but generally lower. This attenuation reflects that our dependent variable (SNAP enrollment) has less scope to rise when almost all people from 0–50% of the FPL already take up SNAP. Panel B assesses enrollment among households with children, as these people are likely not subject to the additional ABAWD work requirements that were relaxed and reimposed during the sample period. Here, we see similar sized, though noisier, effects in this sample. Together, these results and Table 2.1 provide strong evidence of inframarginal effects from the BBCE, as the effect persists across specifications and samples.

Distributional effects. From which portion of the income distribution do inframarginal effects arise? We present treatment effect heterogeneity by household income (Figure B.5). We estimate a version of Equation (2.2), using take-up *rates* instead of log enrollment counts so that the values are more directly comparable across income groups with different base rates. Take-up rates increase most among those earning 130–160% FPL, who are barely ineligible before an expansion. The effect in this group is larger than the largest effect in the inframarginal population, among those earning 100–130% FPL. However, even after the expansion, take-up in the newly eligible group is still much lower than any other group. We also see that the treatment effect size is increasing with household income within the inframarginal sample; however, this may partially reflect that the base take-up rate is much lower among households with relatively more income.

Characterizing compliers. Who is most affected by eligibility expansions? To the extent that inframarginal effects are driven by reductions in barriers to take-up ("ordeals"), they may affect the targeting properties of the expansions (Nichols and Zeckhauser, 1982). If inframarginal effects influence SNAP's screening capacity, we expect the people who join the program after an eligibility expansion to look different on observables than the previously enrolled. On the contrary, we find little evidence that the eligibility threshold affects the characteristics of SNAP enrollees earning 50–115% FPL (Table 2.2). Of the characteristics we can analyze, we only find a significant positive effect on the average poverty level of enrollees. However, the magnitude of these effects is small: increasing the eligibility threshold from 130% FPL to 140% FPL, for example, would imply a 0.07% FPL increase in the average gross income of SNAP recipients. Together, these results suggest that whatever the ordeals behind inframarginal effects, they do not have substantial screening effects.[24]

2.2.4 Interpreting the Magnitude of the Results

We now provide three ways of interpreting the magnitude of the results.

Take-up elasticity. We estimate the elasticity of take-up with respect to the share of the population who is eligible. The elasticity will also play a critical role in the theoretical model.

We employ an instrumental variables approach to estimate this elasticity. The share of the population eligible for SNAP is affected by confounding conditions which also affect the number of people below a certain income level. The eligibility expansions provide plausibly exogenous shocks to the share eligible. Thus we instrument for the log share eligible for SNAP using the state-and-year-specific income cutoff as a ratio of the Federal Poverty Level. The exclusion restriction is that eligibility expansions are not associated with take-up of inframarginal people except through changes in the share eligible.

We return to Equation (2.2) from Section 2.2. We use a log-log specification, with ln(take-up) and ln(share eligible) as the dependent and independent variables, respectively. The estimating equation is:

$$\ln(\text{take-up})_{s,t} = \eta \ln(\text{share eligible})_{s,t} + X'_{s,t}\phi + \delta_s + \gamma_t + \varepsilon_{s,t}, \tag{2.3}$$

where we instrument for ln(share eligible) using the state eligibility threshold as a ratio of the FPL. Here η represents an elasticity rather than a level effect.

We present the IV estimates for the 0–130% sample using all the data (Table 2.4, Panel A) as well as the event-study sample (Panel B). We document a strong first stage: in the full sample, increasing the eligibility threshold by 10% of the FPL increases the share of a state population that is eligible by 7.28% (t-stat = 21.37), with similar results for the

[24]Table 2.2 also shows no evidence of an increase in the share of enrollees whose SNAP certification period is less than 6 months, suggesting that new enrollees also do not have more volatile income.

event-study sample. Our 2SLS estimate in the full sample is $\eta_m = 0.130$ (SE: 0.067); the estimate in the event-study sample is $\eta_m = 0.104$ (SE: 0.077). We also document that simple OLS regressions of log take-up on the log share eligible have the opposite sign, likely due to the omitted variables bias we described above. The full sample estimate is more precise, so we prefer it when used for welfare analysis.

Comparison to inframarginal effects in Medicaid. We now convert our inframarginal effect estimate to the same units as Sacarny et al. (2022) to compare magnitudes. Sacarny et al. (2022) find that about 0.1 previously-eligible children enter Medicaid for every adult who entered Medicaid from the Oregon Health Insurance experiment. To compare to this point estimate, we employ the magnitude of the inframarginal effect among the entire inframarginal population (Table 2.1A).[25] We find that .91 (standard error: 0.57) inframarginal people between 0–130% of the FPL are induced to take up the program for every newly eligible person who takes up the program.

We cannot reject that the treatment effects are equal to those in Sacarny et al. (2022). Even so, our point estimate is that inframarginal effects in this setting are nine times larger than in Sacarny et al. (2022), which warrants discussion. Altogether, we have no reason to expect that inframarginal effects will be of the same magnitude across programs and over time. In this setting, expanding the SNAP eligibility threshold for gross income does not loosen other eligibility criteria (e.g., the net income threshold). These criteria may bind for people with higher incomes. As a result, an eligibility expansion can lead to higher take-up among the inframarginal population without many newly eligible people joining the program.

Comparison to outreach spending. A final way of benchmarking our effects is to compare the take-up from inframarginal effects to the take-up from direct SNAP spending on information and outreach. The SNAP Policy Database contains information on states' outreach spending, but we do not have quasi-random variation in this spending. For an effect of outreach on take-up, we turn to the randomized control trial run by Finkelstein and Notowidigdo (2019), where the authors find that sending mailers to people who are likely eligible for SNAP but not enrolled boosts enrollment. They calculate that their intervention costs about \$20 per additional enrollee induced to join by the outreach

[25] Let the point estimate for the entire inframarginal population from Table 2.1A, Column 1 be $\hat{\eta}_i$. We then estimate a version of Equation (2.2), using the log of the total number of people on the program as the dependent variable (and controlling for the log of the number of people below 130% of the FPL from the CPS). Let the point estimate from this regression be $\hat{\eta}_t$. We then present $\frac{\hat{\eta}_i}{\hat{\eta}_t - \hat{\eta}_i}$, where the denominator represents the increase in the marginal population and the numerator represents the increase in the inframarginal population.

intervention. At this rate, it would cost about $66 million to increase enrollment in the inframarginal population by the same amount as raising the income eligibility threshold from 130% FPL to 200% FPL.[26]

On the one hand, $66 million is a fraction of the total annual spending on SNAP ($70 billion in 2016). On the other hand, it is more than three times what all states combined spent on outreach in 2016 ($17.4 million). Finally, the mechanical cost of raising eligibility goes to program recipients who are newly eligible. But the mechanical cost of outreach does not go to program recipients. To summarize, outreach spending may be an alternative instrument for increasing SNAP take-up, but it is not obviously a better one than increasing information by raising the eligibility threshold.

2.2.5 Cost-effectiveness

We conduct a back-of-the-envelope calculation to compare the mechanical costs of two natural interventions to raise take-up of the inframarginal population by 1 pp: raising the eligibility threshold and raising the benefit size. We find that the methods have similar mechanical costs (Table 2.3). Using the η_m estimated in Equation (2.3), we calculate that to increase take-up by 1 percentage point, an additional 4 pp of the US population would need to be eligible for SNAP. If take-up in the newly eligible population is similar to take-up among people who are just barely eligible (25%, Figure B.5), and the benefit size is similar to the benefit size in this group ($707 per person-month, calculated from the QC data), then this intervention costs an additional $2.2 billion per year. To compare raising the eligibility threshold to the cost of raising take-up by raising the benefit size, we assume that the elasticity of take-up with respect to the benefit size is 0.5 (see Section 2.5.3 for details). To get a 1 pp increase in take-up, the benefit size would need to increase by $56 per year for 44 million SNAP enrollees — costing $2.5 billion per year.

This cost-effectiveness point does not have direct implications for social welfare, since the cost and benefit of each policy instrument also depend on recipients' willingness to pay. However, it is a model-free way to compare the tools.

[26]There were around 44 million SNAP enrollees in 2016. To derive the number of new enrollees from such an increase, we multiply 44 million by the increase in take-up (7.5%) implied by our estimates in Table 2.1 at the modal eligibility threshold increase (130% to 200% FPL). Finally, we multiply this by $20 per additional enrollee to arrive at $66 million.

2.2.6 Robustness

Balance. The identifying assumption in our event study is that there are not other factors besides the eligibility threshold that contribute to inframarginal take-up and coincide with the means test policy change. A related concern is that the states which change their threshold are different from those that do not; note that with our event study framework, internal validity does not require that control and treatment states are similar.

We first test whether states that implement the BBCE bundle the change with other adjustments to SNAP policy. We note that, following Ganong and Liebman (2018), all our regressions control linearly for an index of eight other SNAP policies that occur during the same period (measured by the SNAP Policy Database). As Table 2.1 shows, including this index makes little difference, which gives additional confidence that unobserved policies do not affect the results. Moreover, when the index is separated into its component parts (Column 2), the magnitude of the effect is not diminished. One might nevertheless worry that the eligibility expansions were bundled with informal policies (e.g., flyer campaigns) that the SNAP Policy Database does not measure. To allay this concern, we present a placebo event study, with the SNAP policy index as the dependent variable (Figure B.4A). We find no evidence that the SNAP index increases after the eligibility expansions. Overall, the test is inconsistent with economically material bundling of SNAP policies. Finally, because we control for this index, this objection requires unobserved policies to affect the outcomes even after residualizing by the index.

We next examine whether economic conditions change leading up to the changes in the eligibility threshold. We estimate Equation (2.1) with the log of the CPS counts of the people at 50–115% of the FPL as the dependent variable (Figure B.4B). We find a slight pre-trend in the CPS populations three years before the event, but the effects are modest. We discuss whether including this control affects the results above; it helps alleviate a moderate but insignificant pre-trend in the main event study. Similarly, we estimate Equation (2.1) with the unemployment rate as the dependent variable. Although the unemployment rate appears to grow in advance of the policy, the trends are insignificant (Figure B.4C). Moreover, the time series pattern of the changes in the unemployment rate do not align with our main results: the unemployment rate returns to 0 after 5 years, whereas our main effects persist. That is why when we control for the unemployment rate, this control does not materially affect our results (Figure B.2B versus Figure 2.4b). We conduct two additional tests to address the concern about the unemployment rate changing in advance of the policy. We include a further control for lagged unemployment (Column 3, Table 2.1). We also exclude the Great Recession (Column 4), when unemployment rates had the greatest fluctuation. Our results remain robust.

Our final balance exercise is a standard one: we compare states which did and did not ever change their eligibility threshold (Table B.2). Because our main results use an event study, imbalance in levels in the pre-period is not itself concerning; however, it can still be helpful to understand whether treated states were different from untreated states. In Panel A, we see that states which ever changed their eligibility threshold have significantly higher average family incomes in the pre-period (measured in 2000, the last year before any state changed its eligibility threshold), and marginally significantly higher measures of SNAP access-related policy (the Ganong-Liebman index and SNAP outreach spending). However, Panel B shows that these measures are not strongly associated with the size of the means-test change, which provides suggestive evidence that the policy decision is not driven by these measures.

Measurement error. To mitigate concerns about measurement error, our key empirical fact (inframarginal effects) uses the QC numerators as the dependent variable, as in Table 2.1. Even so, we control for the size of the eligible population, which may be measured imperfectly in the CPS.[27] First, we note the above point that the share eligible does not change with treatment. This pushes against concerns that differential measurement error in the pre- and post-periods drives our results. Appendix B.2.3 presents a simulation that shows that only an implausible amount of measurement error, exactly coinciding with the event and only in treated states, could explain our results.

There may also be measurement error in the timing of the policy implementation.[28] We use data at the annual level in our main specification because we measure the number of people who are eligible from the March CPS, which is only available annually. Moreover, the QC data contain relatively few people at the month-state-income group level. However, BBCE policies can be implemented mid-year. In Figure B.3a, we show our event study using monthly data to estimate Equation (2.2). It looks broadly similar, although the inframarginal response is slightly slower to appear. This reflects the fact that in our main specification, we index policy implementation to the beginning of the first fully treated

[27]We show the main event study with take-up rates on the left-hand side in Figures B.3c and B.3d.

[28]We follow the date of the policy implementation in the SNAP Policy Database. However, the precise implementation date may vary across sources, and the legal implementation date may not coincide with the date that the program actually began accepting people with incomes larger than 130% of the FPL (e.g., if program social workers need to be trained on the new procedures). In practice, measurement error along these lines would merely add noise to the event study.

year.

Other effects of the BBCE. A related concern is that some states grant extra eligibility through the BBCE together with explicit referrals or brochures to SNAP. As a part of the BBCE, states sometimes use the budget from the Temporary Assistance for Needy Families to fund referrals to state services, including SNAP. As Figure 2.4c shows, states which adopted the BBCE but did not expand eligibility did not see similar effects on SNAP enrollment. This placebo test thus constitutes strong evidence that only the eligibility threshold, and not ancillary BBCE-related policies, are responsible for the take-up effect.

The BBCE also waives some rules on the maximum assets that block families in states without BBCE from obtaining SNAP. First, the above placebo study also rejects this concern, since BBCE states that maintain eligibility at 130% of the FPL but do change their asset limits do not exhibit a take-up increase. Second, in practice, these asset rules affect a small number of families. Ganong and Liebman (2018) find asset waivers were responsible for only a small share of increased take-up in recent years. Eslami (2015) finds that 4 percent of inframarginal people who participate in SNAP are eligible only due to state asset eligibility rules.[29] There are a host of such asset waivers, including many not linked to the BBCE. But even assuming all these households were only eligible due to the BBCE, the asset waivers could not explain even half of the inframarginal effects we find.

Two-way fixed effects and negative weights. Concerns about negative weights (Callaway and Sant'Anna, 2020; Sun and Abraham, 2021) are unlikely to apply in our setting, since: (i) there is a large pool of never-treated units, and (ii) we do not have always-treated units. As a check we implement the heterogeneity-robust stacked estimator from Cengiz et al. (2019a) and the Sun and Abraham (2021) estimator. We obtain similar results (Appendix Figure B.6); the Sun and Abraham (2021) estimator delivers somewhat larger results in years 4 and 5 (but the confidence interval safely contains the original point estimates).

Policy salience. An additional concern is that the inframarginal effects arise in our setting because the expansions are *salient* to people, but they are not steady-state responses. First, we show that eligibility expansions boost take-up up to five years after the expansion, so they at least have effects in the medium-term. Second, the event study plots also show that the jump in take-up does not coincide with the expansion but grows over time.

[29] See computation in Ratcliffe et al. (2016).

2.3 Mechanisms: Information and Stigma

Why does the eligibility threshold affect inframarginal take-up? The question relates to a long-standing literature on incomplete take-up of social programs that categorizes barriers to take-up into incomplete information, stigma, and other enrollment costs. Furthermore, the model in the following section will also make clear that the mechanisms matter for welfare analysis.

One hypothesis is that the eligibility change affects stigma around SNAP take-up. For example, it is possible that when SNAP becomes available for relatively wealthier people, SNAP no longer conveys as much of a negative signal. We test this hypothesis using an online experiment in which we exogenously change participants' beliefs about the SNAP means test. A second hypothesis is that changing the eligibility threshold increases the information about the program. For example, because more people are eligible, people can more easily obtain information about how to apply from friends or family. We test this hypothesis by making novel use of USDA survey data on SNAP stigma and information.

We do not emphasize non-stigma enrollment costs as a potential mechanism, because Section 2.2 provides evidence that the eligibility changes did not meaningfully change enrollment costs. For example, we find no differential effect on people with different recertification periods. However, our theoretical model will permit changes in these costs.

2.3.1 Online Experiment: Evidence of Stigma

Here, we present evidence from an online experiment that the eligibility threshold may affect perceived stigma around SNAP take-up.[30]

2.3.1.1 Experiment Design

The objective of the experiment is to induce variation in participants' beliefs about the share of people who are income-eligible for SNAP. In particular, we study how raising people's beliefs about the share eligible affects self-reported stigma.[31] Figure B.9 summarizes the experiment design.

[30]We used the survey provider Lucid; other papers using Lucid include Wood and Porter (2019) and Bursztyn et al. (2020). We ran the experiment in March 2020. The onset of the coronavirus pandemic should not complicate the treatment-control differences via our randomized information provision.

[31]The complete survey instruments are available from Rafkin's website.

Main experiment. Our main experiment was embedded in a question asking respondents to report what share of Americans they thought were income-eligible for SNAP in 2016.[32] On this page of the survey, all respondents were given a truthful hint: *"In 2016, in one of the U.S. states, roughly [X] of the population had low enough income that they could qualify for SNAP."*

X was randomly either 15% or 38%, which were the highest and lowest state-level eligibility shares we see in the administrative SNAP data from 2016. We refer to those participants who saw the 38% hint as those in the "high-share" treatment.

Belief elicitation. After implementing the treatment, we conduct a manipulation check by eliciting people's beliefs about the share of people eligible for SNAP. We asked: *"In 2016, how many out of every 100 people (in all U.S. states) do you think have low enough income that they could qualify...?"*

Auxiliary experiment. Following the belief elicitation, we included an auxiliary randomization: we informed a random subset of participants about the correct share (27%, as per our calculations combining the CPS and the SNAP Policy Database). Depending on their prior beliefs, this treatment (which we call the "belief-correction" treatment) is intended to cause participants to update up or down about the share of people who are eligible for treatment. Unlike the main treatment, the auxiliary belief correction treatment does not have a tight connection to changes in an eligibility threshold.[33] As a result, we relegate discussion of the belief-correction treatment to the Appendix.

Stigma elicitation. We asked respondents to rate their agreement, on a scale from 1 to 9, to a series of eight statements about SNAP: *(1) I would prefer not to use food stamps because I would rather be self-reliant and not accept help from the government; (2) I believe that people should do what they can to avoid being on food stamps; it is better to make it on your own; (3) Most people believe that someone who uses food stamps is just as hard-working as the average citizen; (4) If I used food stamps, I would be concerned that people would treat me disrespectfully at stores; (5) Most people believe that someone who uses food stamps does so because of circumstances outside their control; (6) Most people think less of a person who uses food stamps; (7) Most people who use food stamps would go out of their way to prevent others knowing about their food stamp receipt; (8) If I used food stamps, I would avoid telling other people about it.*

We aggregate the statements into two indices: (i) "first-order stigma," which ask

[32] Reports were incentivized as follows: participants were told at the beginning of the survey that a lottery would be conducted among respondents who answered a factual question correctly, and the winner would have $50 donated to her choice of charity.

[33] We originally included the auxiliary experiment because recent papers, e.g. Bursztyn et al. (Forthcoming), use similar belief corrections to manipulate people's prior beliefs.

respondents about their own attitudes (statements 1, 2, 4, and 8 above), and (ii) "second-order stigma," which asks respondents about others' attitudes (statements 3, 5, 6, and 7).[34] We standardize these outcomes using the mean and standard deviation of the control group and then average the standardized values as in Kling et al. (2007a). We also show the effects on an aggregated index.

Either first- or second-order stigma could play a role in inframarginal effects, depending on the model. If people care about social image and take-up is partly observable, the extent to which others condone or sanction SNAP may affect take-up costs. With first-order stigma, people may have a hedonic aversion to SNAP that does not depend on others' views. Such aversion could easily influence take-up if modeled as a direct take-up cost.

Sample construction and balance. We drop participants who fail either of two pre-registered attention checks, as well as those who did not provide a prior or respond to all stigma questions. Our final sample has 2,131 participants (79% of the original sample). Table B.6 summarizes these sample limitations and confirms that attrition, inattention, and non-response were balanced between treatment and control. Appendix B.1 describes the data cleaning in more detail.

The sample is balanced across the high-share treatment (joint p-value: 0.94) and has a relatively similar composition as the U.S. on average (Table B.4). In some tests, we restrict the sample only to the 512 people below 130% of the FPL, because this subgroup — the inframarginal SNAP sample — is of particular interest for inframarginal effects. Among this subgroup only, a joint F-test suggests experimental imbalance (p-value: 0.02).[35] The experiment was randomized but not stratified, and any imbalance in this subgroup occurred by chance. To address the lack of balance when studying treatment effects in this subgroup, we present robustness tests that control for available demographics. We stress that the experimental treatment is balanced in the full sample, and we emphasize results from the full sample as a result.

Econometric strategy. In our primary specification, we simply compare the difference in means across treatments:

$$y_i = \beta \mathbb{1}(\text{high})_i + \gamma \mathbb{1}(\text{truth})_i + \varepsilon_i, \tag{2.4}$$

[34] We reverse the scale for questions 3 and 5 so that positive numbers always indicate more stigma.

[35] The most imbalanced covariate is that the high-share treatment is less concentrated in the Northeast region than the low-share treatment (p-value of difference: 0.02).

for individual i, where β represents the coefficient of interest. In robustness exercises, we estimate a version of Equation (2.4) with additional demographic controls. We conduct inference using robust standard errors.

2.3.1.2 Experiment Results

Beliefs about eligibility. The high-share treatment successfully moved beliefs about eligibility (Figure B.10). Both groups report beliefs that are slightly overestimated but reasonable; the mean for the low group is that about 30% of people are eligible, and the mean for the high group is that about 39% are eligible. The raw difference in means is 9.21 pp (SE: 0.80, p-value < 0.001). The standard deviation of beliefs in the control group is 19.8 pp, so the treatment raised the beliefs by a sizable 0.47 standard deviations. Moreover, while the low- and high-share treatments anchored a large fraction of people toward the numbers we provided them (15% and 38%), it also moved beliefs for others throughout the distribution.

Stigma. First, we note that responses to the eight stigma statements are somewhat but not overwhelming correlated (Figure B.11), so each question may contain independent information about the participants' views. A concern is that participants simply anchor to their responses on the first question since the question order was not randomized. In fact, while we find that responses to the second question are relatively correlated with the first question (correlation ≈ 0.65), other questions do not display a large correlation with the first question.

Next, we turn to investigating the treatment effects. Increasing individuals' beliefs about the share of Americans eligible for SNAP decreases their self-reported second-order stigma (Figure 2.5A). Aggregating the results into indices, the high-state treatment reduced second-order stigma by -0.050 standard deviations (SE: 0.027, $p = 0.061$). Effects are larger in magnitude among the 512 participants below 130% FPL (point estimate: -0.109, SE: 0.058, $p = 0.061$).

The treatment effects for second-order beliefs are similar across questions that form the second-order index. In the full sample, the high-share treatment reduces stigma the most in the question about whether most people believe recipients "go out of their way to prevent others knowing about their food stamp receipt." We find larger effects among people who have ever taken up SNAP, men, and Democrats, although treatment effect heterogeneity is not generally significant (Figure B.12). On the other hand, we find positive but statistically insignificant results on first-order stigma (Panel B).

We summarize these results in Table B.7, and we find very similar results when we include demographic controls (Table B.8). Moreover, when we aggregate the second- and first-order stigma results into a combined index, we find no statistical evidence of an average effect on stigma, although the point estimate is negative. The null result is mechanically driven by the null or slightly positive effect on first-order stigma.

2.3.1.3 Experiment Conclusions and Caveats

This experiment provides new empirical evidence on one possible mechanism underlying inframarginal effects. It serves as a useful contribution in its own right. The health literature on inframarginal effects has not provided clean evidence that either information frictions or stigma costs contribute to inframarginal effects. Additionally, evidence about stigma in social welfare programs remains elusive (Currie, 2004; Bhargava and Manoli, 2015). Our experiment suggests key aspects of program design, e.g. the eligibility threshold, indeed have the potential to affect program stigma.

While the experiment suggests that stigma could, in principle, drive inframarginal effects, the evidence we provide is not dispositive. We note several caveats. First, we find no effects on first-order stigma. First-order beliefs about, say, whether one should accept help from the government may represent deep-seated aspects of one's identity. It is therefore not surprising that people's first-order beliefs may be hard to move in a light-touch survey experiment. Second, an important caveat about our design is that we presented the high- and low-share treatments before a belief correction exercise. The belief-correction exercise itself does not provide evidence that the means test affects stigma (see results in Appendix C.4). Third, as with any online experiment, one may worry about external validity. We cannot experimentally manipulate the actual SNAP eligibility threshold, only people's perceptions of it.

Finally, we do not have a measure of whether the intervention affects SNAP take-up; this motivates our next empirical analysis.

2.3.2 Stigma, Information, and Take-Up

In the previous section, we found evidence that the means test affects perceived stigma around SNAP take-up. In this section, we show that the subgroups whose stigma decreases the most in the online experiment do *not* have the largest changes in take-up in the administrative data used in Section 2.2. Instead, those subgroups who appear to have the lowest stigma about SNAP, and those that are least likely to have information about

SNAP, are those that see the largest changes in take-up. Together, this suggests that the means-test affects take-up largely by increasing information availability.

Data. For this exercise, we include data from an additional source: the USDA's Food Stamp Program Access Study (FSPAS) (Bartlett et al., 2004).[36] The USDA's FSPAS involved phone and in-person interviews conducted in 2001 with a reference month of June 2000. Since the analysis of inframarginal effects uses QC data from 1996–2016, the FSPAS data occur toward the beginning of the sample period. We use data from two subsurveys: one of a random sample of approved SNAP applicants, and another of a nationally representative sample of likely eligible nonparticipants.[37]

In both surveys, respondents are asked a series of four questions about their perceived stigma around SNAP; they are also asked a number of questions about the information they have about SNAP. We consider respondents who reported any feelings of stigma to be affected by stigma. We consider any nonparticipants who reported a lack of information about any of three information questions to be affected by information frictions.[38]

The data include demographic information, including gender, age, race, marital status, and number of children. Because we also have these variables (as well as household income) in our online experiment and in the administrative SNAP data, we can compare statistics at the demographic cell level between datasets. Each cell is defined by the gender and age (binned into 18-30 year-olds, 31-65, and 65-100) of the household head; whether or not the household head is a non-Hispanic white; the household composition (married adult with children, unmarried adult with children, or adult(s) without children); and, where available, the income decile of the household when compared to the distribution of incomes in the US Current Population Survey.

Descriptives. Figure B.7 shows the stigma and information statements presented to FSPAS respondents and the share of respondents who agreed with each statement. About 40% of the sample agreed with at least one of the stigma statements, leading us to categorize them as being affected by stigma. Of those who agreed with any stigma statements, almost half (47%) agreed with only one, and another 29% agreed with two. Meanwhile, about 60% of the nonparticipant sample disagreed with any of the

[36]To our knowledge, this is the first academic study of the FSPAS, which the USDA generously shared with us.

[37]Among nonparticipants deemed to be eligible from an initial screener, 96.3% completed the survey. Among applicants randomly sampled from lists provided by SNAP offices, 56.7% of were reached and completed the survey. We analyze a sample of 1,585 respondents who either answered questions about stigma or answered questions about information (and have non-missing weights assigned by the USDA).

[38]These asked whether participants had heard of SNAP; whether they thought they were eligible for SNAP; and whether they knew where to go to get SNAP benefits.

information statements, leading us to categorize them as being affected by information. Finally, we show descriptive statistics by whether we consider the respondent to be affected by stigma (Table B.3). Those who report any stigma are more likely to be white, are on average younger, and are more likely to have children in their household. Notably, those who report any stigma are *more* likely to be enrolled in SNAP.

Results. Next, we study whether demographic cells with many stigma or information types have larger inframarginal effects (the binned scatterplots in Figures 2.6).[39] First, we find that cells with many stigma types have smaller inframarginal effects (panel A), and cells with many information types have larger inframarginal effects (panel B). While the relationships are noisy, we can statistically reject that the slopes are equal to zero at $p < 0.05$. The fact that the cells with many stigma types are statistically *less* likely to have large inframarginal effects is particularly suggestive that inframarginal effects are not driven by stigma. To complete the story, Figure 2.6C presents the correlation between the treatment effect from the online experiment (i.e., the effect of the perceived means test on reported stigma) and the treatment effect from the main analysis (the effect of the means test on take-up). Subgroups with the largest reductions in stigma when the means test increases do *not* have the largest inframarginal effects.

Discussion. Taken together, we find no evidence that stigma contributes to inframarginal effects. We find some suggestive evidence that information is responsible. Nevertheless, the experiment shows that increases in the means test decrease stigma costs. How should we interpret these facts? We use a model to conduct welfare analysis.

2.4 Model

In this section, we develop a model for analyzing optimal eligibility in the presence of inframarginal effects. Our model takes as given that SNAP — or, more generally, any lump-sum, means-tested transfer program — exists. We do not model the optimality of SNAP above and beyond redistribution via an income tax. Instead, we use the model to consider how to determine the share of the population which should be eligible for a redistributive program that has incomplete take-up. We use the model to emphasize the relevance of distinguishing between different mechanisms for the effects in Section 2.2. Our main argument is that whether take-up barriers are consistent with agent

[39]Appendix B.4 gives details about forming these measures. Because these binned scatterplots plot cell-level coefficients estimated with error, we conduct our tests weighting by the inverse of the product of the variances of the coefficients, also discussed in the Appendix.

optimization affects both the incidence and the size of the welfare gains.

2.4.1 Benchmark

We begin by analyzing a benchmark model where take-up responds endogenously to the eligibility threshold, but all consumers optimize. We add optimization failures due to imperfect information in Section 2.4.2.

We start by assuming take-up costs are normatively relevant (i.e., consumers are perfectly rational optimizers with respect to the take-up decision). In the following discussion, we often refer to these costs as "stigma costs," since we are especially interested in the case in which raising the eligibility threshold can reduce stigma and therefore boost take-up. However, the costs refer to any cost that inhibits take-up, e.g. hassle costs. Other possible mechanisms that might be cast as changing costs include transaction costs, for instance if more stores accept SNAP once more people become eligible. If the threshold reduces *uncertainty* about eligibility, that might be either an increase in the expected net benefit or an increase in awareness, depending on the model.

There is a continuum of individual types $\theta \sim F$, which correspond to ability or higher consumption. Types are perfectly observable, but we consider an environment in which the government cannot give a type-specific transfer (e.g., due to political economy or implementation constraints). The government offers a social program with a lump-sum consumption benefit B. The government provides B only to types $\theta < m$, where m is the eligibility threshold (or means test/income cutoff) also chosen by the government. We normalize the distribution of types to be *quantiles* of the distribution used to determine program eligibility (for example, the income distribution), i.e. $F := U[0,1]$.[40]

Denote the welfare weight on type θ by λ_θ, which refers to the welfare weight of quantile θ in the type distribution. For example, λ_0 refers to the weight that the planner places on the lowest-quantile person. We assume that the welfare weights are weakly decreasing in θ.

Assume all people have the same twice continuously differentiable and concave utility function from taking up the benefit, denoted by $u(B)$. Normalize individuals' outside income to be 0 and outside utility to be $u(0) = 0$. We already permit differences in realized consumption utility for each type to enter the planner's problem through λ_θ. We

[40] Note that this normalization is innocuous: it amounts to letting type m simply refer to the m-th quantile of the type distribution. The planner chooses what fraction of people are eligible, rather than the threshold type who is eligible.

can simply redefine a type's welfare weights to capture the different consumption utility that the type experiences.

Individuals choose whether to take up the benefit. We incorporate inframarginal effects by allowing the take-up probability to depend on the eligibility threshold m. In particular, every individual faces a take-up utility cost c, drawn from a continuously differentiable distribution H (which we additionally assume has a finite first moment). We suppose H depends on m, so $H(\cdot|m)$ and $h(\cdot|m)$ are the CDF and PDF of c.

We assume separability between the consumption benefits and take-up cost. Write realized utility as $U(B,c) = u(B) - c$. Then individuals participate in the program if $u(B) - u(0) > c$, i.e. $u(B) > c$.[41] Because $H(u(B)|m)$ is the take-up probability, define $p(B,m) := H(u(B)|m)$. We sometimes suppress arguments and write $p(B,m)$ as p, so that the probability an individual of type θ takes up the program is p_θ. We also assume that each type takes a cost draw from the same distribution, so that $p_\theta = p$.

Labor supply. We assume households' labor supply is fixed: there are no labor supply responses to the threshold. We relax this assumption in Appendix B.7 and show how a general problem with endogenous labor supply nests the key insights in this framework. Assuming fixed labor supply simplifies the framework considerably and permits us to focus on our novel mechanism (inframarginal effects).

Planner's problem. The planner faces a budget constraint T. In our setting — as in, e.g., Finkelstein and Notowidigdo (2019) — the social planner cannot simply set the optimal nonlinear income tax. For instance, the planner in our benchmark model might correspond to a state-level administrator tasked with choosing the parameters of a fixed program budget allocated by Congress. Indeed, such a setting is especially natural with SNAP, where state administrators choose the eligibility threshold but face an exogenous federal income tax.[42]

The planner solves:

$$\max_{B,m} p(B,m) \left(\int_0^m \lambda_\theta u(B) d\theta - \int_0^m \int_{c \leq u(B)} \lambda_\theta c h(c|c < u(B), m) dc d\theta \right) \tag{2.5}$$

[41] Of course, utility is also a function of other consumption. The model takes this consumption as exogenous and normalizes $u(0) = 0$. In Appendix B.7 we show that the model can accommodate different consumption across types, at the cost of notational complexity.

[42] We close the planner's budget constraint by trading off eligibility threshold increases with per-person benefit size decreases. While this is a natural tradeoff to consider theoretically, in practice in SNAP, the benefit schedule and the eligibility threshold are chosen by different decision-makers (federal and state, respectively).

subject to $p(B,m)\int_0^m B d\theta \leq T$ and $m \in [0,1]$.

Let $\eta_m := \frac{\partial p}{\partial m}\frac{m}{p(B,m)}$ be the take-up elasticity with respect to the eligibility threshold. The parameter η_m is the inframarginal effect, represented as an elasticity. We assume throughout that increases in m reduce costs, so $\frac{\partial p(B,m)}{\partial m} > 0$ for all B. For instance, raising the eligibility threshold might decrease stigma costs if stigma directly depends on the share of people who are eligible or take-up, as in Lindbeck et al. (1999).

Define η_B as the elasticity of take-up with respect to the benefit size, B: $\eta_B := \frac{\partial p(B,m)}{\partial B}\frac{B}{p(B,m)}$. Let $\gamma(B,m) := \frac{E[c|c<u(B),m]}{u(B)}$, noting $\gamma(B,m) < 1$. The parameter γ is the expected cost-benefit ratio conditional on take-up. It represents the share of the welfare gain from the benefit dissipated by the cost of taking up the benefit. For instance, if $\gamma = 0.5$, then costs represent half the utility gain (at $u(B)$).

Let $\lambda_{\text{avg}}(m)$ be the average welfare weight up to type m:

$$\lambda_{\text{avg}}(m) := \frac{\int_0^m \lambda_\theta d\theta}{\int_0^m d\theta} = \frac{\int_0^m \lambda_\theta d\theta}{m}.$$

Then the first-order conditions yield the following benchmark:

Proposition 2.1. *At an interior optimum, m and B satisfy:*

$$\overbrace{\frac{\frac{\lambda_m}{\lambda_{\text{avg}}}(1-\gamma)u(B)}{u'(B)}}^{\text{Welfare-weighted WTP of newly eligible}} + \overbrace{\frac{\frac{m}{p(B,m)}\int_0^{u(B)}\frac{\partial H(c|m)}{\partial m}dc}{u'(B)}}^{\text{WTP for lower } c} = \overbrace{\frac{B(1+\eta_m)}{(1+\eta_B)}}^{\text{Fiscal externality}}. \quad (2.6)$$

All proofs are in Appendix B.6.[43] Proposition 2.1 has familiar Baily (1978)-Chetty (2006) logic. At an optimum, the social planner equates the willingness to pay for a higher means test (the left-hand side) to its fiscal externality (the right-hand side). The willingness to pay for a higher means test combines: (i) the (welfare-weighted) utility gains of people who are newly eligible, and (ii) the utility gains from lower costs to previously eligible types who would have enrolled irrespective of the means test. The fiscal cost incorporates two fiscal externalities, one positive and one negative: raising the means test causes higher take-up from reduced costs, but it also causes lower take-up from the lower benefit amount given to each enrollee.

Notably, our model embeds Baily (1978)-Chetty (2006) logic in the context of a *redistributive* program, rather than as an analysis of social insurance against risk. Similar

[43]This statement refers to a necessary but possibly not sufficient condition for an interior optimum. We describe the statement in more detail in Appendix B.6.

intuitions appear regardless because the curvature of the utility function gives the planner a motive to smooth consumption across individuals.

Simple case: the means test does not affect stigma. Equation (2.6) nests the case where there are no inframarginal effects and $\eta_m = 0$. In that case, the planner seeks to equalize welfare-weighted marginal utility across people. Due to the concavity of the utility function, she does not give the entire budget to the lowest type so the solution is interior. On the other hand, as the welfare weight schedule is decreasing, the planner values the marginal utility of the lower types more than that of higher types. The solution will thus depend on the utility function's curvature as well as the schedule of welfare weights.

Stigma costs. As is standard, our optimality condition is governed by an envelope argument: people who take up the program due to a reduction in costs are just indifferent. They impose a fiscal externality because they take up the program, thus reducing how much the planner can transfer to others, but they experience no first-order utility gain. In this setting, the planner has an additional way to raise the utilities of people who always take up the program. She can reduce stigma by raising the eligibility threshold. Since these people are not indifferent, they do experience first-order utility gains. A change in the eligibility threshold itself also has first-order implications for social welfare, as those who are newly eligible enjoy the benefit of the program.

In this way, our model embeds a key trade-off in policies that reduce stigma either as an end goal or incidentally. On the one hand, reducing stigma can give a fiscal externality by raising take-up for people who do not value the program. But people who would take-up anyway will enjoy a first-order gain.

2.4.2 Incorporating Information Frictions

In this section, we present our main optimality condition. We now permit some share of consumers not to optimize. Assume share $s \in [0, 1]$ of consumers are "stigma(-only)" agents who behave as in the previous section. We introduce share $(1 - s)$ of consumers who suffer from optimization frictions: raising the eligibility threshold for these consumers raises take-up because it increases information. We call these consumers "information(-only)" agents. We assume that the probability of being a stigma-only agent is independent of m.

Let the take-up probability for stigma agents be p^s and for information agents be p^i.

For information agents, costs are distributed:

$$c = \begin{cases} \infty, \text{ with probability } 1 - p^i(m) \\ 0, \text{ with probability } p^i(m) \end{cases}, \tag{2.7}$$

for continuously differentiable $p^i(m)$. Put another way, information agents always participate if they know about the program. If they know about the program, the cost they face is 0 and they take it up. An agent's awareness does not depend on her type.

Let η^i_m and η^s_m represent the take-up elasticities with respect to the eligibility threshold for information and stigma agents, respectively. In Appendix B.6, we set up the planner's problem and obtain the following optimality condition:

Proposition 2.2. *At an interior optimum, m and B satisfy:*

$$\frac{\overbrace{\frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i} + \frac{1-s}{p^s}\right)u(B)}^{\text{Welfare-weighted WTP of newly eligible}} + \overbrace{\frac{1-s}{p^s}\eta^i_m u(B)}^{\text{WTP of info agents who now take-up}} + \overbrace{\frac{s}{p^i}\frac{m}{p^s}\int_0^{u(B)}\frac{\partial H(c|m)}{\partial m}dc}^{\text{WTP for lower } c}}{u'(B)}$$

$$= \frac{\overbrace{B\left[\frac{(1-s)}{p^s}\left(\eta^i_m + 1\right) + \frac{s}{p^i}\left(\eta^s_m + 1\right)\right]}^{\text{Fiscal externality}}}{\frac{p^s}{sp^s+(1-s)p^i}s\eta_B + 1}. \tag{2.8}$$

The key difference between Equation (2.8) and Equation (2.6) from the benchmark case is the appearance of the term $\frac{1-s}{p^s}u(B)$ on the LHS. This new expression gives the utility gains of the information-only agents who were previously eligible but learn about and join the program when the eligibility threshold increases. These information agents are *not* subject to an envelope condition like the stigma agents because they do not initially optimize. As a result, they would be willing to pay for the full benefit of the program, now that they know about it. Note that because we assume the distribution of stigma agents is independent of the welfare schedule, there is no welfare weight adjustment to the new inframarginal term. Other differences between the equilibrium conditions constitute simple rescaling factors to adjust for the share of the population that is affected by stigma and information.

If $s = 1$, Equation (2.8) nests Equation (2.6). Moreover, if $s = 0$ so all people are information agents, the distribution of stigma costs captured by γ and $\frac{\partial H}{\partial m}$ no longer enter the expression; in this case, since B does not affect the take-up rate for information agents,

η_B no longer enters the planner's optimality conditions. Then, the information-only case has the especially parsimonious expression:

$$\frac{u(B)\left(\frac{\lambda_m}{\lambda_{\text{avg}}}+\eta_m\right)}{u'(B)} = B\left(1+\eta_m\right). \tag{2.9}$$

The LHS encodes the welfare-weighted WTP for the newly eligible types and the WTP for the inframarginal types who now take-up. The RHS captures the fiscal externality from take-up.

Empirical implementation. In Appendix B.6, we show that a second-order Taylor expansion as in Gruber (1997) gives:

$$1+\frac{1}{2}\rho \approx \frac{\frac{(1-s)}{p^s}\left(\eta_m^i+1\right)+\frac{s}{p^i}\left(\eta_m^s+1\right)}{\left(\frac{p^s}{sp^s+(1-s)p^i}s\eta_B+1\right)\left(\frac{(1-\gamma)s}{p^i}\eta_m^s+\frac{1-s}{p^s}\eta_m^i-\frac{sm}{p^i}\frac{\partial\gamma}{\partial m}+\frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i}+\frac{1-s}{p^s}\right)\right)}, \tag{2.10}$$

for coefficient of relative risk aversion $\rho := -\frac{u''(B)}{u'(B)}B$.[44]

Equation (2.10) is the main condition that we examine empirically. Relative to Equation (2.8), this expression substitutes out the utility function $u(B)$ and derivative of the distribution of stigma costs with respect to the eligibility threshold $\frac{\partial H(\cdot|m)}{\partial m}$, which reduces the number of parametric assumptions we need to make. In their place, we add the risk aversion parameter ρ, which is more familiar to calibrate. The upshot is that we can take Equation (2.10) to the data by estimating η_m^i and η_m^s for a given social program. While estimating separate elasticities by type may seem daunting, Section 2.5 shows how the combination of our empirical approaches yields estimates of these parameters.

2.4.3 Policy Implications

We next derive sufficient conditions for when inframarginal effects unambiguously serve as a force to increase the eligibility threshold. Along the way, we derive an empirical test for whether the eligibility threshold is set suboptimally low. We proceed informally, to emphasize intuition, but present a formal treatment in Appendix B.7.

We study how the *naïve* planner's choice of (B, m) will differ from the *sophisticated* planner's choice. We define the naïve social planner as one who sets policy according to Equation (2.8) but: (i) erroneously believes that inframarginal effects arising from either

[44]We use the coefficient of relative risk aversion in B, evaluated at the sub-utility $u(B)$, since c is just an additive shifter and does not affect curvature.

agent are zero ($\eta^i_m = \eta^s_m = 0$), and (ii) does not realize that the eligibility threshold affects stigma. On the other hand, the sophisticated planner sets policy optimally according to Equation (2.8) and knows the true values of inframarginal effects.

In this section, we hold fixed the parameters $\{p^i, p^s, \lambda_\theta, s, \gamma, \eta_B, u(\cdot)\}$. We state some basic assumptions in Appendix B.7 that rule out edge cases. In the following comparative statics, we assume that the coefficient of relative risk aversion $\rho \geq 1$. We also employ the following non-trivial assumption that merits discussion:

Assumption 2.1. $\frac{\partial \gamma}{\partial m} = \frac{\partial (E[c|c<u(B),m]/u(B))}{\partial m} \leq 0.$

This assumption imposes that the average cost-benefit ratio, conditional on taking up the program, does not rise with a looser eligibility threshold. As m rises, stigma costs fall, which tends to reduce γ. On the other hand, new people may take up the program. Since they are nearly indifferent, they have relatively high draws of c, which raises γ. The assumption is true as long as the mass of just-indifferent people who newly sign up for the program as a result of reduced stigma costs do not raise the cost-benefit ratio more than the reduction in inframarginal stigma costs. For instance, in the case where costs are distributed uniformly, $\frac{\partial \gamma}{\partial m} = 0$.

Assumption 2.1 is sufficient but not necessary. In Appendix B.7 we give a substantially weaker but less concise necessary condition. We also prove that the assumption always holds for costs that are distributed normally or exponentially.

We then arrive at the following proposition.

Proposition 2.3. *Inframarginal effects raise the sophisticated planner's eligibility threshold relative to the naïve planner's eligibility threshold, if stigma types' take-up is less elastic to the threshold than information types' ($\eta^s_m \leq \eta^i_m$).*

Due to the planner's budget constraint, this proposition equivalently implies that, if the same hypotheses hold, the sophisticated benefit size B is smaller than the naïve benefit. The condition $\eta^s_m \leq \eta^i_m$ in Proposition 2.3 is sufficient but not necessary. There exist cases with stigma types who are more elastic than information types where the sophisticated eligibility threshold is larger than the naïve threshold.

Proposition 2.3 implies an empirical test for whether the eligibility threshold is unambiguously too low. The threshold value determining whether the statement is sharp is $\eta^s_m = \eta^i_m$: for any $\eta^s_m \leq \eta^i_m$, the naïve planner unambiguously sets the eligibility threshold too low. Accordingly, testing $H_0 : \eta^s_m > \eta^i_m$ permits the analyst to determine whether the eligibility threshold should optimally rise. If the test fails to reject that

$\eta^s_m > \eta^i_m$ then the threshold may still be set too low. But if the test does reject, then the normative conclusions are unambiguous if one accepts the assumptions in the model. We conduct this test in the following section.

We discuss both possible cases to aid intuition.

Example 2.1 ($\eta^s_m \leq \eta^i_m$). We first consider the case where inframarginal effects are driven by reduced information frictions, i.e. the hypotheses to Proposition 2.3 hold. Then, for a small change in the eligibility threshold, more people who take up capture the full benefit than people who take up and are just indifferent. The naïve planner thus employs a version of Equation (2.8) that unambiguously underestimates the welfare gains of a small increase in the threshold.

Example 2.2 ($\eta^i_m < \eta^s_m$). The policy implications in the second case are not sharp. There exist parameterizations in which the naïve planner sets the eligibility threshold too high or too low. To understand why, note that a high η^s_m introduces two forces. On the one hand, if just-indifferent stigma types are very sensitive to the eligibility threshold and many newly take-up, a small increase in the threshold introduces a large fiscal externality. The naïve planner will ignore this cost, a force pushing the sophisticated planner to lower the threshold. On the other hand, the fact that some stigma types are very sensitive implies that stigma types who always take up will enjoy large reductions in stigma from a small change in the threshold. Put another way, because many people react strongly to the threshold, that implies the threshold has a large effect on stigma. This logic of course requires a connection between the stigma gains from those who always enroll and the stigma gains from those who are just indifferent. That is precisely the role that Assumption 2.1 plays: it gives a sufficient value for how similar the stigma responses between those who always enroll and the indifferent types need to be.

To summarize, in this case, the naïve planner neglects two forces that accrue from reducing stigma. The net contribution of these forces (as well as the gains to information types) is unsigned.

We note that these normative conclusions are *not* sensitive to the share of stigma or information agents s. In fact, as $s \to 0$, Appendix B.7 shows that we can substantially relax Assumption 2.1. In the limit case where $s = 0$, neither of Assumption 2.1 or the condition $\eta^s_m \leq \eta^i_m$ is required at all (as is intuitive, since $\frac{\partial \gamma}{\partial m}$ vanishes from the optimality condition). Thus, the case with $s > 0$ is *conservative* for the model's normative conclusions. If all people are information types, then Proposition 2.3 holds under weaker conditions.[45]

[45]The statement that welfare analysis is conservative if $s = 1$ does *not* mean that the planner should

2.4.4 Discussion of Model Assumptions

Our framework yields a tractable benchmark for welfare analysis that we can take to the data. Even so, it involves several stark assumptions.

Lump-sum benefits. Many social programs, including SNAP, have non-linear benefits schemes that vary based on income and household size. If the planner could give non-linear benefits, she might extend a small benefit to a larger share of people, to reduce stigma costs and boost take-up without incurring as large a fiscal externality. Our model abstracts from this choice, but we view B as representing the (appropriately weighted) average benefit given to inframarginal types. Relatedly, we assume that people are perfectly informed about the benefit to which they are entitled.[46]

Campaigns to inform or destigmatize. Our model does not feature an instrument by which the planner can spread information about SNAP or reduce the stigma of SNAP directly. Even if the planner has other means of spreading information or reducing stigma (i.e., the eligibility expansion is not the most effective way to do so), the model highlights that eligibility expansions could nevertheless affect information and stigma. The planner must choose *some* eligibility threshold for her means-tested program. She must contend with the trade-offs inherent in setting the threshold.

Identical take-up probabilities. If in fact p varies with θ, it is possible to undo some of our normative conclusions. For example, suppose most of the increase in take-up from inframarginal effects is concentrated in types for whom λ_θ is small. Then inframarginal effects can yield a smaller transfer to the types for whom λ_θ is large. A fruitful extension of the model could consider different take-up probabilities.

increase the eligibility threshold by a greater amount as $s \to 0$. It means that Proposition 2.3 holds without Assumption 2.1. Proposition 2.3 deals with infinitesimal changes in the eligibility threshold. Analyzing non-marginal changes requires more structure, which we develop in Section 2.5. Moreover, if $s = 1$ and the reduction in always-takers' stigma costs are large, then that serves as another motive to increase the eligibility threshold.

[46] An alternative model, as in Finkelstein and Notowidigdo (2019), casts information frictions as a noisy (mis)perception of benefits. Even if misperceptions are symmetric, correcting them can still increase take-up in our model, since benefits enter a concave utility function. The welfare implications of this model are different: the utility gain to the newly enrolled inframarginals is bounded above in relation to the size of the misperception, while the previously enrolled do not gain.

2.5 Welfare Analysis

2.5.1 Set-Up

In this section, we combine the model and empirics to show that under reasonable assumptions, inframarginal effects meaningfully increase the optimal eligibility threshold. To make this point, we compare the optimal means test under the naïve social planner, who sets policy optimally but erroneously believes $\eta_m = 0$, to that under a sophisticated social planner who understands that $\eta_m > 0$. The model does not capture every relevant economic force, so we view this welfare analysis as illustrative.

First, we implement the empirical test we proposed in Section 2.4. This test signs whether inframarginal effects should cause the eligibility threshold to rise. We then extend our analysis to quantify how much the inframarginal effects we measure should affect optimal policy. We implement a novel method to quantify the planner's mistake using only local policy analysis. Finally, we impose more structure to make global claims about the optimal eligibility threshold.

2.5.2 Decomposition

A takeaway from our model is that welfare effects of the eligibility threshold depend on the mechanism underlying inframarginal effects. Figure 2.6 suggests that, although increasing the eligibility threshold appears to reduce stigma, this effect does not drive the results in Section 2.2. We use the model and our empirical estimates to decompose inframarginal effects between information and stigma. This decomposition quantifies the mechanisms underlying inframarginal effects and is therefore useful in its own right. Moreover, it gives estimates of η_m^i and η_m^s, the inframarginal take-up elasticities for information and stigma types. With these in hand, we can directly implement our empirical test, proposed in Section 2.4, of whether inframarginal effects should rise.

The key piece of model structure that we leverage is that all agents are either stigma or information types. In that case, it is an identity that:

$$\frac{\partial p}{\partial m} = s \times \frac{\partial p_s}{\partial m} + (1 - s) \times \frac{\partial p_i}{\partial m}, \tag{2.11}$$

and manipulations give:

$$\frac{\partial \ln \left(\text{Number Enrolled}\right)}{\partial m} = s\frac{1}{p}\left(\frac{\partial p_c}{\partial c}\frac{\partial c}{\partial m}\right) + \frac{1}{p}(1-s) \times \frac{\partial p_i}{\partial m} + \frac{\partial \ln \left(\text{Number Eligible}\right)}{\partial m}. \tag{2.12}$$

That is, the increase in take-up after a change in the means test can be decomposed into the increase in take-up among stigma agents (mediated by their change in stigma costs) and the increase in take-up among info agents. We are able to estimate a demographic-cell level version of Equation (2.12) by combining our various datasets. Specifically, we estimate:

$$\frac{\partial \ln(\text{Number Enrolled})}{\partial m}_d = \frac{1}{p}\left(\frac{\partial c}{\partial m}_d s_d\right) \times \beta^s + \frac{1}{p}(1-s)_d \times \beta^i + \epsilon_d, \tag{2.13}$$

which replaces unobserved terms in Equation (2.12) with coefficients to be estimated, β^i and β^s.[47] For each demographic cell d, we estimate $\frac{\partial \ln(\text{Number Enrolled})}{\partial m}_d$ using the QC data and instrumental variables regressions analogous to Equation (2.3). We estimate the cell-level effect of changing eligibility on cost $\frac{\partial c}{\partial m}$ using the second-order stigma results from the experiment. We extract cell-level values of s using the FSPAS.

Noting that $\beta^s = \frac{\partial p^s}{\partial c}$ and $\beta^i = \frac{\partial p^i}{\partial m}$, Equation (2.13) permits us to recover estimates of elasticities $\hat{\eta}^s_m$ and $\hat{\eta}^i_m$. We provide more estimation details (including details of bootstrapping the estimates on both the right- and the left-hand sides of the equation and jointly weighting by cell sizes across datasets) in Appendix B.4.

This exercise yields that the inframarginal effect principally arises from information frictions, rather than stigma (Table 2.5). We are unable to reject the null that $\beta^s = 0$ (Row 1) but robustly reject that $\beta^i = 0$ (Row 2). Combining these parameters with our other empirical estimates, we have $\eta^s_m \approx 0$ (Row 3), depending on the specification, and $\eta^i_m > 0$ (Row 4).

The upshot of conducting the formal decomposition is that it gives the machinery to test $H_0 : \eta^s_m > \eta^i_m$ empirically. We conduct a (conservative) two-sided test of the null that $\eta^s_m = \eta^i_m$, and we reject this null at the 5% significance level in all specifications (Row 6). This implies that $\eta^s_m \leq \eta^i_m$. Therefore, Proposition 2.3 holds: in our setting inframarginal effects imply that the eligibility threshold is too low, assuming the social planner does not presently account for them when setting SNAP's eligibility threshold.

[47]Equation 2.13 also uses the assumption that $\frac{\partial \ln \text{Number Eligible}}{\partial m} = 0$, if our identifying variation is valid; put another way, changing m should not change the number of *inframarginals* who are eligible.

2.5.3 Calibration

Next we calibrate the parameters necessary to quantify the implications of inframarginal effects. Informed by the evidence in the previous section, we henceforth assume that information frictions are the dominant mechanism underlying inframarginal effects.[48] Additionally, for mathematical simplicity, we also assume that at the optimum, $p^i = p^s$. We state the assumptions as follows:

Assumption 2.2. At the planner's solution, $p^i = p^s$ and $\frac{\partial p^s}{\partial m} = 0$.

From Assumption 2.2, we proceed using $\eta^i_m = \eta_m(1-s)^{-1}$. We use the full-sample estimate of the effect of eligibility expansions on the 0–130% take-up rate from Table 2.4.

Now, we discuss calibration of other parameters, summarized in Table 2.6. As is common in welfare analysis, some of the economic primitives have a high degree of uncertainty. As a result, we show robustness to the particular choice of parameter.

Cost-benefit ratio of taking up the program (γ). Although our evidence suggests that information frictions are the dominant mechanism underlying inframarginal effects, reductions in stigma costs among eligible people who would have taken up SNAP regardless can still confer welfare gains when the means test increases. Thus, we need some assumptions on stigma costs among stigma agents. We choose a conservative assumption which simplifies the analysis: $c \sim U[0, \overline{A}(m)]$ for $\overline{A} > u(B)$.[49] We then have that $\gamma = \frac{1}{2}$, as $E[c|c < u(B), m] = \frac{1}{2}u(B)$.

With this assumption, $\frac{\partial \gamma}{\partial m} = 0$: no welfare gains accrue to inframarginal stigma agents. We see this assumption as therefore being conservative: if inframarginal effects also deliver utility to inframarginal types who already take up the program, then that only increases the planner's motive to raise the eligibility threshold. Intuitively, this assumption lets us avoid needing to compute enrollees' willingness to pay for reduced stigma — an interesting and policy-relevant parameter that future work should explore.

Take-up elasticity with respect to benefit size (η_B). The SNAP benefits schedule B is set nationally. As a result, we cannot use an event-study design to estimate η_B. Instead, we collect estimates of the typical elasticity of take-up with respect to benefit size for related programs. Krueger and Meyer (2002) review papers estimating η_B for UI and

[48]Note that this is different from saying that all agents are information agents; we continue to allow some share of the population $s \in [0, 1]$ to face stigma costs, but we assume that these agents' take-up decision does not respond to changes in the means test.

[49]Economically, this assumption posits that: (i) changing m does not change the shape of the cost distribution, and (ii) there exist people for whom the take-up cost exceeds the utility gain.

worker's compensation and conclude that, for these programs, η_B ranges from 0.3 to 0.6. We choose $\eta_B = 0.5$ as a sensible midpoint and show robustness to other values.[50]

Other parameters. We use $\rho = 3$ as a benchmark.[51] For the share of stigma types s in the population — informed by our analysis of the FSPAS — we use $s = 0.4$ as a benchmark. For the means test m, we take the population-weighted average of states' share eligible across years (accounting for varying eligibility thresholds) to obtain $m^* = 0.27$ in 2016. For the take-up probability p^*, our data from 2016 suggest the take-up probability is $p^* = 0.53$.[52]

2.5.4 Local Policy Analysis

A standard problem with conducting empirical analysis of social optimality conditions is that if one rejects that the optimality condition exactly holds, it is difficult to estimate the magnitude of the planner's mistake. For concreteness, imagine one has data to statistically reject that the LHS and RHS of a standard Baily (1978)-Chetty (2006) condition exactly coincide in the analysis of a given social insurance program. Is the planner's mistake large or small? A typical approach is to impose structure so that the researcher can extrapolate agents' behavior away from equilibrium; we will take this approach in the next section. First, we propose a new method for estimating the size of the planner's mistake that uses only the local optimality conditions. The advantage of this approach is that it does not require extra parametric structure. The disadvantage is that it does not permit making policy recommendations like what the optimal eligibility threshold should be.

In short, we estimate the magnitude of the naïve planner's mistake by studying the implied value of ρ that would be required to make the current means test optimal (when inframarginal effects are present). We establish that, if the planner assumes $\eta_m = 0$ but otherwise optimizes according to the theory, she will treat people as if they are much less

[50]Kroft (2008) also cites the Krueger and Meyer (2002) review and uses $\eta_B = 0.5$. Auray and Fuller (2020) is an example of a recent paper that finds a similar η_B in later years. In their data from 2002–2015, $\eta_B = 0.63$ (SE: 0.23), where η_B is the elasticity of UI take-up with respect to the replacement rate.

[51]Chetty and Finkelstein (2013) note that this parameter is notoriously difficult to calibrate, but review other papers that test values of $\rho \in [1, 4]$ (e.g., Gruber, 1997).

[52]This number is below the number the USDA reports because our denominator includes some people who are not eligible for SNAP due to work requirements, asset thresholds, or other tests; moreover, it is not clear that the USDA number includes people with incomes above 130 if they live in states with an eligibility threshold beyond 130. We use the number for illustrative purposes in this exercise, but the results are not sensitive to adjusting the equilibrium p^*.

risk averse than they really are.[53]

Our approach is as follows. We assume some ground-truth value of ρ, say $\rho = 3$. Consider the naïve planner who chooses m and B to solve Equation (2.10) assuming $\eta_m = 0$ and that $p^i = p^s$. Given $\eta_m \neq 0$, what is the implied $\tilde{\rho}$ that keeps the optimality condition equated? We use the following algorithm:

1. Obtain inverse-optimum weights: Assuming $\eta_m = 0$, solve for $\lambda_m/\lambda_{\text{avg}}$ that satisfies Equation (2.10).

2. Obtain implied $\tilde{\rho}$: Given the inverse-optimum weights $\lambda_m/\lambda_{\text{avg}}$ and *true* value of η_m, solve for the $\tilde{\rho}$ that satisfies Equation (2.10).

Intuitively, because the planner ignores η_m, she treats people "as-if" they have risk aversion $\tilde{\rho}$, when they really have risk aversion ρ. Put another way, there is some value of $\tilde{\rho}$ that satisfies the optimality condition even under the (incorrect) assumption that $\eta_m = 0$. We focus on the value of $100 \times \frac{\tilde{\rho}-\rho}{\rho}$, which is a measure of the *bias* in the coefficient of relative risk aversion.

Results. We show that the magnitude of the planner's mistake can be substantial for the range of η_m that we estimate (Figure 2.7). The x-axis plots values of η_m. On the y-axis, we plot the bias in the "as-if" risk aversion parameter relative to the true risk aversion parameter, assuming $\rho = 3$. If $\eta_m = 0$, there is no bias: the naïve and sophisticated solutions coincide by construction. As η_m grows, the bias rises; for $\eta_m = 0.05$, the bias is about 10%. For our primary estimate of $\eta_m = 0.13$, we find that the bias can be quite large: the naïve planner's solution will treat people as if they are about 30% less risk averse than they really are. We show robustness to parameterizations in Appendix B.5.

Intuitively, the planner who ignores inframarginal effects transfers too much to inframarginal types who already take up the program. She overvalues inframarginal types' marginal utility and undervalues the gain in utility from those who would take up the program if she raised the eligibility threshold. As a result, she optimizes as if the coefficient of relative risk aversion were smaller than it really is.

MVPF. Another approach to gauging the size of the naïve planner's mistake with limited parametric assumptions is the Marginal Value of Public Funds (MVPF) (Hendren and Sprung-Keyser, 2020). We study the MVPF of an eligibility expansion in Appendix

[53]We use the standard interpretation of ρ as risk-aversion. However, it also corresponds to the planner's unweighted valuation of transferring B to someone who is ineligible from someone who takes up (has a benefit of B). To see this, note that $\rho = -B\frac{u''(B)}{u'(B)} \approx \frac{u'(0)-u'(B)}{u'(B)}$.

B.5. We document the potential for the naïve planner to have substantial bias in her estimate of the MVPF.

2.5.5 Global Policy Analysis

In this section, we impose structural assumptions to extrapolate take-up probabilities and welfare weights away from what we observe in equilibrium. Appendix B.5 provides details on our parameterizations. With these assumptions, we numerically invert Equation (2.10) to solve for the optimal m^{opt} and B^{opt} as a function of η_m.

2.5.5.1 Results

We first ask how much larger is m^{opt} relative to today's $m^* = 0.27$. Because we use the inverse optimum approach to calibrate the welfare weights, Proposition 2.3 guarantees that the optimal m^{opt} exceeds today's m^*: $m^{\text{opt}} - m^* > 0$.

We present the percent increase in m^{opt} relative to m^*, i.e. percent increase $:= 100 \times \frac{m^{\text{opt}} - m^*}{m^*}$. This value represents the percent increase in the optimal eligibility threshold relative to today's threshold. Because the threshold is measured in terms of the share eligible, it equivalently represents the percent increase in the share of people who should be eligible relative to today.

We present our estimates of the percent increase in m as a function of η_m (Figure 2.8A) for both $s = 0.4$ (black line) and $s = 0.8$ and $s = 0$ (gray dashed lines). By construction, if $\eta_m = 0$, we find the optimal eligibility threshold coincides with today's threshold. As η_m rises, the optimal m rises too. At our preferred value of $\eta_m = 0.13$, about 13% more people should be eligible than are eligible today.

We also show the optimal take-up rate (Figure 2.8B). Because of our social planner's fixed budget, the optimal take-up rate is not monotonic: increases in m require decreases in B, which, through η_B, decrease take-up. At some point, however, take-up falls enough that those on the program are granted larger B, and take-up begins to rise again. This dynamic does not exist for $s = 0$ since η_B has no effect for information agents.

While Panel A and Proposition 2.3 show that the sophisticated planner will expand eligibility beyond today's m, Panel B highlights that we cannot conclude that take-up today is suboptimally low.[54] The naïve planner erroneously believes take-up will fall *more than it actually would* for a small increase in m because she does not account for η_m

[54] Note that take-up depends on the benefit size as well as the eligibility threshold, so the higher threshold does not necessarily imply higher take-up on net.

and only accounts for η_B. She therefore sets m too low. We show similar conclusions for $\rho \in \{1, 2, 4\}$ (Figure B.15A) and $\eta_B = 0.3$ (Figure B.15B). Notably, the magnitude of ρ does not have a large effect on the percent change in m. In this setting, η_m is much more important than ρ.

The planner has a fixed budget, so the increase in the optimal eligibility threshold and small change in optimal take-up rates imply that the optimal benefit is decreasing in the inframarginal effect, even for non-local changes (Figure B.16). For various s, at our preferred estimate of η_m, the optimal benefit is 5–10% lower than the current optimum.

A weakness of our numerical approach is that we assume that $\rho = 3$ both in today's equilibrium and also at an optimum, but that third-order utility terms vanish (in order that Equation (2.10) holds). We conduct a second exercise where we assume a quadratic utility function that imposes that $\rho = 3$ at today's B (Figure B.17).[55] This exercise gives similar results, although the magnitudes of the increase in eligibility are attenuated because risk aversion changes rapidly for quadratic utility.

2.6 Conclusion

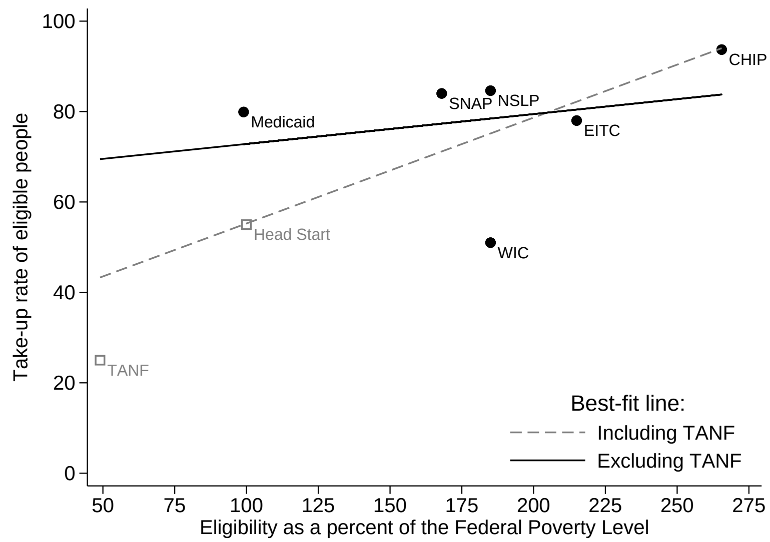
This paper documents the existence of inframarginal effects in SNAP. We find that the inframarginal effects arise from increased information after states relax eligibility thresholds, but our online experiment also finds that relaxing eligibility thresholds can reduce stigma. We develop a general model for incomplete take-up of social welfare programs when the planner can control program eligibility. We apply our model to SNAP and assess the implications for the optimal eligibility threshold given the inframarginal effects. Because the information mechanism dominates, inframarginal effects unambiguously increase optimal SNAP eligibility.

All social programs, even universal ones, make some determination about eligibility. This threshold is often chosen by the planner and thus is not an exogenous feature of the policy environment. As a result, our normative insights have applications in many areas in public economics. When inframarginal effects are present, our theoretical framework highlights that they may serve as a motive to raise the eligibility threshold. Future work could enrich the model to include a larger set of policy instruments and more heterogeneity in individual responses.

[55] Together with $u(0) = 0$, this assumption yields that utility is: $u(B) = -(B-k)^2 + k^2$ for $k := \frac{\rho+1}{\rho} B^*$.

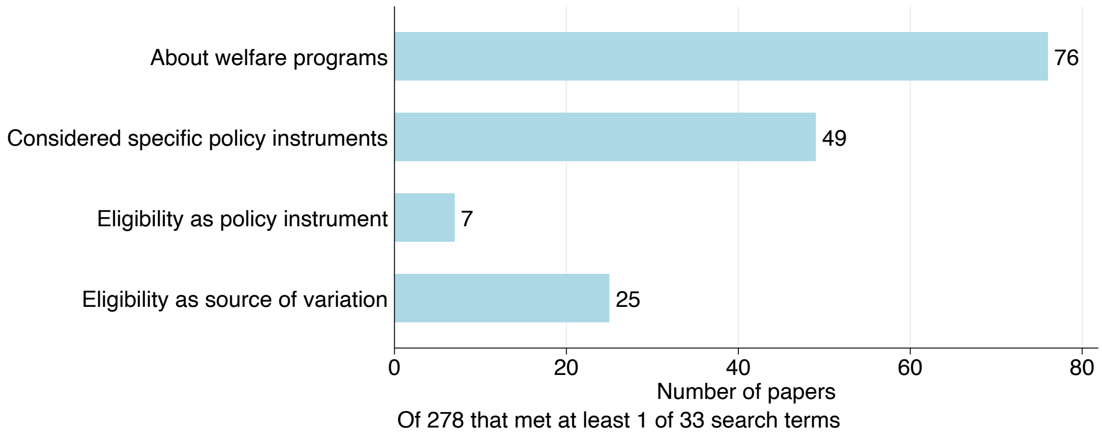
2.7 Figures

Figure 2.1: Eligibility Thresholds and Program Take-Up



The figure shows income eligibility thresholds as a percent of the Federal Poverty Level (FPL) for the largest U.S. means-tested social programs against estimates of their national take-up rates, compiled from different sources. We plot TANF and Head Start in a separate series because eligibility and take-up rates for these programs are particularly difficult to estimate; see Appendix B.1 for information on constructing these data. Take-up rates are estimated out of the eligible population for each program. In programs with different eligibility thresholds per state, the level plotted is the population-weighted average of those thresholds. The SNAP take-up rate displayed here is higher than that used in our paper because the USDA uses a more involved and restrictive method for assessing eligibility than we do; our empirical results are not affected by a denominator that is too large. Where the eligibility threshold is defined in dollars (e.g., EITC, TANF), the figure shows the threshold as in terms of percent of the FPL for a family of three. Some programs (e.g., WIC, TANF) are restricted to certain subgroups in addition to imposing income thresholds — for example, families only — or have additional requirements. Given Head Start's capacity constraints, additional assumptions were made to estimate a take-up rate. These are also documented in Appendix B.1.

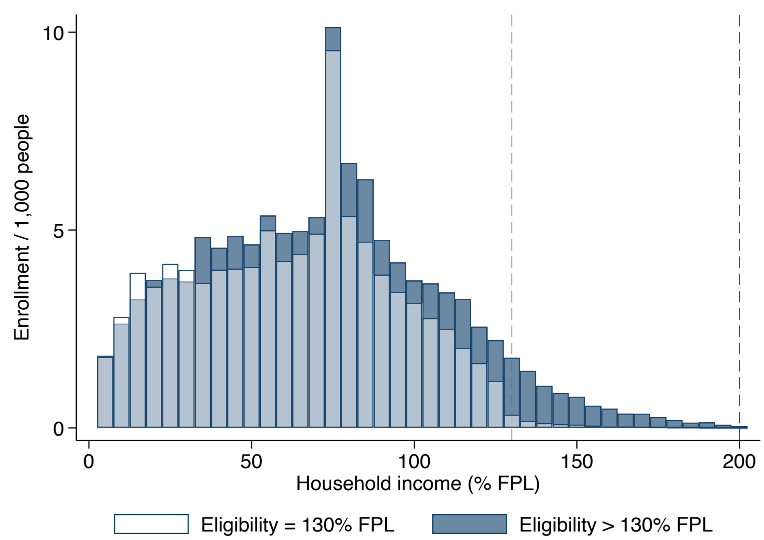
Figure 2.2: Literature Review: *AER* and *QJE* papers about Eligibility Criteria



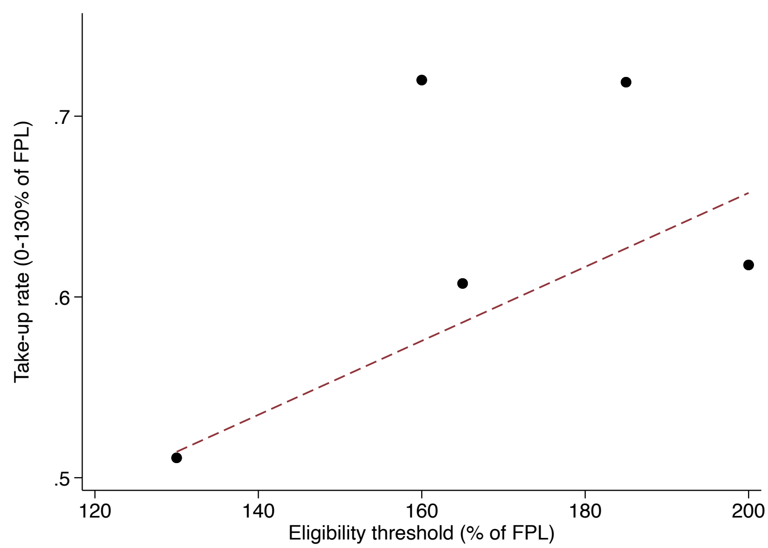
The figure presents the results from our literature review of papers in the *Quarterly Journal of Economics* (2010–2019) and the *American Economic Review* (2010–2018). Appendix B.1 provides details about the sampling frame. The first row shows the total number of papers that we concluded were about welfare programs, after reading the abstract and introduction. The second row shows the number of papers that considered instruments with which the planner could enact optimal policy, e.g. the benefit size or duration. The third row shows the number of papers that considered the eligibility threshold as an instrument with which the planner could enact optimal policy. The fourth row shows the number of papers that use the eligibility threshold as a source of variation with which the authors estimated a treatment effect for the program.

Figure 2.3: Descriptive Evidence of Higher Inframarginal Enrollment with Expanded Eligibility

(a) Enrollment by Household Income and Eligibility Threshold



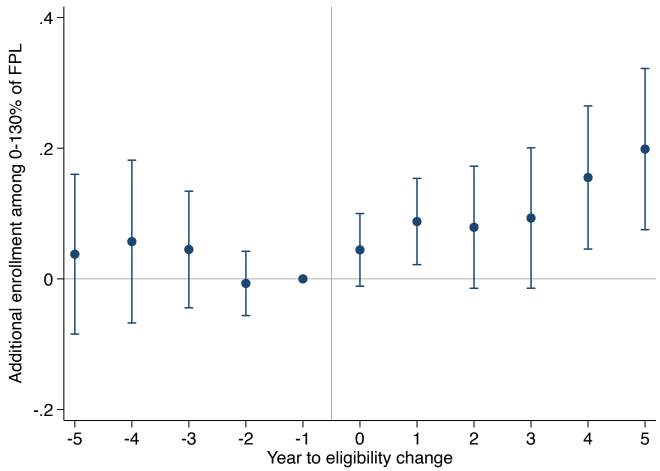
(b) Take-up by Eligibility Threshold



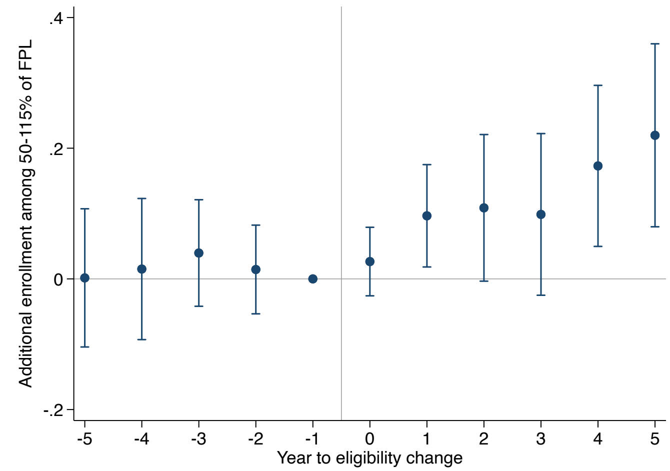
This figure presents the relationship between the eligibility threshold and SNAP take-up and enrollment. Panel A shows SNAP enrollment per 1,000 people in states and years where the eligibility threshold is 130% of the Federal Poverty Level (FPL) versus above 130% (and up to 200%). Each bar takes the number of people in the USDA Quality Control data whose household income is in each income bin, divided by the total population (i.e., all people, with any household income) in all state-years with the indicated eligibility regime. The data are limited to the sample we use in the main event study, and household income is top-coded at 200% FPL. Panel B shows average take-up among those earning 0–130% of the FPL in states with each eligibility threshold observed in the data. The USDA Quality Control data provide estimates of the numerator for the outcome (take-up counts, by state-year), and the Current Population Survey data provide estimates of the denominator (total counts of individuals within this sample).

Figure 2.4: Event Study of Changes to Eligibility Threshold

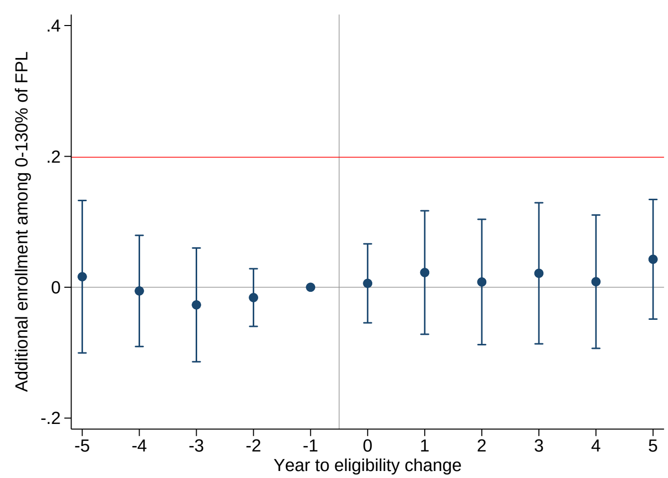
(a) Sample: 0 to 130% of FPL



(b) Sample: 50 to 115% of FPL



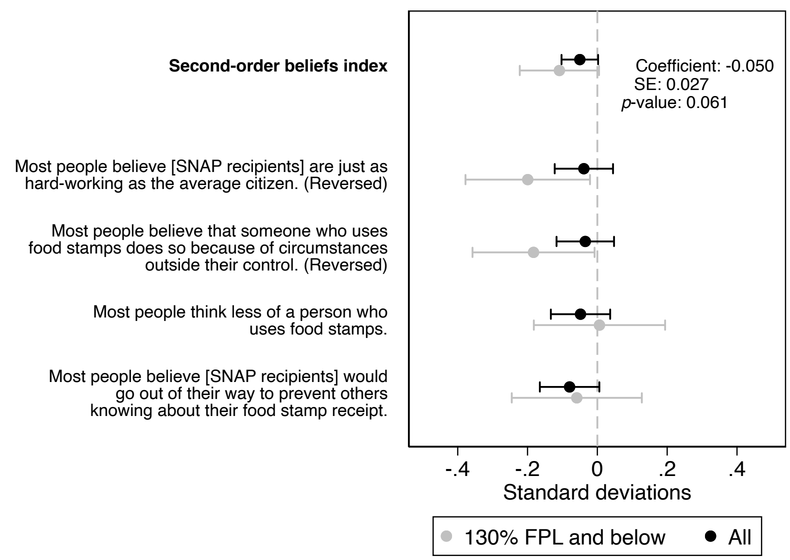
(c) Placebo: BBCE States that Do Not Expand Eligibility



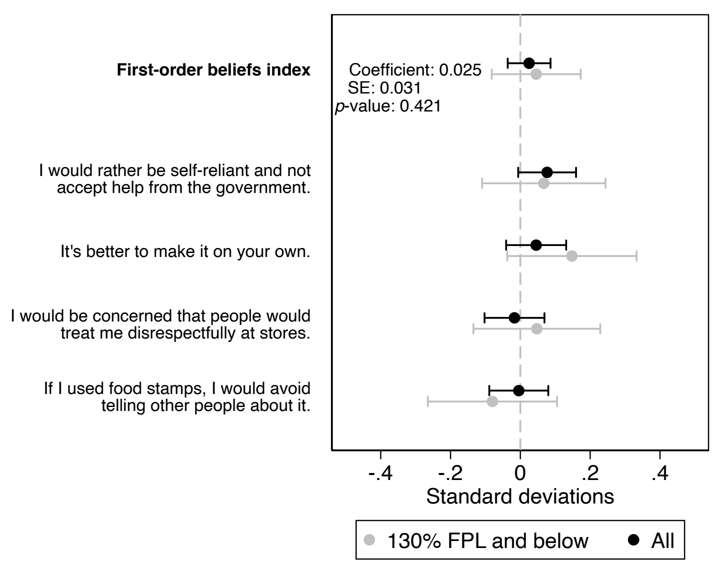
This figure presents the event-study estimate of η (Equation (2.1)), the effect of the eligibility rate on inframarginal take-up. Panel A presents results for the sample of individuals from 0–130% of the Federal Poverty Level (FPL); Panel B presents results for 50–115% of the FPL. Panel C presents a placebo event study, using the nine states that adopt the Broad Based Categorical Eligibility policy but do not expand eligibility (see Section 2.2). The red line in Panel C plots the 5-year point estimate from Panel A. The minimum eligibility in all states is 130% of the FPL. Standard errors are robust to heteroskedasticity and clustered by state.

Figure 2.5: Effect of High-Share Treatment on Stigma

(a) Second-order beliefs



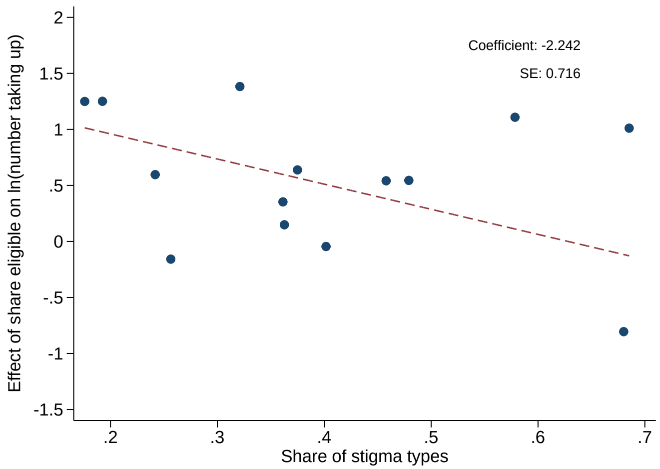
(b) First-order beliefs



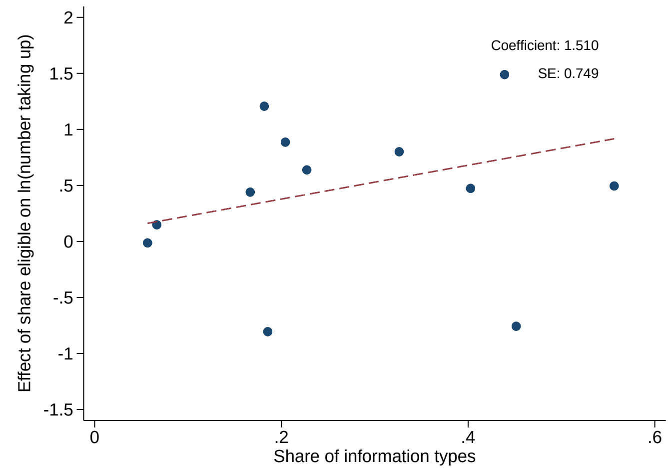
This figure presents results from the online experiment; it shows the effect of the "high-share" treatment (where respondents were randomly given a hint that *increased* their reported beliefs about the share of Americans who are eligible for SNAP) on agreement with each statement in the stigma instrument (Equation (2.4)). Outcomes marked with "(Reversed)" were reverse-coded so that for all items, a higher score indicates more stigma. The coefficients correspond to a reduced-form (intent-to-treat) estimate and do not account for the amount by which the treatment moved people's beliefs about the share of Americans who are eligible for SNAP. Each outcome is in units of standard deviations, and the indices average the set of outcomes displayed in each panel. Bars plot 95% confidence intervals.

Figure 2.6: Inframarginal Effects Heterogeneity by Demographic Cell

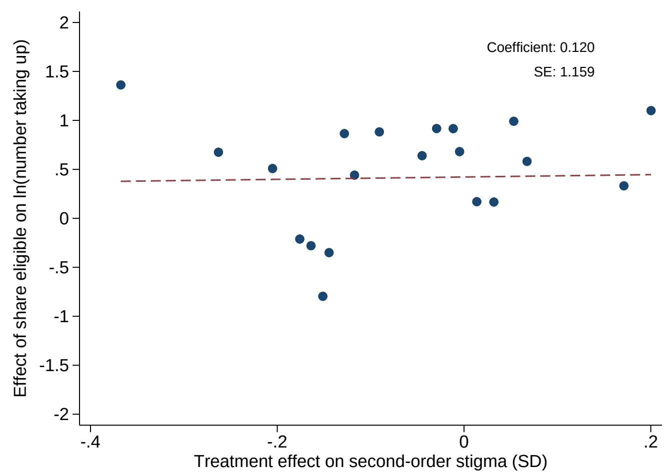
(a) Share Stigma Type vs. Inframarginal Effects



(b) Share Information Type vs. Inframarginal Effects

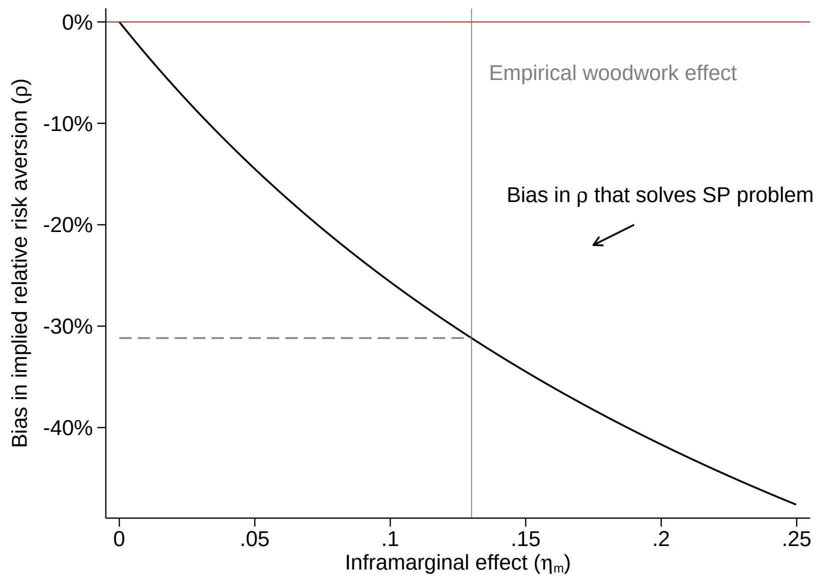


(c) Experiment Treatment Effect vs. Inframarginal Effects



Panels A and B show the correlation between the subgroup-specific inframarginal effects with the share of respondents in the USDA FSPAS survey who reported (A) any stigma and (B) less than complete information. Panel C shows the correlation between the subgroup-specific inframarginal effects and the subgroup-specific treatment effect in the online experiment. Subgroups are defined by household head age bin, gender, and race/ethnicity (non-Hispanic white vs. other), as well as by their household composition and income decile in the national distribution. Estimates are weighted by the inverse of the product of the variances of the cell-level coefficients; see Appendix B.4 for details.

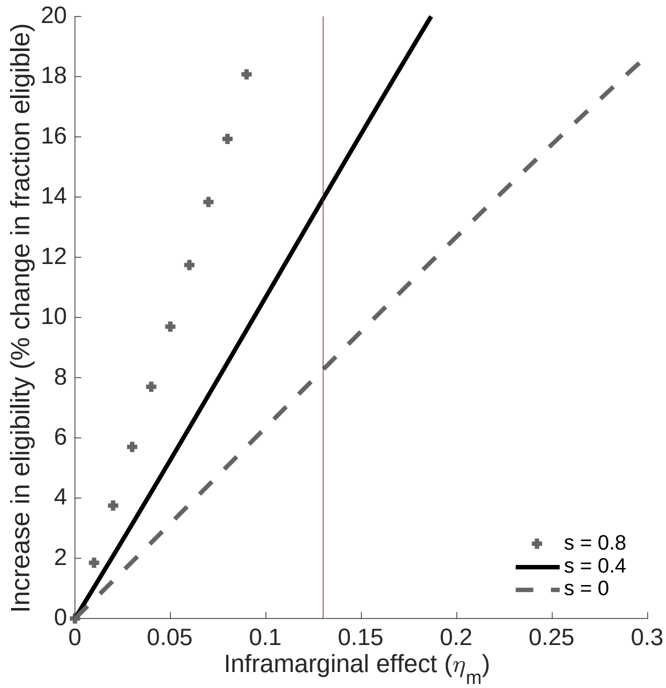
Figure 2.7: Naïve Planner's Biased Risk Aversion



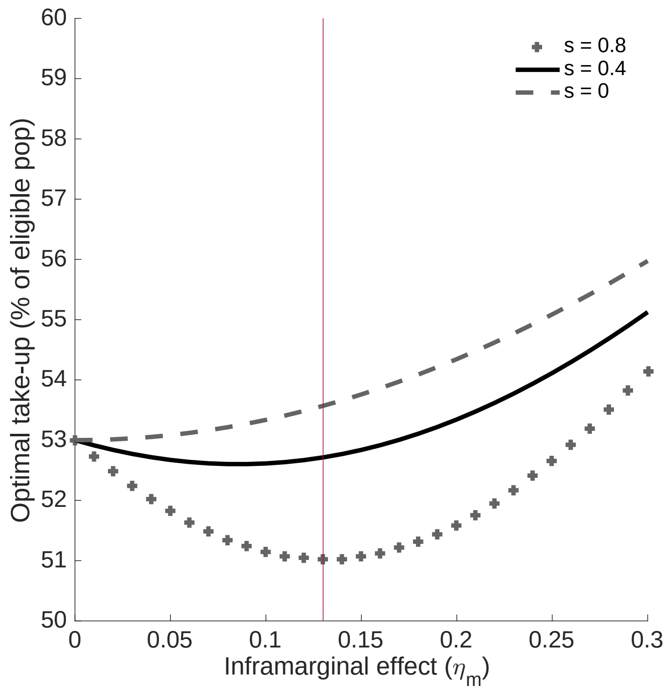
This figure shows the percent bias between the planner's "as-if" risk aversion ($\tilde{\rho}$) and the ground-truth risk aversion (ρ) (black line). Negative numbers indicate that the planner is behaving as if people are less risk averse than they really are. Panel A plots the bias as a function of the inframarginal effect; the vertical gray line plots the empirical inframarginal effect presented in Table 2.6. Panel B fixes η_m at the empirical inframarginal effect from Table 2.6 and varies s, the share of stigma agents.

Figure 2.8: Numerical Simulations: Optimal Eligibility Threshold and Take-Up

(a) Optimal Eligibility Threshold vs. Inframarginal Effects



(b) Optimal Take-Up Rate vs. Inframarginal Effects



This figure shows the results from our numerical simulation exercise: it presents the change in the percent of people who are eligible relative to current policy if the planner were to acknowledge inframarginal effects (Panel A) and the optimal take-up rate (Panel B), as a function of the inframarginal effect η_m, using our preferred optimality condition (Equation (2.10)). Auxiliary parameters are set according to the values in Table 2.6.

2.8 Tables

Table 2.1: Estimates of the Inframarginal Effect

	(1) Main estimate	(2) Extra controls	(3) Waivers, lag unemp.	(4) Excludes recession	(5) Weighted	(6) Avg of coefficients	(7) All data
Panel A. 0–130% FPL							
Income limit (% FPL) / 100	0.085	0.087	0.074	0.086	0.082	0.076	0.091*
	(0.056)	(0.055)	(0.055)	(0.059)	(0.072)	(0.064)	(0.048)
Panel B. 50–115% FPL							
Income limit (% FPL) / 100	0.107**	0.112**	0.097*	0.114**	0.116*	0.103*	0.121**
	(0.051)	(0.054)	(0.050)	(0.053)	(0.064)	(0.056)	(0.047)
Observations	705	705	680	628	705	705	1071
N states	45	45	45	45	45	45	51

This table shows the effect of the eligibility threshold on log enrollment among the inframarginal population (0–130% FPL in Panel A and 50–115% FPL in Panel B). Column 1 estimates Equation (2.2), and the following columns present various extensions to show robustness. Column 2 separates the Ganong-Liebman policy index into separate indicators. Column 3 includes a control for the previous year's unemployment rate in each state and a control for the population-weighted average number of months a state had ABAWD work requirement waivers in effect. Column 4 excludes years 2008–2011, during the Great Recession. Column 5 weights observations by the state-year population. Column 6 presents the difference between the average pre- and post-period event study coefficients. Finally, Column 7 estimates Equation (2.2) using all the data available instead of only the event study sample. Standard errors are robust to heteroskedasticity and clustered by state.*, **, and *** indicate $p < 0.1$, 0.05, and 0.01, respectively.

Table 2.2: Effects on Demographic Composition (50–115% FPL)

	(1) Female	(2) Black	(3) Age	(4) Has child	(5) Avg net income	(6) % FPL	(7) Certification $\leq$ 6 mo.
Income limit (% FPL) / 100	-0.001	0.059	0.391	-0.002	-28.557	0.732**	0.013
	(0.004)	(0.064)	(0.420)	(0.010)	(20.033)	(0.299)	(0.105)
Baseline mean	0.59	0.22	28.94	0.71	817.41	79.62	0.40
Observations	705	705	705	705	705	705	705
R^2	0.70	0.81	0.85	0.84	0.89	0.70	0.67

This table presents results from estimating the effect of the SNAP eligibility threshold on the composition of enrollees earning 50–115% FPL. The columns present estimates of Equation (2.2) with the indicated outcome variable. The independent variable is the eligibility threshold as a ratio of the Federal Poverty Level, so that increasing by 1 corresponds to increasing the eligibility threshold from, e.g., 130% FPL to 230% FPL. In each column, we use the specification described in Equation (2.2), where the outcome is indicated by the column header: Column 1 shows the effect of the eligibility threshold on the fraction of the 50–115% FPL enrollee sample who are female, and so on. "Baseline mean" refers to the average of the outcome indicated by the column in state-years where the eligibility threshold is 130% FPL. Outcomes are calculated using the USDA's Quality Control (QC) data, limiting the data to households earning 50-115% FPL. Standard errors are robust to heteroskedasticity and clustered by state. *, **, and *** indicate $p < 0.1$, 0.05, and 0.01, respectively.

Table 2.3: Cost-Effectiveness Calculation

	Eligibility threshold	**Benefit size**
Change required for 1 pp take-up increase	3.9 pp	\$56 per person-year
1. Number of people affected	12 million (newly eligible)	44 million (enrolled)
2. Take-up among affected people	25%	100%
3. Cost per person-year	\$707	\$56
4. **Mechanical cost of intervention** (= Row 1 $\times$ Row 2 $\times$ Row 3)	\$2.2 billion	\$2.5 billion

This table shows the cost-effectiveness of increasing take-up by raising the means test versus by increasing the benefit size. Let m be the share of the U.S. population eligible for SNAP, B be the benefit size per person, and p be the take-up probability. The top row shows the required change in the instrument (m or B) to achieve a one percentage-point increase in take-up. We calculate this row by noting that $\eta_m = \frac{dp}{dm}\frac{m}{p}$ and rearranging to solve for dm when $dp = 0.01$ (and likewise for B). The remaining rows show the mechanical cost to the program (without including the costs incurred by inframarginal effects) of changing these instruments. When using the means test m to increase take-up, 12 million more people become eligible, but we estimate only 25% of those would take-up. When using the benefit size B, benefits are increased for all program participants. The cost per person uses averages from the QC data. The final row of the table shows the total mechanical cost for each policy tool, which multiplies rows 1-3.

Table 2.4: Estimates of the Take-up Elasticity with Respect to Eligibility Cutoff (η_m)

	OLS	IV		
		First Stage	Reduced Form	2SLS
Panel A. All data				
ln(Share eligible)	-0.105*			0.130*
	(0.060)			(0.067)
Income limit (% FPL) / 100		0.728***	0.094*	
		(0.034)	(0.048)	
Observations	1071	1071	1071	1071
Panel B. Event study sample				
ln(Share eligible)	-0.153**			0.104
	(0.069)			(0.077)
Income limit (% FPL) / 100		0.756***	0.079	
		(0.038)	(0.057)	
Observations	705	705	705	705

This table presents estimation results for η_m, the elasticity of take-up with respect to the share of the population who are eligible, controlling for the covariates included in Equation (2.3). We estimate this elasticity using the eligibility threshold as an instrument for the share of residents in a state who are eligible for SNAP. The first column shows results from a naïve OLS regression of ln(take-up) on ln(share eligible). The second column presents the first stage — the coefficient from a regression of ln(share eligible) on the eligibility threshold as a % of the Federal Poverty Level (FPL). The third column, the reduced form, gives the relationship between the eligibility threshold and ln(take-up). The final column gives the 2SLS estimate, our final estimate for η_m. Standard errors are robust to heteroskedasticity and clustered by state. ** and *** indicate $p < 0.05$ and 0.01, respectively.

Table 2.5: Decomposition: Stigma vs. Information

	(1)	(2)	(3)	(4)	(5)
1. Estimate of β_s	-1.211	6.686	0.580	-1.740	20.256
	(2.060)	(2.047)	(2.916)	(2.137)	(3.745)
2. Estimate of β_i	0.750	1.284	0.921	0.988	1.515
	(0.241)	(0.228)	(0.337)	(0.218)	(0.483)
3. Estimate of η_m^s	0.077	-0.434	-0.040	0.109	-1.316
	(0.151)	(0.150)	(0.213)	(0.156)	(0.274)
4. Estimate of η_m^i	0.593	0.952	0.677	0.780	1.038
	(0.090)	(0.085)	(0.126)	(0.082)	(0.181)
5. N cells	80	80	80	80	80
6. p-value for $H_0 : \eta_m^s = \eta_m^i$	0.006	< 0.001	0.008	< 0.001	< 0.001
Weights:	QC, Exp.	QC, FSPAS	Exp.	QC	FSPAS

This table shows the result of the formal decomposition exercise described in Section 2.5.2. Standard errors and p-values are formed from 99 bootstraps. Appendix B.4 describes the estimation and weighting procedures.

Table 2.6: Summary of Parameters for Welfare Analysis

Parameter	Description	Primary Value	Range of Reasonable Values	Source
$\boldsymbol{\eta_m}$	**Take-up elasticity with respect to eligibility threshold (inframarginal effect)**	**0.1**	$\mathbf{[0.02, 0.24]}$	**Table 2.4 (and 90% CI)**
η_B	Take-up elasticity with respect to benefit size	0.5	$[0.3, 0.6]$	Krueger and Meyer (2002)
ρ	Coefficient of relative risk aversion	3	$[1, 4]$	Chetty and Finkelstein (2013)
s	Share of stigma-only types	0.4	$[0, 1]$	Food Stamp Program Access Study
γ	Cost-benefit ratio, conditional on take-up	0.5		Uniform costs assumption
$\frac{\partial \gamma}{\partial m}$	Change in cost-benefit ratio, conditional on take-up	0		Uniform costs assumption
m^*	Eligibility threshold (share eligible)	0.27		QC and CPS data
p^*	Take-up rate (all eligible)	0.53		QC and CPS data
$\lambda_m / \lambda_{\text{avg}}$	Ratio of marginal to inframarginal welfare weights	0.427		Inverse-optimum approach

This table summarizes the parameters used in the welfare analysis. We note the preferred value and source, but also show robustness to the range of values. The uniform costs assumption implies that γ and $\frac{\partial \gamma}{\partial m}$ are precisely 0.5 and 0, respectively.

Chapter 3

Preferences for Rights

This chapter is coauthored with Aviv Caspi and Julia Gilman.[1]

3.1 Introduction

Political debates over public provision of in-kind assistance often invoke "rights." Advocates for universal health care appeal to the "right to health care."[2] Since 2017, 17 cities and four states in the U.S. passed "right to counsel" policies, which give free lawyers to defendants in eviction cases that are not covered by the 6th Amendment.[3] The United Nations recognizes rights as far-reaching as the "right to enjoy the benefits of scientific progress."

Despite rights' central role in philosophy and political science, economists often restrict attention to "welfarist" allocative preferences, either for individuals or social welfare

[1]We are grateful to Abhijit Banerjee, Amy Finkelstein, Jacob Goldin, Daniel E. Ho, Jim Poterba, and Frank Schilbach for their guidance and support. For helpful suggestions, we thank Jenna Anders, Peter Andre, Doug Bernheim, John Conlon, Esther Duflo, Peter Hickman, Muriel Niederle, Ted O'Donoghue, Alex Rees-Jones, Ian Sapollnik, Jesse Shapiro, Adam Solomon, Pierre-Luc Vautrey, and particularly Shakked Noy, along with seminar participants at Cornell, MIT, Stanford, and Wharton. We are grateful for funding from The Institute of Consumer Money Management (ICMM) Pre-doctoral Fellowship on Consumer Financial Management, awarded through the National Bureau of Economic Research; Stanford Law School's John M. Olin Program in Law and Economics; the National Science Foundation Graduate Research Fellowship under Grant No. 1122374; the MIT Shultz Fund; and Frank Schilbach. This study was approved by MIT's Committee on the Use of Humans as Experimental Subjects under protocol #2112000534 and pre-registered at the AEA RCT Registry under AEARCTR-0012065.

[2]When advocating for the Affordable Care Act, President Barack Obama said, "Health care is not a privilege for the fortunate few — it is a right" (Obama, 2013).

[3]The four states are CT, MD, MN, and WA. The cities include New York City, Newark, San Francisco, and St. Louis. The American Civil Liberties Union argues, "Tenants' right to legal representation in eviction cases is a civil liberties issue, a gender justice, racial justice, and economic justice issue" (ACLU, 2022).

functions. Welfarist allocative preferences depend only on recipients' utilities.[4] That is, a welfarist person i values allocating in-kind good $y = (y_1, \ldots, y_J)$ among the J others in society only because the good instrumentally enters their utility. Then i's allocative utility is $v_i(y) = f(u_1(y_1), \ldots, u_J(y_J))$ for some function $f(\cdot)$ which aggregates others' utilities $u_j(\cdot)$. This class of welfarist allocative utilities nests standard redistributive motives, Becker (1974)-type altruism, and paternalism. Welfarist reasons to value rights could include that constitutional rights constrain harmful despots, or that behaving as if there are rights increases cooperation (Dal Bó et al., 2010).

Valuing rights intrinsically, and not because they instrumentally enter others' utilities, is "non-welfarist." If i has non-welfarist preferences for rights, her allocative utility is $v_i(y) = f(u_1(y_1), \ldots, u_J(y_J), \phi_i(y))$. Such preferences depend on how allocating y affects rights $\phi_i(y)$, holding fixed how y affects others' utilities. This class of non-welfarist allocative utilities includes frameworks like in Tobin (1970) or Sen (1985).

We study preferences for rights, asking two questions. First, to what extent do people have preferences for rights? Second, how do these preferences for rights influence redistributive choices? We answer these questions by conducting allocation experiments with participants in online samples. In several of the experiments, more than half of participants exhibit behaviors that are consistent with preferences for rights, but which are dominated if participants were exclusively welfarist. Those who do exhibit these behaviors also make more universal (less targeted) redistributive choices when recipients differ in need.

Whether people actually have non-welfarist preferences is an empirical question — and one with significant implications for economics. First, economists (should) care about the basic science of measuring allocative/redistributive preferences, as they are key to analyzing the optimal allocation of scarce resources (Hausman and McPherson, 1993). Second, these preferences seemingly motivate recent policy changes like right to counsel programs. Yet it is unclear if advocates appeal to rights when they just mean that certain in-kind goods are very instrumentally valuable to recipients. In that case, rights would not be an extra rationale for in-kind transfers beyond welfarist arguments (e.g., Currie and Gahvari, 2008). Third, in an influential paper, Kaplow and Shavell (2001) argue that non-welfarist Social Welfare Functions (SWFs) do not satisfy the Pareto principle. In part

[4] We use the term "allocative preferences" to mean preferences about the allocation of goods to *ex ante* identical others. We use the term "redistributive preferences" to mean preferences about allocation of goods to others whose need differs (e.g., they have different incomes). We use "welfarist"/"non-welfarist" rather than "individualistic"/"non-individualistic" but either applies here.

due to this point, economists often use welfarist SWFs with no weight on rights. However, as Kaplow and Shavell (2001) themselves note (p. 285), one can sidestep their argument if rights enter individuals' utility functions directly. The force of Kaplow and Shavell (2001) thus depends on the empirical prevalence of non-welfarist preferences.[5]

We use a simple framework to define non-welfarist preferences and derive testable predictions (Section 3.2). In the framework, a social planner or Spectator has utility over allocations in society. Allocative utility is separable in a welfarist component that aggregates others' utilities and a non-welfarist component that has a reference-dependent form (Kőszegi and Rabin, 2006). The reference point is the Spectator's views about how society should be (e.g., "everyone should have access to a given good"). If upholding normative views about society matters more for rights than other goods, Spectators place greater value on the non-welfarist component. Because the loss domain is especially costly, the framework predicts "anti-targeting" — that is, the Spectator redistributes rights goods more universally than non-rights goods.

Non-welfarist preferences are hard to study empirically. Once a right legally exists, everyone has it. Little variation remains to identify preferences for the right. For this reason, it is difficult to measure willingness to pay for freedom of speech, or for lawyers in criminal cases which are covered by the 6th Amendment. Among rights that do not legally exist, it is not obvious how to separate non-welfarist and welfarist preferences, particularly with observational methods.

We overcome these challenges by fielding laboratory experiments that give tight control over the economic environment (Section 3.3). Although the framework employs special functional forms, the experiments give nonparametric tests. Participants ($N = 1{,}800$ Spectators) face incentivized choices about allocating goods to low-income households. We experimentally vary what the goods are. We contrast goods related to rights (treatment) with goods that are instrumentally valuable but less related to rights (control). The "rights goods" are lawyers for tenants facing eviction and health care — two goods that feature in controversial policy debates about rights. The "benchmark goods" are YMCA memberships and bus passes. Because Spectators may believe that benchmark goods relate to rights, comparing rights goods versus benchmarks yields a lower bound on preferences for rights. We implement some of Spectators' choices over lawyers and benchmarks by partnering with a nonprofit that assists tenants facing

[5] If i herself does not value rights, but others do, and i's allocative utility aggregates others' utilities, then i may be a welfarist who still cares about rights.

eviction.[6]

Our first three experiments test for preferences for features of rights (Section 3.4). These experiments examine Spectators' allocations to *ex ante* identical recipients. Experiment 1 studies "inalienability," or the idea that there is a discrete harm from taking a right away. We inform Spectators that a lottery allocated a good to one recipient. Spectators can save money for the nonprofit's future tenant programming by rerunning the lottery, which may take from one recipient and give to another. We elicit Spectators' willingness to pay (WTP) to preserve the lottery, where WTP is measured as the amount of money saved that is required for Spectators to be indifferent to rerunning the lottery. We inform Spectators — and emphasize with confirmation checks — that neither recipient would know about the initial allocation. As recipients' utility only depends on the final allocation, welfarist Spectators should not pay to preserve the lottery.

Yet Spectators do pay to preserve the lottery, even though doing so has no welfarist payoff and actively destroys surplus. Even when the lottery involves the benchmark goods, 46% of Spectators have positive WTP to preserve the lottery. This result may reflect non-welfarist preferences for, say, procedural justice. Allaying concerns that high levels among the benchmarks merely reflect elicitation errors, we find that these preferences are significantly more concentrated among Spectators who face lotteries over rights goods. With rights goods, 53% of Spectators exhibit a positive WTP (s.e. of difference: 2.5 pp; p-value $= 0.002$). Preferences not to rerun the lottery are similar for health care and lawyers.

Experiment 2 studies "dignity of choice." Consider a welfarist who is certain that a recipient would choose $\$y$ in cash over a lawyer. This welfarist should never be willing to pay to let the recipient choose between the two rather than giving them $\$y$ directly. Experiment 2 elicits Spectators' beliefs that recipients choose $\$y$ in cash over the good. Focusing on Spectators who are certain that the recipient chooses cash, we elicit Spectators' WTP, again measured in dollars saved for future tenant programs, to provide choice rather than $\$y$ directly. Thus, Experiment 2 trades off instrumental costs of providing choice with possible non-instrumental benefits.

Spectators have positive WTP for choice even when certain that the choice will not be exercised. Similar to Experiment 1, we find evidence of non-welfarist preferences even for the benchmark goods. Among Spectators who are at least 90% sure the recipient will

[6] Health care choices are always hypothetical. We test the importance of incentives by randomizing a share who see lawyers and benchmarks into identical hypothetical framing as with health care. We reject even small differences in behavior due to lack of incentives for health care allocations.

choose cash, 39% are still willing to pay for choice, meaning they burn surplus if the choice is not exercised. These preferences are again stronger for rights goods, where 57% of these Spectators have positive WTP for choice (s.e. of difference: 2.8, $p < 0.001$).[7] Preferences for choice are stronger for lawyers, but both goods are statistically distinguishable from the benchmarks.

Experiment 3 studies "egalitarianism." Suppose z% of people in society can access a good. Classical welfarists' WTP for providing additional rights goods to those beyond the z% does not depend on z.[8] Non-welfarists' WTP may depend on z, if, for instance, allocating the good ensures all of society gets the good (i.e., z is close to 100%). We inform Spectators that (randomized) z out of 10 tenants already receive a lawyer. We elicit WTP to provide goods to the $(z+1)$th. Spectators have higher WTP to provide rights goods when doing so ensures that all recipients receive the good ($z = 9$). But the increase in WTP is not statistically distinguishable between the benchmark and rights goods. We conclude that we detect evidence only of inalienability and dignity of choice, and focus on these features in the rest of the paper.[9] Interpreted through the lens of the framework, we find that a large share of Spectators have non-welfarist utility with reference points over provision (Experiment 1) or letting people choose for themselves (Experiment 2).

Having found evidence of preferences for rights, we next consider their implications for Spectators' redistributive choices, their magnitudes relative to welfarist preferences, and their relationship to support for in-kind transfers (Section 3.5). Experiment 4 measures redistributive choices and targeting. Spectators choose how to allocate goods among 10 anonymous recipients with varying need, as indicated by their incomes. Spectators choose between giving the good to everyone, or the good plus cash to people with lower incomes. These choices hold the total redistributive budget fixed. For instance, we price lawyers at \$500 per recipient and fix the budget at \$5,000. Spectators can give 10 lawyers

[7] Willingness to pay is 0.51 s.d. higher (s.e.: 0.11, $p < 0.001$) among those who express 100% certainty. Additionally, we contrast the welfarist value of choice, which is the probability the choice is exercised (obtained via belief elicitations) times the value of choice if exercised (obtained in a separate experiment). This welfarist value of choice is lower than the WTP for choice for 31% of Spectators allocating rights goods and 24% of benchmarks (s.e. of difference: 2.2).

[8] More broadly, non-classical welfarists' WTP may depend on z if they have inequity aversion (Fehr and Schmidt, 1999). However, inequality aversion should not be differential across rights versus benchmark goods.

[9] That the rights goods are not differential to the benchmark in Experiment 3 could be: (i) because Spectators are egalitarian in the domains of bus pass or YMCA provision; or (ii) because they are welfarists with inequality-averse preferences (which are equal across goods). Elicitation errors or confusion are less likely to explain the result, as they cannot explain the higher valuations across all goods when z rises. We view our interpretation as conservative, but note that Experiment 3 does not reject that egalitarianism is present for benchmarks and rights goods alike.

to tenants facing eviction; or give the poorest tenant a lawyer and $4,500; or give each of the poorest two tenants a lawyer and $2,000; and so on.

Spectators target rights goods more universally than benchmarks, and preferences for rights may explain why. Spectators are 17 pp (64% of the benchmark mean; s.e.: 2 pp) more likely to provide rights goods to all 10 recipients than to provide benchmarks or cash universally. Propensity to "anti-target" (provide goods universally) is highly correlated with non-welfarist behaviors in Experiment 1 and 2. This result confirms the framework's predicted relationship between normative reference points and redistribution.

To additionally quantify the importance of non-welfarist preferences, we next compute the share of people with welfarist versus non-welfarist preferences. Spectators are at least partially welfarist if their allocation decisions depend on the good's instrumental benefits to recipients. To identify welfarists, we conduct an information-provision experiment that shocks Spectators' beliefs about the instrumental benefits of the rights goods. They then have the choice of revising their initial allocation in the targeting task (Experiment 4).

We find that 26% have exclusively non-welfarist preferences, which is only slightly less than the 31% who have partially welfarist preferences. The remaining share cannot be unambiguously classified. This result challenges the prevailing approach in welfare economics of ignoring how people value rights, at least in the domains of health care or lawyers. It suggests that welfarist Social Welfare Functions can still value rights.

We conclude by showing how preferences for rights correlate with political preferences and support for government policies involving in-kind provision. We find that preferences for rights are uncorrelated with political preferences — we do not merely pick up liberals, for instance. They negatively correlate with income. Preferences for rights predict support for Right to Counsel but not universal health care, perhaps because health care is more politicized.

Related Literature. Philosophers, political scientists, and economists have questioned exclusive focus on utility. Economists have proposed non-welfarist frameworks like Tobin (1970)'s specific egalitarianism, Rawls (1971)'s primary goods, Sen (1985)'s capabilities approach, and Saez and Stantcheva (2016)'s generalized social marginal welfare weights.[10] Gasparini and Pinto (2006) consider theoretical properties of non-welfarist social preferences. These philosophical frameworks motivate our empirical tests, which contribute to several literatures in economics.

[10] An important contribution by Holmes and Sunstein (2000) situates rights within a cost-benefit framework and emphasizes the costs of public provision. We measure the potential benefits to weigh against such costs.

First, we add to a literature in behavioral economics that considers potentially non-welfarist preferences like fairness and moral concerns (e.g., Rabin, 1993; Fehr and Schmidt, 1999; Bénabou and Tirole, 2011). We build off Polman (2012), who studies demand to change initial allocations for others, but who cannot identify preferences for rights.[11] Bartling et al. (2014a) and Bobadilla-Suarez et al. (2017) find subjects value choice for themselves, but do not study how subjects value choice for others. Andreoni et al. (2020) study fairness in the presence of uncertainty and find subjects apply deontological principles when allocating lottery tickets. Unlike these studies, we manipulate rights versus benchmark goods, toward detecting preferences for rights.

Second, we build on the behavioral and experimental literature on redistributive decisions and their determinants (Levitt and List, 2007b; List, 2007; Cappelen et al., 2013, 2020).[12] Recent work pushes beyond standard redistributive preferences to study the determinants of paternalism (Ambuehl et al., 2021; Bartling et al., 2023), but still embeds these views within welfarist frameworks. We stress different non-welfarist considerations and conduct experiments to isolate them.

Third, we contribute to the public-finance literature on in-kind benefit programs. Due to Kaplow and Shavell (2001)'s criticism, economists rarely appeal to rights to justify providing assistance in-kind. Instead, economists focus on various classical rationales (reviewed in Currie and Gahvari, 2008) or non-classical (paternalistic) appeals to internalities. These rationales are still welfarist, even if they involve non-classical preferences or paternalism, because they view social welfare as exclusively depending on experienced utilities in society (Chetty, 2015) or individuals' choices made in welfare-relevant domains (Bernheim and Rangel, 2009).[13] We document that individual preferences for rights are empirically prevalent. As a result, SWFs can both value rights and aggregate

[11]First, in some of Polman (2012)'s experiments, the recipients know about the initial allocation, so the Spectators may still be welfarists who aggregate others' loss aversion. Second, several of the experiments are explicitly framed as thought experiments because they involves willingness to pay for a non-quantifiable outcome (e.g., the recipient's "ability to get dates"). Third, several of the experiments measure Spectators' willingness to pay to improve the outcome or stop the outcome from getting worse. If the Spectators perceive recipients' utility as being concave in the outcome, then higher WTP for stopping the outcome from getting worse is consistent with welfarist preferences.

[12]Like Fisman et al. (2007), we study how bystanders trade off efficiency for redistributive considerations, but focus on willingness to pay for ensuring rights rather than altruistic giving. Like Alatas et al. (2012), we quantify how people target in-kind goods, but document a different phenomenon — "anti-targeting" of rights goods — and propose explanations. Like Charité et al. (2022), we consider how non-classical forces affect Spectators' choices for others.

[13]See Bernheim (2016) and Bernheim and Taubinsky (2018b) for discussions of behavioral welfare analysis.

preferences in a welfarist manner consistent with Kaplow and Shavell (2001).[14] Liscow and Pershing (2022) conduct survey experiments to decompose hypothetical demand for in-kind redistribution. Relative to this work, we conduct experiments that quantify preferences for rights without relying on stated attitudes about rights, and we show how these preferences influence real redistributive decisions.[15]

Finally, we add to political-economy research that considers economic justifications for rights or liberal institutions more broadly (North, 1991; Acemoglu et al., 2005; Mialon and Rubin, 2008). Experimental work has considered how institutions can affect cooperation (Dal Bó, 2014; Dannenberg and Gallier, 2020). Relative to this work, we consider whether people value a particular institution (right to health care or counsel) per se.

3.2 Conceptual Framework

3.2.1 Intuition

In a simple benchmark, the social planner (in our setting, a Spectator in an allocation experiment) redistributes toward the poor. Suppose the Spectator has an exogenous budget to distribute among a population with heterogeneous initial endowments. The Spectator maximizes welfare by allocating goods to those with the highest marginal utility, optionally weighted by social welfare weights. Absent too much complementarity between the good and the endowment, diminishing marginal utility over wealth causes targeting toward those with low endowments.

If the Spectator redistributes relative to a reference point, that can reduce this targeting motive. Suppose the Spectator has loss aversion relative to a reference point for some individuals. Allocating to any individual in the loss domain achieves a higher marginal welfare gain than allocating to those already above the reference point. This motive pushes the Spectator to allocate goods in a "flatter" manner. We model rights as affecting the Spectator's reference point, which then changes how the Spectator targets in-kind provision.

Appendix C.3 provides details and proofs.

[14]Kaplow (2022) discusses the use of welfarist and non-welfarist SWFs in economics.

[15]One of Liscow and Pershing (2022)'s treatments makes the right to in-kind goods salient and studies hypothetical support for in-kind or cash redistribution by the government. They find that the salience framing does not affect redistributive choices. This result may come from people already having preferences for rights in both treatment and control, attenuating the salience treatment: 60% of Liscow and Pershing (2022)'s participants say that rights at least partially drove their choices.

3.2.2 Setup and standard targeting

Set-up. Consider Spectator i's preferences for allocating goods to others $j \in \{1, \ldots, J\}$. Allocative utility is welfarist over an allocation of goods $y \equiv (y_1, \ldots, y_J)$ if it can be expressed using the form:

$$v_i(x, y;\ u_j(\cdot)) = f_i\left(\left\{u_j(x_j, y_j)\right\}_j\right), \tag{3.1}$$

where f_i aggregates others' utilities $u_j(\cdot)$, y_j is the consumption bundle offered to person j, and $x \equiv (x_1, \ldots, x_J)$ captures other aspects of utility or endowments (e.g., income). The subscript j in $u_j(\cdot)$ stresses that social welfare depends on j's experienced utility. Welfarist allocative utilities depend only on the realized utility in society. Utilitarian, redistributive, Rawlsian (maximin), paternalistic, and many other commonly used allocative utilities in economics are welfarist.[16]

Equation (3.1) does not meaningfully restrict recipients' utilities. Recipients' utilities may themselves be non-classical (e.g., reference-dependent).[17]

To simplify notation, we consider a benchmark with homogeneous utility functions[18] and where allocative utility is additively separable in recipients' utilities:

$$v(x, y; \gamma) = \sum_{j=1}^{J} \gamma_j u(x_j, y_j) \tag{3.2}$$

for exogenous welfare weights $\gamma \equiv (\gamma_1, \ldots, \gamma_J)$.

Standard Targeting. Consider allocating m indivisible goods among the population of J recipients. Fixing the vector of welfare weights, the optimal allocation $\{y_j^*\}$ solves

$$\max \sum_{j=1}^{J} \gamma_j u\left(x_j, y_j^*\right), \text{ such that } \sum_j y_j^* \leq m. \tag{3.3}$$

[16]Note that allocative utilities and Social Welfare Functions are distinct. Allocative utilities refer to individuals in society's preferences to allocate to others. SWFs capture how the social planner aggregates preferences of individuals in society. SWFs may be welfarist and value rights if individuals themselves are non-welfarist.

[17]The formulation does exclude altruism among recipients, as their utilities depend only on their own bundle. That is, for simplicity, recipients' utilities depend only on (x_j, y_j). We could easily generalize Equation (3.1) and what follows to allow u_j to depend on the vector (x, y) instead.

[18]In our setting, as in many allocation problems, the Spectators have no information about (heterogeneous) preferences of the recipients.

In this benchmark, the Spectator gives the goods to the $k \leq m$ people where the social marginal welfare gains are largest.[19] If y is not too complementary with x, then the social marginal welfare gain is maximized by providing to those with smaller endowments and higher welfare weights.

3.2.3 Reference points and kinked utility

We now let allocative utility depend on welfarist and non-welfarist components:

$$v_i = f_i\left(\left\{u_j(x_j, y_j)\right\}_j,\ \phi_{iy}(x, y)\right). \tag{3.4}$$

Here, $\phi_{iy}(x, y)$ corresponds to rights that enter the Spectator's utility, separately from how (x, y) influence others' utilities. We let ϕ be indexed by y to stress that different goods could have different non-welfarist utilities. As our experiments detect non-welfarist preferences in the context of rights, we often call the non-welfarist utilities in this study "preferences for rights."

We impose homogeneity in recipients' utilities and parameterize the ϕ function with a form of Kőszegi and Rabin (2006)-type reference dependence:

$$v_i(x, y; \gamma, r) = \sum_{j=1}^{J} \gamma_j\big(u(x_j, y_j) + \eta_y \phi(y_j \mid r_{ij})\big) \tag{3.5}$$

$$\phi(y_j \mid r_{ij}) = \begin{cases} u(x_j, y_j) - u(x_j, r_{ij}) & \text{if } y_j > r_{ij} \\ \lambda\left[u(x_j, y_j) - u(x_j, r_{ij})\right] & \text{if } y_j \leq r_{ij} \end{cases}. \tag{3.6}$$

Here, $r_i \equiv (r_{i1}, \ldots, r_{iJ})$ corresponds to Spectator i's reference points for allocating each $y_j \in y$. The subscript i emphasizes that r_i is Spectator i's reference point for allocating to another person j, and not a reference point that enters j's utility. If η_y and $\lambda > 1$, the Spectator i experiences loss aversion over recipient j's utility relative to i's reference point on good y. Notice that this formulation remains agnostic about whether u_j is itself reference-dependent. The case with $\eta_y = 0$ nests welfarist allocative utility.

[19] Notice that the Spectator can give multiple of the good to the same person. Formally, the Spectator provides goods such that, for all $j, k \leq J$, $\gamma_j\left(u\left(x_j, y_j^*\right) - u\left(x_j, y_j^* - 1\right)\right) \geq \gamma_k(u\left(x_k, y_k^* + 1\right) - u\left(x_k, y_k^*\right))$. A similar result obtains with continuous goods.

3.2.4 Targeting with kinked utility

We introduce a notion of flatness when the Spectator allocates discrete goods.

Let $\mathcal{J}_i$ represent the set of j such that optimal $y_j^* > 0$ under allocative utility v_i. Allocation $\mathcal{J}_2$ is *weakly flatter* than $\mathcal{J}_1$ if and only if $\mathcal{J}_1 \subseteq \mathcal{J}_2$.

Put another way, a set of recipients $\mathcal{J}_2$ is weakly flatter than $\mathcal{J}_1$ if all individuals who receive at least one good under $\mathcal{J}_1$ also receive at least one under $\mathcal{J}_2$. There may be an individual in $\mathcal{J}_2$ who is not in $\mathcal{J}_1$.

We compare optimal allocations when allocative utilities are exclusively welfarist (Equation 3.2) versus have non-welfarist components (Equation 3.5). Put the optimal set of recipients $\tilde{\mathcal{J}}$ as the one that maximizes welfarist allocative utility for a given set of welfare weights, utility functions, and endowments and places no value on non-welfarist utility (that is, $\eta_y = 0$). Put the optimal set of recipients $\mathcal{J}^*$ as the one that maximizes allocative utility for the same set of welfare weights, utility functions, and endowments but which also places value on non-welfarist utility with $r_{ij} = 1$ for all j, $\lambda > 1$, and $\eta_y > 0$. Then the Spectator with non-welfarist allocative utility exhibits "anti-targeting," in the following sense:

Proposition 3.1. *The set of recipients with non-welfarist allocative utility $\mathcal{J}^*$ is weakly flatter than the set of recipients without non-welfarist allocative utility $\tilde{\mathcal{J}}$.*

If allocative utility depends on the Spectator's reference points, then the Spectator allocates goods in a flatter manner. Loss aversion gives a kink in welfare around the reference point which pushes toward more universal provision. This link between reference-dependent allocative utility and redistribution structures our empirical tests.

Connection to Rights. We view rights as mapping onto reference points that the social planner might value. Experiments 1–3, which test for preferences for rights, give joint tests of welfarist allocative utility. They test $H_0 : \eta = 0$ and $\lambda = 1$ and $r_j = 0$, for a particular r in the experiment. Experiment 4 tests the model implication in Proposition 3.1.

The reference points in our model could relate to any feature of provision. Some may relate to rights. Others could relate to concerns about procedural justice. The model nests welfarist allocative preferences but also lets the social planner evaluate an outcome with respect to rights.

For instance, suppose the Spectator views "freedom to choose a lawyer" as a right. In this model, the Spectator's utility over granting freedom to choose has a reference-dependent form. Our experiments test for reference-dependent allocative utility over providing rights. They also test whether there is more reference-dependent allocative utility when providing goods associated with rights versus other goods.

3.3 Experiment Overview

3.3.1 Sample and Design Overview

Sample and Design Overview. We recruit $N = 1{,}800$ participants from Prolific, a widely used online platform for survey experiments (Appendix C.4.1). As is common in Prolific studies, participants are higher-income, younger, and more educated than in the U.S. in general (Table 3.1).

Figure 3.1 presents the experiment flow. We randomize participants ("Spectators") into one of four goods at the start. They then complete four experiments, each with the same good, before answering several questions about political preferences.[20] To be included in the study, participants needed to pass at least two out of three attention checks throughout the survey (Appendix C.4.3). 7% fail one of three and are included in the sample. We routinely include comprehension checks after providing task instructions but before doing the elicitation. Participants passed all comprehension checks at rates exceeding 85%. When they fail them, we correct the participants before the exercise.

We ran the experiment on September 11–12, 2023. The survey took 19 minutes on average. We paid $6 for participating, which is 56% above Prolific's suggested wage for a 19-minute survey.

Set-up and Goods. We inform participants early in the survey that they will face allocation choices on behalf of a nonprofit in Memphis, Tennessee, which assists tenants facing eviction. All participants are informed that the clients are those facing eviction. We make this choice so that all between-good comparisons hold fixed the need and financial situation of the recipients.

The four goods are: attorneys, who can provide legal assistance to tenants; one year of fully subsidized health care at urgent care; a bus pass containing $350 of prepaid fare; and an annual membership at the local YMCA, which can provide child care and wellness services. All goods except health care can be purchased for about $350. Our

[20] Half the benchmark good participants were randomized into doing Experiment 4 with cash instead of their assigned good.

main tests compare “rights goods” (attorneys and health care) to “benchmark goods” (bus and YMCA).

The choice of goods is important but challenging. There are many potential rights goods and comparison goods. Once a right legally exists, the experimenter cannot easily manipulate endowing the right in the lab. Even if the experiment involves hypothetical choices, if the right has no analog in the market, participants may have difficulty forming views about willingness to pay for the right. For instance, even a hypothetical choice about willingness to pay for a lawyer may be more ecologically valid than a hypothetical choice about willingness to pay for free speech. We pick rights goods relevant to active policy debates about in-kind transfers and where appeals to rights are common in the public discourse. We pick comparison goods that are clearly valuable to low-income recipients, but to which Spectators are unlikely to attach special rights.

The comparison between rights and benchmark goods nets out potential elicitation errors. Otherwise, levels of behaviors in the experiments risk conflating inattentiveness that causes random clicking on Prolific with non-welfarist preferences.

However, comparing rights to benchmark goods is a conservative exercise that likely leads us to understate the extent of non-welfarist preferences. At one extreme, Spectators could value the “right to property,” and therefore exhibit non-welfarist preferences for *any* potential benchmark good. In that case, the comparison between rights and benchmark goods might not be distinguishable from zero. Less extreme versions of this preference, in which Spectators value the “right to transit” or the “right to exercise”, would attenuate our results but perhaps not to zero.

Connection to Framework. The experiments yield nonparametric tests. We do not require any functional form assumptions to detect preferences for rights or test whether these preferences correlate with redistributive choices. The special functional forms in Section 3.2 are useful insofar as they explain why preferences for rights could influence redistributive choices.

Taking the functional forms in Section 3.2 seriously, the levels of non-welfarist behaviors in the experiments test the joint hypothesis $H_0 : \eta_y = 0$ and $\lambda = 1$ and $r = 0$. The differences between rights and benchmark goods give a notion of whether reference-dependent preferences are “stronger” when allocating rights versus benchmark goods.

For instance, suppose $r = r' > 0$ and $\lambda = \lambda' > 1$ for all goods. Such reference-dependent allocative utility even for benchmark goods could represent either a true allocative preference, or capture as-if preferences that Spectators exhibit due to elicitation error or inattention. Then, our experiments test whether $\eta_{\text{rights}} \neq \eta_{\text{benchmarks}}$.[21] Under this interpretation, our tests of $H_0 : \eta_{\text{rights}} - \eta_{\text{benchmarks}} = 0$ are conservative for testing $H_0 : \eta_{\text{rights}} = 0$, as long as $\eta > 0$ for all goods. If $\eta = 0$, this framework continues to nest welfarist allocative utility.

Ethics. Spectators provide informed consent and face no risk of harm. Through the nonprofit partner, we only allocate goods or cash that have the potential to help tenants. Tenants may also decline the offer of assistance. As all tenants are needy, and there is not enough funding to give all tenants assistance, providing assistance based on the actual choices from some Spectators is a reasonable way of targeting. Indeed, the allocation choices in this study are similar to those in real-world political decisions about means testing, as well as community targeting studies (Alatas et al., 2012) and related work.

3.3.2 Incentives

A feature of this paper is that we incentivize choices for all goods except health care. We use the strategy method, informing Spectators that there is a chance their choices will be implemented. To implement choices, we provide legal assistance, YMCA memberships, and bus passes via a nonprofit partner in Memphis. Participants are informed and must confirm that they face real choices which could affect allocations for needy recipients. We introduce another incentive by telling all participants (including in the health-care treatment) that the study results could influence the nonprofit's future programming. We also incentivize belief elicitations by paying participants if they are accurate. See Appendix C.4 for details on all incentives.

We embedded several tests to study whether lack of incentives affects results for health care. We find no reasons to be concerned (Section 3.5.4). That said, we still view incentivization as an important part of this paper. It was unclear that incentives would have small effects *ex ante*, and some readers may (reasonably) have been skeptical of results if all elicitations were hypothetical.

3.3.3 Main and Secondary Elicitations

We conduct four main experiments, testing for features of rights (Experiments 1–3, Section 3.4) and redistributive preferences (Experiment 4, Section 3.5). We also present

[21] There are other interpretations of the experiments. They could also test whether $r_{\text{rights}} \neq r_{\text{benchmarks}}$ or $\lambda_{\text{rights}} \neq \lambda_{\text{benchmarks}}$. Either way, the framework gives a way of organizing and interpreting non-welfarist behaviors.

several secondary experiments/elicitations (pre-registered as such). The first directly elicits (bounds on) the Spectator's indifference point between providing the good and giving cash to an anonymous recipient, using a multiple price list and the strategy method. We interpret this elicitation as a Spectator's willingness to pay WTP_i for the good and use this value in extra tests throughout. The second secondary experiment is an information-provision experiment that tests for welfarism (Section 3.5).

3.3.4 Specification, Balance, and Attrition

Specification. Our statistical tests follow the forms:

$$y_i = \beta_0 + \beta \text{Right}_i \ (+X_i\delta) + \varepsilon_i \tag{3.7}$$

$$y_i = \beta_0 + \beta_l \text{Lawyer}_i + \beta_h \text{HealthCare}_i \ (+X_i\delta) + \varepsilon_i \tag{3.8}$$

where β is the effect of being a rights good on an outcome y_i in Equation (3.7), and β_h and β_l are the effects for lawyers and health care respectively in Equation (3.8). We pool both benchmarks for power. Our main specifications omit controls X_i and compare raw means between rights goods and benchmarks, but robustness checks include them. We use robust standard errors for inference.

Balance and Attrition. Demographics are balanced across rights treatments versus benchmarks (Table 3.1, joint $p = 0.441$).[22] Attrition rates were 4% (Table C.2).

3.4 Features of Rights (Experiments 1–3)

We present the design and results for each experiment in turn.

3.4.1 Experiment 1: Inalienability

3.4.1.1 Design

The logic behind Experiment 1 is that welfarists should care only about the ultimate allocation of a good. Suppose a good first is assigned to one person, then is transferred to an *ex ante* identical person, and neither person is aware of the transfer. The welfarist

[22]Table C.1 further disaggregates balance across possible treatments and finds $p = 0.857$ for the lawyers treatment against benchmarks and $p = 0.129$ for the health care treatment against benchmarks.

should not care that the good was transferred. Non-welfarists may dislike transferring goods from one person to another because it requires removing the good from someone who has, in some sense, received it.

Experiment 1 hews closely to this idea. We tell Spectators that Recipient B was assigned a good in a lottery. Spectators have the choice of rerunning the lottery, which has the chance of taking the good from Recipient B and giving it to Recipient A. If they rerun the lottery, the Spectator saves $\$x$ for future programming at the nonprofit. We tell the Spectator that the money saved will assist other tenants. We use a multiple price list to find (bounds on) the point at which Spectators are indifferent between saving $\$x$ and preserving the lottery result. We refer to I_i, the midpoint of the elicited bounds, as willingness to pay (WTP) to preserve or not rerun the lottery, in units of the dollars saved for future programming. If $I_i > 0$, the Spectator is willing to burn I_i of surplus to satisfy a non-welfarist preference.

As the units of I_i are unintuitive, we normalize I_i so that the pooled benchmarks have mean 0 and standard deviation 1. We then compare I_i among the rights goods to the benchmarks. We also study the propensity to have a positive WTP to preserve the lottery (i.e., we form $\mathbb{1}(I_i > 0)$ where I_i is unnormalized).

3.4.1.2 Results

Spectators have 0.3 s.d. higher WTP I_i to preserve the lottery for rights goods than for benchmarks (Figure 3.2, s.e. of difference: 0.05). Converted back to units of money saved for future programming, participants' WTP is \$20 higher for rights goods than benchmarks. Inspecting the propensity to pay anything, we find that 53% of participants have positive WTP to preserve the lottery for rights goods, 8 pp higher than benchmarks (s.e. of difference: 2.5). The differences are slightly larger for lawyers than rights goods in the continuous WTP measure, but similar for the extensive margin.

Quotations from free-response questions support our interpretation of these results. For instance, one participant wrote: "It's the principle of the matter. Even if the tenants wouldn't know, you'd know." Another wrote, "If I re-ran the lottery it would feel like I was removing the lawyer from the first winner, and it would feel wrong."

Yet a remarkable 46% of Spectators exposed to benchmark goods still have a positive WTP to preserve the lottery. Why? One explanation is that non-welfarist preferences are present in the allocation of any good, and are just stronger for lawyers and health care. As one participant who saw a benchmark good wrote: "I decided to keep the lottery results for each trial because rerunning the lottery and taking away the original winners' YMCA seems very unfair." Such preferences could owe to a normative respect for procedural justice, for instance. If Spectators have non-welfarist preferences for benchmark goods, comparing rights to benchmark goods implies the true extent of non-welfarist preferences may be closer to the levels who have a positive WTP (i.e., 51% on average).

We aggregate these results in Table 3.2, which presents outcomes for willingness to pay I_i, an indicator for having a positive WTP, and an indicator for having the maximum WTP that we elicit.

Interpretation and Connection to Framework. Our test cannot be explained as Spectators being welfarists who care about recipients' loss aversion or their endowment effect. Loss aversion models require the recipient to be aware of the initial allocation, such that the recipient can form a reference point. We take significant steps to inform and remind Spectators that recipients will not know about the initial allocation. The experiment text says: "Remember that the **tenants will not know that the lottery was rerun**. They will just learn the final result, and the ultimate allocation will be anonymous" (emphasis in experiment text). We also include a confirmation check that asks participants whether the recipients will know who was originally supposed to receive the good. 98% of participants get the question right. For these participants, we reiterate: "**That is correct**. Tenants will only learn the final result of the lottery." We correct the 2% of participants who get the question wrong: "**That is incorrect**. Tenants will only learn the final result of the lottery."

While this experiment does identify non-welfarist utility, whether it identifies "preferences for rights" of the form in Section 3.2 is more debatable. Viewed through our framework, Spectators place meaningful weight on the reference point of initial allocation of rights goods ($\eta_{\text{rights}} - \eta_{\text{benchmarks}} > 0$). We see "inalienability" as mapping to the reference point $r = 1$, which applies to all goods. The value placed on this reference point is stronger for rights than benchmarks.[23]

Still, we cannot reject alternate models. For instance, if Spectators feel guilt or responsibility only if they change others' allocations, then they may not want to intervene.

We cannot rule out these interpretations entirely, but note two points. First, alternate explanations must account for a difference across rights versus benchmark goods. It is not clear why Spectators feel more responsible for intervening in rights goods. Second, guilt about intervention would also imply non-welfarist allocative utility of the form in

[23]Another interpretation is that $r = 0$ for benchmarks, which only amplifies the preferences for rights that obtain with lawyers or health care.

the introduction.

Robustness: Valuation of the Good. Another concern is that we merely identify a behavioral phenomenon in which Spectators do not like to shuffle valuable goods from one recipient to another. According to this view, our results owe to the fact that lawyers are more valuable than bus passes (say). Such a preference would be still non-welfarist, as switching the goods does not affect recipients' utilities. Nevertheless, it may affect interpretation of our results as a true preference for rights per se.

We reject this concern by controlling for fixed effects in the Spectator's valuation for giving the good directly, WTP_i (Table C.3). Intuitively, this test compares Spectators who find rights and benchmark goods equally valuable with respect to cash. We continue to find that Spectators facing the rights good are more likely to pay to preserve the lottery.

3.4.2 Experiment 2: Dignity of Choice

3.4.2.1 Design

Design and Main Measure. The logic behind Experiment 2 is that welfarists only value providing recipients with the ability to choose insofar as the choice might be exercised. If welfarists are completely sure a recipient always chooses (a) over (b), then their willingness to pay to provide a choice between (a) and (b), versus giving (a) directly, is zero.[24]

Experiment 2 begins by eliciting Spectators' beliefs about the probability that a recipient facing the choice of $\$y$ versus the good would choose the good, which we denote as p. Then, Spectators face the choice of: (i) providing $\$y$ to the tenant directly and saving $\$x$ for future programming for the nonprofit, versus (ii) giving the recipient the choice between $\$y$ and the good. We elicit (bounds on) the value of $\$x$ that makes Spectators indifferent between (i) and (ii).

We compare C_i, the midpoint of these bounds, among rights and benchmark goods. We focus on C_i as p approaches 1. When $C_i > 0$ and the Spectator has high beliefs, the Spectator burns surplus to let the recipient choose. As in Section 3.4.1, we normalize C_i so that it is mean 0, standard deviation 1 among the pooled benchmark goods.

Our test requires conditioning on people with high beliefs p, which could lack power. To increase Spectators' beliefs, we provide all Spectators with truthful information from

[24]Welfarists may value providing unexercised choices if they project intrinsic values of decisions (Bartling et al., 2014b; Lenk, n.d.) onto recipients. Analogous to procedural justice concerns in Experiment 1, this may explain high levels of WTP among benchmark goods. However, intrinsic values of choices cannot account for differential WTP between benchmark and rights goods, as we find.

a randomly selected pilot sample. The information says that all tenants in the pilot sample chose cash over the good (see Appendix C.4.5 for details on the treatment). This information raises power by increasing the number of Spectators with high beliefs.[25]

Second Measure. We also form a second measure of a non-welfarist willingness to pay for choice. Welfarists value providing the choice of good g and cash $\$y$ versus cash as:

$$C_i^w = p(-i \text{ chooses g}) \times E[f(u_{-i}(g)) - f(u_{-i}(y)) \mid -i \text{ chooses } g]. \tag{3.9}$$

This expression says that welfarists value choice at their WTP to provide the good, times the probability of exercising the choice. We obtain the value $E[f(u_{-i}(g)) - f(u_{-i}(y)) \mid -i \text{ chooses } g]$ by eliciting Spectators' willingness to provide g versus cash to a recipient. We assume small selection on gains (supported by a direct test below), such that

$$E[f(u_{-i}(g)) - f(u_{-i}(y))] \approx E[f(u_{-i}(g)) - f(u_{-i}(y)) \mid -i \text{ chooses } g]. \tag{3.10}$$

We study the effect of rights goods on Δ_i, the difference between actual WTP for choice C_i and welfarist implied WTP for choice C_i^w:

$$\Delta_i(t) := \mathbb{1}(C_i - C_i^w - t > 0). \tag{3.11}$$

Setting tolerance $t = 0$ lets us examine whether the elicited WTP for choice is exactly equal to the welfarist WTP for choice. We focus on $t \gg 0$, to conservatively account for trembles (e.g., imperfect ability to scale WTP by beliefs) and selection on gains.[26]

The advantage of the second measure relative to the first measure is that it does not require us to condition on having high beliefs. The disadvantage is that we lack a principled way to choose t. If t is too small and perceived selection on gains is large, the

[25] Conditioning on beliefs could, in theory, affect experimental balance. Randomization does not guarantee that participants who have high beliefs p for one good have the same potential outcomes as those who have high beliefs for a different good. First, randomization is not required for this test. Any positive WTP as $p \to 1$ (or difference across goods) still indicates non-welfarist preferences. Second, Table C.4 shows that balance persists conditioning on beliefs. Third, our second measure of the value of choice is not subject to this concern.

[26] Setting $t > 0$ also accounts for minor elicitation differences between C_i and C_i^w. In particular, C_i is WTP in units of dollars of future programming for the nonprofit. C_i^w is WTP in units of dollars of money provided directly to that tenant. Crucially, no matter how large we set t, we find differences in $\Delta_i(t)$ across goods. Relatedly, both C_i and C_i^w are subject to top-coding in the multiple-price list. To be conservative and push against finding a large $\Delta_i(t)$, we top code the maximum direct WTP at \$1,500, whereas we top code C_i at \$950. Note that top coding across does not introduce bias unless differential by good.

test is invalid. We use $t = \$250$ and show robustness to this decision.

3.4.2.2 Results

Spectators exhibit differential preferences for the dignity of choice among rights goods relative to benchmark goods (Figure 3.3A). Focusing on rights goods (blue series), we reassuringly find that WTP C_i is decreasing in beliefs that recipients will choose the good over cash. This negative relationship reflects that choice is less valuable if it is unlikely to be exercised.

However, for every bin of beliefs, willingness to pay for rights goods is higher than for benchmark goods. Among Spectators with beliefs larger than 0.9, WTP is 0.5 sd (s.e.: 0.05) higher for pooled rights goods than benchmarks. Converted back to units of money for future programming, Spectators are willing to pay \$395 on average for pooled rights goods (\$223 for benchmarks).[27] Spectators have differentially higher WTP for lawyers than health care, but both differ from the benchmarks (see formal tests in Table C.5). These differences persist even if we condition on having beliefs larger than 0.95 (farthest right whiskers).

Rights good versus benchmark differences also persist on the extensive margin, when we examine having a positive WTP for choice at all (Figure 3.3B). For instance, among Spectators who think there is at least a 95% chance that the recipient will choose cash, 52% have positive WTP for choice with a rights good (versus 37% who provide a benchmark good; s.e. of difference: 3.5 pp).

As in Experiment 1, Spectators exposed to the benchmarks still exhibit high levels of non-welfarist preferences (Figure 3.3, orange series). These levels may reflect welfarist or non-welfarist valuations of giving choice even in non-rights cases (Bartling et al., 2014a). The difference between benchmarks and rights goods rule out elicitation errors, so we stress differences to be conservative.

Quotations again support our interpretation of the results. A Spectator seeing lawyers who had the maximum WTP wrote, "I think the tenant has a right to choose what assistance to accept." Another wrote, "The tenant has a right to choose, no matter what the monetary consequences."

As one way of summarizing these accounts, we ask Spectators a qualitative question about why they made their decision in the experiment. The share of Spectators who say

[27] The scales of the elicitation in Experiments 1 and 2 differ, since providing choice could have large instrumental benefits if, say, lawyers help tenants win an eviction case.

that recipients have the right to choose when facing a rights good is about 40%, which is 13.2 pp (48%) more likely than with benchmarks (Table C.6, Column 5).[28] Reassuringly, the share of Spectators whose self-described motivations include the right to choose is only somewhat smaller than the share who are willing to pay for choice (40% versus 52%).

Formal tests reinforce these results (Table 3.3). Pushing our experiment to its logical conclusion, the effect on the overall WTP is large and highly significant even if we examine only the 454 Spectators who say there is a 100% chance the recipient will choose cash (Column 9).[29] The effect on an indicator for the extensive margin attenuates if we consider only those with posteriors of 100% (to 7.2 pp, s.e.: 4.7). The gap between the overall WTP and extensive margin results is driven by a large effect of rights goods on willingness to pay the maximum to ensure choice (Columns 5 and 8). That is, more Spectators appear to value providing choice very highly for rights goods than for benchmarks.

The second measure of dignity of choice corroborates the primary measure (Figure 3.3C). We find that rights goods have larger $\Delta_i(t)$ for all tolerances between \$0 and \$500. For instance, focusing on a tolerance of \$250, we find that Spectators exposed to rights goods are 6.4 pp more likely to have a WTP for choice that exceeds their instrumentalist WTP by \$250 or more (s.e.: 1.7). The value $\Delta_i(t)$ is guaranteed to decrease in t. But the difference in $\Delta_i(t)$ for rights versus benchmark goods remains large as a share of benchmark goods' $\Delta_i(t)$.

Testing Selection on Gains. A concern is that we do not account for selection on gains. Suppose Spectators believe that recipients who choose g over \$$y$ benefit substantially from it, but also believe that most recipients choose \$$y$ over g and would not benefit from g. Then Spectators may: (i) have a low average WTP for lawyers; (ii) have a high WTP for choice.

Our primary tests above address this concern. Selection on gains is relevant only if p is mismeasured, since selection on gains still vanishes from welfarist's utility as $p \to 1$.

[28]After eliciting WTP, we ask participants "Which of the following reasons motivated your choice(s)? Select all that apply." Options included: "I thought anyone who would choose the [good] would really want it"; "I did not think anyone would choose the [good] in reality"; "Saving is my priority"; "All tenants should be entitled to the choice of a [good]" (the right to choice option); and "None of the above."

[29]Our incentive scheme rewards people equally if they had posteriors of 96–100%, and we find large effects on the extensive margin if we focus on posteriors of 95% or above (Figure 3.3B). We can detect an effect on the extensive margin limiting to Spectators with 100% posteriors if we use the machine learning method of Chernozhukov et al. (2018) to select controls (Table C.6). Moreover, the extensive margin is still distinguishable for lawyers versus benchmarks among Spectators with 100% posteriors (Table C.5, Column 7). Spectators with 100% posteriors are 11 pp more likely to select that people have a right to choice when doing the experiment with a rights good (Table C.6, Column 4).

However, it is reasonable to worry that Spectators' beliefs are mismeasured due to noise, lack of numeracy, or elicitation issues.

We embed another test to directly examine selection on gains. In particular, we randomize the value of the bundle $y \in \{\$200, \$300\}$. Intuitively, randomizing the bundle traces a supply curve to provide choice. If this supply curve is upward sloping, holding beliefs about the share who choose the good over $\$y$ constant, that suggests selection on gains. In fact, we find no evidence that this supply curve is upward sloping in $\$y$. Appendix C.4.6 explains formally how this subexperiment tests for selection on gains.

Connection to Framework. To view this experiment through the lens of our framework, consider the choice itself as a good y. Suppose also that $r = 1$ for providing choice, that $\lambda = \lambda'$ for all goods, and that u_j does not itself depend on choice. Then our experiments test $H_0 : \eta_{\text{rights}} = \eta_{\text{benchmarks}}$, which is conservative for testing $H_0 : \eta_{\text{rights}} = 0$ as long as $\eta_{\text{benchmarks}} > 0$. Here, η's relate to non-welfarist concerns over providing choice.

One complication is if u_j depends on choice. Conditioning on $p \to 1$ intends to restrict to the subset of people for whom choice is not instrumentally valuable. But if recipients value choice even when choice is not exercised (Bartling et al., 2014a), then welfarist Spectators may also value choice. Differencing with respect to benchmarks still provides a valid test of $H_0 : \eta_{\text{rights}} = \eta_{\text{benchmarks}}$ if: (i) u_j is additively separable in the intrinsic value of choice and other parts of utility, and (ii) the intrinsic value of choice is equal for goods of equal value.

3.4.3 Experiment 3: Egalitarianism

3.4.3.1 Design

The logic behind Experiment 3 is that welfarists' utility should not depend on the share of people in society who already get a good. To be concrete, suppose z out of 10 people get a lawyer regardless. Welfarists' WTP to provide the $(z + 1)$th recipient a lawyer should not vary with z differentially for rights goods.

We operationalize this idea by informing participants that z out of 10 *ex ante* identical and anonymous recipients were selected to receive lawyers. We then elicit participants' willingness to pay to provide the $(z + 1)$th person with a lawyer. In this case, the outside option is a donation to a food bank.[30] We randomize $z \in \{1, 5, 9\}$.

[30] Had the outside option been "saving for future programs" as in Experiments 1–2, then choices in Experiment 3 could never be egalitarian. Tenants later would not be assisted. Put another way, "future

We estimate the following difference-in-differences specification:

$$y_i = \delta_0 \text{Right}_i + \delta_1 \mathbb{1}(z_i = 9) + \beta_0(\text{Right}_i \times \mathbb{1}(z_i = 9)) + \varepsilon_i \tag{3.12}$$

$$y_i = \delta_0 \text{Lawyer}_i + \delta_1 \mathbb{1}(z_i = 9) + \delta_2 \text{HC}_i + \beta_l(\text{Law}_i \times \mathbb{1}(z_i = 9)) + \beta_h(\text{HC}_i \times \mathbb{1}(z_i = 9)) + \varepsilon_i. \tag{3.13}$$

The coefficients of interest are β_0, β_h, and β_l.

The difference-in-differences specification addresses an important concern that inequity aversion (Fehr and Schmidt, 1999) generates a higher willingness to pay for *any* good if $z_i = 9$. One way of thinking about this experiment is that it essentially examines whether inequity aversion differs by good. As another test of inequity aversion, we augment Equations (3.12) and (3.13) with a control $g(\text{WTP})_i$, which is a flexible function of the directly elicited WTP to provide the good. As inequity aversion still depends on realized utilities, controlling for the value $g(\text{WTP})_i$ isolates the non-instrumental role of differential z_i.

3.4.3.2 Results

We find no evidence of differentially egalitarian preferences for lawyers or health care compared to benchmark goods (Figure C.3). We find that across all goods, rights and benchmark, preferences become more egalitarian as z rises. This null result can be interpreted several ways. First, non-welfarist egalitarian preferences may extend to all four goods, including benchmarks. Alternatively, as there is no differential egalitarianism across rights versus benchmarks, these tests cannot reject the presence of welfarist but inequity-averse preferences. Elicitation errors or inattention are a less persuasive explanation for the spike at $z = 9$ as they are unlikely to differ at $z = 9$ versus $z = 5$ and $z = 1$.[31]

Egalitarian preferences were *ex ante* reasonable to examine. Equal rights are a fundamental tenet of liberalism. However, they are also challenging to manipulate in the lab. We cannot control the share of people in society who have access to the good. One explanation for the null result could be that Spectators internalize that, no matter their choice, many people will still lack lawyers or health care.

programs" would have raised the denominator from 10 to an unknown number.

[31] Formal tests of the difference-in-differences — including or excluding a control for the direct WTP for the good — also fail to detect evidence of differential egalitarianism (Table C.7). If anything, we find that the coefficient is negative (and significant at $p < 0.05$ with the WTP control), which implies more differential egalitarianism for the *benchmarks*.

Multiple Hypothesis Corrections. As we find results consistent with our hypotheses in two of three experiments, we perform multiple hypothesis corrections for inalienability, dignity of choice, and egalitarianism (Table C.8). Romano-Wolf adjusted p-values for the continuous WTP measures in Experiments 1–2 remain significant at $p < 0.001$.

3.5 Implications for Targeting and Political Preferences

Having found evidence of preferences for rights, we now turn to their implications. First, we show that they correlate with redistribution decisions, using a novel redistribution experiment that is suitable for this setting. Second, we use this task to quantify the share of people with welfarist versus non-welfarist preferences. Finally, we consider preferences for rights and support for in-kind provision. As we find evidence only for the features of inalienability and dignity of choice, we focus on how these correlate with the outcomes of interest.

3.5.1 Anti-Targeting (Experiment 4)

3.5.1.1 Design

It is not trivial to measure redistributive preferences over indivisible goods where recipients only benefit from provision at the extensive margin. Suppose there are 10 people who need lawyers; they can be uniquely sorted by income (i.e., no ties); there are $\ell < 10$ lawyers; and no one benefits from multiple eviction lawyers. Anyone with progressive redistributive preferences gives the lawyers to the ℓ poorest people. Thus, we cannot simply ask Spectators how they would allocate ℓ lawyers among 10 people.

Experiment 4 introduces smoothness into the problem as follows. We truthfully tell Spectators that 10 tenants with annual household incomes ranging from \$0 to \$36,000, in increments of \$4,000, have applied for assistance. Spectators may give all tenants the good g, again randomized across the four goods. Alternatively, Spectators may give the poorest $\ell \in \{1, 2, \ldots, 9\}$ people the good as well as cash. The value of the cash is decreasing in ℓ. Thus, Spectators face a trade-off between (i) giving more money and the good to fewer, needier households, versus (ii) less money and the good to more households, where the marginal household is less needy.

To ensure that every choice considers the same budget, we fix the total redistributive budget for this choice at B. Good g's price is $p_g = B/10$. Any money not spent on “purchasing” the good is divided equally among tenants who receive the good. The Spectator faces a choice of giving $\frac{B-p_g\ell}{\ell}$ dollars and the good g to ℓ tenants, or all 10 tenants the good. We use multiple price lists to identify the number of tenants at which a Spectator is indifferent between giving the good to everyone and money plus the good to fewer recipients.

As an example for one good, we price lawyers at \$500 and consider a total redistributive budget $B = \$5{,}000$. First, Spectators choose between giving (i) lawyers to everyone, versus (ii) five tenants a lawyer and \$500 in cash each. If they choose (i), they face the choice of giving lawyers to everyone versus six tenants a lawyer and \$333 in cash each. We iterate on these questions until we find R_i, the Spectator's preferred value, for $R_i \in \{1, \ldots, 10\}$.

Design Considerations. It is important to choose the price of each good (equivalently, the budget) correctly. To see why, suppose p_{lawyer} were \$1 and p_{bus} were \$100. Then, since not much cash can be redistributed by giving lawyers to fewer people, and lawyers may be very effective, most Spectators would likely choose to give 10 tenants the lawyer. We price health care at \$600, lawyers at \$500, YMCA at \$300, and bus passes at \$250. We selected these prices to be the median of pilot WTP elicitations. Notice that making rights goods more expensive pushes toward allocating them less universally. This choice is conservative for our ultimate conclusions. As one check, we find they are similar but not identical to the direct WTP we elicit for each good.

In addition to conducting this exercise with rights and benchmark goods, we also randomize some of the Spectators assigned to benchmark goods into doing this exercise with cash, at a total budget of $B = \$5,000$. Spectators in this elicitation choose the value ℓ at which they are indifferent between giving B/ℓ in cash to ℓ people or $B/10$ to 10 people.

3.5.1.2 Results

Spectators are more likely to “anti-target” — that is, give goods universally (to all 10 tenants) with rights goods than benchmarks (Figure 3.4). Pooling lawyers and health care, 43% of Spectators who allocate rights goods anti-target, compared to 26% of Spectators allocating benchmarks or cash (s.e. of difference: 2.3 pp). Both lawyers and health care are significantly different than the benchmarks and cash. Spectators are more likely to anti-target with lawyers than health care. Lawyers are 26.3 pp more likely to be anti-targeted than benchmarks or cash (s.e. of difference: 2.7 pp), whereas health care is 7.1 pp more likely to be anti-targeted (s.e. of difference: 2.6 pp). Table 3.4 aggregates these tests and shows similar results with R_i, a continuous measure of the number of tenants allocated assistance.

One should not interpret these results as suggesting that Spectators would give goods universally no matter the recipient population. All 10 recipients are quite needy. However, when facing the same group of needy tenants, Spectators' targeting preferences are flatter when distributing rights goods versus benchmarks and cash.

3.5.1.3 Non-Welfarist Preferences and Anti-Targeting

Our framework predicts that preferences for rights lead to flatter allocation of goods (Proposition 3.1). Thus we expect that Spectators who demonstrate preferences for rights in Experiments 1 and 2 provide rights goods to more recipients in Experiment 4.[32]

We find support for this prediction (Figure 3.5). WTP for preserving the lottery (I_i) and choice (C_i) are both significantly higher among Spectators providing goods universally. Table 3.5 presents bivariate and multivariate regressions of anti-targeting on I_i and C_i. Multivariate regressions of anti-targeting on both WTPs reveal that C_i is more predictive than I_i when considered jointly. The predictiveness of C_i on Spectators' anti-targeting propensity persists when considering each right separately. Further supporting this result, the propensity to exhibit non-welfarist preferences for benchmarks in Experiment 1 and 2 also predicts anti-targeting of benchmarks (Figure C.5).

Addressing Objections. As noted above, a key concern about this exercise is whether we set the "price" correctly for rights goods versus benchmark. We purposefully set a price for each good that is conservative with respect to generating anti-targeting of rights goods — that is, Spectators could give more cash to the poorest if they chose not to anti-target lawyers and health care. Despite these efforts, the concern remains reasonable, as WTP in the experiment for lawyers exceeds that from pilots. In particular, average WTPs for YMCA, bus, health care, and lawyers are: \$328, \$373, \$507, and \$765 respectively.

To further test this point, we control for Spectators' elicited WTP (Table C.9). Rights still predict anti-targeting propensity, although the effect attenuates modestly in the pooled sample. Lawyers still predict anti-targeting even when we control for WTP, even though lawyers are the sole good where elicited WTP exceeds the implied price (which would push toward more universal provision). Moreover, the correlation between I_i and

[32]The framework does not map literally onto this experiment because Proposition 3.1 is based on providing any number of a one-dimensional good. Experiment 4 provides a good with two dimensions, cash and the single right or benchmark. Allocations change both dimensions. The fixed budget in Experiment 4 constrains allocations in a similar way to the fixed number of goods in Proposition 3.1.

C_i and universal provision persists even with controls for direct WTP (Table C.10). Indeed, controlling for direct WTP is a highly conservative exercise here, as preferences for rights may generate high direct WTPs, so controlling for direct WTP risks being a "bad control" (Angrist and Pischke, 2009).

A related objection involves paternalism. If Spectators believe low-income participants will misuse cash, they may prefer universal provision of any good. Rights goods versus benchmark differences in universal provision address this concern.

3.5.2 Quantifying Welfarist and Non-Welfarist Preferences

We embedded a sub-experiment into Experiment 4 to identify potential welfarists. By doing so, we can compare the share of non-welfarists to welfarists.

Design. The idea behind the experiment is that welfarists change redistributive choices based on surprising information about the instrumental effects of providing a good. For instance, if welfarists learn that lawyers are ineffective, they should be less inclined to provide them rather than cash. In this experiment, we give truthful information about the efficacy of lawyers and health care. Then we ask Spectators if they want to change redistributive choices based on the information. We label Spectators as "welfarist" if they do revise redistributive choices when this information conflicts with beliefs.

We implement this design as follows (Appendix C.4.7 gives full details). Spectators assigned to lawyers or health care are randomized into seeing information that the good is effective or ineffective. There is no equivalent experiment for the benchmark goods. For health care, we show either a positive or null treatment effect about health care from the Oregon Health Insurance Experiment (Allen et al., 2013; Baicker et al., 2013). For lawyers, we show either a large or small treatment effect from an ongoing RCT of providing lawyers to tenants facing eviction in Memphis, TN (Caspi and Rafkin, 2023). Before giving information, we elicit prior beliefs about the efficacy of lawyers and health care.

After providing information, we let Spectators choose whether to revise their targeting choice in Experiment 4. (The anti-targeting results in Section 3.5.1 all report Spectators' *initial* choices.) In particular, we ask Spectators assigned to lawyers: "Previously, you made choices distributing a limited budget across hiring lawyers and giving tenants cash. Given this information, would you like to revise any of your choices?" Spectators who say they want to revise their choice then do the same targeting elicitation from Experiment 4. The set-up is similar for health care.

Our goal is to label Spectators as welfarist or non-welfarist. We focus on Spectators who had beliefs about lawyers/health care efficacy that disagree with information we provided them in the treatment. Among these Spectators, we label them as *welfarist* if they revise their targeting decision.[33] Among the same group of Spectators, we label them as *non-welfarist* if they do not revise their targeting decision and do exhibit either positive WTP not to rerun the lottery or positive WTP for choice. Because Spectators may have positive WTP for choice for instrumental/welfarist reasons (e.g., selection on gains), robustness tests restrict to Spectators who also have high beliefs that the tenant will choose cash in Experiment 2 ($p \to 1$, in the notation of Section 3.4.2).

The way we label Spectators is conservative. Anyone who revises targeting in Experiment 4 is labeled as welfarist, even if they also make non-welfarist choices in Experiments 1 or 2. We only label Spectators as non-welfarist if they *both* forgo the chance to make a welfarist revision to targeting in Experiment 4 *and* make a non-welfarist decision in one of Experiments 1 or 2.

Results. Upon receiving information, updates are fairly rare. About 40% choose to update their allocation if they receive information about lawyers, and less than 30% update if they receive information about health care (Figure C.4). Conditional on updating, Spectators tend to update in the direction of the information shown (e.g., they provide more lawyers if they get positive information about lawyers).

Despite conservative classification choices, we observe comparable magnitudes of welfarism and non-welfarism (Figure 3.6A).[34] Restricting only to Spectators who do the experiments with rights goods, 81% exhibit preferences for rights in at least one of Experiments 1–2, and 39% exhibit preferences for rights in both experiments. Meanwhile, 31% are welfarist, meaning that they revise targeting choices after receiving information. If we label people as having preferences for rights only if they make non-welfarist choices in both Experiments 1–2 and do not make a welfarist choice, we find that 26% have non-welfarist preferences.

We therefore decisively reject that non-welfarist preferences are not quantitatively

[33] The beliefs we elicited for lawyers exactly correspond to the information provided in the treatment. Because it was difficult to elicit beliefs that were identical to the information we provided for health care, we label people as having priors that exceed the information if their prior about how health care vouchers increase the percent of tenants with improved health outcomes 1 year later exceed the analogous percent increase for lawyers (80% for high information, 20% for low information) (Appendix C.4).

[34] The figure restricts to a constant sample of Spectators whom we could have observed as welfarist. That is, if the information does not conflict with priors, then we cannot identify Spectators either way and they are not included.

meaningful. In fact, we find that they are just 18% less prevalent than welfarist preferences, even though assuming welfarist preferences is by far the norm in welfare economics.[35]

Inattention. A natural concern is that we mislabel Spectators as being non-welfarist or unclassified if they actually just fail to update due to inattention. Measurement error from inattention is undoubtedly present, but unlikely to change our conclusions that non-welfarist preferences are prevalent. First, we designed the belief updating task to require an active choice to update or not. Participants must select either: "**Yes**, I would like to revise my choices and give more people lawyers"; "**Yes**, I would like to revise my choices and give more people cash"; or "**No**, I would not like to revise my choices." It is not obvious that inattentive participants would choose not to update versus choose to update. In fact, inattention might work in the other direction, by leading true non-welfarists not to exhibit preferences for rights in both Experiments 1 and 2.

Second, participants are attentive overall (Section 3.3). This concern thus requires attention to lapse at precisely this elicitation and essentially nowhere else.

Finally, as our results are large in magnitude, an implausible amount of measurement error is required to undo them entirely. Suppose a full 50% of the Spectators whom we label as non-welfarists are actually inattentive welfarists (who pass two other attention checks). Even then, the share of non-welfarists to welfarists would still be quantitatively meaningful (about 30%).

Correlations with Anti-Targeting. Having categorized Spectators as welfarist and non-welfarist, we now return to whether these preferences predict anti-targeting. While Figure 3.5 suggests behaviors in Experiments 1–2 predict anti-targeting, it is not guaranteed that non-welfarist preferences, which also require not updating based on relevant information, are still predictive. Yet this concern is unfounded (Figure 3.6B, see Table C.11 for standard errors and hypothesis tests). 32% of anti-targeters are non-welfarist versus 23% of non-anti-targeters (p-value of difference $= 0.006$). Moreover, we also find a sharp drop in welfarist preferences among anti-targeters (19% versus 38%, p-value of difference < 0.001).[36] These correlations cast doubt on random elicitation errors as explaining the

[35]The difference of 4.8 pp (s.e.: 2.5, $p = 0.058$) indicates that welfarist preferences are more common than this stringent classification of non-welfarist preferences, but not overwhelmingly so.

[36]The result that non-welfarism is more prevalent among those who anti-target is mostly robust to alternative definitions and dropping those with low posteriors in Experiment 2 (Table C.11), and vice-versa with welfarism. Results are driven by lawyers, and are not significant for non-welfarist preferences if we focus on health care alone (Panels C–D). Some correlations between non-welfarism and targeting attenuate with different definitions of non-welfarism (Columns 1–2). This attenuation relative to Figure 3.5 is caused by: (1) focusing on the extensive margin (those with I_i or C_i larger than 0), as continuous measures are

large share of non-welfarists, since noise would not predict other choices.

3.5.3 Support for In-Kind Redistribution and Heterogeneity

Support for In-Kind Redistribution. We conclude the study by asking Spectators if they support (i) "right to counsel" policies that provide lawyers to tenants facing eviction, (ii) rent control, and (iii) universal health care. We then regress support for these policies on exhibiting preferences for rights in both Experiments 1–2. For right to counsel and rent control, we conduct this exercise among Spectators who did the experiments with lawyers. For universal health care, we conduct the exercise among Spectators who did the experiments with health care.

Rights preferences predict support for right to counsel and rent control, but not universal health care (Figure 3.7). For right to counsel and rent control, the relationship survives adding a control for whether the Spectator believes the policy would be effective, as well as for whether the person is a liberal.[37] Thus, preferences for rights, at least in the context of providing lawyers, predict policy support on top of welfarist/instrumental views about whether policy will help people. One explanation for why preferences for rights could be less predictive of support for universal health care is that this issue is more politicized.

As a final test, we ask Spectators if they agree there is a right to several types of in-kind goods. Strongly agreeing there is a right to in-kind goods like food, education, and housing is robustly correlated with having rights preferences (Figure C.7). The one exception is agreement with the view that there is a right to a lawyer in criminal cases.[38]

Demographic Heterogeneity. If the people who have preferences for rights were mostly rich, then non-utilitarian SWFs that aggregate preferences might still place a small value on preferences for rights. To the contrary, Spectators with preferences for rights are, if anything, less likely to be rich or well-educated (Figure C.6). We regress an indicator for expressing non-welfarist preferences in both Experiments 1–2 on demographics. Households with incomes larger than $60,000 per year have significantly smaller preferences for rights, driven by their choices over allocating lawyers.

robustly correlated with anti-targeting (Table C.9), and (2) the fact that the sample of people whom we can unambiguously label as welfarist or non-welfarist is smaller.

[37] We elicit beliefs about policy efficacy by asking whether people in the U.S. would be on average worse or better off with the policy.

[38] One explanation is that Spectators may view that question as a factual matter. It is the only good we ask about which is actually guaranteed in the U.S.

Other correlations are small. Notably, we find no correlation between having preferences for rights and being a self-reported liberal. Thus, heterogeneity does not support the hypothesis that preferences for rights reflec polarization or preferences among Democrats vs. Republicans.

3.5.4 Robustness Checks

Incentives. As health care was not incentivized, we embedded two complementary tests to see how much incentives might matter (see details in Appendix C.4.2). First, in our WTP elicitation where we ask Spectators to choose between giving a recipient cash or the good, we randomly assign half of the participants assigned to the three incentivized goods to have explicitly hypothetical framing. We reject even small effects of incentives for this elicitation (Table C.13).

Second, we randomize benchmark goods into being incentivized throughout all experiments (as in lawyers) or not incentivized (as in health care). In particular, we randomize benchmark goods into receiving identical language as those who see health care. Minor parts of the introduction to the study were different for health care to ensure truthfulness regarding incentivization. Comparing benchmark participants who see the health care language to those who see the lawyer language therefore jointly tests for the importance of incentives and the other language differences. We find no effect on behaviors among benchmark participants (Table C.12). As a consequence, Table C.12 also shows that rights goods versus benchmark goods differences also persist using only incentivized or only unincentivized benchmark participants.

The advantages of the first test are that: (i) it provides evidence that behaviors about allocating a treatment good (lawyers) are not affected by incentives, and (ii) the only difference in wording pertains to hypothetical versus not. The advantage of the second test is that it provides direct evidence about incentives for our main experiments, but only among the control group, and it bundles additional wording differences. We did not randomize incentives among lawyers in the main experiments because we did not know *ex ante* that incentives would have small effects, and we did not want to reduce power if they proved important. Together, these tests and results contribute to a conflicted literature on the importance of incentives in laboratory experiments (Andersen et al., 2011; Charness et al., 2021; Danz et al., 2022).

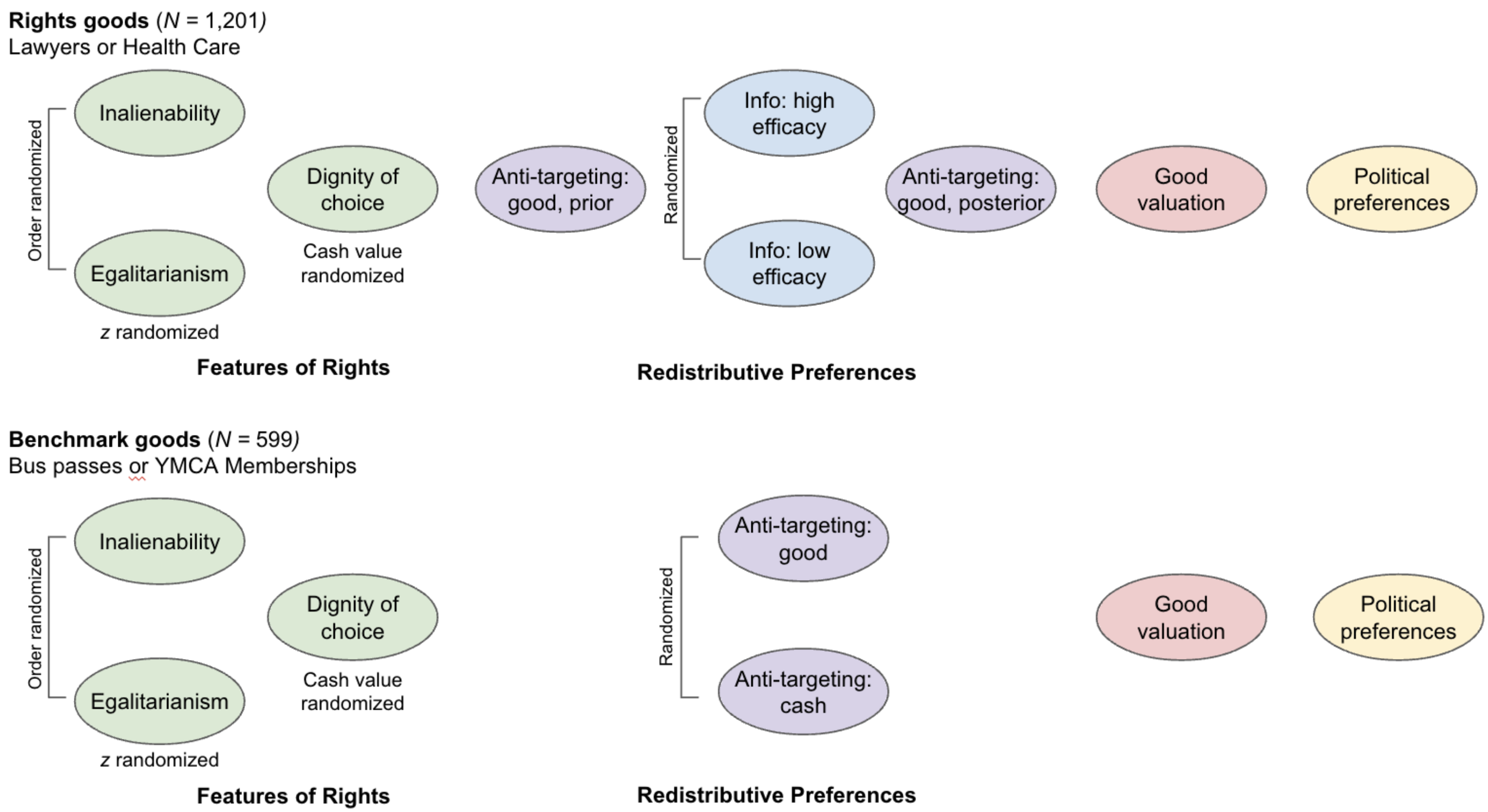
Other Tests. Results do not change if we use double/debiased machine learning (Chernozhukov et al., 2018) to select among the demographic controls in Table 3.1 (Table C.14).

3.6 Conclusion

Using several experiments, we document that preferences for rights are almost as common as the welfarist preferences that are the default in welfare economics. These preferences for rights are correlated with preferences to provide in-kind assistance universally. Our results may salvage the rights-based justification for in-kind assistance. Social Welfare Functions that depend only on the preferences of individuals in society still value rights in the domains of eviction defense lawyers and health care. Our experimental techniques are portable to other settings with potential non-welfarist preferences, for instance right to shelter or sustenance.

3.7 Figures

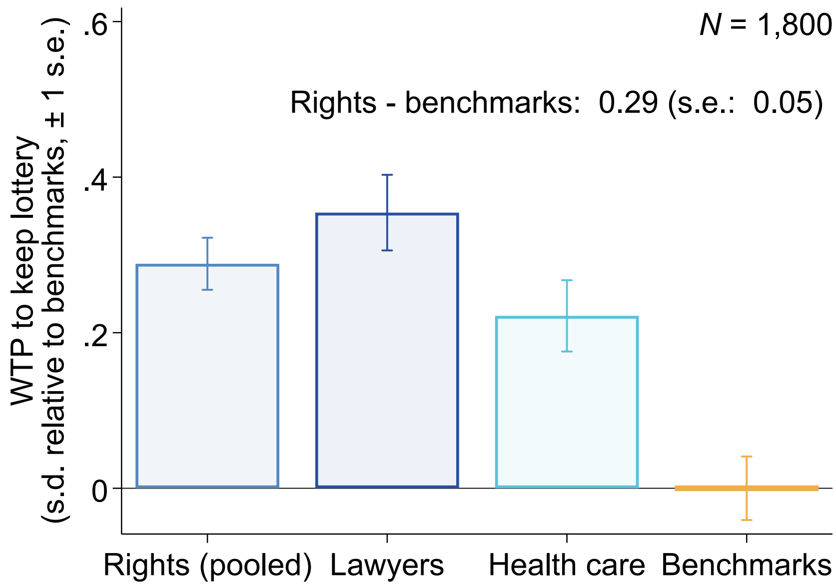
Figure 3.1: Survey Flow



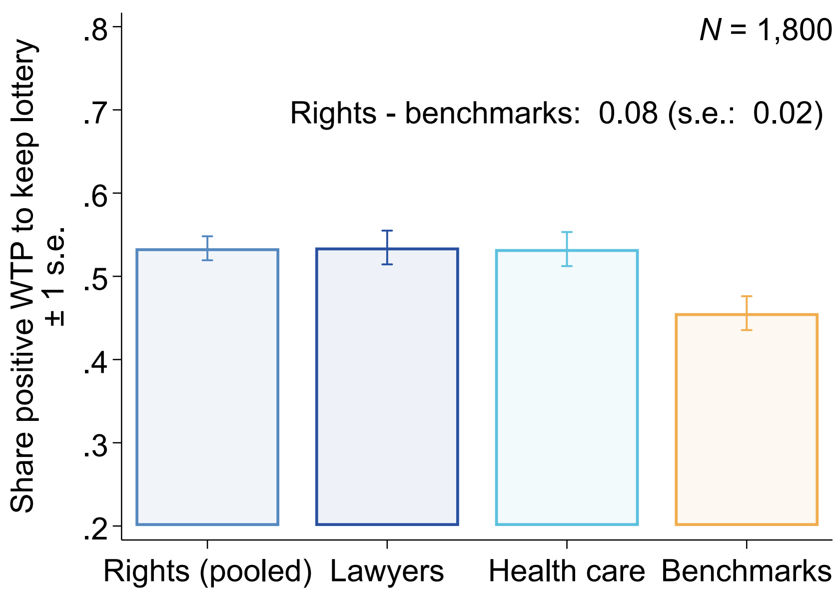
Note: This figure presents the flow of the survey, differentiating between those who saw rights goods (2/3) and those who saw benchmark goods (1/3). Within each arm, participants were randomly shown one of the two potential goods. After introductory information, participants first completed experiments related to features of rights. The order of *Inalienability* and *Egalitarianism* was randomized, followed by *Dignity of Choice*. Next, participants answered questions about *Anti-Targeting*; those assigned to rights goods also saw an information treatment. Finally, we elicited participants WTP for the good they were assigned, demographic information, and political preferences. Participants were required to pass two out of three attention checks.

Figure 3.2: Experiment 1: Inalienability

(a) WTP to keep lottery

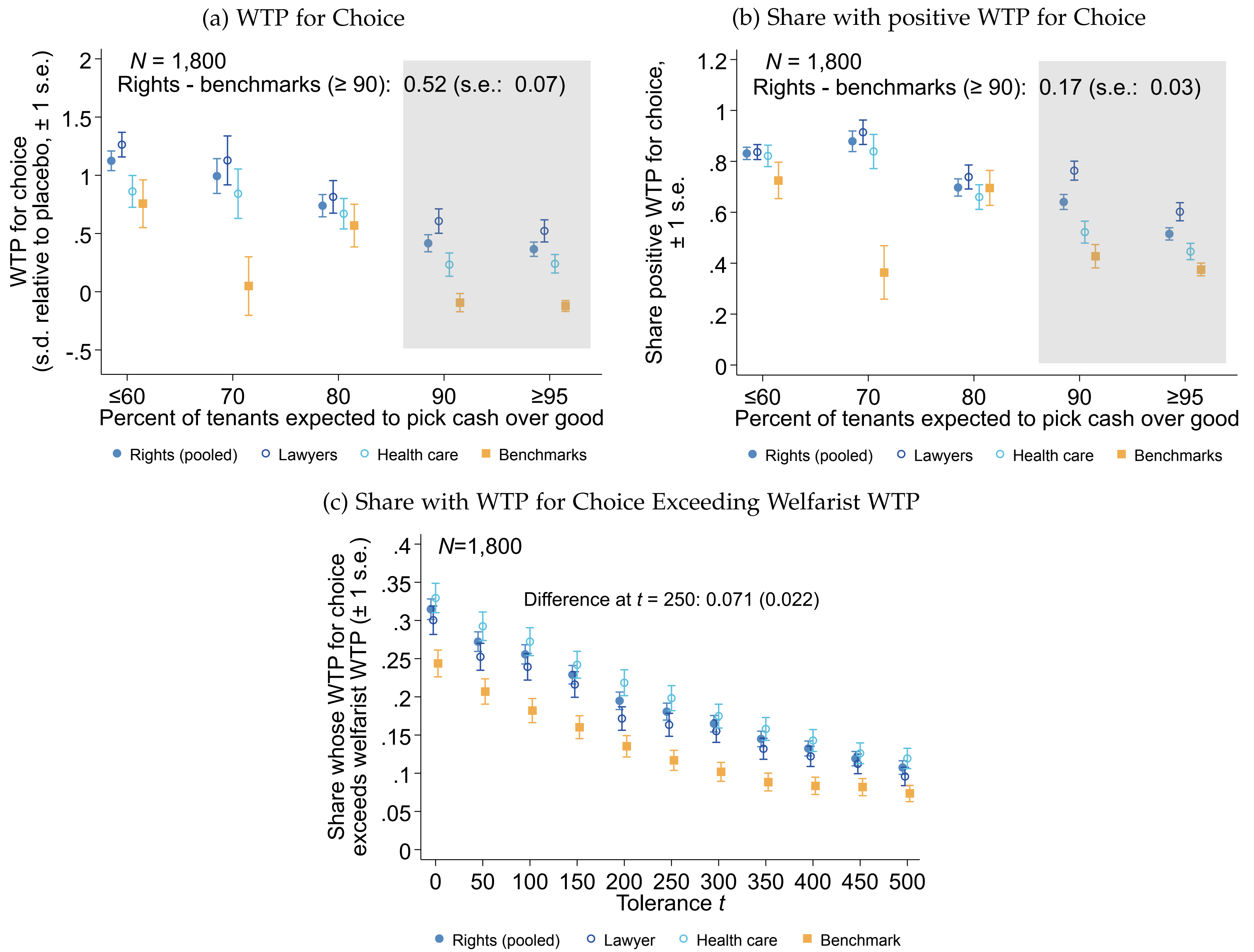


(b) Share with positive WTP to keep lottery



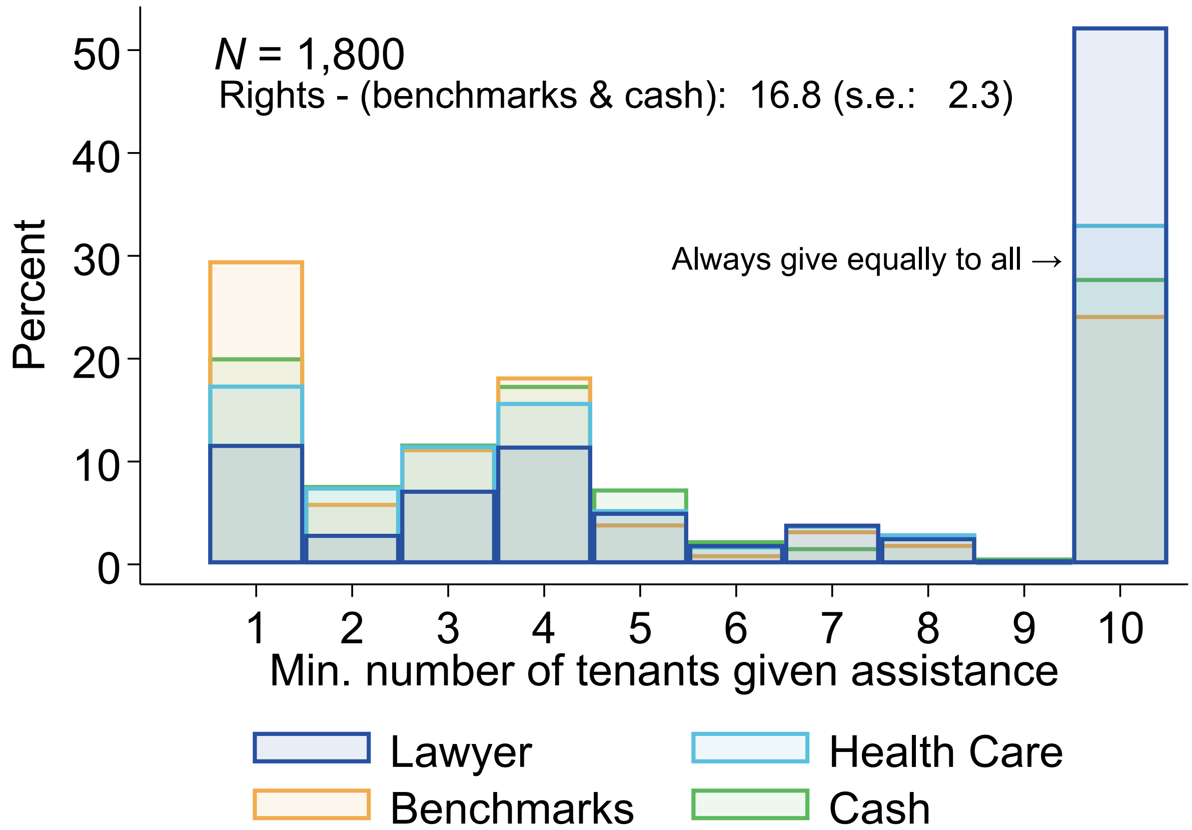
Note: This figure shows participants' decisions about keeping or rerunning a lottery in order to save money for future programs in Experiment 1 (Section 3.4.1). In panel A, we show WTP to keep the initial lottery result in standard deviations relative to the benchmark goods. The far left blue bar shows both rights goods pooled, while the next two bars show results for lawyers and health care disaggregated. Panel B shows the share of respondents with positive WTP to keep the lottery by good. Both panels include ± 1 standard errors. See Table 3.2 for detailed regression results.

Figure 3.3: Experiment 2: Dignity of Choice



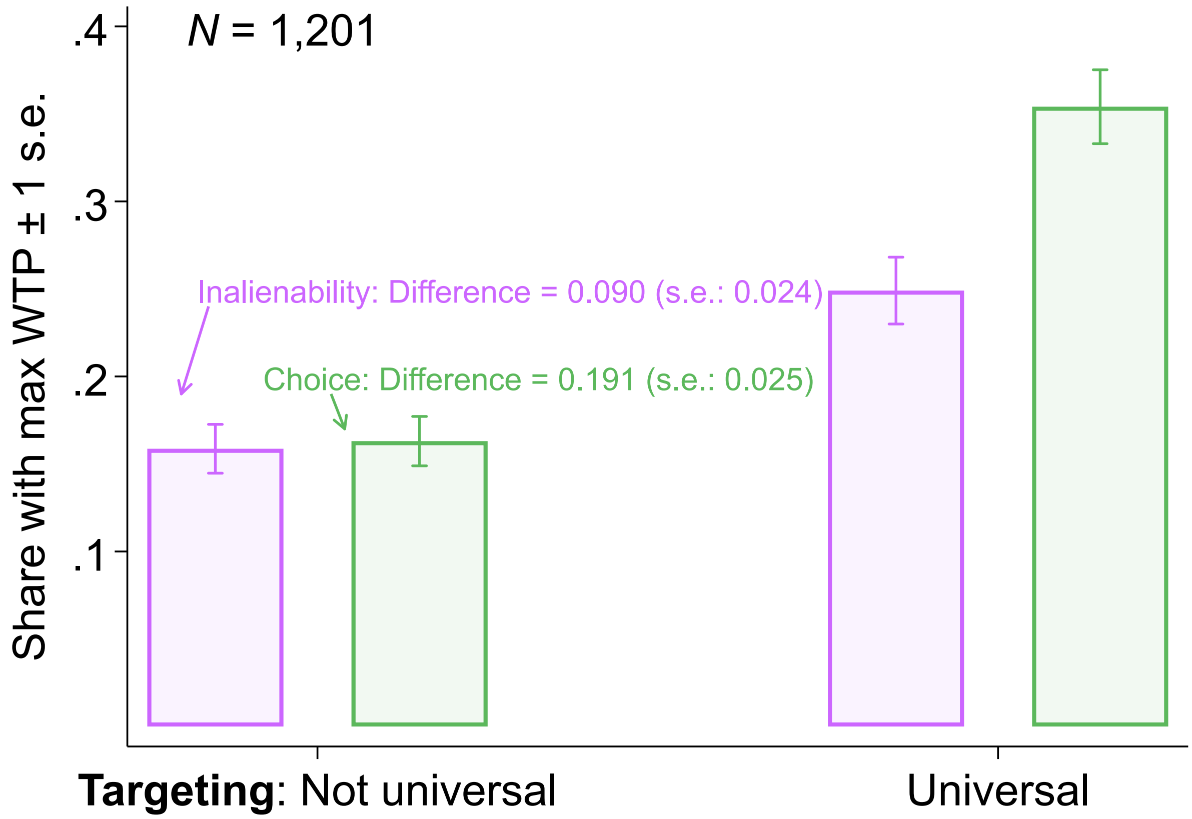
Note: This figure shows participants' decisions about giving tenants a choice between cash and a good, versus saving money for future programs in Experiment 2 (Section 3.4.2). Panel A plots participants' WTP to give tenants a choice in standard deviations relative to the benchmark goods (±1 s.e.). We show this for different posterior beliefs about the percent of tenants expected to pick cash over the good. The shaded gray area emphasizes those with high posteriors (≥ 90%). See Table 3.3 for detailed regression results. Panel B plots the same for the share with positive WTP. Panel C shows the share of participants whose WTP for choice exceeds their welfarist WTP (Equation 3.11) as a function of the tolerance t.

Figure 3.4: Anti-Targeting



Note: This figure shows the distribution of the minimum number of tenants given assistance in Experiment 4 (Section 3.5.1). For example, if a participant chose to give two tenants the good and cash, but to give everyone a lawyer rather than giving one tenant the good and cash, the minimum number of tenants given assistance is two. If the minimum is 10, the participant always chose to give the good to everyone. The blue series show distributions for lawyers and health care, the orange series shows the distribution for benchmark goods, and the green series shows the distribution for cash; half of participants who saw benchmark goods throughout the survey were asked about distributing cash in this experiment. See Table 3.4 for detailed regression results.

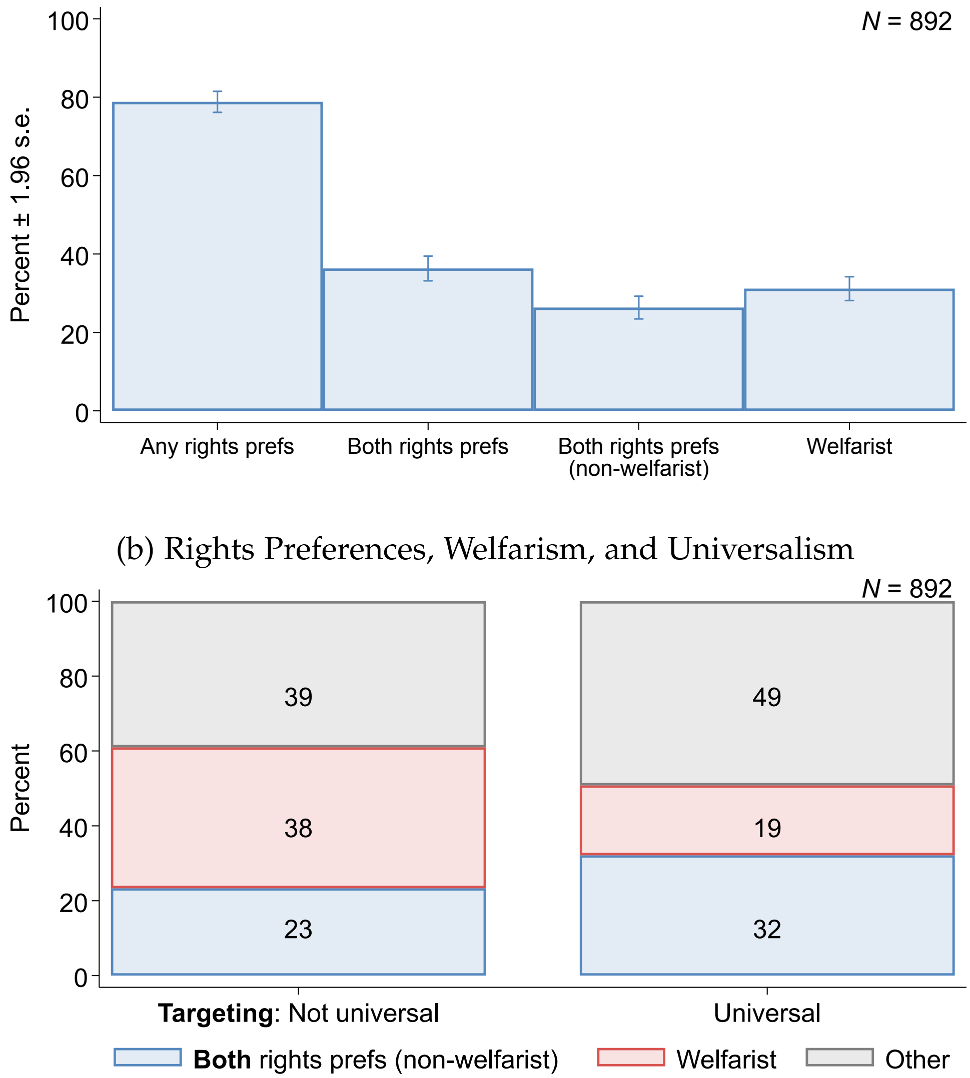
Figure 3.5: Correlations Between Universalism and Features of Rights



Note: This figure shows the share of participants with the maximum possible WTP in the inalienability (pink) and dignity of choice (green) experiments. Shares are split by whether participants distributed the good universally in Experiment 4. The sample here is restricted to those who saw rights goods (lawyers or health care). See Table 3.5 for detailed regression results.

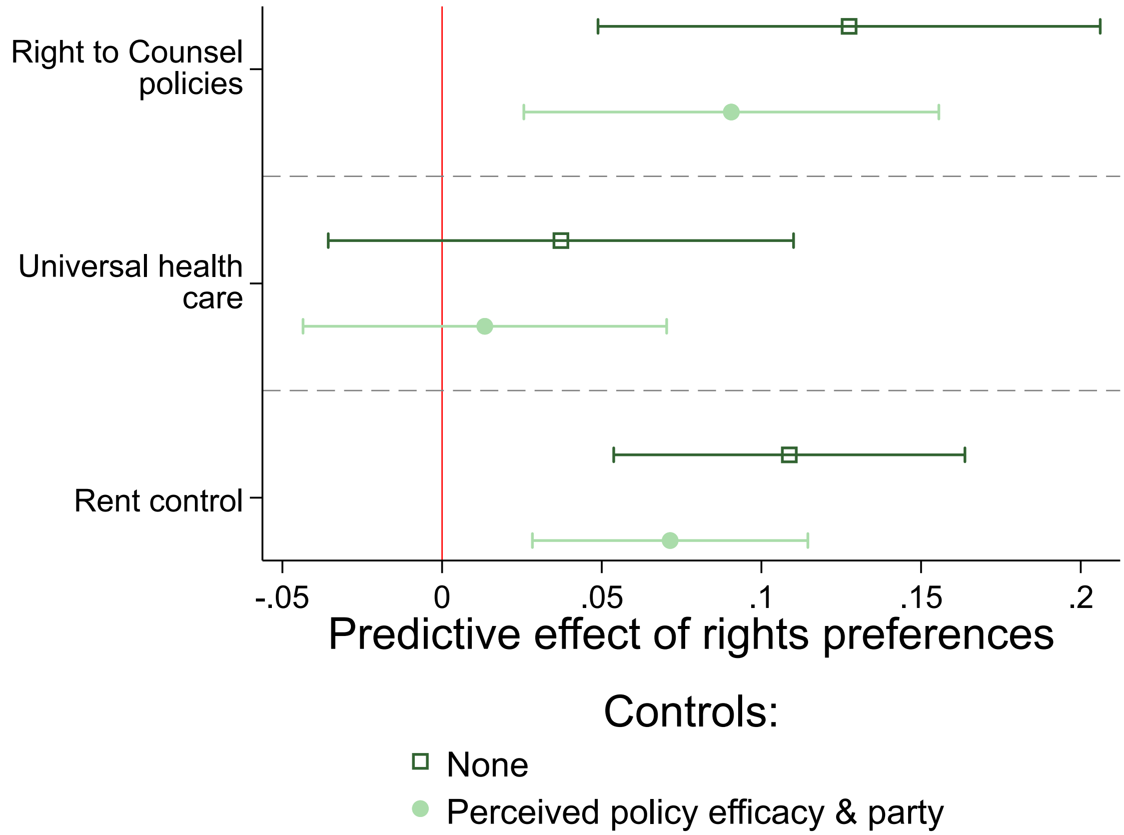
Figure 3.6: Rights Preferences and Welfarism

(a) Prevalence



Note: Panel A presents the prevalence of rights based preferences and welfarism according to several definitions. First, *Any rights prefs* includes those that have a positive WTP in either Experiment 1 or 2. Second, *Both rights prefs* refers to those who with positive WTP in *both* Experiment 1 and 2. Third, *Both rights prefs (Non-Welfarist)* adds the restriction of non-welfarism. *Welfarist* participants, fourth, are those who change their choices about targeting when information about efficacy of lawyers or health care disagrees with their priors. Panel B plots the share of *Both rights prefs (Non-Welfarist)*, *Welfarist*, and *Other* participants among those who did and did not distribute the good universally in Experiment 4. See Table C.11 for detailed regression results. Both panels restrict to a constant sample of participants who saw rights goods, and further to those we can classify as welfarist or non-welfarist based on their prior beliefs.

Figure 3.7: Political Preferences



Note: This figure shows the predictive effect of rights preferences—having positive WTP in both Experiment 1 and 2—on support for specific policies. We show results for *Right to Counsel policies* among those who did the experiment with lawyers, *Universal health care* among those who did the experiment with health care, and *Rent control* among those who did the experiment with either lawyers or health care. Estimates shown with ± 1.96 standard errors.

Table 3.1: Demographics and Balance

	(1) U.S.	(2) Experimental sample	(3) Rights -Benchmarks
White non-Hispanic	0.66	0.71	-0.01
	[0.47]	[0.45]	(0.02)
Income > 60k	0.24	0.48	-0.03
	[0.43]	[0.50]	(0.02)
Less than Bachelor's	0.67	0.45	-0.00
	[0.47]	[0.50]	(0.02)
Female	0.51	0.51	-0.02
	[0.50]	[0.50]	(0.02)
Less than age 40	0.33	0.56	0.00
	[0.47]	[0.50]	(0.02)
Liberal		0.57	0.05
		[0.49]	(0.02)
Legal case without a lawyer		0.16	0.01
		[0.36]	(0.03)
Urgent health issue without HC		0.31	-0.02
		[0.46]	(0.03)
F-statistic			.996
p-value			0.437
Observations	2,624,206	1,800	1,800

Note: This table shows the composition of our experimental sample relative to all U.S. adults (18+, a requirement on Prolific) in the 2021 ACS (Ruggles et al., 2023). Column (3) show differences in participants assigned to lawyers and health care compared to those assigned to benchmarks from an OLS regression. The *F*-statistic is from a joint test of significance for the listed demographic variables. Brackets show standard deviations. Parentheses show robust standard errors.

Table 3.2: Tests of Inalienability

	(1) Non-zero WTP (= 1)	(2) Max WTP (= 1)	(3) WTP (s.d.)	(4) WTP (s.d.)	(5) WTP (s.d.)
Rights good (= 1)	0.078	0.080	0.289		
	(0.025)	(0.017)	(0.053)		
	[0.002]	[0.000]	[0.000]		
Lawyers (= 1)				0.354	
				(0.064)	
				[0.000]	
Health care (= 1)					0.222
					(0.061)
					[0.000]
Raw mean (benchmarks)			58.4	58.4	58.4
Raw s.d. (benchmarks)			69.9	69.9	69.9
Mean (benchmarks)	0.456	0.117	-0.000	-0.000	-0.000
Observations	1,800	1,800	1,800	1,205	1,194

Note: This table shows the effects of being assigned to a rights good on three measures of WTP for inalienability (Equations 3.7 and 3.8). Columns (1)-(3) pool lawyers and health care, while column (4) shows results for lawyers and column (5) shows results for health care. Parentheses show robust standard errors. In columns (3)-(5), WTP is reported in standard deviations relative to the benchmark goods. Parentheses show robust standard errors. Brackets show *p*-values.

Table 3.3: Tests of Dignity of Choice

				Posterior ≥ 0.9			Posterior $= 1$		
	(1) Non-zero WTP (= 1)	(2) Max WTP (= 1)	(3) WTP (s.d.)	(4) Non-zero WTP (= 1)	(5) Max WTP (= 1)	(6) WTP (s.d.)	(7) Non-zero WTP (= 1)	(8) Max WTP (= 1)	(9) WTP (s.d.)
Rights good (= 1)	0.171	0.123	0.495	0.166	0.139	0.518	0.072	0.153	0.516
	(0.025)	(0.018)	(0.057)	(0.030)	(0.020)	(0.067)	(0.047)	(0.033)	(0.109)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.122]	[0.000]	[0.000]
Posterior	-0.442	-0.244	-1.046						
	(0.044)	(0.047)	(0.135)						
	[0.000]	[0.000]	[0.000]						
Raw mean (benchmarks)			223.0			223.0			223.0
Raw s.d. (benchmarks)			277.4			277.4			277.4
Mean (benchmarks)	0.432	0.092	0.000	0.432	0.092	0.000	0.432	0.092	0.000
Observations	1,800	1,800	1,800	1,058	1,058	1,058	459	459	459

Note: This table shows the effects of being assigned to a rights good on three measures of WTP for dignity of choice (Equation 3.7). Columns (1)-(3) show results for the whole sample controlling for posterior beliefs about the percent of tenants who will choose cash over the good. Columns (4)-(6) restrict to those with posteriors beliefs greater than or equal to 90% and columns (7)-(9) restrict to those with posteriors of 100%. In columns (3), (6) and (9), WTP is reported in standard deviations relative to the benchmark goods. Parentheses show robust standard errors. Brackets show *p*-values.

Table 3.4: Tests of Anti-Targeting

	(1) Universal (= 1)	(2) No. Tenants	(3) Universal (= 1)	(4) No. Tenants	(5) Universal (= 1)	(6) No. Tenants
Rights good (= 1)	0.044	0.458				
	(0.024)	(0.187)				
	[0.067]	[0.015]				
Lawyers (= 1)			0.116	1.007		
			(0.034)	(0.252)		
			[0.001]	[0.000]		
Health care (= 1)					0.013	0.215
					(0.026)	(0.204)
					[0.621]	[0.293]
WTP for good FE	✓	✓	✓	✓	✓	✓
Mean (benchmarks)	0.260	4.855	0.260	4.855	0.260	4.855
Observations	1,800	1,800	1,205	1,205	1,194	1,194

Note: This table shows the effects of being assigned to a rights good on choices in the anti-targeting experiment (Equations 3.7 and 3.8). The outcome columns (1), (3) and (5) is an indicator for whether the participant provided the good universally. The outcome columns (2), (4) and (6) is the minimum number of tenants the participant distributed to, which ranges between 1 (preferring to giving the poorest tenant the good and cash) and 10 (always preferring universal provision). Columns (1)-(2) pool lawyers and health care, while columns (3)-(4) shows results for lawyers and column (5)-(6) shows results for health care. Parentheses show robust standard errors. Brackets show p-values.

Table 3.5: Tests of Correlations with Universalism

	Dep. Var.: Universal (= 1)			
	(1)	(2)	(3)	(4)
Panel A. Rights goods (pooled)				
WTP for inalienability (s.d.)	0.031			0.006
	(0.012)			(0.012)
	[0.013]			[0.626]
WTP for dignity of choice (s.d.)		0.086		0.085
		(0.011)		(0.011)
		[0.000]		[0.000]
Mean WTP (s.d.)			0.098	
			(0.014)	
			[0.000]	
Panel B. Lawyers				
WTP for inalienability (s.d.)	0.031			-0.002
	(0.017)			(0.017)
	[0.067]			[0.911]
WTP for dignity of choice (s.d.)		0.094		0.095
		(0.015)		(0.015)
		[0.000]		[0.000]
Mean WTP (s.d.)			0.100	
			(0.019)	
			[0.000]	
Panel C. Health care				
WTP for inalienability (s.d.)	0.021			0.009
	(0.018)			(0.018)
	[0.238]			[0.612]
WTP for dignity of choice (s.d.)		0.057		0.056
		(0.016)		(0.016)
		[0.000]		[0.001]
Mean WTP (s.d.)			0.069	
			(0.021)	
			[0.001]	
Mean	0.428	0.428	0.428	0.428
Observations	1,201	1,201	1,201	1,201

Note: This table shows the effect of WTP for inalienability, WTP for dignity of choice, and mean WTP on universal provision in the anti-targeting experiment. All WTPs are in terms of s.d. relative to the benchmark goods. The outcome is an indicator for whether the participant distributed the good universally. The sample is those who did the experiment with rights goods: Panel A shows pooled results, while Panels B and C show lawyers and health care, respectively. Parentheses show robust standard errors. Brackets show p-values.

Appendix A

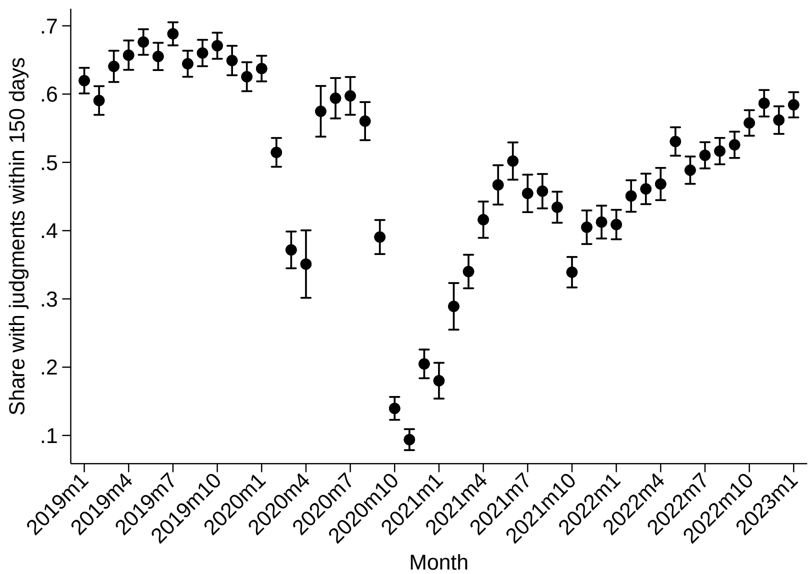
Appendix to Chapter 1

A.1 Additional Figures and Tables

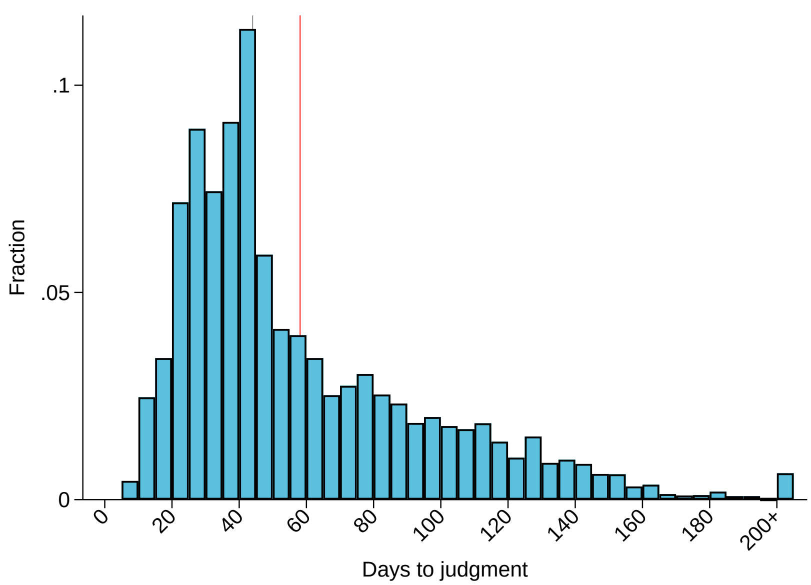
A.1.1 Figures

Figure A.1: Eviction Prevalence: Descriptive Statistics

(a) Share with Judgments over Time



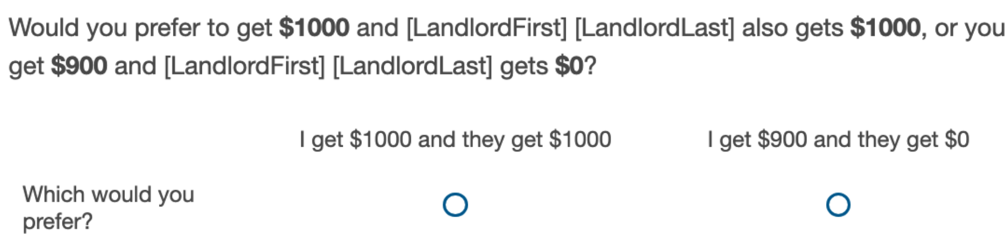
(b) Time between Filing and Judgment



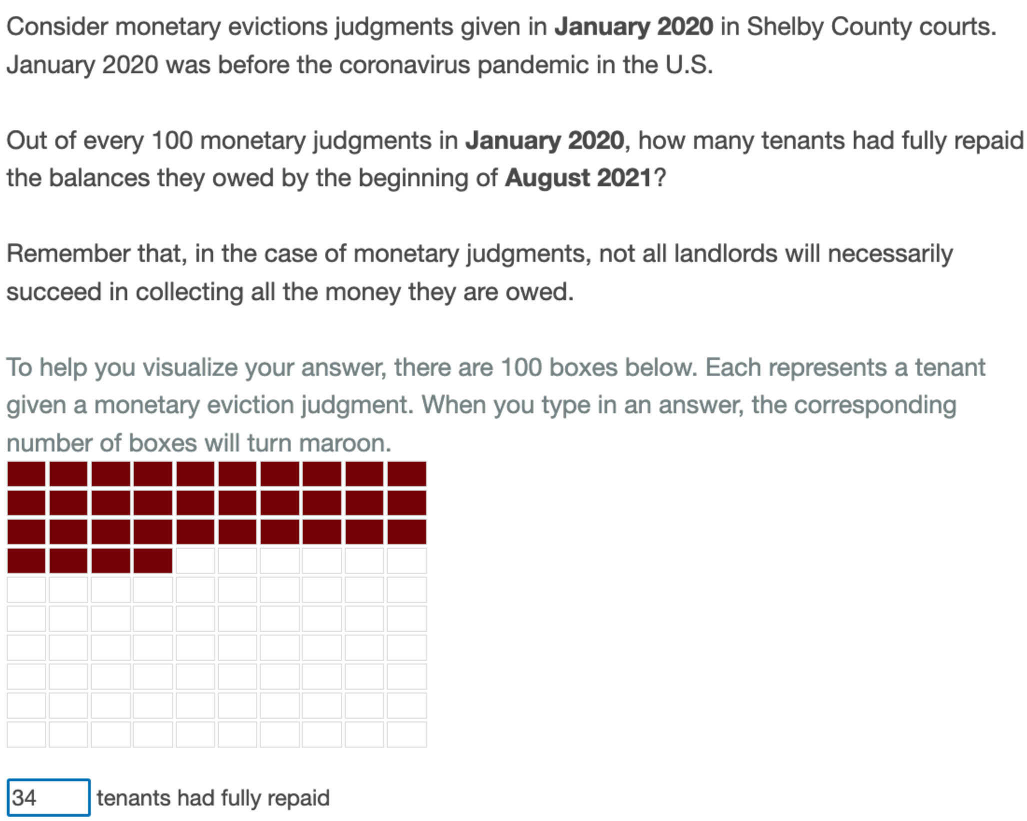
Note: Panel A shows the share of filings that result in judgments within 150 days over time. Panel B shows the distribution of time elapsed between filings and judgments, if a judgment occurs.

Figure A.2: Experiment Screenshots

(a) Example Elicitation: Dictator Game (Tenants)



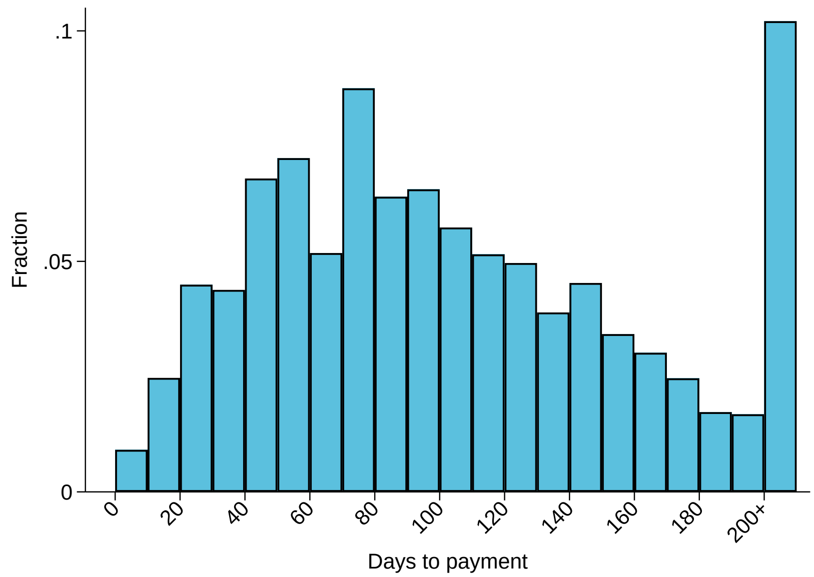
(b) Example Elicitation: Belief Elicitation (Landlords)



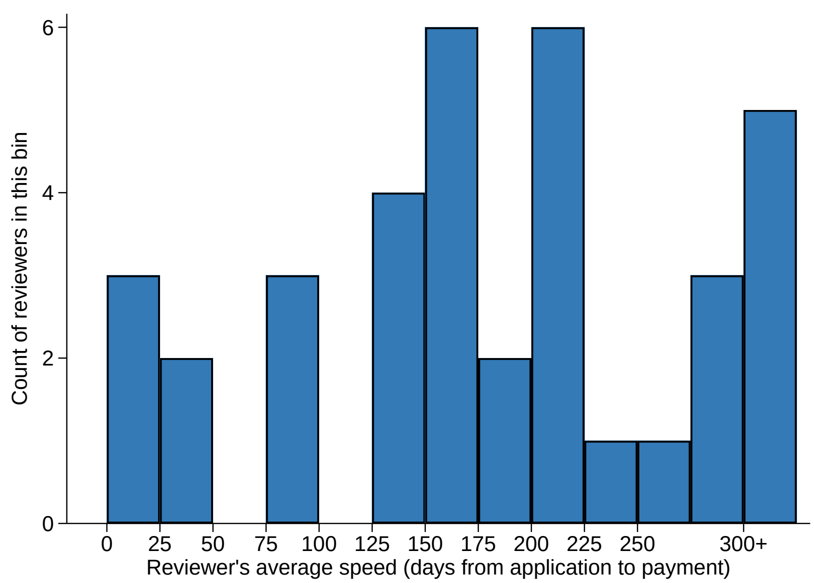
Note: Panel A shows a screenshot of one question asked in a multiple price list in the DG played among tenants. Stakes are $1,000. The elicitation iterates between questions of this type until we find the participant's indifference point. Panel B shows a screenshot of how we elicit average beliefs among landlords. The subsequent screen is a confirmation check.

Figure A.3: ERAP Sample Statistics

Panel A: Filings, Judgments, and Applications



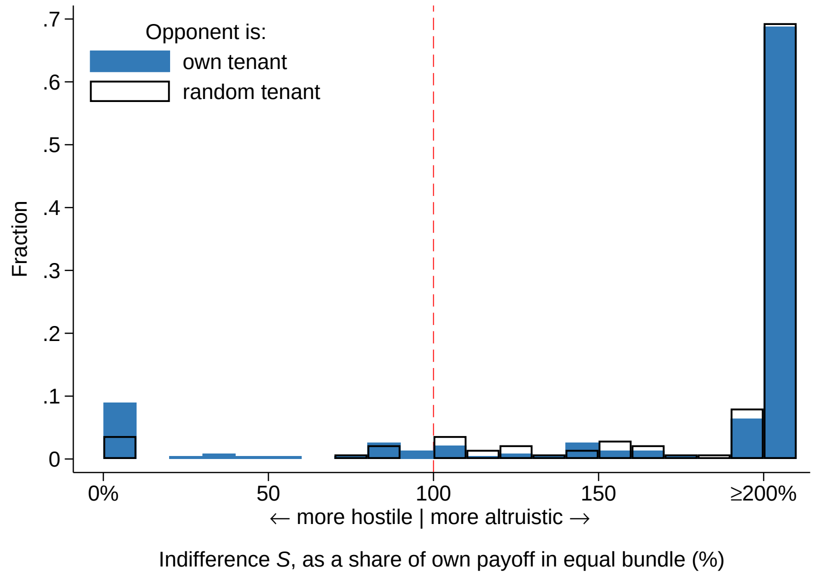
Panel B: Average Time from Application to Payment, by Reviewer



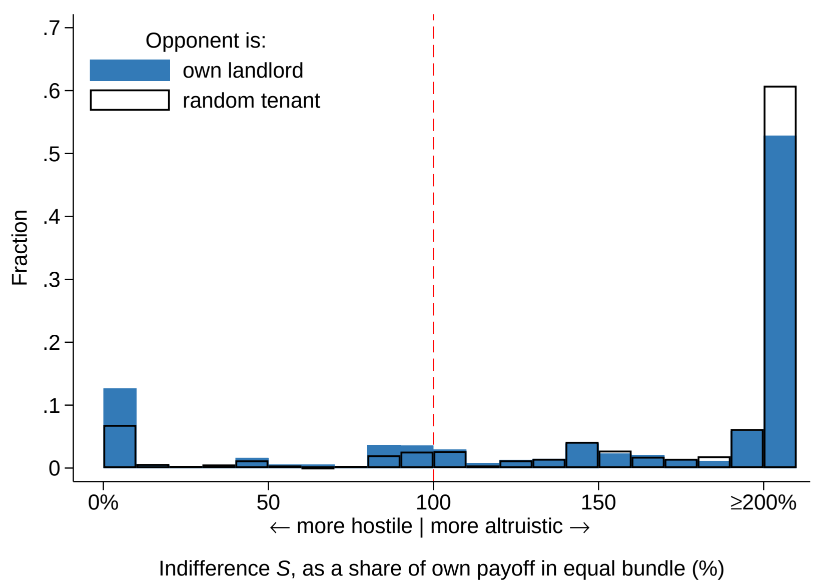
Note: Panel A shows the distribution of time until payment, conditional on application. Panel Bok shows the number of reviewers who take the indicated number of days to review cases, on average.

Figure A.4: Behavior in Modified Dictator Game: Histograms

(a) Landlord participants ($N = 371$)



(b) Tenant participants ($N = 1{,}808$)

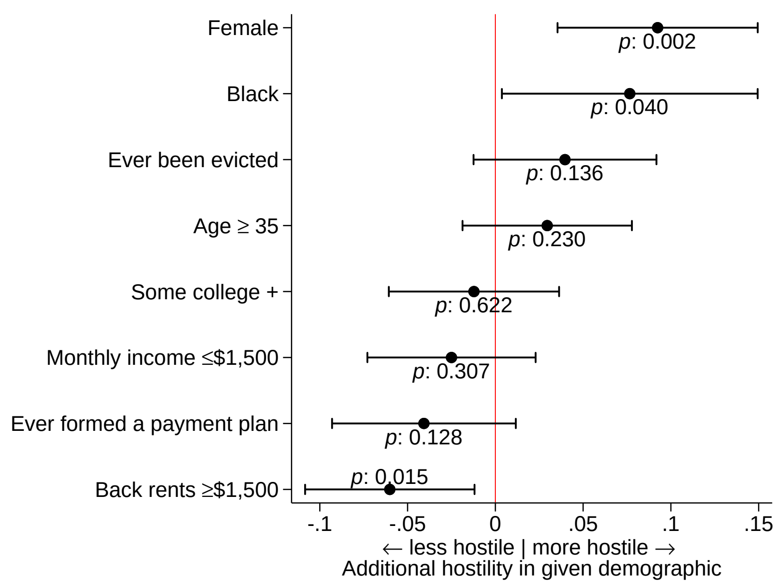


Note: This figure presents histograms of the distribution of the indifference point $S(x)$ in the dictator game for landlord (Panel A) and tenant (Panel B) participants, rescaled as a percentage of x. $S(x)$ represents the point at which a participant is indifferent between the bundle (\$$s$ self, \$0 other) and (\$$x$ self, \$$x$ other). If $S(x) < x$, then the player is hostile; if $S(x) > x$, then the player is altruistic. The horizontal axis presents $100 \times S(x)/x$.

Figure A.5: Dictator Game: Heterogeneity

(a) Landlords

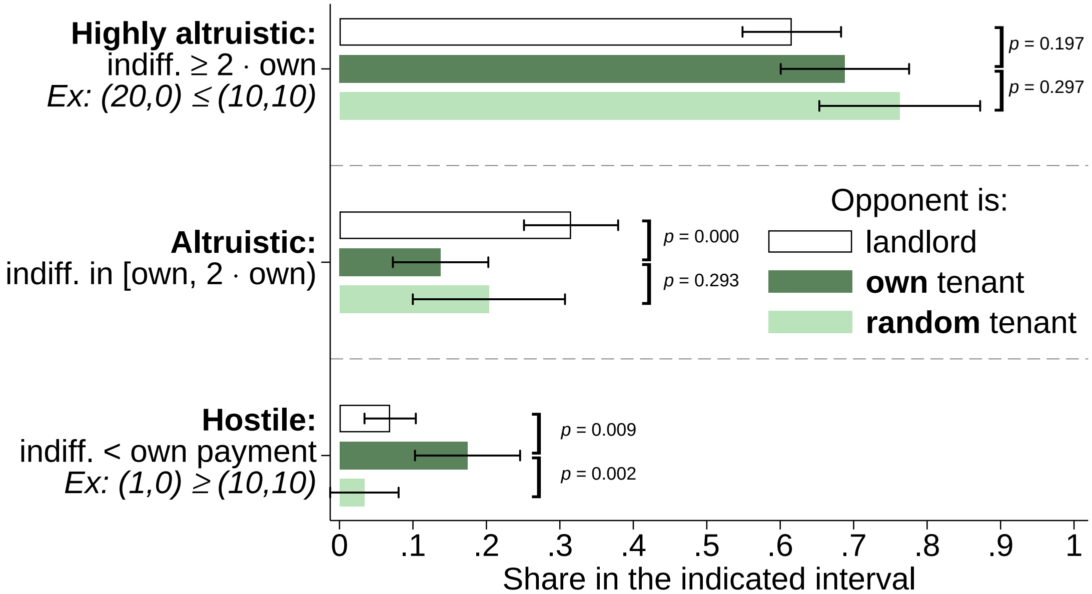
(b) Tenants



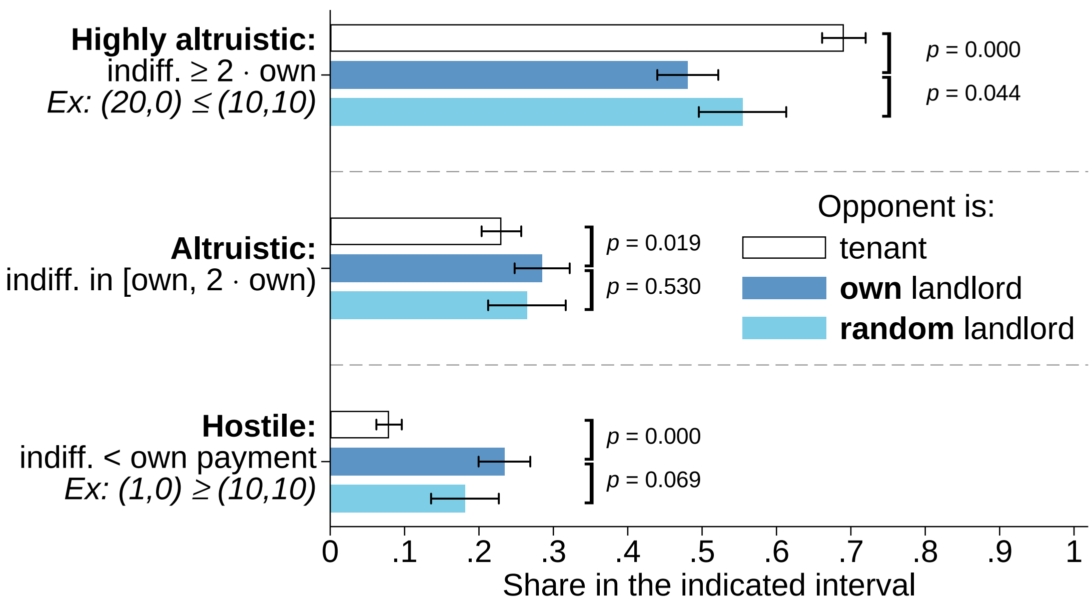
Note: This figure shows heterogeneity in our measures of hostility among landlords (Panel A) and tenants (Panel B). We elicit the point $S(x)$ at which a participant is indifferent between the bundle (\$$s$ self, \$0 other) and (\$$x$ self, \$$x$ other). If $S(x) < x$, then the player is hostile; if $S(x) > x$, then the player is altruistic. In Panel A, we interact our measure of hostility toward own tenants with the indicated demographic. In Panel B we interact our measure of hostility toward own landlords with the indicated demographic. Whiskers show 95-percent confidence intervals.

Figure A.6: Behavior in Modified Dictator Game: First DG Only

(a) Landlord participants

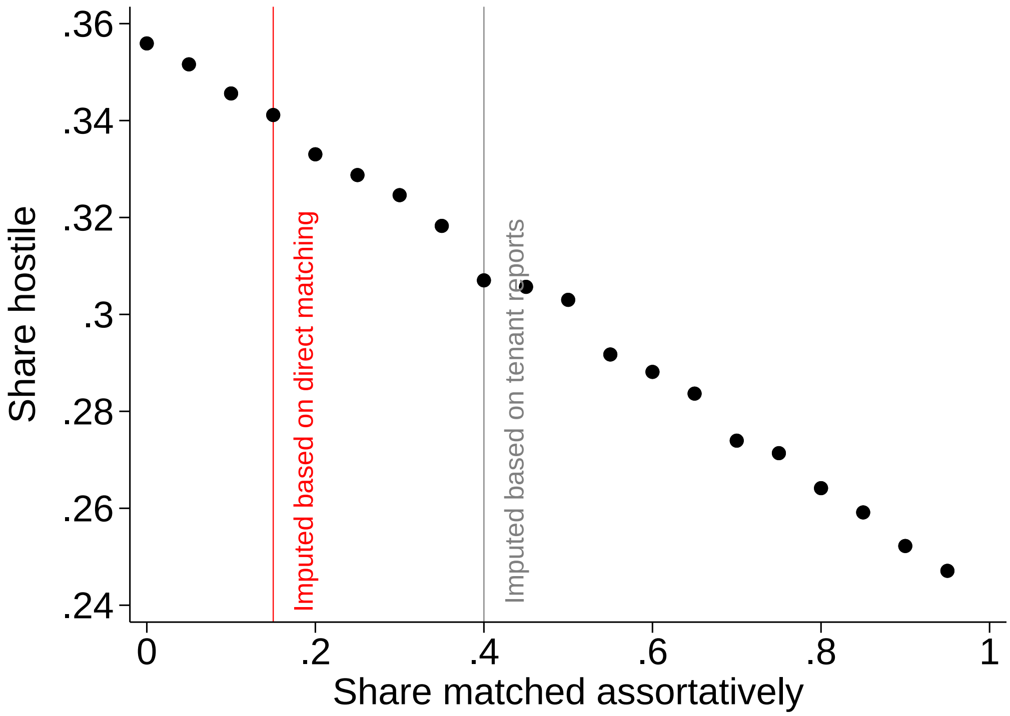


(b) Tenant participants



Note: This figure summarizes the results from the Dictator Games among landlord participants (Panel A) and tenant participants (Panel B). The figure is identical to Figure 1.2 except it uses only the first instance that each participant plays the DG. See Table A.17 for sample sizes and additional details.

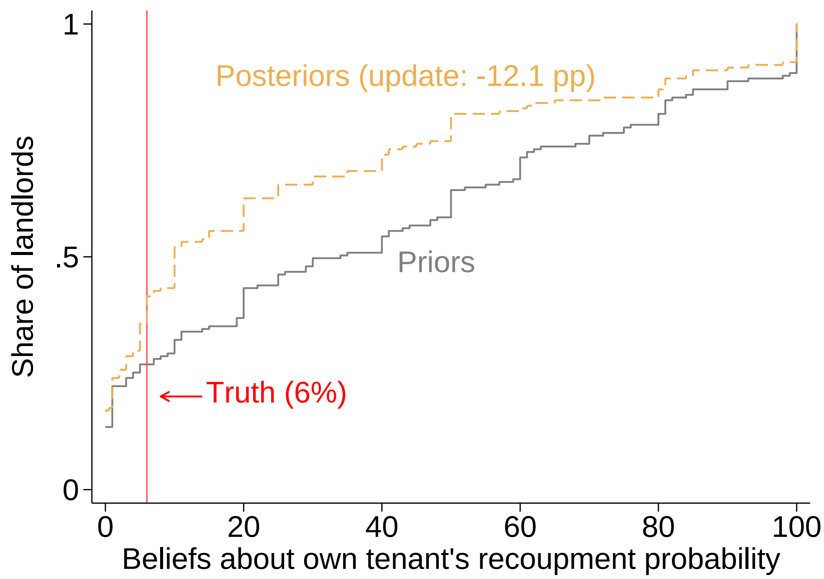
Figure A.7: Assortative Matching: Simulations



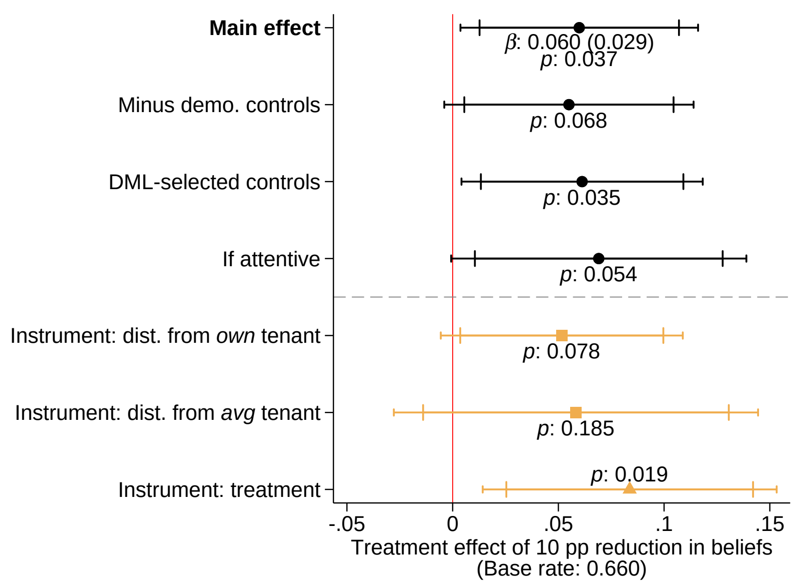
Note: We produce this figure by the process in Section A.5.2. Each point represents the mean of 25 simulations, where we block re-simulate the entire process. The vertical lines indicate the amount of assortative matching that generates the coefficients described in Section A.5.2 in the simulated sample.

Figure A.8: Posteriors and IV: Landlord Information

(a) Posteriors about recoupment



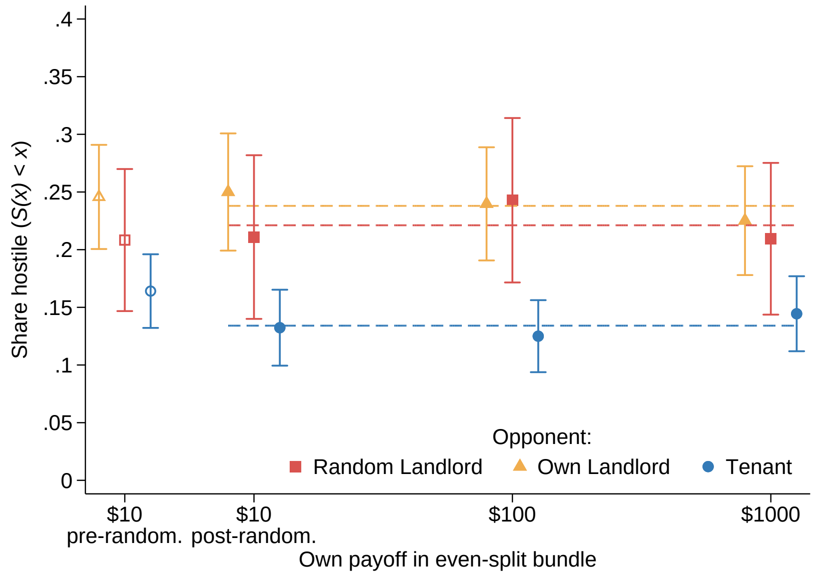
(b) IV: requests materials



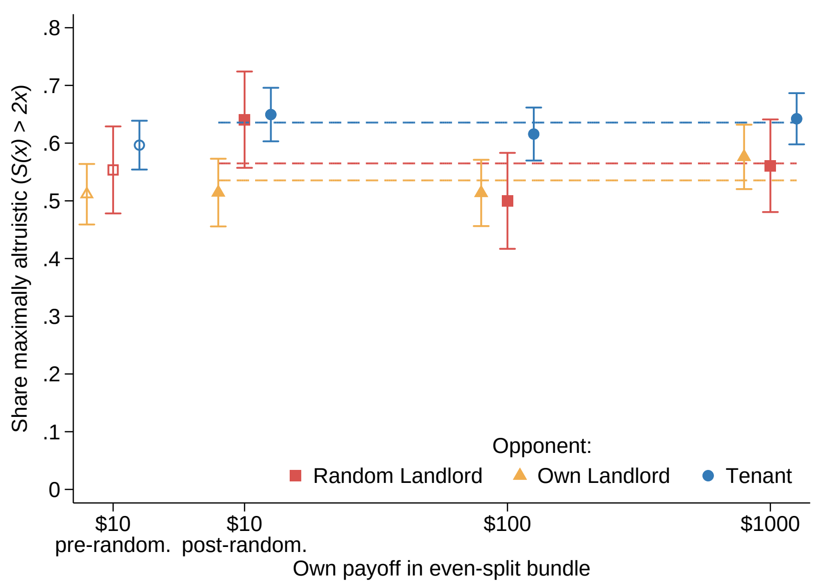
Note: Panel A includes only treated individuals and shows the prior and posterior beliefs about the treatment effect. Panel B shows the effect of beliefs on the outcome, instrumenting for the belief update.

Figure A.9: The effect of stakes on tenant hostility and altruism

(a) Hostility



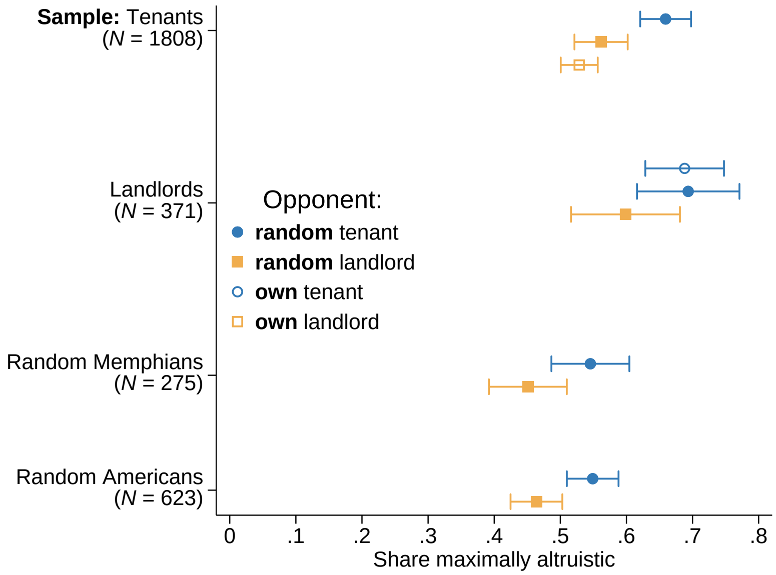
(b) High altruism



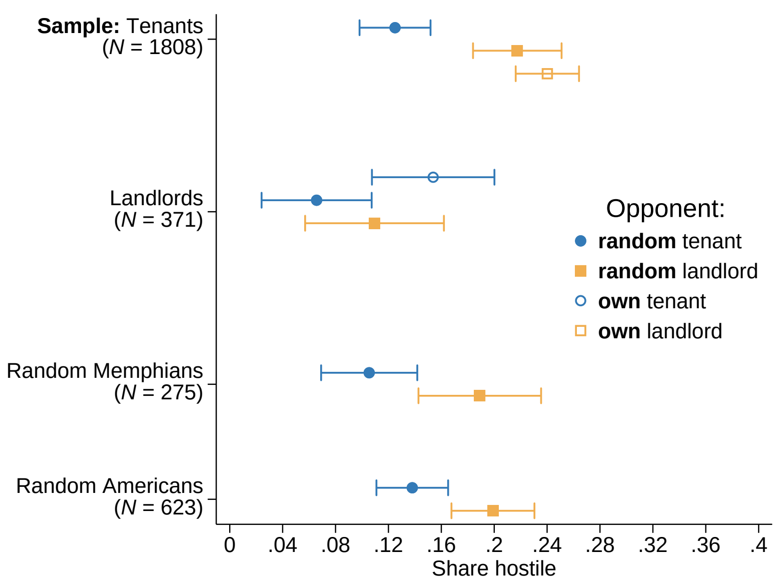
Note: This figure presents the effect of randomizing stakes on tenant behavior in the Dictator Game. We elicit bounds on the the point $S(x)$ at which a participant is indifferent between the bundle (\$$s$ self, \$0 other) and (\$$x$ self, \$$x$ other). If $S(x) < x$,then the player is hostile; if $S(x) > x$, then the player is altruistic; if $S(x) > 2x$, then the player is highly altruistic. We randomize $x \in \{\$10, \$100, \$1000\}$ for tenants starting March 27, 2022. Prior to March 27, $x = 10$ for all tenants. We show the share hostile when the opponent is the tenant's own landlord (orange series), random landlord (red series), and random tenant (blue series). Because the elicitation changed slightly once we randomize stakes (Appendix C.4), we disaggregate the data for $x = 10$ in the two leftmost points to show whether these subtle survey changes affected behavior when the stakes were the same.

Figure A.10: Tenants, Landlords, Random Memphians, and Random Americans

(a) High altruism

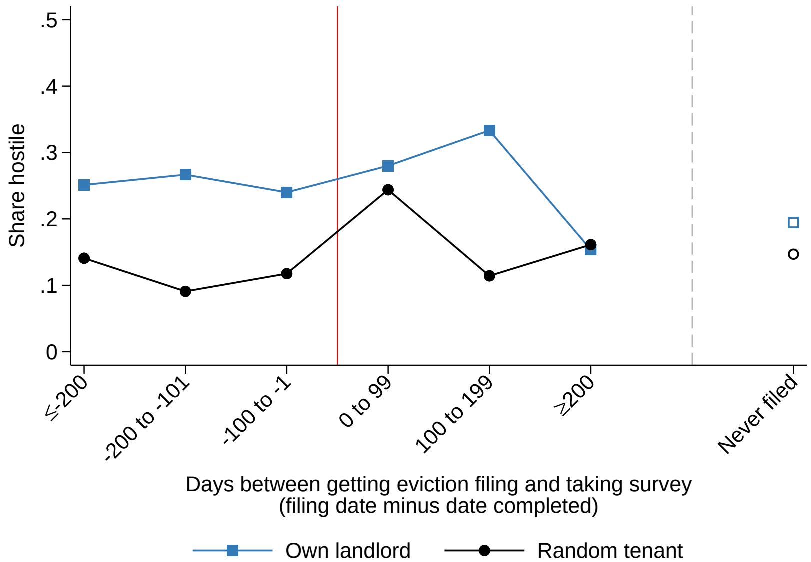


(b) Hostility



Note: This figure summarizes the share highly altruistic (Panel A) or hostile (Panel B) across four samples: the landlord sample and the tenant sample, as well as a sample of random Memphis residents and random Americans. We obtain the Memphis and American samples from survey provider Lucid. We elicit bounds on the the point $S(x)$ at which a participant is indifferent between the bundle (\$$s$ self, \$0 other) and (\$$x$ self, \$$x$ other). If $S(x) < x$,then the player is hostile; if $S(x) > x$, then the player is altruistic; if $S(x) > 2x$, the player is highly altruistic. Among the Memphis and American samples, we elicit $S(x)$ when the opponent is a random unnamed landlord or a random unnamed tenant only, to avoid identifying research subjects.

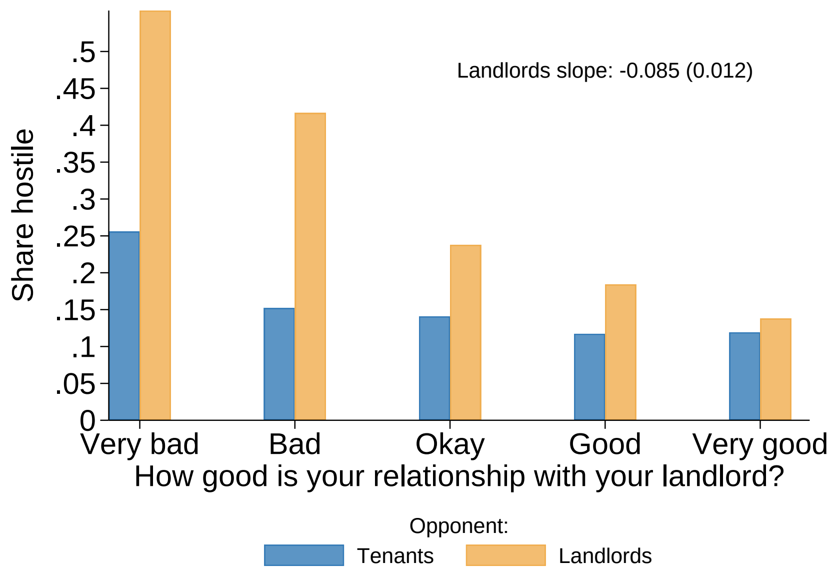
Figure A.11: Tenant Hostility by Period between Filing and Survey



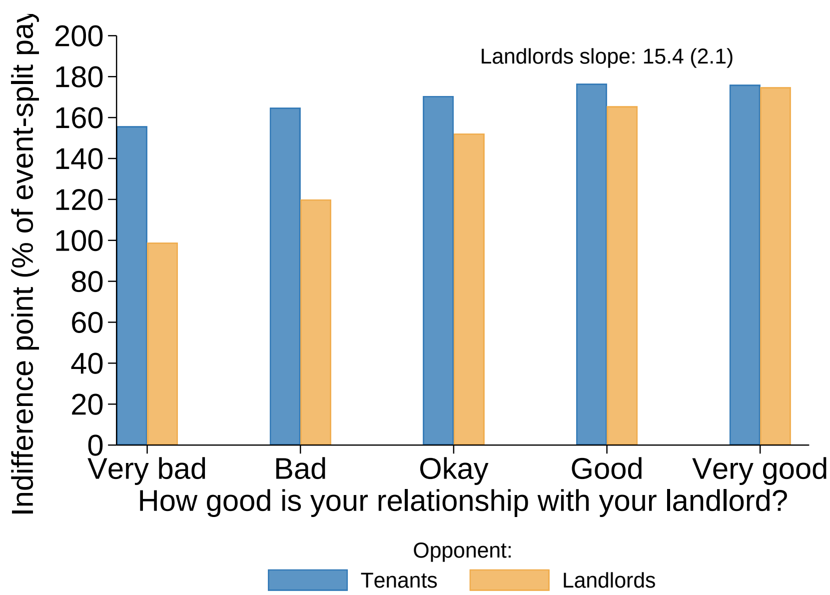
Note: The figure shows tenant hostility by the number of days until a filing, using the sample in Section 1.6.1.

Figure A.12: Tenants' Self-reported Relationship with Landlord and Hostility and Indifference Points

(a) Hostility

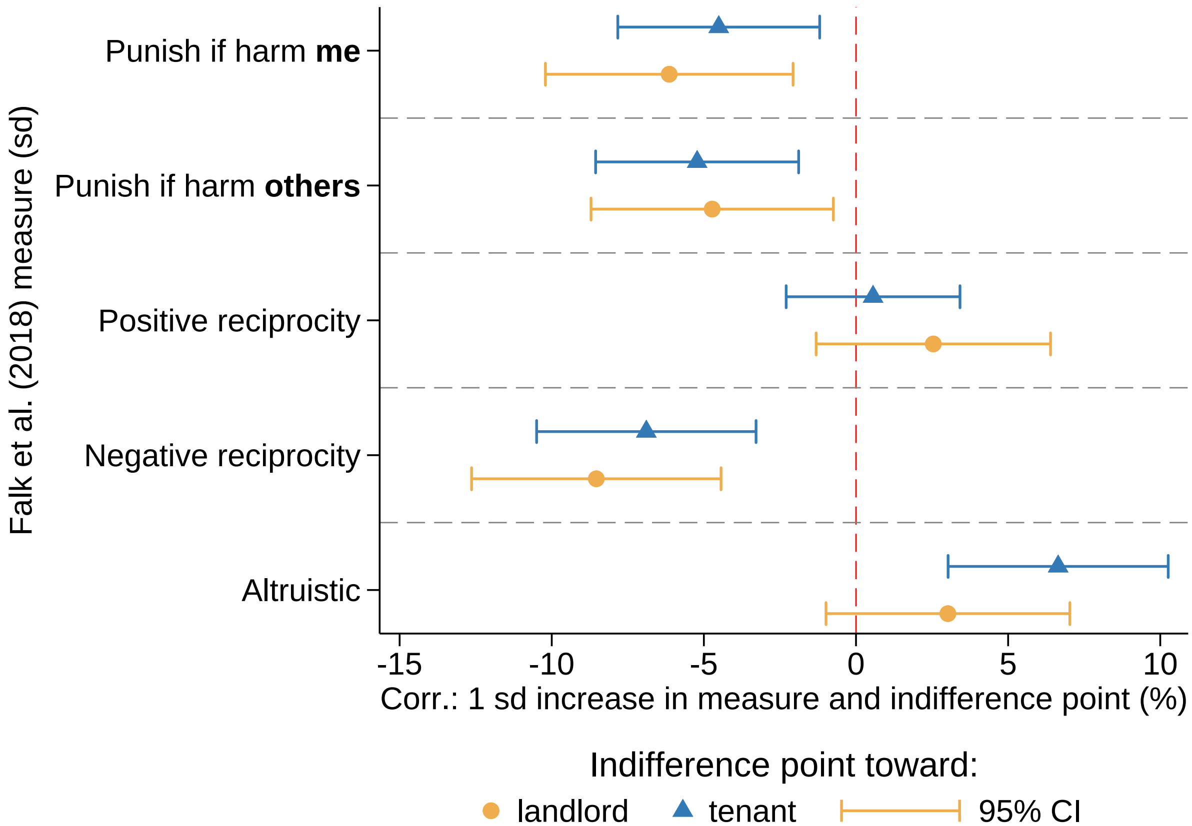


(b) Indifference point $S(x)$



Note: This figure shows the relationship between a Likert scale measuring relationship with the tenant's landlord and their behavior in the DG. The blue bars show the share of tenants who are hostile toward their own landlord, cut by their response on a question about their relationship with their landlord. The orange bars show the share of tenants who are hostile toward a random tenant, cut by their response to the question about their own landlord. $S(x)$ represents the point at which a participant is indifferent between the bundle (\$$s$ self, \$0 other) and (\$$x$ self, \$$x$ other). If $S(x) < x$, then the player is hostile; if $S(x) > x$, then the player is altruistic. Panel A shows the share hostile. Panel B shows the indifference point.

Figure A.13: Correlation Between Falk et al. (2018) Questions and Modified Dictator Game

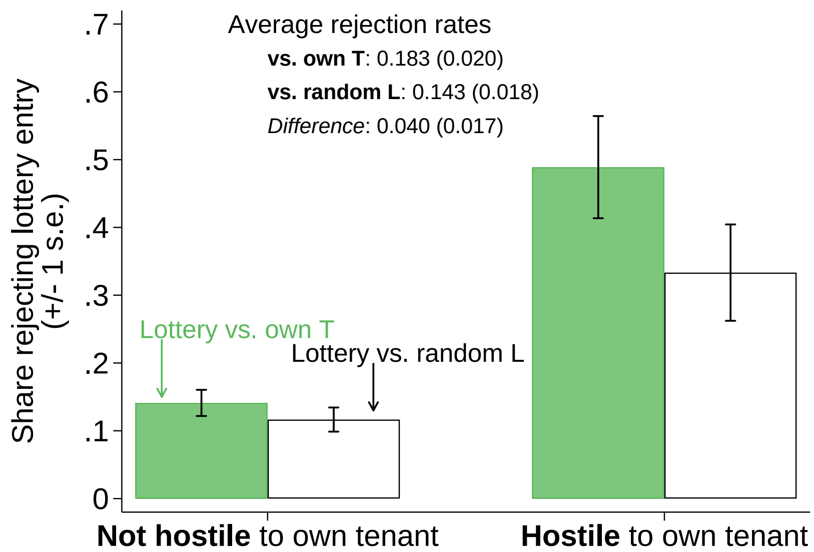


Note: This figure shows the correlation between the survey measures adapted from the Global Preferences Survey (Falk et al., 2018) and tenants' indifference points in the DG toward random tenants and landlords. We pool both own and random landlord opponent. The indifference point $S(x)$ corresponds to the value at which the tenant participant is indifferent between the bundle (\$$s$ self, \$0 other) and (\$$x$ self, \$$x$ other). If $S(x) < x$, then the player is hostile; if $S(x) > x$, then the player is altruistic. The questions from the Falk et al. (2018) survey are Likert scales (from 0 to 10) ask:

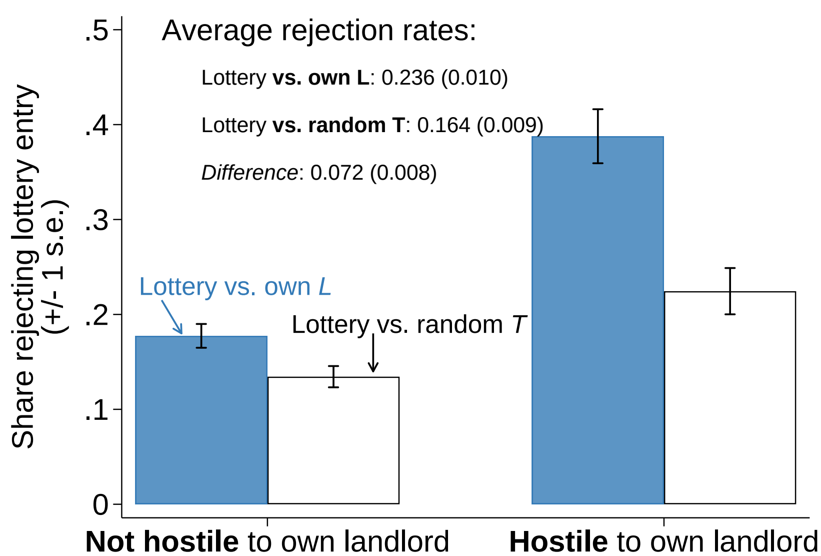
- *Punish if harm me:* "How willing are you to punish someone who treats **you** unfairly, even if there may be costs for you?"
- *Punish if harm others:* "How willing are you to punish someone who treats **others** unfairly, even if there may be costs for you?'
- *Positive reciprocity:* "When someone does me a favor, I am willing to return it."
- *Negative reciprocity:* "If I am treated very unjustly, I will take revenge at the first occasion, even if there is a cost to do so."
- *Altruistic:* "How willing are you to give to good causes without expecting anything in return?"

Figure A.14: Hostility: Lottery outcome

(a) Landlords

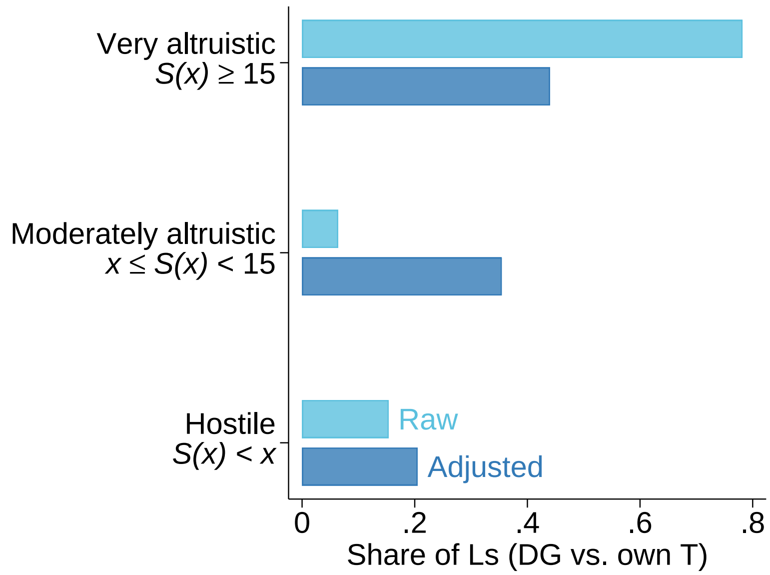


(b) Tenants



Note: This figure presents behaviors in the simple lottery task. The task asks landlords whether they would like to enroll their own tenant or a random landlord in a lottery to win a gift card. The task asks tenants whether they would like to enroll their own landlord or a random tenant in a lottery to win a gift card. The task mentions that the participant's response will be anonymous. It is costless for the participant to enroll the opponent. The text on the figures shows unconditional average rejection rates of enrolling the opponents among the entire sample. The bars show the correlation between behaviors in the lottery task and the DG. The bars limit the sample to the two-thirds that play the Dictator Game against their own tenant (among landlords) or own landlord (among tenants).

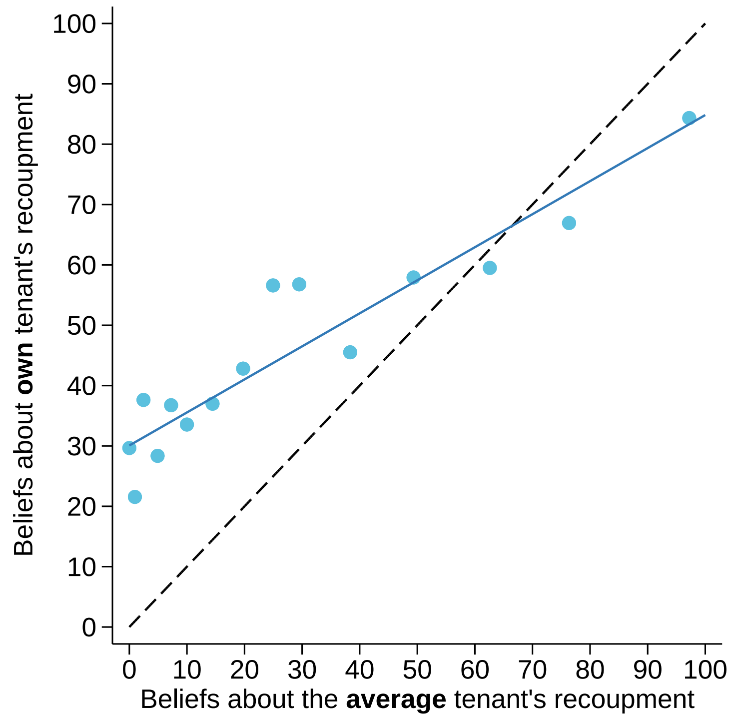
Figure A.15: Social Preferences: Rescaling Landlords' Altruism



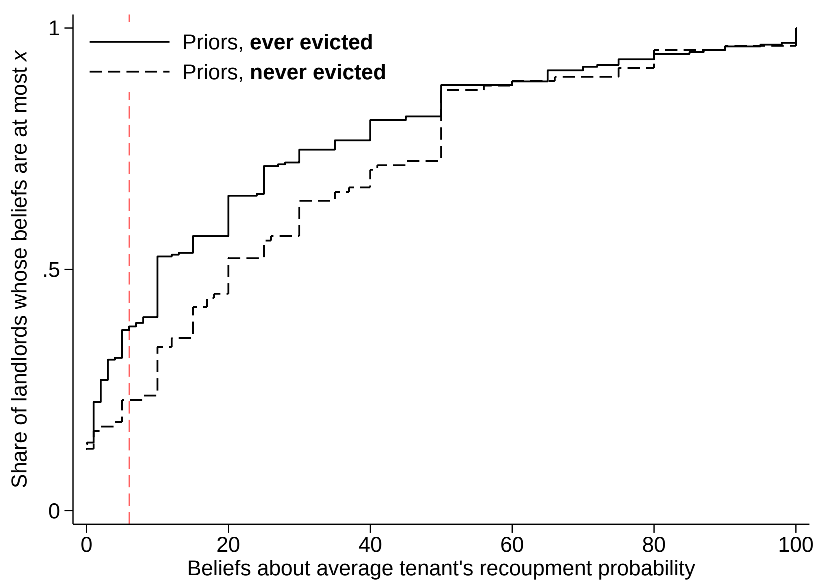
Note: This figure replaces the landlord's indifference point S_{Li}/x as $S_{Li}/x - p_{Li}$, where p_{Li} is the share of back rents they expect to recoup in an eviction. We see p_{Li} as an upper bound on how much money they can expect to recoup from a tenant if they endow them with money. This test therefore adjusts for potential beliefs landlords may have about the amount they could recoup in back rents. Mechanically, no $S_{Li}/x - p_{Li}$ can exceed 2 for $p_{Li} > 0.05$, since the maximally altruistic DG choice we elicit is $(20, 0)$ versus $(10, 10)$ for landlords. Thus we recategorize the most altruistic group into "very altruistic" ($S(x) > \$15$) rather than "highly" as in the main text.

Figure A.16: Landlord beliefs: heterogeneity

(a) Correlation between Own and Average

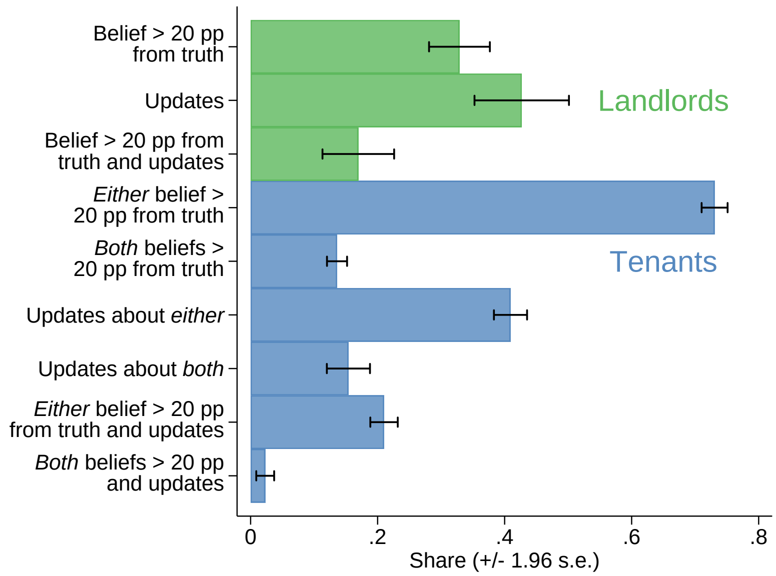


(b) Landlord Experience



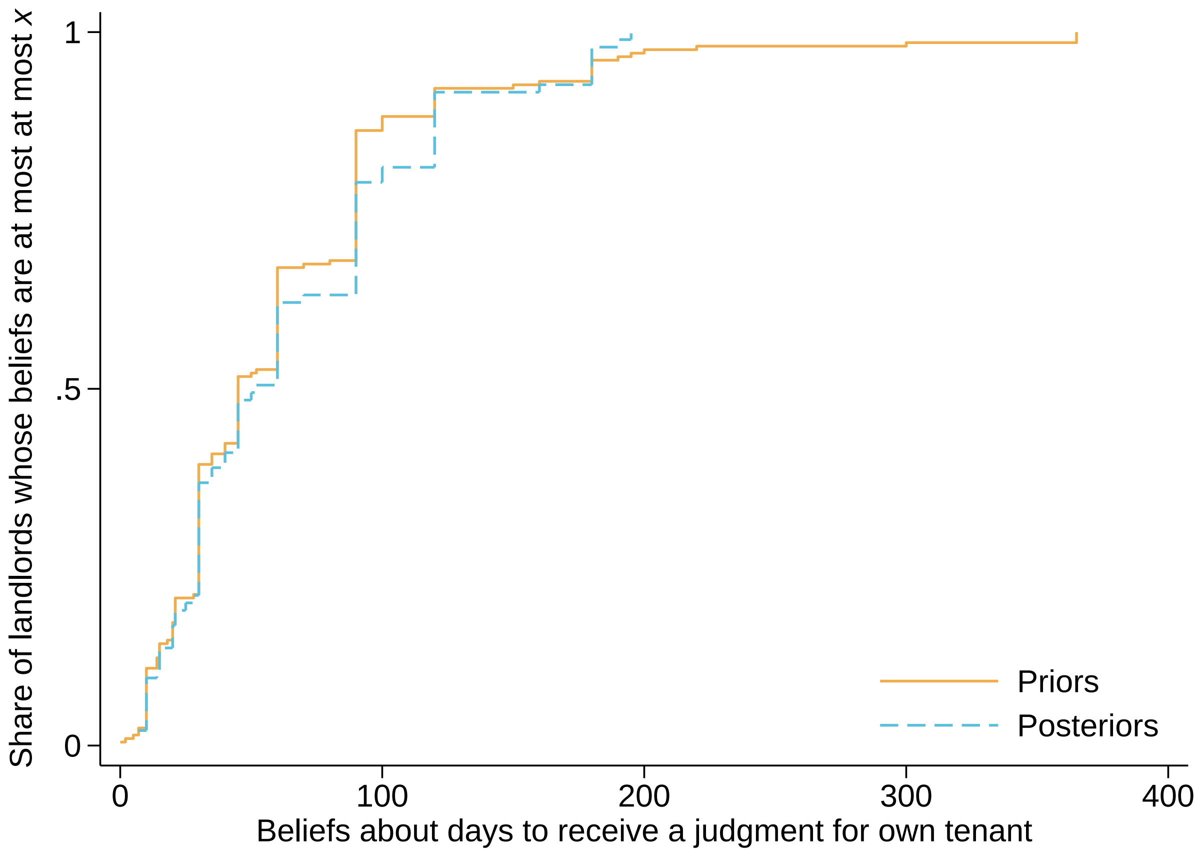
Note: Panel A shows a binned scatterplot of the correlation between a landlords' beliefs aobut their *own* tenant's recoupment probability and their beliefs about the *average* tenant. The dashed line is the 45 degree line. Panel B shows the cumulative distribution functions of landlord landlords' prior beliefs about tenants' recoupment probabilities, cut by whether the landlord reports having evicted a tenant before. The red dashed line shows the true value (6 percent).

Figure A.17: Aggregating Tests for Misperceptions



Note: This figure shows the share of tenants and landlords who have misperceptions. We show misperceptions for any fact where we also provided an information correct. Bars show different samples, since we only compute the share who update among those who are exposed to information. For the tenant bars about *both* beliefs, we restrict to the sample of tenants who see both information treatments.

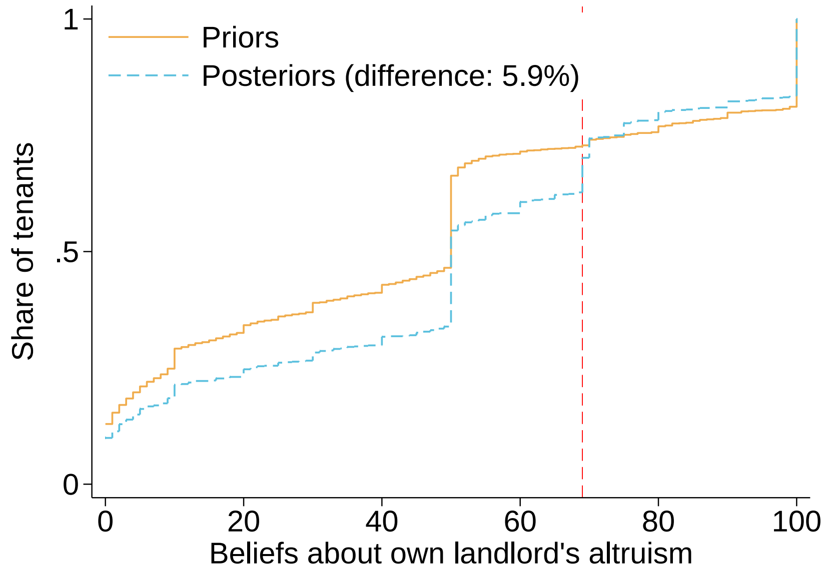
Figure A.18: Placebo: information about recoupment and beliefs about days to receive judgment for own tenant



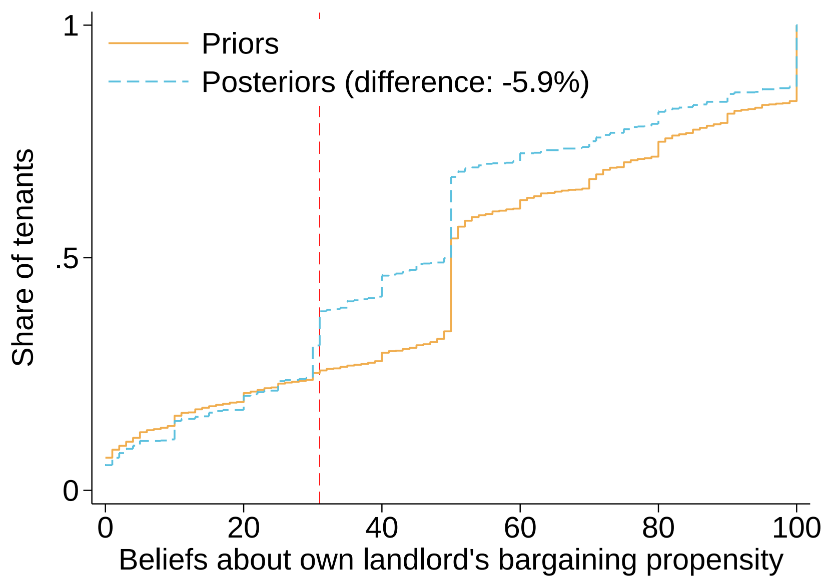
Note: This figure shows a placebo test of the cumulative distribution functions of landlord beliefs about court delays in receiving a judgment for their own tenant. The orange line shows prior beliefs, elicited before providing information; the blue dashed line shows posterior beliefs, elicited after providing information. This figure constitutes a placebo test for the exclusion restriction that providing information about average recoupment probabilities only affects beliefs about own recoupment probabilities.

Figure A.19: Tenant Average Belief Updates

(a) Treatment effect: beliefs about landlord altruism



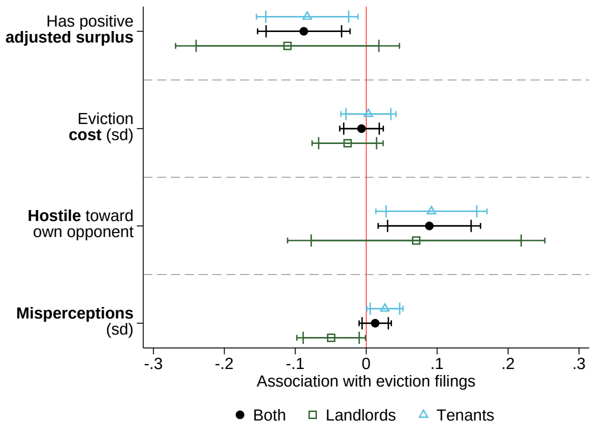
(b) Treatment effect: beliefs about landlord bargaining



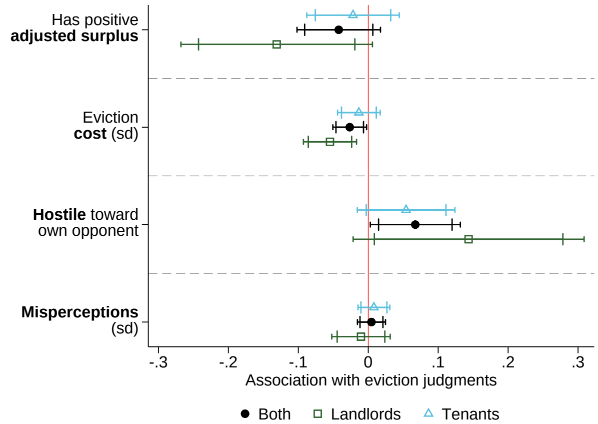
Note: This figure shows belief updates about landlord altruism and bargaining. We only show the sample of people who received each information. Treatments were cross-randomized.

Figure A.20: Model Validations: Robustness

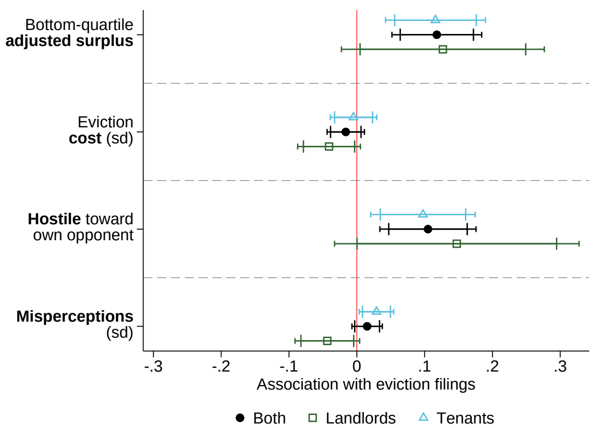
(a) Adjusted Surplus Predicts Eviction Filings: Controls



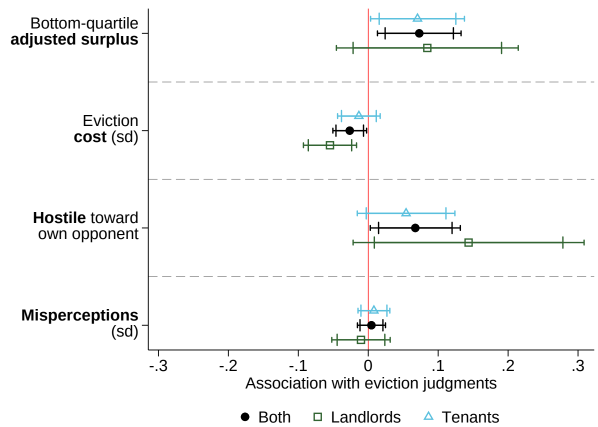
(b) Adjusted Surplus and Eviction Judgments



(c) Adjusted Surplus Predicts Eviction Filings: Bottom-Quartile Surplus



(d) Adjusted Surplus and Eviction Judgments: Bottom-Quartile Surplus

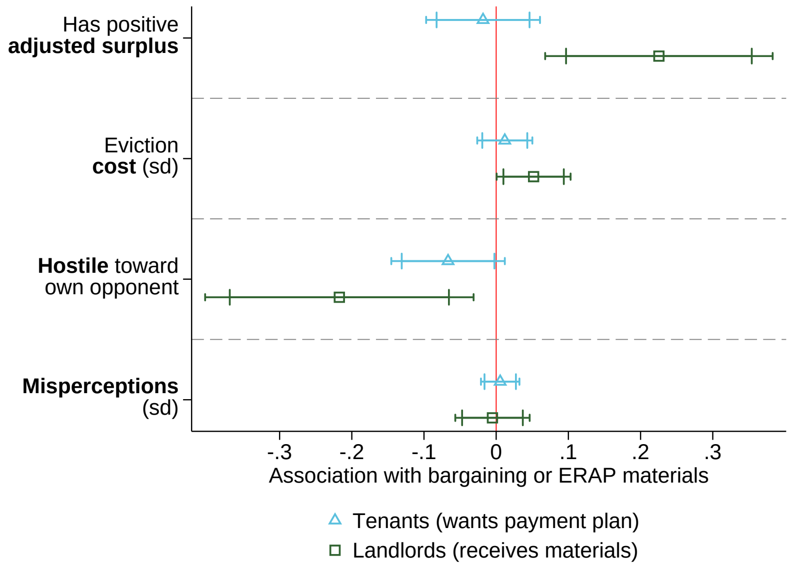


Note: Panel A presents estimates of $\hat{\beta}$ from Equation (1.10), where the covariate X_i is either positive adjusted surplus, a proxy of eviction cost, hostility, or misperceptions. Appendix A.5 gives details on how these are formed. Relative to Figure 1.5A, it includes the following controls:

- Landlords. Gender, race (indicators for White or Black), education (indicators for less than high school or some college), occupation indicators (landlord or property manager), landlord size indicators (small (1–2) or medium (3–5)), an indicator for passing an attention check, and linear controls for: age, tenure, rent, and experience.
- Tenants. Indicators for being Black, female, having less than high school, being employed, indicators for passing each attention check, and linear controls for monthly rent and monthly income.

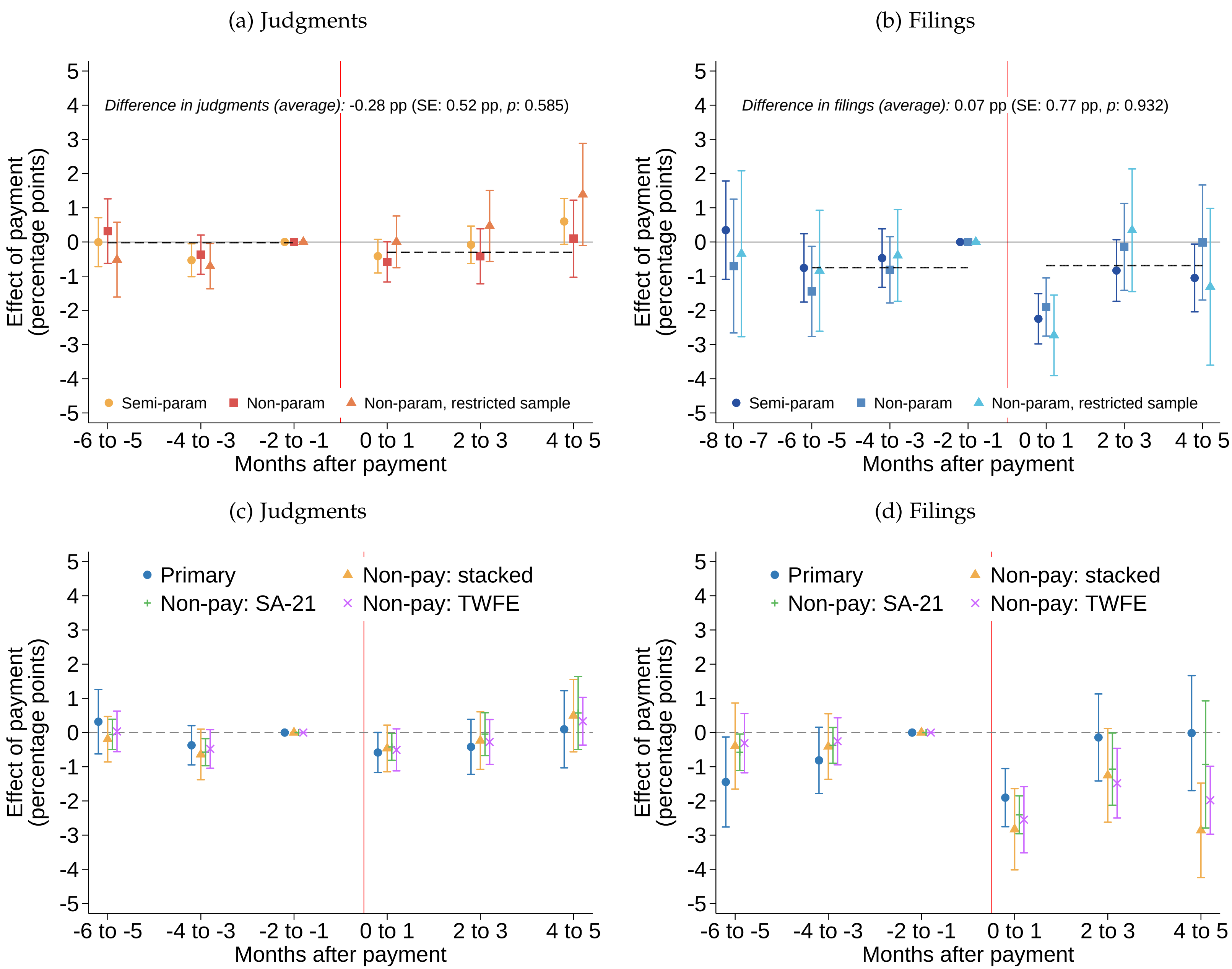
Panel B shows the effects on eviction judgments. Panels C is the same as Figure 1.5A, but the first coefficient presents $\hat{\beta}$ where the covariate X_i is an indicator for whether the individual has bottom-quartile surplus. Panel D shows the association with judgments using this notion of surplus.

Figure A.21: Adjusted Surplus and Take-up/Bargaining



Note: Panel A presents estimates of $\hat{\beta}$ from Equation (1.10), where the covariate X_i is either positive adjusted surplus, a proxy of eviction cost, hostility, or misperceptions. Appendix A.5 gives details on how these are formed. Panel B splits the data by having above- or below-median misperceptions and hostility. It presents mean eviction filings. The outcome for landlords is whether they wish to receive ERAP materials and for tenants is whether they wish to form a payment plan, as in the information experiment (Section 1.5).

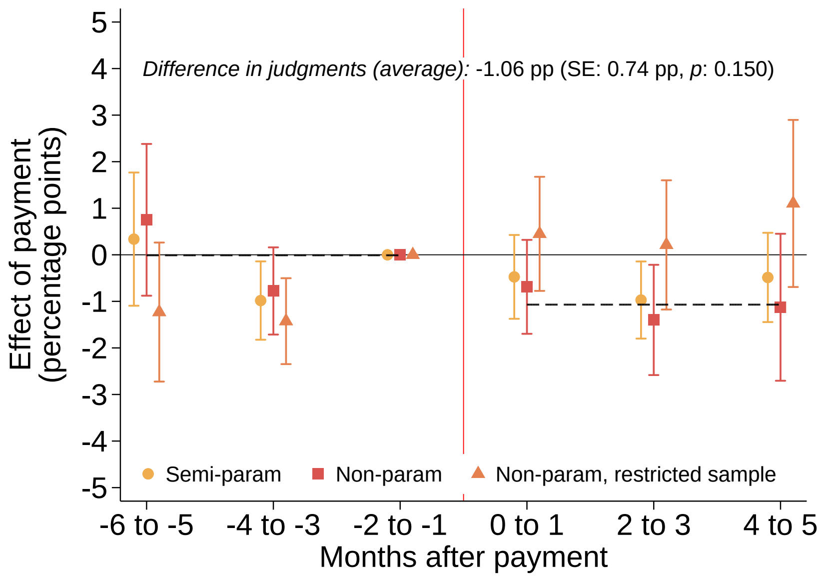
Figure A.22: Effects of Rental Assistance on Eviction Judgments and Filings



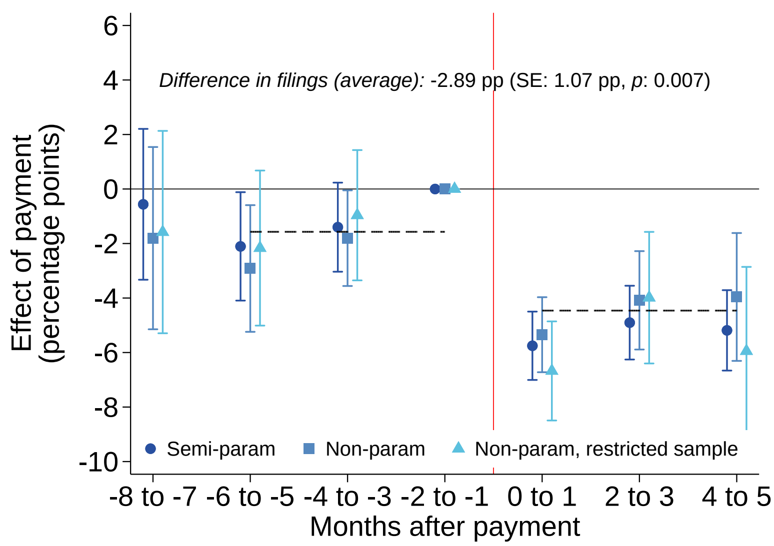
Note: Table A.24 gives sample sizes. Panels A and B show the effect on judgments and filings from Equation (A.83). The semi-parametric specification sets $\sigma_s = 0$ for all s. The figure's point estimates come from the non-parametric specification on the full sample. The restricted sample drops households who apply with eviction notices or shutoffs, who were eligible to be expedited. Panels C and D show the alternative design which compares to non-paid households (Section A.5.6.3). The primary specification in Panels C and D refer to the estimates from Equation (A.83), nonparametric specification. The other estimates come from Equations (A.84) or (A.85), nonparametric specification. SA-21 refers to estimates from Sun and Abraham (2021), using non-paid households as a control group. TWFE refers to estimates from a Two-Way Fixed Effects specification. The primary estimates are estimated on the microdata. They are clustered at the household level. The non-pay coefficients are estimated on data collapsed to the payment-period by application-period by calendar-period level. They are weighted by the number of observations and clustered at the payment period (TWFE, Sun and Abraham (2021)) or payment period by dataset level (stacked).

Figure A.23: Effects of Rental Assistance on Eviction Judgments and Filings (Reweighted)

(a) Judgments



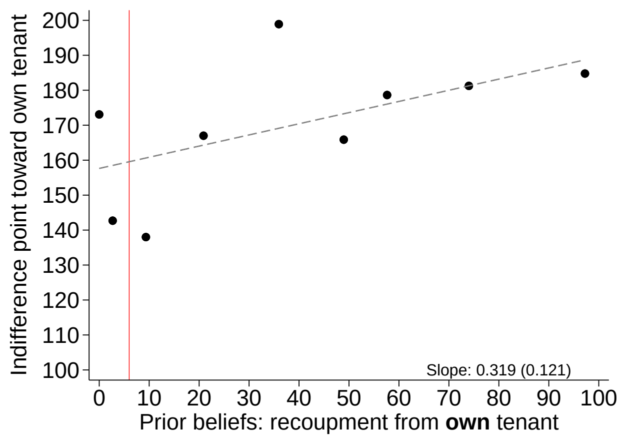
(b) Filings



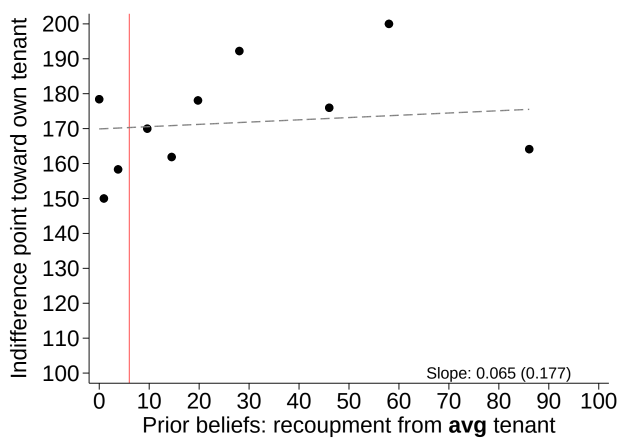
Note: This figure shows the treatment effect on judgments (Panel A) and filings (Panel B) from Equation (A.83). The semi-parametric specification sets $\sigma_s = 0$ for all s. The nonparametric specification is exactly as in Equation (A.83). The restricted sample drops households who apply with eviction notices or shutoffs, who were eligible to be expedited. Relative to Figure A.22, we weight the estimates using the procedure in Appendix A.5. The figure's point estimates come from the non-parametric specification on the full sample. We cluster at the household level.

Figure A.24: Correlation between Misperception and Altruism among Landlords

(a) Own tenant recoupment



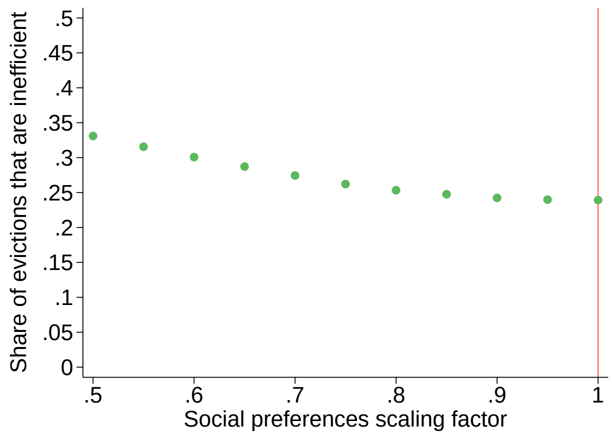
(b) Average tenant recoupment



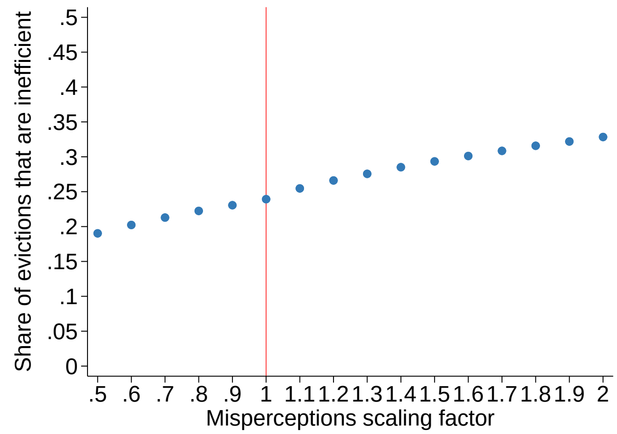
Note: This figure shows binned scatterplots of landlords' prior beliefs (horizontal axis) and landlord behavior toward their own tenant in the Dictator Game (vertical axis). Panels A and B show prior beliefs about the probability of recouping back rents from the landlord's own tenant or the average tenant.

Figure A.25: Measurement Error: Simulations

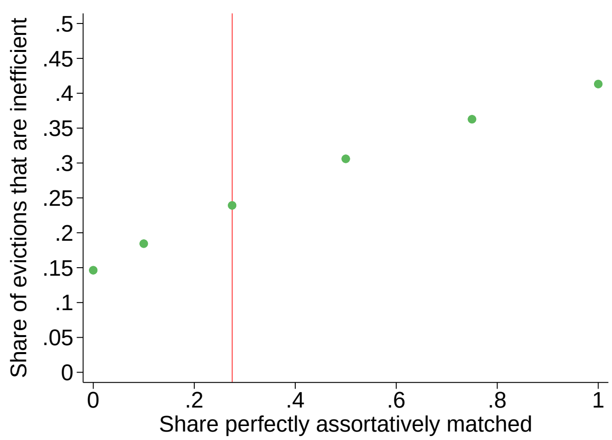
(a) Social Preferences



(b) Misperceptions

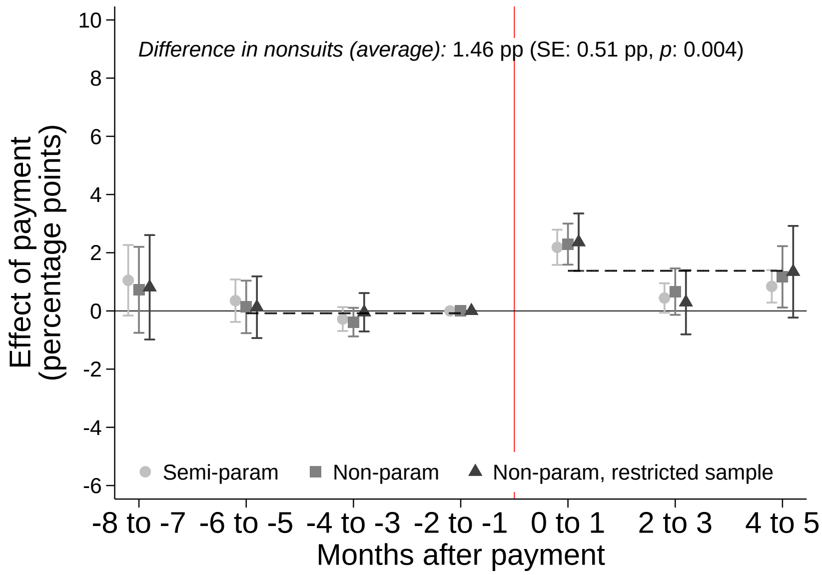


(c) Misperceptions



Note: This figure presents the share of evictions that we estimate are inefficient if we rescale social preferences (Panel A) or misperceptions (Panel B). In Panel A, we replace social preferences $\hat{a}_{ji} := sa_{ji}$ for $s < 1$. In Panel B, we replace tenant beliefs about bargaining and landlord beliefs as: $\hat{p}_{ji} := \max\{s, 1\}p_{ji} + (1-s)p_{\text{true}} + \mathbb{1}(s > 1)(s-1)p_{ji}$ for various $s \in \mathbb{R}$, where p_{true} is the true value (e.g., 0.06 for landlords' beliefs). For tenant beliefs about landlord altruism, we replace tenant beliefs as: $\hat{p}_{Ti} := \max\{s, 1\}sp_{Ti} + (1-s)p_{\text{true}} - \mathbb{1}(s > 1)(s-1)p_{Ti}$. This procedure raises misperceptions for $s > 1$ and explains the kink at 1 in Panel B. After applying this scaling factor for beliefs, we then apply the same procedure to obtain the misperceptions that we input into the estimation. In Panel C, we simulate different shares of households who are perfectly assortatively matched (see Appendix A.5.2). All panels use the estimated and calibrated values of other parameters as in our primary specification. The vertical red lines show the primary estimate.

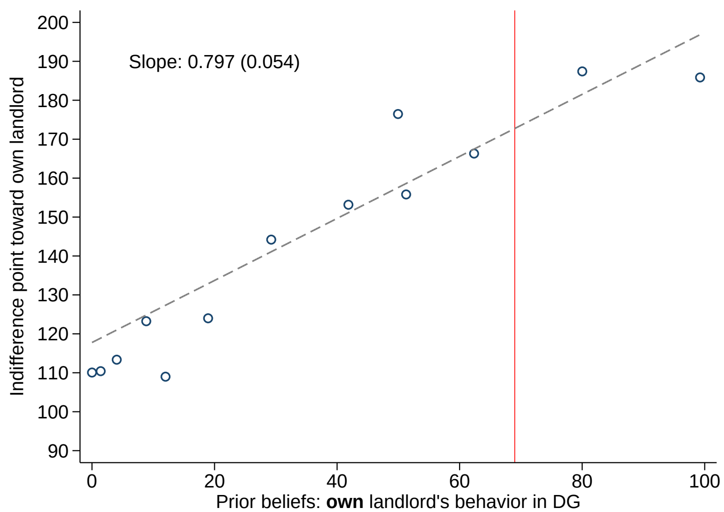
Figure A.26: ERAP Effect on Non-Suits



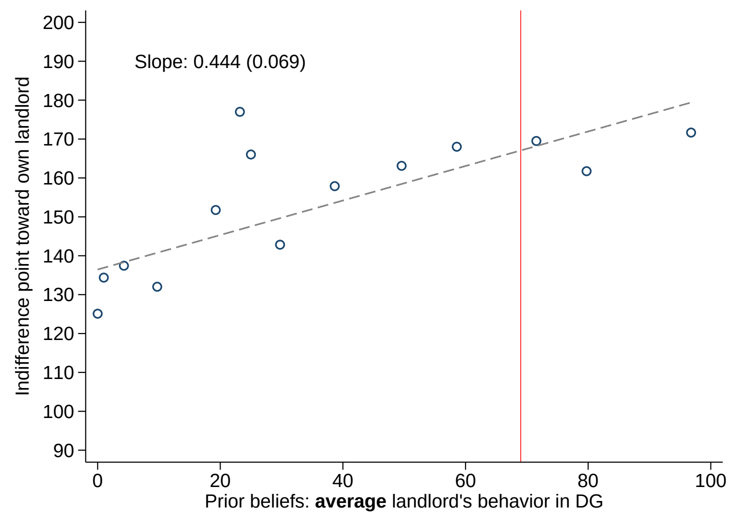
Note: This figure shows the effect of Memphis/Shelby County's ERAP on non-suits (explicit withdrawals from the court system) using the primary design (Equation A.83).

Figure A.27: Correlation between Misperceptions and Altruism among Tenants

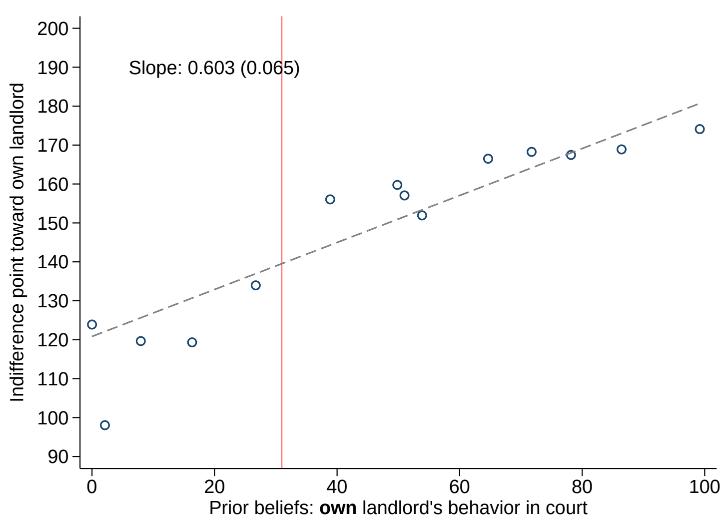
(a) Own landlord altruism



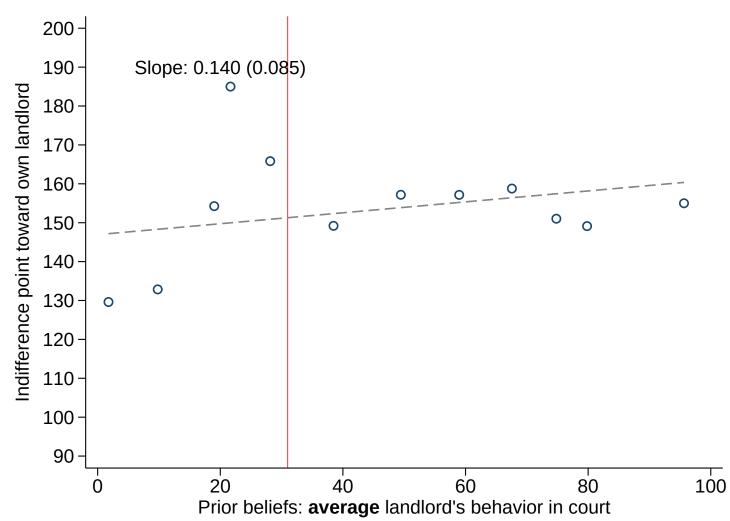
(b) Average landlord altruism



(c) Own landlord bargaining



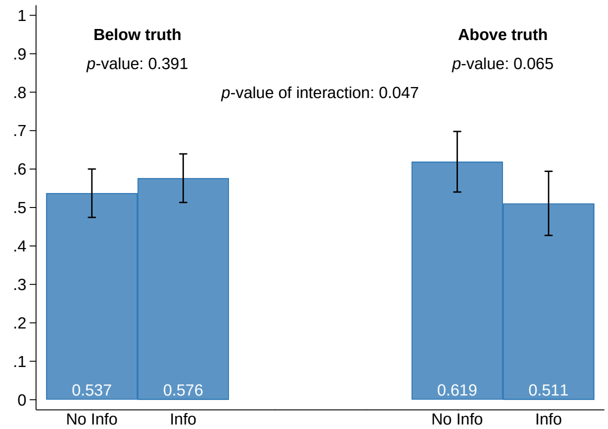
(d) Average landlord bargaining



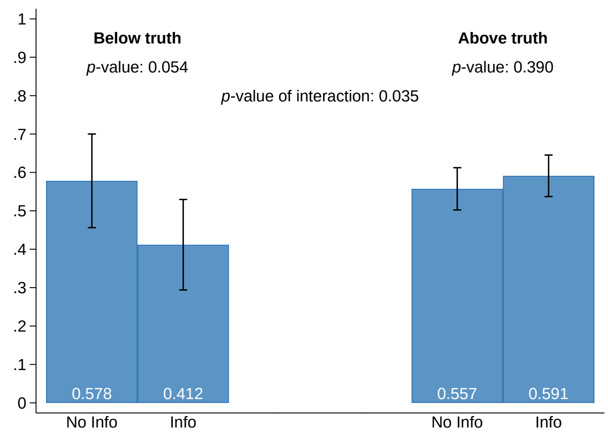
Note: This figure shows binned scatterplots of tenant prior beliefs (horizontal axis) and tenant behavior toward their own landlord in the Dictator Game (vertical axis). Panels A and B show prior beliefs about own and average altruism. Panels C and D show prior beliefs about own and average bargaining behavior. The red vertical line indicates the truth about the average.

Figure A.28: Information is More Effective among Tenants with Strong Relationships

(a) Altruism treatment



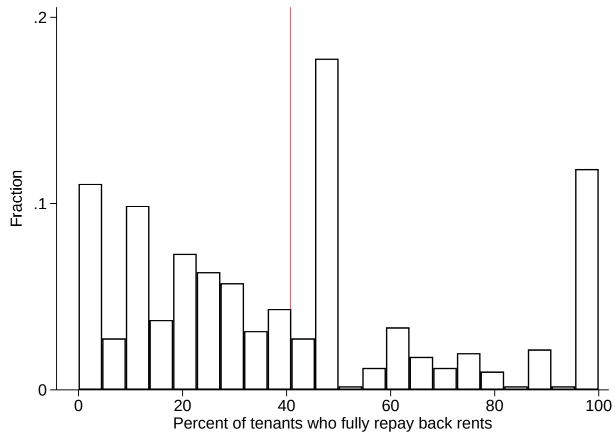
(b) Bargaining treatment



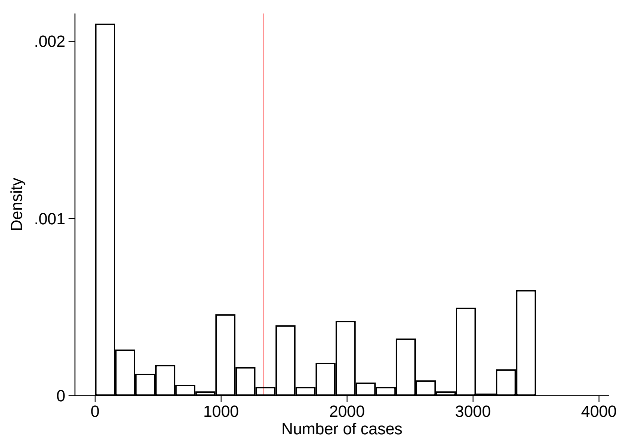
Note: Panels A and B present versions of Figure 1.4C and D, limiting only to tenants with high degrees of altruism toward their own landlords. In particular, we keep the tenants who prefer (x self, x landlord) to ($2x$ self, 0 landlord).

Figure A.29: Beliefs about Eviction: Memphis Sample

(a) Beliefs about Percent of Tenants who Repay Money Judgments

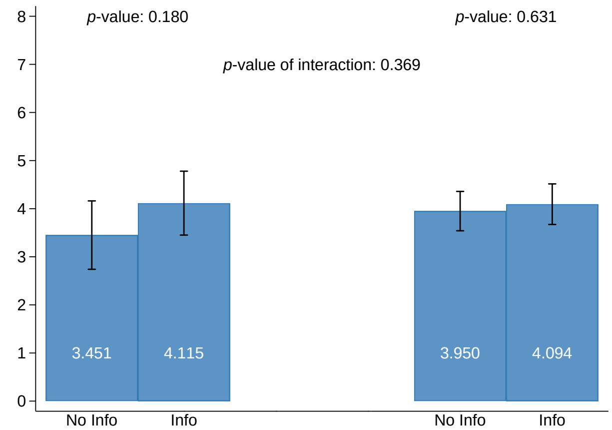


(b) Beliefs about Number of Cases Processed in Eviction Court



Note: Panels A and B present beliefs about the eviction process, elicited among the Memphis sample. The vertical lines represent means. The true values are 6 (Panel A) and 54 (Panel B).

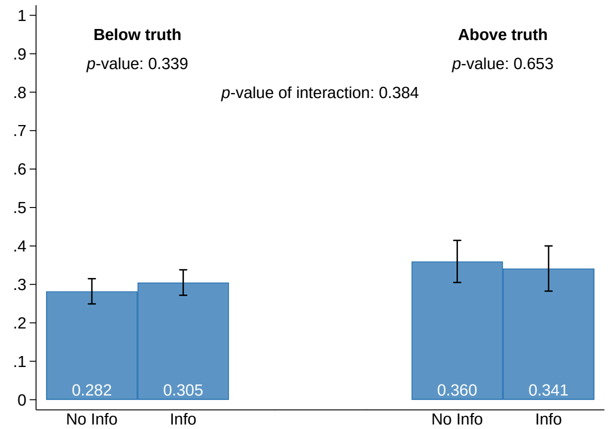
Figure A.30: Treatment Effect of Bargaining on WTP for Information about Altruism



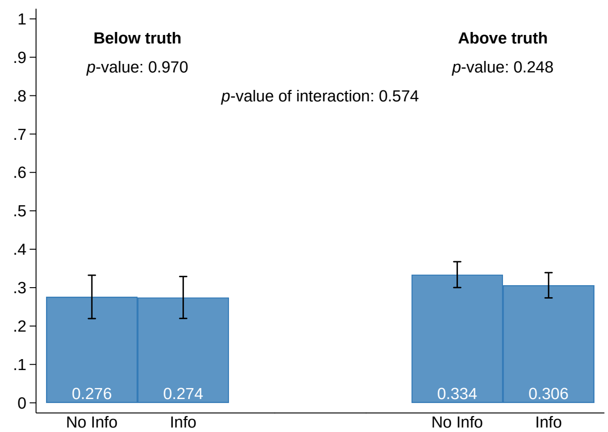
Note: This figure shows intent-to-treat effects of the bargaining information treatment among tenants on willingness to pay for information about the share of landlords who had the highest possible indifference point in the DG, in the landlord sample. The max WTP that could be consistently reported was $8.

Figure A.31: Treatment Effect of Information on Repayment Rate in Payment Plan

(a) Altruism treatment



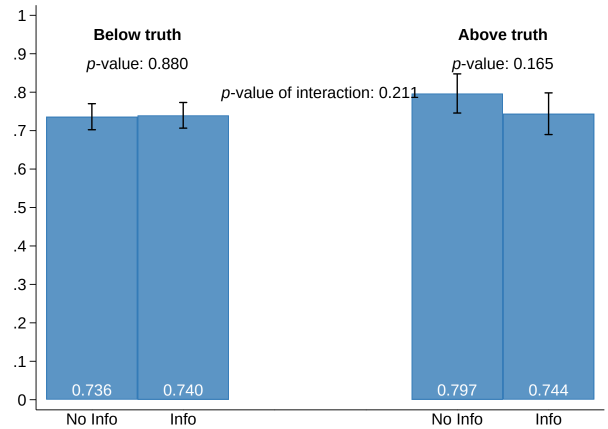
(b) Bargaining treatment



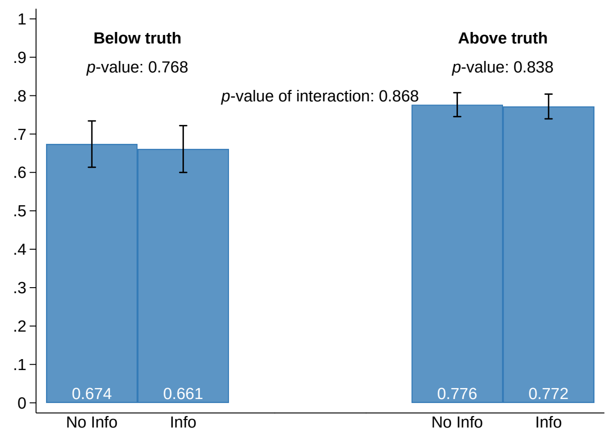
This figure shows intent-to-treat effects of the bargaining and altruism information treatments among tenants on offered repayment rates in the payment plan. The repayment shares are 0 if they do not want a payment plan but were offered the chance to form one.

Figure A.32: Treatment Effect of Information on Hypothetical WTP to Move

(a) Altruism treatment



(b) Bargaining treatment



This figure shows intent-to-treat effects of the altruism and bargaining information treatments on tenants' willingness to accept $1000 versus move. This question was asked of all tenants.

Table A.1: Behavior in Dictator Game: Only Tenants After March 30

	(1) Hostile	(2) Indifference point	(3) Highly hostile	(4) Highly altruistic
A. Landlord sample (N = 371)				
1. Own Tenant	0.154***	171.5***	0.090***	0.688***
$N = 234$	(0.024)	(4.2)	(0.019)	(0.030)
	[0.000]	[0.000]	[0.000]	[0.000]
2. Random Tenant	0.066***	182.5***	0.036**	0.693***
$N = 137$	(0.021)	(4.0)	(0.016)	(0.040)
	[0.002]	[0.000]	[0.025]	[0.000]
3. Random Landlord	0.119***	174.3***	0.038***	0.623***
$N = 371$	(0.017)	(2.8)	(0.010)	(0.025)
	[0.000]	[0.000]	[0.000]	[0.000]
4. Own Tenant – Random Tenant	0.088***	-11.0*	0.053**	-0.005
(Row 1 – Row 2)	(0.032)	(5.8)	(0.025)	(0.050)
	[0.006]	[0.060]	[0.031]	[0.914]
5. Random Tenant – Random Landlord	-0.053**	8.2**	-0.001	0.071**
(Row 2 – Row 3)	(0.024)	(4.0)	(0.017)	(0.036)
	[0.028]	[0.042]	[0.941]	[0.048]
6. Own Tenant – Random Landlord	0.035	-2.8	0.052***	0.065**
(Row 1 – Row 3)	(0.024)	(4.0)	(0.017)	(0.030)
	[0.150]	[0.485]	[0.003]	[0.030]
B. Tenant sample (N = 1,102)				
7. Own Landlord	0.249***	152.4***	0.132***	0.523***
$N = 742$	(0.016)	(2.7)	(0.012)	(0.018)
	[0.000]	[0.000]	[0.000]	[0.000]
8. Random Landlord	0.225***	156.9***	0.106***	0.561***
$N = 360$	(0.022)	(3.7)	(0.016)	(0.026)
	[0.000]	[0.000]	[0.000]	[0.000]
9. Random Tenant	0.131***	172.6***	0.059***	0.631***
$N = 1,102$	(0.010)	(1.8)	(0.007)	(0.015)
	[0.000]	[0.000]	[0.000]	[0.000]
10. Own Landlord – Random Landlord	0.024	-4.5	0.027	-0.038
(Row 7 – Row 8)	(0.027)	(4.6)	(0.020)	(0.032)
	[0.371]	[0.325]	[0.194]	[0.232]
11. Random Landlord – Random Tenant	0.094***	-15.7***	0.047***	-0.070***
(Row 8 – Row 9)	(0.023)	(3.8)	(0.017)	(0.025)
	[0.000]	[0.000]	[0.006]	[0.006]
12. Own Landlord – Random Tenant	0.119***	-20.2***	0.073***	-0.108***
(Row 7 – Row 9)	(0.017)	(2.7)	(0.012)	(0.017)
	[0.000]	[0.000]	[0.000]	[0.000]

Note: See notes to Table A.16 for description of our altruism and hostility measures. Parentheses show robust standard errors. Brackets show p-values. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table is identical to Table A.16 except Panel B only includes tenants who participate after all changes to DG wording. All tenant participants in Panel B see language that stresses anonymity in the context of the DG, and see a confirmation check about anonymity. Panel A is the same as in Table A.16.

A.1.2 Tables

Table A.2: Landlord sample balance: random landlord treatment

	Own tenant	Random tenant	p-value
Age	49.3	47.8	0.390
Missing age	3.0	3.6	0.730
Female	67.1	54.0	0.012
White	32.9	29.9	0.553
Black	59.0	56.9	0.702
Has ever evicted	68.4	74.5	0.216
HS or less	17.1	10.2	0.070
Some college	27.4	27.7	0.936
Landlord	61.5	62.8	0.814
Property manager	29.5	27.0	0.611
Tenant tenure (months)	32.7	30.7	0.495
Tenant rent (monthly $)	829.0	778.4	0.198
Missing units	3.0	2.2	0.646
1–2 units	26.5	26.3	0.963
3–5 units	24.8	23.4	0.757
Experience (years)	13.0	13.2	0.866
Attentive	61.5	66.4	0.347
Information treatment	47.4	43.8	0.499
Joint F-test p-value			0.233
Observations	137	234	

Observations denote the total number of observations; some demographics are missing for a small number of observations. The joint p-value is from the joint test when all non-missing.

Table A.3: Landlord sample balance: information treatment

	No information	Information	p-value
Age	49.9	47.4	0.117
Missing age	2.0	4.7	0.147
Female	59.0	66.1	0.162
White	28.5	35.7	0.140
Black	60.0	56.1	0.454
Has ever evicted	64.0	78.4	0.002
HS or less	18.0	10.5	0.042
Some college	23.5	32.2	0.063
Landlord	65.5	57.9	0.133
Property manager	23.5	34.5	0.019
Tenant tenure (months)	31.6	32.3	0.792
Tenant rent (monthly $)	805.3	816.2	0.776
Missing units	3.0	2.3	0.696
1–2 units	29.5	22.8	0.146
3–5 units	27.0	21.1	0.184
Experience (years)	13.1	13.0	0.927
Attentive	62.5	64.3	0.717
Random treatment	38.5	35.1	0.499
Hostile to T	0.1	0.1	0.934
Hostile to L	0.1	0.1	0.292
Tenant indifference (% of own payoff)	175.8	175.3	0.934
Landlord indifference (% of own payoff)	172.3	176.6	0.447
Priors (own recoupment)	46.4	39.9	0.079
Priors (avg recoupment)	26.2	22.2	0.155
Priors (own days)	6664.0	6043.3	0.299
Priors (avg days)	1621.7	1355.2	0.051
Uncertainty: average tenant	51.0	48.0	0.560
Uncertainty: own tenant	53.0	46.2	0.193
Joint F-test p-value			0.011
Observations	171	200	

Observations denote the total number of observations; some demographics are missing for a small number of observations. The joint p-value is from the joint test when all non-missing. We include experimental outcomes from the Dictator Games and prior beliefs that are elicited before the treatment.

Table A.4: Tenant sample balance: random landlord treatment

	Own landlord	Random landlord	p-value
Black	90.4	89.7	0.633
Female	83.0	87.0	0.030
Age	35.5	35.5	0.948
HS or less	43.0	46.1	0.216
Ever payment plan	56.5	58.0	0.561
Ever overdue rent	85.4	86.5	0.533
Ever evicted	32.4	35.1	0.261
Back rent	2032.8	2010.4	0.904
Monthly rent	890.7	879.7	0.502
Monthly income	2653.6	2117.0	0.059
Employed	60.0	58.6	0.547
Paid by ERAP	54.8	55.1	0.899
Attentive	91.5	89.2	0.116
Attentive (alt. measure)	33.9	32.7	0.614
Treatment (altruism)	52.5	46.7	0.023
Treatment (bargaining)	48.9	49.1	0.935
Joint F-test p-value			0.285
Observations	584	1224	

Observations denote the total number of observations; some demographics are missing for a small number of observations. The joint p-value is from the joint test when all non-missing.

Table A.5: Tenant sample balance: information treatment (altruism)

	No information (altruism)	Information (altruism)	p-value
Own L bargain	53.8	53.9	0.937
Avg. L bargain	46.8	46.6	0.869
Own L altruism	45.2	45.7	0.753
Avg. L altruism	39.3	39.5	0.884
Uncertain: own L bargain	0.4	0.4	0.911
Uncertain: own L altruism	0.4	0.4	0.601
Black	90.7	89.7	0.484
Female	84.4	84.2	0.869
Age	35.4	35.5	0.751
HS or less	44.2	43.7	0.825
Ever payment plan	56.7	57.2	0.832
Ever overdue rent	86.9	84.6	0.161
Ever evicted	32.5	34.1	0.464
Back rent	1994.9	2055.5	0.728
Monthly rent	898.4	876.2	0.149
Monthly income	2419.0	2540.1	0.649
Employed	59.8	59.3	0.844
Paid by ERAP	54.2	55.6	0.542
Hostile to T	13.0	15.5	0.124
Hostile to L	22.3	24.3	0.320
Indiff. for T ($S(x)$)	172.0	169.9	0.448
Indiff. for L ($S(x)$)	155.7	154.4	0.687
Attentive	89.8	91.7	0.167
Attentive (alt. measure)	31.8	35.2	0.127
Random landlord	34.8	29.8	0.023
Treatment (bargaining)	50.5	47.5	0.208
Joint F-test p-value			0.662
Observations	915	893	

Observations denote the total number of observations; some particular demographics are missing for a small number of observations. The joint p-value is from the joint test when all non-missing. We include experimental outcomes from the Dictator Games and prior beliefs that are elicited before the treatment.

Table A.6: Tenant sample balance: information treatment (bargaining)

	No information (altruism)	Information (altruism)	p-value
Own L bargain	54.2	53.5	0.646
Avg. L bargain	46.2	47.3	0.356
Own L altruism	45.1	45.8	0.676
Avg. L altruism	39.1	39.8	0.605
Uncertain: own L bargain	0.4	0.4	0.233
Uncertain: own L altruism	0.4	0.4	0.969
Black	90.0	90.4	0.783
Female	85.5	83.1	0.162
Age	35.6	35.4	0.621
HS or less	43.6	44.4	0.746
Ever payment plan	55.9	58.0	0.401
Ever overdue rent	85.1	86.3	0.465
Ever evicted	33.8	32.7	0.617
Back rent	1868.6	2188.9	0.067
Monthly rent	881.4	893.1	0.444
Monthly income	2566.7	2390.3	0.508
Employed	59.9	59.3	0.790
Paid by ERAP	57.0	52.7	0.064
Hostile to T	14.0	14.6	0.730
Hostile to L	24.7	21.8	0.139
Indiff. for T ($S(x)$)	170.2	171.6	0.627
Indiff. for L ($S(x)$)	151.7	158.6	0.040
Attentive	90.8	90.7	0.979
Attentive (alt. measure)	32.9	34.2	0.548
Random landlord	32.2	32.4	0.935
Treatment (altruism)	52.1	49.1	0.208
Joint F-test p-value			0.704
Observations	886	922	

Observations denote the total number of observations; some particular demographics are missing for a small number of observations. The joint p-value is from the joint test when all variables are non-missing. We include experimental outcomes from the Dictator Games and prior beliefs that are elicited before the treatment.

Table A.7: Experimental Attrition

Panel A: Tenant Survey Attrition

Attrition Funnel	N	% of total consenting	% of total completing demos
1. All who consent	4,440	100.0	·
2. Complete demographics	2,502	56.4	100.0
3. Complete altruism	2,402	54.1	96.0
4. Complete prior beliefs	2,175	49.0	86.9
5. Complete survey	1,929	43.4	77.1
Difference in attrition by treatment (conditional on finishing demographics)			
Random landlord in DG			-0.026
			(0.018)
			[0.147]
Altruism info treatment			0.020
			(0.017)
			[0.236]
Bargaining share info treatment			-0.013
			(0.017)
			[0.437]
Joint p across treatments			[0.252]

Panel B: Landlord Survey Attrition

Attrition Funnel	N	% of total consenting	% of total completing demos
1. All who consent	708	100.0	·
2. Complete demographics	620	87.6	100.0
3. Complete altruism	565	79.8	91.1
4. Complete prior beliefs	448	63.3	72.3
5. Complete survey	404	57.1	65.2
Difference in attrition by treatment (conditional on finishing demographics)			
Random landlord in DG			0.035
			(0.040)
			[0.380]
Info treatment			0.020
			(0.017)
			[0.236]
Joint p across treatments			[0.394]

Brackets indicate p-values. The total sample size is not the same as the text because it includes several drops, e.g. for tenants who complete the survey twice. The own versus random tenant treatment occurs between items 2 and 3. The information treatment occurs between items 4 and 5. The joint p value stacks the two treatments using seemingly unrelated regression.

Table A.8: Landlord–tenant assortative matching (names)

(a) Landlord indifference point

	(1) Tenant: indifference point	(2) Maximal altruism	(3) Hostile	(4) Maximal hostile
Landlord indifference	16.1	0.084	-0.058	-0.054
	(11.2)	(0.053)	(0.078)	(0.049)
	[0.158]	[0.121]	[0.458]	[0.283]
Observations	471	471	471	471

(b) Landlord hostility

	(1) Tenant: indifference point	(2) Maximal altruism	(3) Hostile	(4) Maximal hostile
Landlord hostile	-26.4	-0.13	0.085	0.088
	(20.8)	(0.099)	(0.15)	(0.098)
	[0.212]	[0.204]	[0.577]	[0.377]
Observations	471	471	471	471

Note: This table shows the degree of assortative matching between landlords and tenants. Panel A regresses the tenant Dictator Game outcomes on the landlord's indifference point in the DG. Panel B regresses the tenant DG outcomes on the landlord's hostility in the DG. We use landlords–tenant pairs whom we could match (Appendix A.5.

Table A.9: Tenant perceptions about assortative matching

	(1) Tenant: extreme altruism
Highly altruistic to own L	0.24***
	(0.019)
	[0.000]
Constant	0.32***
	(0.013)
	[0.000]
Observations	1224

Note: This table regresses tenant propensity to be highly altruistic (i.e., $S(x) > 2x$) in the DG on tenant beliefs about her own landlord's behavior in the DG.

Table A.10: Tenant hostility and free response sentiment

	(1) Hostile: own landlord	(2) Random landlord	(3) Tenant
VADER score (sd)	-0.0811***	-0.0396	-0.00292
	(0.0217)	(0.0295)	(0.0135)
Observations	437	214	651

Note: This table presents the relationship between tenant hostility in the Dictator Game and their free response sentiment. The free response question was: Tenants play a Dictator Game (DG) against a landlord. We elicit $S(x)$, the value at which the tenant participant is indifferent between the bundle (\$$s$ self, \$0 other) and (\$$x$ self, \$$x$ other). If $S(x) < x$, then the player is hostile; if $S(x) > x$, then the player is altruistic. The VADER score represents a measure of sentiment from text responses, developed by Hutto and Gilbert (2014). A higher VADER score indicates more positive sentiment. The free response question was only elicited starting on March 27, 2021, and it was optional. The question was: "Do you have any thoughts about [landlord name] that you want to share?"

Table A.11: Landlords Robustness

Panel A: Hostility

	(1) Raw	(2) If attentive	(3) Controls	(4) Full controls	(5) DS lasso
Random tenant	-0.0882***	-0.105***	-0.0847**	-0.0816**	-0.0763***
	(0.0318)	(0.0389)	(0.0339)	(0.0333)	(0.0287)
Observations	371	235	371	371	371
p	0.00582	0.00763	0.0131	0.0150	0.00777

Panel B: Indifference

	(1) Raw	(2) If attentive	(3) Controls	(4) Full controls	(5) DS lasso
Random tenant	11.02*	14.03*	10.27*	9.662*	9.705**
	(5.842)	(7.294)	(5.757)	(5.740)	(4.900)
Observations	371	235	371	371	371
p	0.0600	0.0556	0.0755	0.0933	0.0477

Note: This table presents robustness for the Dictator Game (DG) in the landlord experiment. Panel A focuses on the difference between hostility for own vs. random tenants (Row 4 of Table A.16), whereas Panel B focuses on the difference between the indifference points $S(x)$. Column 1 corresponds to Row 4, Column 1 of Table A.16. Column 2 limits to attentive landlords who correctly answer the "teal" attention check (Appendix A.3.6). Column 3 adds a vector of demographic controls for: behavior in the DG toward tenants, age, experience, landlord size, race, gender, education occupation, tenant's tenure in the appartment, rent, attentiveness, and date of completing the survey. Column 4 adds prior beliefs and landlord reports of tenant's tenure in the appartment/rent to the controls in Column 3. We separate them since they were collected after the DG randomization treatment and therefore could, in principle, be affected by it. Column 5 shows the effect using post-double-selection Lasso to select controls (Belloni et al., 2014). The observations is the total number of individuals, since each participant plays against one landlord.

Table A.12: Tenants Robustness

Panel A: Hostility

	(1) Raw	(2) If attentive (one q)	(3) If attentive (both q's)	(4) Controls	(5) DS lasso
Random tenant	-0.0975***	-0.1000***	-0.0961***	-0.0990***	-0.0991***
	(0.0130)	(0.0134)	(0.0223)	(0.0130)	(0.0139)
Observations	3032	2780	1002	3032	3032
p	0.000	0.000	0.000	0.000	0.000

Panel B: Indifference

	(1) Raw	(2) If attentive (one q)	(3) If attentive (both q's)	(4) Controls	(5) DS lasso
Random tenant	17.40***	18.16***	18.82***	17.54***	17.51***
	(2.110)	(2.189)	(3.650)	(2.111)	(2.337)
Observations	3032	2780	1002	3032	3032
p	0.000	0.000	0.000	0.000	0.000

This table presents robustness for the Dictator Game (DG) in the tenant experiment. Panel A focuses on the difference between hostility for random landlords versus random tenants, whereas Panel B focuses on the difference between the indifference points $S(x)$. Column 1 corresponds to Row 11, Column 1 of Table A.16. Column 2 limits to attentive tenants who correctly either attention check (Appendix A.3.6). Column 3 limits to tenants who pass both attention checks. Column 4 adds a vector of controls for: prior beliefs (about own and average, across both beliefs, as well as uncertainty), demographics: indicators for Black, female, less than HS, as well as linear controls for age; economic status: having ever formed a payment plan, ever having overdue rents, ever having been evicted, back rents, monthly rent, monthly income, an employment indicator, and self-reports about having been paid by ERAP; and indicators for passing either attention check. Column 5 uses post-double-selection Lasso to select controls (Belloni et al., 2014). The observations is the total number of DGs. As we have multiple observations per individual, standard errors cluster by individual.

Table A.13: Landlords Information Treatment: Robustness

	(1) Request materials	(2) Notify	(3) Number referrals	(4) Wants offer	(5) Never agree	(6) Breakeven
Information	0.109**	-0.0444	-0.105	0.0225	-0.0293	3.306
	(0.0457)	(0.0426)	(0.168)	(0.0530)	(0.0240)	(2.122)
Joint F-test p-value	0.014					
Observations	371	371	371	202	371	371
Control Mean	0.660	0.800	0.605	0.790	0.0750	81.78
p-value	0.0168	0.297	0.533	0.671	0.222	0.119
MHC-adjusted p-value	0.099	0.624	0.762	0.762	0.624	0.465

Note: This table presents the effect of the information treatment on all outcomes in the landlord experiment. We use post-double-selection Lasso to select demographic controls from the set in Appendix A.5 (Belloni et al., 2014). ERAP to resend the offer to settle back rents and is only available for landlords who received the offer in the first place (204/371 landlords). The joint p-value comes from a joint test of whether all outcomes are equal to zero, stacked using seemingly unrelated regression. The multiple-hypothesis corrected p-values come from the stepwise procedure in Romano and Wolf (2005).

Table A.14: Validation of Dictator Game

	Hostile			Maximally Altruistic		
	(1)	(2)	(3)	(4)	(5)	(6)
Opponent:	Own L.	Random L.	Random T.	Own L.	Random L.	Random T.
Panel A. Index						
Index (SD)	-0.160***	-0.143***	-0.054***	0.144***	0.131***	0.089***
	(0.019)	(0.024)	(0.013)	(0.020)	(0.027)	(0.017)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
N	1224	584	1808	1224	584	1808
Panel B. Natural Language Processing						
Sentiment score (SD)	-0.132***	-0.069	-0.007	0.162***	0.112*	0.052
	(0.039)	(0.055)	(0.025)	(0.045)	(0.063)	(0.035)
	[0.00]	[0.21]	[0.79]	[0.00]	[0.07]	[0.13]
N	1000	487	1487	1000	487	1487
Panel C. Likert Scale						
Likert (SD)	-0.087***	-0.081***	-0.021**	0.070***	0.070***	0.032**
	(0.015)	(0.021)	(0.010)	(0.017)	(0.024)	(0.014)
	[0.00]	[0.00]	[0.05]	[0.00]	[0.00]	[0.02]
N	874	416	1290	874	416	1290
Panel D. Simple Lottery						
Enrolls own landlord in lottery	-0.218***	-0.261***	-0.111***	0.188***	0.227***	0.168***
	(0.032)	(0.044)	(0.022)	(0.033)	(0.046)	(0.027)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Constant	0.409	0.412	0.227	0.384	0.392	0.496
	(0.029)	(0.041)	(0.020)	(0.029)	(0.040)	(0.024)
Enrolls random tenant in lottery	-0.125***	-0.230***	-0.135***	0.130***	0.237***	0.199***
	(0.037)	(0.051)	(0.027)	(0.039)	(0.052)	(0.031)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Constant	0.346	0.406	0.256	0.419	0.368	0.458
	(0.034)	(0.048)	(0.025)	(0.036)	(0.047)	(0.029)
N	1224	584	1808	1224	584	1808

Note: This table presents validations of the Dictator Game. The index averages the outcomes in Panels B–D. The tasks are described in Appendix A.3.5. Brackets show p-values.

Table A.15: Hostility and behavior in simple altruism task: landlords

	Hostile			Maximally Altruistic		
	(1)	(2)	(3)	(4)	(5)	(6)
Opponent:	Own T.	Random T.	Random L.	Own T.	Random T.	Random T.
Enrolls own tenant in lottery	-0.296***	-0.164**	-0.197***	0.168**	0.212*	0.330***
	(0.078)	(0.083)	(0.057)	(0.067)	(0.109)	(0.082)
	[0.00]	[0.05]	[0.00]	[0.01]	[0.05]	[0.00]
Constant	0.395	0.200	0.279	0.485	0.520	0.419
	(0.075)	(0.081)	(0.055)	(0.061)	(0.101)	(0.076)
Enrolls random landlord in lottery	-0.146*	-0.261**	-0.214***	0.198***	0.389***	0.157*
	(0.079)	(0.113)	(0.065)	(0.074)	(0.124)	(0.089)
	[0.06]	[0.02]	[0.00]	[0.01]	[0.00]	[0.08]
Constant	0.278	0.294	0.302	0.453	0.353	0.556
	(0.075)	(0.111)	(0.063)	(0.069)	(0.117)	(0.083)
N	234	137	371	371	137	234

Note: This table presents the analog to Table A.14, Panel D, among landlords who completed the same task. Brackets show p-values.

Table A.16: Behavior in Dictator Game

	(1) Hostile	(2) Indifference point	(3) Highly hostile	(4) Highly altruistic
A. Landlord sample (N = 371)				
1. Own Tenant	0.154***	171.5***	0.090***	0.688***
$N = 234$	(0.024)	(4.2)	(0.019)	(0.030)
	[0.000]	[0.000]	[0.000]	[0.000]
2. Random Tenant	0.066***	182.5***	0.036**	0.693***
$N = 137$	(0.021)	(4.0)	(0.016)	(0.040)
	[0.002]	[0.000]	[0.025]	[0.000]
3. Random Landlord	0.119***	174.3***	0.038***	0.623***
$N = 371$	(0.017)	(2.8)	(0.010)	(0.025)
	[0.000]	[0.000]	[0.000]	[0.000]
4. Own Tenant – Random Tenant	0.088***	-11.0*	0.053**	-0.005
(Row 1 – Row 2)	(0.032)	(5.8)	(0.025)	(0.050)
	[0.006]	[0.060]	[0.031]	[0.914]
5. Random Tenant – Random Landlord	-0.053**	8.2**	-0.001	0.071**
(Row 2 – Row 3)	(0.024)	(4.0)	(0.017)	(0.036)
	[0.028]	[0.042]	[0.941]	[0.048]
6. Own Tenant – Random Landlord	0.035	-2.8	0.052***	0.065**
(Row 1 – Row 3)	(0.024)	(4.0)	(0.017)	(0.030)
	[0.150]	[0.485]	[0.003]	[0.030]
B. Tenant sample (N = 1,808)				
7. Own Landlord	0.240***	153.5***	0.127***	0.529***
$N = 1{,}224$	(0.012)	(2.1)	(0.010)	(0.014)
	[0.000]	[0.000]	[0.000]	[0.000]
8. Random Landlord	0.217***	158.2***	0.103***	0.562***
$N = 584$	(0.017)	(2.9)	(0.013)	(0.021)
	[0.000]	[0.000]	[0.000]	[0.000]
9. Random Tenant	0.143***	170.9***	0.065***	0.624***
$N = 1{,}808$	(0.008)	(1.4)	(0.006)	(0.011)
	[0.000]	[0.000]	[0.000]	[0.000]
10. Own Landlord – Random Landlord	0.023	-4.7	0.024	-0.033
(Row 7 – Row 8)	(0.021)	(3.5)	(0.016)	(0.025)
	[0.279]	[0.183]	[0.130]	[0.187]
11. Random Landlord – Random Tenant	0.075***	-12.7***	0.038***	-0.063***
(Row 8 – Row 9)	(0.018)	(2.9)	(0.013)	(0.020)
	[0.000]	[0.000]	[0.003]	[0.002]
12. Own Landlord – Random Tenant	0.097***	-17.4***	0.062***	-0.096***
(Row 7 – Row 9)	(0.013)	(2.1)	(0.010)	(0.013)
	[0.000]	[0.000]	[0.000]	[0.000]

Note: For each participant, we elicit the point $S(x)$ at which participants are indifferent between the bundle (s self, 0 opponent) and (x self, x opponent). We randomly assign landlords (Panel A) to play against their own tenant (Row 1) or random tenant (Row 2), as well as a random landlord (Row 3). We randomly assign tenants (Panel B) to play against their own landlord (Row 7) or random landlord (Row 8), as well as a random tenant (Row 9). Columns 4–6 and 10–12 show differences between Dictator Game outcomes depending on the opponent. Column (1) shows the share of participants who are "hostile" — i.e., $S(x) < x$. Column (2) shows the normalized value $100 \times S(x)/x$, so that 100 represents that the participant is indifferent between (x, 0) and (x, x). Columns (3) and (4) show the share who are highly hostile or altruistic, respectively: the multiple price list permitted subjects to report $S(x) \in [0, x/10)$ (high hostility) or $S(x) > 2x$ (high altruism). Parentheses show robust standard errors. Brackets show p-values. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.17: Behavior in Dictator Game: First DG Only

	(1) Hostile	(2) Indifference point	(3) Highly hostile	(4) Highly altruistic
A. Landlord sample ($N = 203$)				
1. Own Tenant	0.174***	169.5***	0.092***	0.688***
$N = 109$	(0.037)	(6.4)	(0.028)	(0.045)
	[0.000]	[0.000]	[0.001]	[0.000]
2. Random Tenant	0.034	189.6***	0.034	0.763***
$N = 59$	(0.024)	(5.4)	(0.024)	(0.056)
	[0.159]	[0.000]	[0.159]	[0.000]
3. Random Landlord	0.069***	179.8***	0.025**	0.616***
$N = 203$	(0.018)	(3.2)	(0.011)	(0.034)
	[0.000]	[0.000]	[0.025]	[0.000]
4. Own Tenant – Random Tenant	0.140***	-20.1**	0.058	-0.075
(Row 1 – Row 2)	(0.043)	(8.3)	(0.036)	(0.071)
	[0.001]	[0.017]	[0.114]	[0.296]
5. Random Tenant – Random Landlord	-0.035	9.8	0.009	0.147**
(Row 2 – Row 3)	(0.030)	(6.2)	(0.026)	(0.065)
	[0.237]	[0.118]	[0.722]	[0.025]
6. Own Tenant – Random Landlord	0.105***	-10.3	0.067**	0.072
(Row 1 – Row 3)	(0.041)	(7.1)	(0.030)	(0.056)
	[0.010]	[0.150]	[0.025]	[0.198]
B. Tenant sample ($N = 960$)				
7. Own Landlord	0.234***	153.0***	0.112***	0.481***
$N = 572$	(0.018)	(2.9)	(0.013)	(0.021)
	[0.000]	[0.000]	[0.000]	[0.000]
8. Random Landlord	0.181***	163.4***	0.083***	0.554***
$N = 276$	(0.023)	(3.9)	(0.017)	(0.030)
	[0.000]	[0.000]	[0.000]	[0.000]
9. Random Tenant	0.079***	182.8***	0.026***	0.691***
$N = 960$	(0.009)	(1.5)	(0.005)	(0.015)
	[0.000]	[0.000]	[0.000]	[0.000]
10. Own Landlord – Random Landlord	0.053*	-10.3**	0.029	-0.074**
(Row 7 – Row 8)	(0.029)	(4.8)	(0.021)	(0.037)
	[0.069]	[0.033]	[0.179]	[0.044]
11. Random Landlord – Random Tenant	0.102***	-19.4***	0.057***	-0.136***
(Row 8 – Row 9)	(0.025)	(4.2)	(0.017)	(0.033)
	[0.000]	[0.000]	[0.001]	[0.000]
12. Own Landlord – Random Tenant	0.155***	-29.8***	0.086***	-0.210***
(Row 7 – Row 9)	(0.020)	(3.3)	(0.014)	(0.026)
	[0.000]	[0.000]	[0.000]	[0.000]

Note: See notes to Table A.16 for description of our altruism and hostility measures. Parentheses show robust standard errors. Brackets show p-values. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table is identical to Table A.16 except it only keeps the first instance that each participant plays the DG.

Table A.18: Behavior in Dictator Game: Entropy-Weight Adjusted (Hainmueller, 2012)

	(1) Hostile	(2) Indifference point	(3) Highly hostile	(4) Highly altruistic
A. Landlord sample ($N = 371$)				
1. Own Tenant	0.149***	172.9***	0.085***	0.701***
$N = 234$	(0.024)	(4.2)	(0.019)	(0.031)
	[0.000]	[0.000]	[0.000]	[0.000]
2. Random Tenant	0.072***	183.0***	0.040**	0.712***
$N = 137$	(0.023)	(4.4)	(0.018)	(0.041)
	[0.003]	[0.000]	[0.025]	[0.000]
3. Random Landlord	0.125***	173.7***	0.040***	0.627***
$N = 371$	(0.018)	(3.0)	(0.011)	(0.027)
	[0.000]	[0.000]	[0.000]	[0.000]
4. Own Tenant – Random Tenant	0.077**	-10.1*	0.045*	-0.011
(Row 1 – Row 2)	(0.033)	(6.1)	(0.026)	(0.052)
	[0.021]	[0.099]	[0.083]	[0.831]
5. Random Tenant – Random Landlord	-0.053**	9.3**	0.000	0.085**
(Row 2 – Row 3)	(0.026)	(4.4)	(0.018)	(0.038)
	[0.043]	[0.033]	[0.982]	[0.026]
6. Own Tenant – Random Landlord	0.025	-0.7	0.045**	0.074**
(Row 1 – Row 3)	(0.025)	(4.1)	(0.018)	(0.032)
	[0.322]	[0.857]	[0.012]	[0.019]
B. Tenant sample ($N = 1,808$)				
7. Own Landlord	0.233***	154.4***	0.126***	0.535***
$N = 1,224$	(0.012)	(2.1)	(0.010)	(0.015)
	[0.000]	[0.000]	[0.000]	[0.000]
8. Random Landlord	0.226***	156.8***	0.102***	0.547***
$N = 584$	(0.018)	(3.1)	(0.013)	(0.022)
	[0.000]	[0.000]	[0.000]	[0.000]
9. Random Tenant	0.141***	171.4***	0.063***	0.628***
$N = 1,808$	(0.009)	(1.4)	(0.006)	(0.012)
	[0.000]	[0.000]	[0.000]	[0.000]
10. Own Landlord – Random Landlord	0.006	-2.4	0.025	-0.011
(Row 7 – Row 8)	(0.022)	(3.7)	(0.016)	(0.026)
	[0.780]	[0.521]	[0.135]	[0.666]
11. Random Landlord – Random Tenant	0.085***	-14.6***	0.039***	-0.081***
(Row 8 – Row 9)	(0.019)	(3.1)	(0.014)	(0.021)
	[0.000]	[0.000]	[0.004]	[0.000]
12. Own Landlord – Random Tenant	0.091***	-17.0***	0.064***	-0.092***
(Row 7 – Row 9)	(0.013)	(2.2)	(0.010)	(0.014)
	[0.000]	[0.000]	[0.000]	[0.000]

Note: See notes to Table A.16 for description of our altruism and hostility measures. Parentheses show robust standard errors. Brackets show p-values. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. We compute weights by comparing tenant participants to non-participants based on observables at application. For landlords, we compare the tenant named on their survey to the randomly selected tenant of non-participants (as in Table 1.2).

Table A.19: Behavior in Dictator Game: All Participants, Prior to Attrition

	(1) Hostile	(2) Indifference point	(3) Highly hostile	(4) Highly altruistic
A. Landlord sample ($N = 565$)				
1. Own Tenant	0.166***	169.8***	0.092***	0.671***
$N = 368$	(0.019)	(3.4)	(0.015)	(0.025)
	[0.000]	[0.000]	[0.000]	[0.000]
2. Random Tenant	0.061***	184.7***	0.030**	0.721***
$N = 197$	(0.017)	(3.2)	(0.012)	(0.032)
	[0.000]	[0.000]	[0.014]	[0.000]
3. Random Landlord	0.126***	173.9***	0.041***	0.628***
$N = 565$	(0.014)	(2.3)	(0.008)	(0.020)
	[0.000]	[0.000]	[0.000]	[0.000]
4. Own Tenant – Random Tenant	0.105***	-14.9***	0.062***	-0.050
(Row 1 – Row 2)	(0.026)	(4.7)	(0.019)	(0.040)
	[0.000]	[0.002]	[0.002]	[0.219]
5. Random Tenant – Random Landlord	-0.065***	10.8***	-0.010	0.092***
(Row 2 – Row 3)	(0.020)	(3.4)	(0.014)	(0.030)
	[0.002]	[0.002]	[0.453]	[0.002]
6. Own Tenant – Random Landlord	0.040*	-4.1	0.052***	0.043*
(Row 1 – Row 3)	(0.021)	(3.4)	(0.014)	(0.025)
	[0.054]	[0.233]	[0.000]	[0.081]
B. Tenant sample ($N = 2{,}402$)				
7. Own Landlord	0.260***	150.0***	0.141***	0.518***
$N = 1{,}621$	(0.011)	(1.9)	(0.009)	(0.012)
	[0.000]	[0.000]	[0.000]	[0.000]
8. Random Landlord	0.228***	156.4***	0.110***	0.552***
$N = 781$	(0.015)	(2.5)	(0.011)	(0.018)
	[0.000]	[0.000]	[0.000]	[0.000]
9. Random Tenant	0.157***	169.3***	0.073***	0.622***
$N = 2{,}402$	(0.007)	(1.3)	(0.005)	(0.010)
	[0.000]	[0.000]	[0.000]	[0.000]
10. Own Landlord – Random Landlord	0.032*	-6.4**	0.031**	-0.034
(Row 7 – Row 8)	(0.019)	(3.1)	(0.014)	(0.022)
	[0.087]	[0.040]	[0.028]	[0.114]
11. Random Landlord – Random Tenant	0.071***	-12.9***	0.037***	-0.070***
(Row 8 – Row 9)	(0.016)	(2.5)	(0.011)	(0.017)
	[0.000]	[0.000]	[0.001]	[0.000]
12. Own Landlord – Random Tenant	0.103***	-19.3***	0.068***	-0.104***
(Row 7 – Row 9)	(0.012)	(1.9)	(0.009)	(0.012)
	[0.000]	[0.000]	[0.000]	[0.000]

Note: This table is the same as Table A.16, but includes all participants from either the landlord survey or tenant survey prior to attrition. The table includes all observations that complete the module, including potential duplicates who take the survey, attrit, and then re-take from another unique link.

Table A.20: Belief Updating and Hostility

Panel A: Landlords

	(1) 1(updates)	(2) Update (signed)	(3) 1(updates)	(4) Update (signed)
Information treatment	0.422***	-12.68**	0.431***	-15.02***
	(0.129)	(4.925)	(0.130)	(4.843)
Treat × indiff.	-0.0000958	0.0165	-0.000138	0.0266
	(0.000709)	(0.0269)	(0.000714)	(0.0259)
Indifference: tenant			-0.0000116	0.00278
			(0.0000280)	(0.00489)
Observations	234	234	234	234
Controls			✓	✓

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Panel B: Tenants

	(1) 1(updates)	(2) 1(updates)	(3) 1(updates)	(4) 1(updates)
Info	0.374	0.340	0.374	0.340
	(0.017)	(0.017)	(0.017)	(0.017)
	[0.000]	[0.000]	[0.000]	[0.000]
Info × highly hostile	-0.068	-0.093	-0.068	-0.089
	(0.051)	(0.043)	(0.051)	(0.043)
	[0.181]	[0.033]	[0.181]	[0.041]
Observations	1808	1808	1808	1808
Outcome	Barg.	Alt.	Barg.	Alt.
Controls			✓	✓
p: joint test	.	0.0446	.	0.0535

Note: This table presents whether landlords and tenants update their beliefs and whether the magnitude of belief updates is related to hostility. For landlords in Panel A, the update is about tenants' recoupment probabilities, interacted by whether they are hostile toward their own tenants. The sample size is the 234 landlords who played the Dictator Game against their own tenant. For tenants in Panel B, the update is about landlords' bargaining probabilities and altruism, interacted by whether they are highly hostile toward landlords. We define high hostility in the notes to Table A.16. Columns 1 and 3 show the effects on tenant beliefs about bargaining. Columns 2 and 4 show the effects on tenant beliefs about landlord altruism. The joint tests in Columns 2 and 4 show joint tests of the two interaction terms in Columns 1–2 and 3–4, respectively.

Table A.21: Landlord heterogeneity: Information and Hostility

	(1) 1(receives materials)	(2) 1(receives materials)
Belief update	-0.0176***	-0.00108
	(0.00578)	(0.00418)
Update × hostile	0.0182**	
	(0.00769)	
Update × indifference point		-0.785*
		(0.422)
Observations	234	234

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the differential effect of the information treatment on receiving materials. We interact the instrumental variables specification in Figure 1.3 with hostility toward own tenant (Column 1) or the landlord indifference point toward the tenant (Column 2). Rows 1 and 2 present separate specifications. The interaction term shows the additional effect of behavior in the dictator game on the relationship between beliefs and behaviors. There are 234 participants because 234 landlords out of 374 play the Dictator Game against their own tenant.

Table A.22: Policy Evaluation Balance Table: Demographics

Demographic	Above Median Payment Time	Below Median Payment Time	Difference
Age	36.75	35.77	-0.98
			(0.45)
Female	0.76	0.75	-0.01
			(0.01)
Black	0.91	0.91	-0.00
			(0.01)
Disabled	0.11	0.11	-0.00
			(0.01)
Household size	2.4	2.3	-0.1
			(0.0)
Employed	0.40	0.42	0.03
			(0.02)
Household monthly income†	1478	1587	109
			(56)
Monthly rent†	850	871	21
			(10)
Back rent owed at application†	4500	3756	-744
			(182)
N	2,037	2,044	

Note: This table shows a regression of pre-payment eviction filing or judgment rates on an indicator for having below-median time between the date the case was created and payment. Columns 2 and 3 include controls for week of application payment and calendar time. Regressions restrict to the sample of people with all non-missing demographics, which is why the total N at the bottom is not the same as in Table 1.1. †: shows medians and differences from quantile regressions.

Table A.23: Balance Table: Changes in Filing/Judgments Rates

	(1) Filing	(2) Judgment	(3) Filing	(4) Judgment
1(below med. time)	0.00320*	-0.000179	-0.00127	-0.00183
	(0.00166)	(0.00106)	(0.00222)	(0.00129)
Calendar week FE	Yes	Yes	Yes	Yes
Application and payment week FE	No	No	Yes	Yes
Case-period obs.	18968	18968	18968	18968
Cases	4742	4742	4742	4742

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents regressions of demographics on an indicator for having below-median time between the date the case was created and payment.

Table A.24: ERAP Treatment Effects

	(1)	(2)	(3)
A. Treatment Effects (pp / 6 months)			
1. Average judgments	-0.85	-3.19	0.63
	(1.56)	(2.22)	(0.69)
	[0.585]	[0.150]	[0.362]
2. Average filings	0.20	-8.67	-6.13
	(2.32)	(3.21)	(1.57)
	[0.932]	[0.007]	[0.000]
3. 0–1 month filings	-5.71	-16.04	-8.48
	(1.30)	(2.11)	(1.82)
	[0.000]	[0.000]	[0.000]
B. Interpretation			
4. Maximum simulated effect: judgments	-4.71	-8.31	-4.77
5. Maximum simulated effect: filings	-11.13	-20.12	-16.44
6. Fiscal cost per judgment (point estimate)	644,310	172,412	·
7. Fiscal cost per filing (point estimate)	·	63,409	89,622
8. Fiscal cost per judgment (95% lower bound)	140,496	72,981	759,349
9. Fiscal cost per filing (95% lower bound)	126,527	36,750	59,621
N paid	4,742	4,742	4,742
N non-paid	·	·	4,905
Design:	Only-paid (non-parametric)	Only-paid, reweighted (non-parametric)	Non-paid (stacked)

Note: This table shows treatment effects, across empirical strategies, of ERAP payment on judgments and filings. Rows 1 and 2 show the cumulative effect over six months; that is, they multiply the average two-month effect by three. Row 3 shows the average effect in the first two months, multiplied by three; that is, it shows the effect in the same units as Rows 1–2. Rows 1 and 2 subtract the pre-period mean whereas Row 3 just shows the event-study coefficient for the first two-month period. The maximum simulated effect on judgments and filings (Rows 4–5) replace judgments or filings as 0 if in the treated group. To recover the treatment effects presented in the figures, divide estimates in Rows 1–2 by 3 (as those show per-period effects, and these show cumulative effects). Rows 6–9 present the fiscal cost of a deferred judgment or filing, based on the average payment made in the sample. Rows 6–9 are empty when the fiscal cost is infinite. Column 1 shows aggregates from the main specification (Equation A.83). Column 2 shows reweighted estimates. Column 3 shows estimates from the alternative design with non-paid applicants (Equation A.85). Parentheses show standard errors. Brackets show p-values.

Table A.25: Efficient, Inefficient, and Repugnant Evictions

Panel A. Eviction	
Total eviction %	19.1
pp efficient	14.6
pp inefficient	4.6
pp from hostility	3.0
Panel B. Bargaining	
Total bargaining %	80.9
pp efficient	80.9
pp from altruism	13.5
pp inefficient	0.0

Note: This table shows the share of evictions that are efficient and inefficient. Inefficient evictions are those that would not occur without misperceptions. "Repugnant" evictions are those caused by hostility.

Table A.26: Match to Moments

Moments	**Estimated Value**	**Targeted Value**
ERAP mean judgment rate (unconditional)	0.158	0.091
Treatment effect on judgments (unconditional)	-0.016	-0.009
Landlord take-up rate	0.668	0.648
Payment plan rate (proportion of back rent)	0.322	0.315
Interaction: a_{Li} and take-up	0.444	0.450
Interaction: p_{Li} and take-up	0.273	0.284
Interaction: $a_{Li} \times p_{Li}$ and take-up	0.214	0.219
Interaction: a_{Ti} and take-up	0.149	0.185
Interaction: $\tilde{p}_{Ti}$ and take-up	-0.057	-0.056
Interaction: $a_{Ti} \times \tilde{p}_{Ti}$ and take-up	-0.022	-0.031
Interaction: d_i and take-up	717	690
Mean: judgments (from tenants)	0.180	0.238
Interaction: a_{Ti} and judgment	0.058	0.110
Interaction: $\tilde{p}_{Ti}$ and judgment	-0.034	-0.045
Interaction: $a_{Ti} \times \tilde{p}_{Ti}$ and judgment	-0.008	-0.016
Interaction: d_i and judgment	404	526
Mean: judgments (from landlords)	0.180	0.161
Interaction: a_{Li} and judgment	0.096	0.064
Interaction: $\tilde{p}_{Li}$ and judgment	0.077	0.065
Interaction: $a_{Li} \times \tilde{p}_{Li}$ and judgment	0.045	0.040
Interaction: Takes_i and judgment	0.115	0.091
Interaction: $a_{Li} \times \text{Takes}_i$ and judgment	0.072	0.049
Interaction: $\tilde{p}_{Li} \times \text{Takes}_i$ and judgment	0.051	0.043
Interaction: $a_{Li} \times \tilde{p}_{Li} \times \text{Takes}_i$ and judgment	0.034	0.027

Notes: This table displays the match to the moments (see full list in Appendix A.2.1.2). The table displays unweighted values in their natural units. In estimation, all moments except the first two "macro moments" are scaled by the inverse standard deviation of residuals so units are comparable. Moments labeled with "Interaction" mean that we are displaying the mean value of the first variable interacted with the second variable.

Table A.27: Section 1.7 Robustness

Change to quantitative model	(1) Efficient eviction rate	(2) Inefficient eviction rate	(3) Repugnant eviction rate	(4) % TOT if no L altruism
1. Primary estimate (no changes)	**14.6**	**4.6**	**3.0**	**-58.8**
2. ERAP TOT: 50% of judgments & filings	16.3	4.5	3.0	-66.0
3. Exclude landlord correlations	18.4	4.1	2.9	-58.1
4. Exclude tenant correlations	10.5	4.8	3.3	-51.6
5. More evictions: 25% judgment rate	16.3	4.6	3.0	-57.0
6. Different take-up proxy	13.6	6.2	3.2	-50.5
7. Different bargaining moment	14.3	4.6	3.1	-55.4
8. Costlier ERAP: k_L^P = $1,000	13.1	6.5	3.3	-56.3
9. Cheaper ERAP: k_L^P = $0	15.1	3.9	2.9	-55.9
10. Correlation: k_{Li} and p_{Li}	12.5	4.5	3.1	-61.1
11. Assume tenants have correct beliefs	16.0	2.6	1.6	-53.4
12. Use beliefs about the average	14.4	4.3	2.9	-55.3
13. Use identity weight for ERAP moments	16.8	4.6	3.0	-65.3

Notes: Rows 1 and 10 are estimated on the same 200 bootstraps as in the main analysis. Other rows are estimated on 51 bootstraps and involve re-estimating parameters. Columns 1–3 show the efficient, inefficient, and repugnant eviction rates ($\times$100), similar to Table A.25. To compute the percent inefficient, divide Column 2 by Column 2 + Column 1. Column 4 shows the effect of ERAP on evictions if altruistic landlords are simulated not to be altruistic, similar to Row 4 of Table 1.4 (dividing Column 3 by Column 5 of that table). Row 2 of the Table A.27 shows the effects if ERAP had a 50% TOT. Rows 3–4 show the effects if we exclude landlord or tenant correlations (Appendix C.3). Row 5 simulates the effects if evictions are more common, conditional on being in the sample. The purpose of this row is to examine the consequence of using depressed Covid-era judgment rates on our conclusions. Row 6 changes the take-up proxy to be 1 if and only if the landlord requested materials in the experiment. Row 7 changes the tenant bargaining moment to rely on an indicator for wanting a payment plan, rather than a continuous variable for the amount repaid (Appendix C.3). Rows 8–9 vary the ERAP cost parameter k_L^P. Row 10 permits correlation between k_{Li} and p_{Li} that roughly matches the empirical correlation between these. In particular, we simulate above-median p_{Li} as having 0.2 standard deviation higher k_{Li}, where standard deviations are measured from the simulated distribution of $\varepsilon_{i1} - \varepsilon_{i2}$. Row 11 assumes tenants have perfect beliefs ($p_{Ti} = 0.06$). Row 12 uses the beliefs about the average in the experiment, rather than beliefs about own landlord or tenant. Row 13 uses the identity weight matrix to weight the ERAP "macro" moments.

Table A.28: Associations between Hostility and 311 Calls

Panel A: Landlords

	(1) Has 311 call	(2) Has 311 call	(3) Has 311 call	(4) Has 311 call
Hostile	0.0522	0.160		
	(0.0864)	(0.0991)		
Highly hostile			-0.0131	0.0392
			(0.104)	(0.117)
Constant	0.353***	0.307***	0.361***	0.329***
	(0.0302)	(0.0379)	(0.0295)	(0.0372)
Observations	289	180	289	180
Sample	All	Own	All	Own

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Panel B: Tenants

	(1) Has 311 call	(2) Has 311 call	(3) Has 311 call	(4) Has 311 call
Hostile	0.0340	0.0285		
	(0.0303)	(0.0364)		
Highly hostile			0.113***	0.116**
			(0.0408)	(0.0484)
Constant	0.404***	0.407***	0.399***	0.400***
	(0.0142)	(0.0173)	(0.0132)	(0.0161)
Observations	1544	1049	1544	1049
Sample	All	Own	All	Own

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Panels A and B shows the relationships between hostility in the DG and matched 311 calls. The sample excludes landlords and tenants in Shelby County but not Memphis, since only Memphis provides public-access 311 data. "All" and "Own" refer to the samples against which the participant played the DG.

A.2 Model Appendix

A.2.1 Quantitative Model Details (Appendix to Section 1.7)

A.2.1.1 Model Set-up

The landlord's payoffs V_{Li}^{k} are:

$$\begin{aligned} V_{Li}^{\text{NoERAP,Evict}} &= (1-a_{Li})p_{Li}d_i - a_{Li}k_T - k_L + \varepsilon_{i1} && \text{(A.1)} \\ V_{Li}^{\text{NoERAP,Bargain}} &= (1-a_{Li})n_i^*(d_i) + \varepsilon_{i2} && \text{(A.2)} \\ V_{Li}^{\text{ERAP,Evict}} &= d_i - k_L^P - a_{Li}k_T - k_L + \varepsilon_{i3} && \text{(A.3)} \\ V_{Li}^{\text{ERAP,Bargain}} &= d_i - k_L^P + (1-a_{Li})n_i^*(0) + \varepsilon_{i4}. && \text{(A.4)} \end{aligned}$$

Nash Bargaining yields the following solutions for bargaining payoffs (see proofs in Appendix B.6):

$$n_i^*(d_i, \varepsilon_{i2}, \varepsilon_{i1}; \beta) = \underbrace{\beta\left(p_{Li}d_i - \frac{k_L + k_{Ti}a_{Li} + \varepsilon_{i2} - \varepsilon_{i1}}{1-a_{Li}}\right)}_{\beta\,\times\,\text{altruism-adjusted outside option of landlord}} + \underbrace{(1-\beta)\left(p_{Ti}d_i + \frac{k_{Ti} + (k_L + \varepsilon_{i2} - \varepsilon_{i1})a_{Ti}}{1-a_{Ti}}\right)}_{-(1-\beta)\,\times\,\text{altruism-adjusted outside option of tenant}} \tag{A.5}$$

$$n_i^*(0, \varepsilon_{i4}, \varepsilon_{i3}; \beta) = \beta\left(0 - \frac{k_L + k_{Ti}a_{Li} + \varepsilon_{i4} - \varepsilon_{i3}}{1-a_{Li}}\right) + (1-\beta)\left(0 + \frac{k_{Ti} + (k_L + \varepsilon_{i4} - \varepsilon_{i3})a_{Ti}}{1-a_{Ti}}\right) \tag{A.6}$$

for tenant bargaining parameter β. The bargaining payoff $n_i^*(\cdot) \in \mathbb{R}$ represents a transfer from tenants to landlords when positive. We typically suppress dependence on the shocks and bargaining power for readability.

Let $\mathcal{E}^{\text{NoERAP}}$ be an indicator that is 1 if and only if

$$k_L + k_T + \varepsilon_{i2} - \varepsilon_{i1} \leq \frac{(p_{Li} - p_{Ti})d_i}{\bar{a}_i} \tag{A.7}$$

is satisfied. Let $\mathcal{E}^{\text{ERAP}}$ be an indicator that is 1 if and only if

$$k_L + k_T + \varepsilon_{i4} - \varepsilon_{i3} \leq 0 \tag{A.8}$$

is satisfied. Then the landlord's maximization problem is:

$$V_{Li} = \begin{cases} \max\left[V_{Li}^{\text{NoERAP,Evict}}, V_{Li}^{\text{ERAP,Evict}}\right] & \text{if } \mathcal{E}^{\text{NoERAP}} = 1 \text{ and } \mathcal{E}^{\text{ERAP}} = 1 \\ \max\left[V_{Li}^{\text{NoERAP,Bargain}}, V_{Li}^{\text{ERAP,Evict}}\right] & \text{if } \mathcal{E}^{\text{NoERAP}} = 0 \text{ and } \mathcal{E}^{\text{ERAP}} = 1 \\ \max\left[V_{Li}^{\text{NoERAP,Evict}}, V_{Li}^{\text{ERAP,Bargain}}\right] & \text{if } \mathcal{E}^{\text{NoERAP}} = 1 \text{ and } \mathcal{E}^{\text{ERAP}} = 0 \\ \max\left[V_{Li}^{\text{NoERAP,Bargain}}, V_{Li}^{\text{ERAP,Bargain}}\right] & \text{if } \mathcal{E}^{\text{NoERAP}} = 0 \text{ and } \mathcal{E}^{\text{ERAP}} = 0. \end{cases} \tag{A.9}$$

Thus, based on the realization of the shocks, the landlord solves a maximization problem of whether to take up ERAP. She accounts for the fact that if she does or does not, the eviction decision is guaranteed.

Discussion of Model and the Memphis ERAP. The model abstracts from ERAP institutional details in several ways. First, if the landlord rejected ERAP, the tenant could still obtain a direct payment. The model assumes the landlord completely ignores this possibility. Second, the model assumes that ERAP pays just d_i but not additional months of rent. The months of rent were supposed to compensate the landlord for delays in receiving ERAP. Thus, we assume ERAP paid just d_i. Third, the model does not account for utilities payments or legal assistance that ERAP could provide.

A.2.1.2 Estimation

Method of Simulated Moments/Generalized Indirect Inference. We use Method of Simulated Moments (MSM). We draw a vector of shocks to form simulated moments, which we stack into $\hat{m}(X_i;\theta)$. We solve:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left[\hat{m}(X_i;\theta)' W \hat{m}(X_i;\theta) \right] \tag{A.10}$$

for weight matrix W.

We use a Generalized Indirect Inference (GII) procedure to match moments (Bruins et al., 2018). MSM can often yield nonsmooth problems. Small changes in the parameter values can cause agents to choose a different bundle, and then simulating behavior given that choice will exhibit discrete differences. In our context, small changes in parameter values can generate "spiky" regions where slightly adjusting parameters induces large changes in conditional probabilities. A natural solution is to smooth the indicator into a probability so that the problem can be fed into standard optimization tools.[1] Bruins et al. (2018) provide a smoothing technique which is popular in problems of this form. We sketch our use of this algorithm below.

We use the following recipe:

1. Form data from the experiment Ξ_i.
2. Using the data from the experiment, form model-implied estimates of probabilities in each of the four end states.
3. Employ the Generalized Indirect Inference algorithm of Bruins et al. (2018):
    - Draw a value of type-I extreme value shocks ε_{ik} for landlord payoffs $k \in 1, \ldots, 4$.
    - Fix an initial smoothing value λ_0.

[1] Global optimization tools did not converge reliably in our case, as the problem is sufficiently nonsmooth.

- For each landlord payoff k, form smoothed simulated choice probabilities $\mathcal{T}_{ik}(\lambda_0) \in [0,1]$ as:

$$\mathcal{T}_{ik}(\lambda_0) = \Phi\left(\lambda^{-1}\left(V_{ik}(X_i, \varepsilon_{ik}; \theta)\right)\right), \tag{A.11}$$

  for logistic CDF $\Phi(\cdot)$.
- Stack simulated moments $\hat{m}(X_i;\theta)$ and solve:

$$\hat{\theta}_0 = \arg\min_{\theta\in\Theta}\left\{\hat{m}(X_i;\theta)' W \hat{m}(X_i;\theta)\right\}. \tag{A.12}$$

- Form $\lambda_j = \rho\lambda_{j-1}$ for $\rho \in (0,1)$. Repeat previous steps with λ_j until $\|\theta(\lambda_j) - \theta(\lambda_{j-1})\| <$ Tol for some norm $\|\cdot\|$.

The basic logic of GII in this setting is that, instead of simulating a discrete take-up choice based on model-implied utilities, let "as-if probabilities" $\mathcal{T}_i$ follow a logit structure over those utilities, and let the logit be characterized by scale parameter λ. These take-up probabilities are "as-if" because they are formed using the logit structure, but they do not have a direct model-based interpretation as probabilities since take-up is deterministic in the model. For a given $\lambda > 0$, the problem is now smooth. Then, iterate on λ until the model converges. As $\lambda \to 0$, $\mathcal{T}_i$ approaches an indicator.

In practice, to implement GII, we initialize at a coarse enough smoothing parameter such that probabilities are fairly uniform over the unit interval. We find an initial valuue using Matlab's `patternsearch` algorithm, which is designed for global optimization, to choose a candidate minimum value. We then iterate on the minimization using `patternsearch` and $\rho = 1/1.05$ until convergence.

We do not smooth the constraints (Equations A.7–A.8) but tests indicate that doing so yields similar results.

Because all our parameter estimates use a matching procedure that involves randomness, we present the average value of the parameter across bootstraps in Table 1.3. Our counterfactuals estimate behavior using the value of θ obtained for the given bootstrap and then averages.

We form standard errors by bootstrapping the entire procedure, including matching the landlord–tenant pairs (below). We bootstrap the treatment effect moments and experimental data separately.[2] Standard errors represent the standard deviation of bootstraps.

Preparing the Experimental Data for Use in the Quantitative Model. The dataset includes the 903 tenants with positive back rents at the time of the tenant survey who played the DG against their own landlord, and the 199 landlords who played the DG against their own tenant.

We document how we prepare the variables that enter the estimation.

[2]We do not bootstrap the mean judgment rates from the ERAP evaluation, so standard errors do not account for this (small) source of sampling variation.

- Altruism a_{Li}, a_{Ti}. We read these directly off the DG experiments, putting $a_{ji} := (S(x) - x)$. To ensure that $\bar{a}_i$ is well-defined, we top-code $|a_{ji}|$ at 0.95.

- Landlord beliefs p_{Li}. We read this off the priors in the experiment questions about landlord recoupment. We use beliefs about own tenant, not average tenant.

- Tenant beliefs p_{Ti}. We do not observe tenant beliefs about paying back a judgment directly. We did not ask tenants whether they thought they could abscond in a court proceeding because we were concerned that this could harm landlords by making tenants more likely to abscond. We form a measure of tenant beliefs:

$$\tilde{p}_{Ti} := -\frac{1}{2}\left(\left(p'_{1i} - 0.31\right) + \left(0.69 - p'_{2i}\right)\right) + 0.06, \tag{A.13}$$

  where p'_{1i} and p'_{2i} are beliefs about own landlord's filing and altruism behaviors, respectively. The logic behind this measure is that if tenants have perfect beliefs about these forces, then we impute them as having perfect beliefs about repayment ($p_{Ti} = 0.06$). If tenants are pessimistic about landlords' altruism, we impute them as having optimistic beliefs (from their perspective) about whether they will pay back, and similarly if tenants are optimistic that landlords will drop a filing. We use beliefs about own landlords, not average landlords.

  Several assumptions are required for this imputation to be valid. First, we assume that beliefs scale linearly in percentage points across the measures. Second, our measure of $\tilde{p}_{Ti}$ may be negative. As noted in footnote 14, we can recast the p_{Li} and p_{Ti} values as scalar shifters of the expected value of reclaiming d_i if bargaining fails. For instance, if $p_{Ti} < 0$, tenants think they will get more money by going to court than by bargaining. Such a perspective is plausible if, for instance, they can reside in the unit while the court process is ongoing and they do not expect to pay a judgment.

  As these are strong assumptions, we also conduct the exercise where we simulate $p_{Ti} = 0.06$ (Table A.27, Row 11). This generally gives lower levels of inefficiency. We also conduct an exercise where we use beliefs about the average landlord or tenant, rather than own landlord or tenant and find similar results (Row 12).

- Back rent d_i. We use the tenant's value of back rents that they report in the experiment.

- Landlord take-up (ERAP$_i$). We estimate landlord take-up from the experiment as follows. Take-up is 1 if they indicate that they wish to receive a rental contract offer from ERAP. We then replace take-up equal to 0 if the landlord declines to receive materials about ERAP or if they decline to have ERAP notify their tenant when future opportunities to apply are available. The advantage of this measure is that it uses several of the take-up proxies. In a robustness check, we impute take-up as 1 if and only if the landlord chooses to receive ERAP materials (Table A.27, Row 6). This moves the share of inefficient evictions to around 30% and share repugnant to around 50%.

- Judgments (Judgment_i). We use landlord and tenant judgments from links to court data (Section 1.6). As we randomly match landlords and tenants (see below), we have different values for whether the landlord links or tenant links to the court data are matched to a judgment (Judgment_{Li} and Judgment_{Ti}), but only one simulated value for a judgment based on the model.

- $n_i^*(d_i)$. We use the tenant's offer of how much back rent to pay back in a payment plan. Using these data requires the assumption that tenants' first offers exactly correspond to the Nash offer. We transform the data so that the variable n_i^*, used in moments below, is the fraction that the tenant offers to repay the landlord as a share of the total back rents that she owed. The variable n_i^* is censored between 0 and 1, since the payment plan activity did not let the tenant demand payment from the landlord or pay more than she owed. We accordingly also censor the model-implied bargaining offer.[3]

Other Data Preparation Details. For power, we do use experiment participants from both the legal and non-legal sides of the program and from people who moved after they applied to ERAP. We do not use participants who did DGs with random landlords or tenants.

We do not observe matched landlord–tenant pairs, since they participate in the experiments separately. As a benchmark, we draw landlord and tenant pairs to match at random. Given that we see moderate assortative matching on altruism, we simulate this assortative matching by replacing the landlord altruism as equal to the tenant altruism for 27.5% of pairs matched at random (Appendix A.5.2). Since there are more tenants than landlords, we resample landlords at random within the landlord sample.

Moments. Our moments are:

1. ERAP's treatment effect on judgments, where

$$\text{TOT} - \left(\mathbb{E}[\hat{\text{Judgment}}_{Li}|\text{ERAP}_{Li}] - \mathbb{E}[\hat{\text{Judgment}}_{Li}|\text{NoERAPExist}_{Li} \text{ and ERAP}_{Li}]\right) = 0. \tag{A.14}$$

   Here TOT is the Treatment Effect on the Treated from Section 1.6.2, the first expectation is the mean judgment rate among landlords whom we simulate would take up, and the second expectation is the mean judgment rate if ERAP did not exist, among landlords whom we simulate would take up. We use the treatment effects on judgments under the nonparametric specification (Appendix A.5) and bootstrap the treatment effects.

2. The mean jugment rates among landlords who take up ERAP.

$$p - \mathbb{E}[\hat{\text{Judgment}}_{Li}|\text{ERAP}_{Li}] = 0, \tag{A.15}$$

   where p comes from the ERAP data, and the second term come sfrom the experiment. We do not bootstrap the mean judgment rates.

[3]There is slight abuse of notation when we write moments below: n_i^* and its predicted value correspond to the (censored) fraction offered to repay, not the continuous bargaining offer.

3. The take-up rate among landlords, which we calibrate using data from the landlord survey.
$$\mathbb{E}\left[\text{ERAP}_{Li} - \hat{\text{ERAP}}_{Li}\right] = 0. \tag{A.16}$$

4. The bargaining offer made by tenants in the payment plans: outcome:
$$\mathbb{E}\left[n^*_{Li}(d_i) - \hat{N}^*_{Li}(d_i)\right] = 0. \tag{A.17}$$

5. Simulated moment conditions $E\left[X_i\left(Y_i - \hat{Y}_i\left(X_i, \varepsilon; \theta\right)\right)\right]$ where Y_i is an outcome and $\hat{Y}_i\left(X_i, \varepsilon; \theta\right)$ is a predicted outcome:

$$\mathbb{E}\left[a_{Li}\left(\text{ERAP}_{Li} - \hat{\text{ERAP}}_{Li}\right)\right] = 0 \tag{A.18}$$
$$\mathbb{E}\left[p_{Li}\left(\text{ERAP}_{Li} - \hat{\text{ERAP}}_{Li}\right)\right] = 0 \tag{A.19}$$
$$\mathbb{E}\left[a_{Li}p_{Li}\left(\text{ERAP}_{Li} - \hat{\text{ERAP}}_{Li}\right)\right] = 0 \tag{A.20}$$
$$\mathbb{E}\left[\text{Judgment}_{Li} - \hat{\text{Judgment}}_i\right] = 0 \tag{A.21}$$
$$\mathbb{E}\left[a_{Li}\left(\text{Judgment}_{Li} - \hat{\text{Judgment}}_i\right)\right] = 0 \tag{A.22}$$
$$\mathbb{E}\left[p_{Li}\left(\text{Judgment}_{Li} - \hat{\text{Judgment}}_i\right)\right] = 0 \tag{A.23}$$
$$\mathbb{E}\left[a_{Li}p_{Li}\left(\text{Judgment}_{Li} - \hat{\text{Judgment}}_i\right)\right] = 0 \tag{A.24}$$
$$\mathbb{E}\left[a_{Ti}\left(n^*_{Li}\left(d_i\right) - \hat{N}^*_i\left(d_i\right)\right)\right] = 0 \tag{A.25}$$
$$\mathbb{E}\left[\tilde{p}_{Ti}\left(n^*_{Li}\left(d_i\right) - \hat{N}^*_i\left(d_i\right)\right)\right] = 0 \tag{A.26}$$
$$\mathbb{E}\left[a_{Ti}\tilde{p}_{Ti}\left(n^*_{Li}\left(d_i\right) - \hat{N}^*_i\left(d_i\right)\right)\right] = 0 \tag{A.27}$$
$$\mathbb{E}\left[d_i\left(n^*_{Li}\left(d_i\right) - \hat{N}^*_i\left(d_i\right)\right)\right] = 0 \tag{A.28}$$
$$\mathbb{E}\left[\text{ERAP}_{Li}\left(\text{Judgment}_{Li} - \hat{\text{Judgment}}_i\right)\right] = 0 \tag{A.29}$$
$$\mathbb{E}\left[a_{Li}\text{ERAP}_{Li}\left(\text{Judgment}_{Li} - \hat{\text{Judgment}}_i\right)\right] = 0 \tag{A.30}$$
$$\mathbb{E}\left[p_{Li}\text{ERAP}_{Li}\left(\text{Judgment}_{Li} - \hat{\text{Judgment}}_i\right)\right] = 0 \tag{A.31}$$
$$\mathbb{E}\left[a_{Li}p_{Li}\text{ERAP}_{Li}\left(\text{Judgment}_{Li} - \hat{\text{Judgment}}_i\right)\right] = 0 \tag{A.32}$$
$$\mathbb{E}\left[\text{Judgment}_{Ti} - \hat{\text{Judgment}}_i\right] = 0 \tag{A.33}$$
$$\mathbb{E}\left[a_{Ti}\left(\text{Judgment}_{Ti} - \hat{\text{Judgment}}_i\right)\right] = 0 \tag{A.34}$$
$$\mathbb{E}\left[\tilde{p}_{Ti}\left(\text{Judgment}_{Ti} - \hat{\text{Judgment}}_i\right)\right] = 0 \tag{A.35}$$
$$\mathbb{E}\left[a_{Ti}\tilde{p}_{Ti}\left(\text{Judgment}_{Ti} - \hat{\text{Judgment}}_i\right)\right] = 0 \tag{A.36}$$
$$\mathbb{E}\left[d_i\left(\text{Judgment}_{Ti} - \hat{\text{Judgment}}_{Li}\right)\right] = 0. \tag{A.37}$$

We choose these moment conditions for the following reasons. We randomly match landlord and tenant pairs. As a result, we only include in X_i the variables that we observe in the experiment that provides the outcome Y_i. For instance, consider the take-up moments. We only include landlord values in X_i because the take-up moments come from the landlord take-up choice. As tenants are matched randomly, we should not use values for tenant variables in X_i. We also include the interactions of altruism and beliefs because Section 1.2 suggests this interaction is important.

For the judgment moments, notice that we have different observed judgments based on whether the landlord or the tenant receives a judgment in the court data. We use the landlord-side demographics with the landlord-side observed judgment, and vice-versa for tenants.[4]

We weight the "micro" moments by the inverse of their variance. In order to precisely match the "macro" moments (Equations A.14 and A.15), we weight the treatment effect by 10 and the mean by 2. We use these heuristic weights because the treatment effect is naturally estimated less precisely than the means from the experimental data, but it is valuable to match precisely. Results are highly similar if we weight these moments by the identity matrix (Table A.27, Row 13).

Welfare Impact Calculation. The welfare impact is defined as follows:

$$W_i = \begin{cases} (1+a_{Ti})\left(\varepsilon_{i3}-\varepsilon_{i1}-k_L^P\right) \text{ if evicts with ERAP, evicts without ERAP} \\ (1+a_{Ti})\left(\varepsilon_{i4}-\varepsilon_{i2}-k_L^P\right) \text{ if bargains w/ ERAP, bargains w/o ERAP} \\ (1+a_{Ti})\left(\varepsilon_{i4}-\varepsilon_{i1}-k_L^P\right)+(1+a_{Ti})k_L+(1+a_{Li})k_T \text{ if bargains w/ ERAP, evicts w/o ERAP} \\ (1+a_{Ti})\left(\varepsilon_{i3}-\varepsilon_{i2}-k_L^P\right)-(1+a_{Ti})k_L-(1+a_{Li})k_T \text{ if evicts w/ ERAP, bargains w/o ERAP.} \end{cases} \tag{A.38}$$

The version that does not normatively respect altruism is identical but sets the altruism parameters equal to zero. Notice that the welfare impact excludes any intrahousehold transfer or the direct transfer (which comes at a fiscal cost of $1 for each dollar transferred). We do not scale the fiscal transfer by tenant altruism in either calculation.

A.2.1.3 Extension: Two-Period Model, Concave Utility, Nash Bargaining

We extend the model to add dynamic elements as in Dávila (2020) (in the setting of dischargeable bankruptcy), as well as more flexible utility and Nash Bargaining.

In period 0, the tenant decides whether or how much rent to pay of the balance d_0. In period 1, she either bargains over the remaining balance or is evicted. There is a random endowment shock in period 1 that is unknown to the tenant in period 0. Eviction takes place if and only if it is more desirable than bargaining for both landlords and tenants. The fundamental tradeoff is between consumption-smoothing motives and the fact that choosing to pay little rent today can make eviction or bargaining more likely tomorrow, depending on the realization of the state.

[4] Note that Equations (A.21) and (A.33) will have the solver choose to interpolate between the two intercepts, as we have two values in the data for whether a match gets a judgment.

Tenants have well-behaved, strictly increasing and concave utility functions $u(\cdot)$. Landlords are risk-neutral. Tenants and landlords have altruism parameters $a_T, a_L \in (-1, 1)$ which scale payoffs and have the same interpretation as in the main text. We suppress the relationship-specific heterogeneity i.

Tenants' period-0 flow utility is:

$$u(c_0) := u(n_0 - d(1 - a_T)), \tag{A.39}$$

where c_0 is total consumption, n_0 is an exogenous period-0 endowment, and $d \in [0, d_0]$ is the amount of rent that the tenant chooses to pay of the total balance owed d_0. Notice that we put altruism as scaling consumption *within* the utility function.

In period 1, tenants get an exogenous and random period-1 endowment $s \sim F$, where the endowment s is weakly increasing in the state s and F has bounded or unbounded support $[\underline{s}, \bar{s}]$. If $s \geq d_0 - d$, tenants can then pay the full remaining balance and consume

$$c_1 = s - (d_0 - d)(1 - a_T). \tag{A.40}$$

Tenants are either evicted or bargain in period 1. Eviction or bargaining takes place over the residual amount that the tenant has not paid, $d_0 - d$.

In an eviction (i.e., an eviction judgment), tenants consume $j_T(x)$ and landlords consume $j_L(x)$, where we often explicitly notate these payoffs as depending on $x \in [0, d_0 - d]$:

$$u(s - j_T(x)) := u(s - \underbrace{(p_T x(1 - a_T) + k_T + (1 - a_T)k_L)}_{:= j_T(x)}) \tag{A.41}$$

$$j_L(x) := p_L x(1 - a_L) - k_L - (1 - a_L)k_T, \tag{A.42}$$

where p_T and p_L denote tenant and landlord beliefs as in the main text.

If landlords and tenants bargain, tenants pay landlords the asymmetric Nash solution $b^*(x)$. The solution arises from maximizing the following problem:

$$b^*(x) := \underset{b(x)}{\arg\max} \Big(u(s - (1 - a_T)b(x)) - u(s - j_T(x)) \Big)^{\beta} \Big(b(x)(1 - a_L) - j_L(x) \Big)^{1-\beta}, \tag{A.43}$$

for tenant bargaining power β. Taking the first-order condition and rearranging, the solution to this problem is implicitly defined by:

$$b^*(x) = \frac{j_L}{1 - a_L} + \left(\frac{1-\beta}{\beta}\right)\left(\frac{u(s - (1 - a_T)b^*(x)) - u(s - j_T)}{u'(s - (1 - a_T)b^*(x))}\right), \tag{A.44}$$

which nests the solution in Equation (A.5) if tenant preferences are linear.

Bargaining is therefore possible if and only if:

$$(1 - a_L)b^*(x) \geq j_L(x) \tag{A.45}$$

$$u(s - b^*(x)(1 - a_T)) \geq u(s - j_T(x)), \tag{A.46}$$

which implies that bargaining occurs if and only if:

$$\frac{\Delta p x}{\bar{a}} \leq k_L + k_T, \tag{A.47}$$

where $\Delta p := p_L - p_T$ just as in the main text. This is exactly the same condition as the main text, therefore recovering Proposition 1.3. We write $\mathcal{B}(x) = 1$ if $\Delta p x / \bar{a} < k_L + k_T$ and $\mathcal{B}(x) = 0$ otherwise.

For discount factor δ, the tenant's ex-ante utility function is:

$$U(d) = u(n_0 - d(1-a_T)) + \delta \Bigg(\int_{\underline{s}}^{d_0 - d} u(s - b^*(s)(1-a_T))\mathcal{B}(s) dF + \int_{\underline{s}}^{d_0-d} u(s - j_T(s))(1-\mathcal{B}(s)) dF + \int_{d_0-d}^{\bar{s}} u(s - b^*(d_0-d)(1-a_T))\mathcal{B}(d_0-d) dF + \int_{d_0-d}^{\bar{s}} u(s - j_T(d_0-d)))(1-\mathcal{B}(d_0-d)) dF \Bigg). \tag{A.48}$$

Solution. We now characterize the tenant's solution. Define $\hat{d} = d_0 - \frac{\bar{a}}{\Delta p}(k_L + k_T)$. Observe that a monotonicity condition holds, in that

$$\mathcal{B}(x) = 1 \implies \mathcal{B}(y) = 1 \text{ for all } y < x. \tag{A.49}$$

Thus, $\mathcal{B}(d_0 - d) = 1 \implies \mathcal{B}(s) = 1$ for all $s < d_0 - d$. This allows us to simplify the problem into only a few cases:

Case 1: The optimal $d^* \geq \hat{d}$. In this case, the tenant always bargains, no matter the draw of the state. Then, she solves the problem of maximizing the function:

$$U(d) = u(n_0 - d(1-a_T)) + \delta \left(\int_{\underline{s}}^{d_0-d} u(s - b^*(s)(1-a_T)) dF + \int_{d_0-d}^{\bar{s}} u(s - b^*(d_0-d)(1-a_T)) dF \right). \tag{A.50}$$

The solution d_1, if it exists, is implicitly characterized by the Euler Equation:

$$u'(n_0 - d_1(1-a_T)) = \delta \int_{d_0-d_1}^{\bar{s}} b'^*(d_0-d_1) u'(s - b^*(d_0-d_1)(1-a_T)) dF, \tag{A.51}$$

noting that the $(1-a_T)$ terms cancel from both sides.

Case 2: The optimal $d^* < \hat{d}$. In this case, the tenant gets a judgment for a sufficiently high draw of the state. Then, we use the condition that

$$\mathcal{B}(s) = 1 \iff s \leq \frac{\bar{a}}{\Delta p}(k_L + k_T). \tag{A.52}$$

Therefore, the tenant solves the problem of maximizing the function:

$$U(d) = u(n_0 - d(1-a_T)) + \delta\Bigg(\int_{\underline{s}}^{\frac{\bar{a}}{\Delta p}(k_L+k_T)} u(s - b^*(s)(1-a_T))dF + \int_{\frac{\bar{a}}{\Delta p}(k_L+k_T)}^{d_0-d} u(s - j_T(s))dF + \int_{d_0-d}^{\bar{s}} u(s - j_T(d_0-d))dF \Bigg). \tag{A.53}$$

This solution d_2, if it is exists, is implicitly characterized by the Euler Equation:

$$u'(n_0 - d_2(1-a_T)) = p_T\delta \int_{d_0-d_2}^{\bar{s}} u'(s - j_T(d_0-d_2))dF, \tag{A.54}$$

noting that $j'_T = p_T(1-a_T)$.

There are also potential corner solutions at $d = 0, d = d'$ or $d = d_0$. To solve the problem, the tenant checks which of $d^* \in \{0, d_0, d', d_1, d_2\}$ maximizes her utility $U(d)$.

Discussion. The tenant has several choices. First, she could pay all her rent ($d = d_0$). Second, she could pay less than all her rent but enough to guarantee she will not be evicted ($d = d_1$ or $d = d'$). Third, she could pay less than all her rent but pay enough that she has a chance of not being evicted if she does not have enough money in the next period that eviction is unprofitable for the landlord ($d = d_2$). Last, she could pay nothing ($d = 0$).

The main set-up in the text is nested in the above problem when the tenant chooses $d^* = 0$ and strictly prefers this choice to an interior d^*. Thus, we recover the model in the main text if $\delta \to 0$ or $n_0 \to -\infty$, as then the tenant consumes her full endowment in period 0 and whether she bargains is based only on the realization of the state s.

Notice that concave tenant utility does not change the fundamental bargaining solution in Equation (A.47). In fact, permitting concave landlord utility over both court costs and bargaining would also yield the same solution. Of course, the equation is not completely generic, as putting court costs *outside* the concave utility function can change it, but at a minimum the parametric expression for this threshold is not an artifact of linearity alone.

A.2.1.4 Proofs

Proof of Proposition 1.3. We prove Proposition 1.3, from which Propositions 1.1 and 1.2 follow as special cases in which $a_{Li} = a_{Ti} = 0$ and $\Delta p_i = 0$. For an offer to be made and accepted, it must satisfy both the tenant and landlord's participation constraints:

$$-(1-a_{Ti})o_i \geq -p_{Ti}d_i(1-a_{Ti}) - k_{Ti} - k_{Li}a_{Ti} \qquad \text{(Tenant constraint)}$$
$$(1-a_{Li})o_i \geq p_{Li}d_i(1-a_{Li}) - k_{Li} - k_{Ti}a_{Li}. \qquad \text{(Landlord constraint)}$$

Such an offer o_i exists if and only if:

$$p_{Li}d_i - \frac{k_{Li} - k_{Ti}a_{Li}}{1-a_{Li}} \leq p_{Ti}d_i + \frac{k_{Ti} + k_{Li}a_{Ti}}{1-a_{Ti}}. \tag{A.55}$$

$$\iff \frac{k_{Li} + k_{Ti}a_{Li}}{1 - a_{Li}} + \frac{k_{Ti} + k_{Li}a_{Ti}}{1 - a_{Ti}} \geq \Delta p_i d_i \tag{A.56}$$

$$\iff (k_{Li} + k_{Ti}a_{Li})(1 - a_{Ti}) + (k_{Ti} + k_{Li}a_{Ti})(1 - a_{Li}) \geq \Delta p_i d_i (1 - a_{Ti})(1 - a_{Li}) \tag{A.57}$$

$$\iff (k_{Li} + k_{Ti})(1 - a_{Li}a_{Ti}) \geq \Delta p_i d_i (1 - a_{Ti})(1 - a_{Li}) \tag{A.58}$$

$$\iff k_{Li} + k_{Ti} \geq \frac{\Delta p_i d_i}{\bar{a}_i} \tag{A.59}$$

meaning that eviction occurs if and only if

$$k_{Li} + k_{Ti} < \frac{\Delta p_i d_i}{\bar{a}_i}, \tag{A.60}$$

as desired.

Proof of Proposition 1.4. The landlord's ERAP participation constraint is

$$d_i - k_{Li}^P \geq \max\left\{(1 - a_{Li})o_i, V_{Li}^E\right\} \tag{A.61}$$

where o_i is the settlement offer they expect to receive from the tenant, which will leave them indifferent between informal settlement and formal eviction, and V_{Li}^E is the payoff from eviction. This is

$$d_i - k_{Li}^P \geq (1 - a_{Li})\left[p_{Li}d_i - \frac{k_{Li} + a_{Li}k_{Ti}}{1 - a_{Li}}\right]. \tag{A.62}$$

Dividing through by $(1 - a_{Li})$ yields

$$p_{Li}d_i \leq \frac{k_{Li} + a_{Li}k_{Ti}}{1 - a_{Li}} + \frac{1}{1 - a_{Li}}\left[d_i - k_{Li}^P\right]. \tag{A.63}$$

We then subtract $p_{Ti}d_i$ from both sides:

$$(p_{Li} - p_{Ti})d_i \leq \frac{k_{Li} + a_{Li}k_{Ti}}{1 - a_{Li}} + \frac{1}{1 - a_{Li}}\left[(1 - p_{Ti}(1 - a_{Li}))d_i - k_{Li}^P\right]. \tag{A.64}$$

Then simplify terms, and also add and subtract the tenant-side cost expressions from the right-hand side:

$$\Delta p_i d_i \leq \bar{a}_i(k_{Li} + k_{Ti}) - \frac{k_{Ti} + a_{Ti}k_{Li}}{1 - a_{Ti}} + \frac{1}{1 - a_{Li}}\left[(1 - p_{Ti}(1 - a_{Li}))d_i - k_{Li}^P\right]. \tag{A.65}$$

Rearranging, we obtain:

$$\Delta p_i d_i \leq \bar{a}_i(k_{Li} + k_{Ti}) + \frac{1}{1 - a_{Li}}\left[(1 - p_{Ti}(1 - a_{Li}))d_i - \frac{1 - a_{Li}}{1 - a_{Ti}}(k_{Ti} + a_{Ti}k_{Li}) - k_{Li}^P\right]. \tag{A.66}$$

Divide through by $\bar{a}_i$ to get that

$$\frac{\Delta p_i d_i}{\bar{a}_i} \leq k_{Li} + k_{Ti} + \frac{1}{\bar{a}_i(1-a_{Li})}\left[(1 - p_{Ti}(1-a_{Li}))d_i - \frac{1-a_{Li}}{(1-a_{Ti})}\left(k_{Ti} + a_{Ti}k_{Li}\right) - k_{Li}^P\right]. \tag{A.67}$$

Then the landlord takes up if and only if:

$$k_{Li} + k_{Ti} + w_i \geq \frac{\Delta p_i d_i}{\bar{a}_i} \tag{A.68}$$

where

$$w_i = \frac{1}{\bar{a}_i}\Bigg(\underbrace{\frac{d_i - k_{Li}^P}{1-a_{Li}}}_{\substack{\text{Altruism-adjusted}\\ \text{net ERAP payment}}} - \underbrace{\left(p_{Ti}d_i + \frac{k_{Ti} + a_{Ti}k_{Li}}{1-a_{Ti}}\right)}_{\text{Outside option adjustment}}\Bigg). \tag{A.69}$$

Proof of Equations (A.5) and (A.6). Suppress dependence on i. Given altruism, Nash bargaining solves:

$$n^* = \arg\max_n \left(-\left(1-a_T\right)n - \mu_T\right)^\beta \left(\left(1-a_L\right)n - \mu_L\right)^{(1-\beta)} \tag{A.70}$$

where μ_L and μ_T represent the parties' outside options inclusive of their own altruism. Then taking logs, the first-order condition is:

$$-\frac{\beta(1-a_T)}{-(1-a_T)n^* - \mu_T} + \frac{(1-\beta)(1-a_L)}{\left(1-a_L\right)n^* - \mu_L} = 0. \tag{A.71}$$

From here, rearrangement gives:

$$n^* = \frac{\beta\mu_L}{1-a_L} - \frac{\mu_T(1-\beta)}{1-a_T}. \tag{A.72}$$

To recover the equations, note that if the landlord does not take up,

$$\mu_L \equiv p_L(1-a_L)d_i - k_L + k_{Ti}a_{Li} + \varepsilon_{i2} - \varepsilon_{i1}, \tag{A.73}$$

and

$$\mu_T \equiv -p_T(1-a_T)d_i - \left(k_{Ti} + (k_L + \varepsilon_{i2} - \varepsilon_{i1})a_{Ti}\right), \tag{A.74}$$

and make a similar substitution if the landlord does take up.

A.3 Experiment Details

A.3.1 Recruitment Details

Landlord Survey. We limit the sample to landlords with valid own and tenant contact information. We asked each landlord participant questions about her tenant, whose name we link from the tenant application, selecting a reference tenant at random when a landlord is linked to multiple tenant applicants.

Tenant Survey. Because tenants may have moved between applying for ERAP and taking the survey, we ask tenants for information about their current landlord, which we use in the survey elicitations involving their landlord. We did not limit tenants to people who were paid. We also require valid own and landlord contact information. We include any tenants who applied or started applying before February 13, 2022.

Both Surveys. We contacted experiment participants with a Memphis/Shelby County ERAP email address and logo, conferring legitimacy to our outreach, in addition to an MIT logo and disclosure of our institutional affiliation.

A.3.2 Design Details

Multiple Price List for Dictator Game. We elicit indifference points S using a multiple price list-style elicitation and the strategy method. For each participant, we ask whether she prefers Bundle $A = (\$0.9x, \$0)$ to $B = (\$x, \$x)$. If she prefers B to A, we present another Bundle $A' = (\$1.5x, \$0)$ and ask her preferences over A' versus B. We vary the value of Bundle A to be as large as $(\$2x, \$0)$. We repeat a version of this elicitation until we obtain her switching point. We adopt this method because it is easier for subjects to understand than asking about their indifference point directly. We assume that if a participant prefers A to B for a given s, she will also prefer A to B for $s' > s$; our multiple price list elicitation uses a binary search-type technique to ask about progressively narrower choices between Bundles A and B.[5]

Multiple Survey Completions. Several individuals take the experiment multiple times, as indicated by the name they report in the experiment (and for tenants, other information such as phone number or address). We drop the second instance. There is no incentive to lie about one's name on the survey since participants were paid even if they took the survey twice.

Information About Own Tenant Or Own Landlord. For the landlord survey, we use information about the tenant applicant associated to ERAP to automatically populate throughout the survey. In the tenant survey, their current landlord may differ from the landlord they applied to ERAP with. For instance, they may have applied in April 2021

[5]Because we only allow A to take whole-dollar values, we obtain upper and lower bounds on S_ℓ, and these bounds have a width of $\$0.1x$. Where appropriate, we assume that the indifference point lies halfway between the bounds. For instance, if a landlord participant prefers B to A if $s = 12$ but A to B if $s = 13$, we assume $S_\ell = 12.5$.

and be taking the survey in May 2022, so they may have moved in the interim. Early in the survey, we ask the tenant for their landlord's name and populate the response in the subsequent elicitations.

When landlords play against random, unnamed tenants, we indicate that the tenant recipient be a tenant of another landlord participant, and similarly for tenants playing against random, unnamed landlords.

A.3.3 Text for Belief Elicitations

A.3.3.1 Landlords

The information provision was:

> *Of all monetary eviction judgments rendered in Shelby County Courts in January 2020, about 6 out of 100 cases had fully repaid their balances by the beginning of August 2021.*

To elicit beliefs about the average, we ask landlords:

> *Consider monetary evictions judgments given in January 2020 in Shelby County courts. January 2020 was before the coronavirus pandemic in the U.S.*
>
> *Out of every 100 monetary judgments in January 2020, how many tenants had fully repaid the balances they owed by the beginning of August 2021?*

To elicit beliefs about their own tenant, we ask landlords:

> *Imagine the courts gave you a monetary eviction judgment for [TenantName] today.*
>
> *We are asking you to make a prediction about what would happen in this scenario.*
>
> *What do you think is the percent chance that [TenantName] would repay the judgment to you, in full, by May 2023?*

We elicit prior beliefs after conducting the Dictator Game. We reject meaningful priming or order effects: we regress prior beliefs about the landlord's own and average tenant repayment on whether landlords play the DG against their own or a random tenant and find no association ($p = 0.91$ for own and $p = 0.93$ for average).

A.3.4 Incentives

A.3.4.1 Landlord survey

Fixed Payments. All payments were made in the form of Amazon gift cards send to participants' emails. Survey participants who complete the survey were paid $20. One participant was randomized to win a bonus of $500, which we advertised to increase participation.

Beliefs. We incentivized two belief elicitations:

1. Prior beliefs about recoupment probability of average tenants. We paid according to the quasi-quadratic function:

$$\text{belief bonus} = \max(0, 22 - 22 \times (\text{truth} - \text{response})/44)^2) + 3, \tag{A.75}$$

rounded to the nearest dollar, where LSC data indicate the truth was 6.

2. Prior beliefs about court delays. We paid according to the quasi-quadratic function:

$$\text{belief bonus} = \max(0, 22 - 22 \times (\text{truth} - \text{response})/3500)^2) + 3, \tag{A.76}$$

rounded to the nearest dollar, where LSC data indicate the truth was 54 cases.

We randomize 20% of participants to be paid for one belief. We choose which belief at random with probability 0.5.

We informed participants that they would maximize their payment if they reported beliefs that were closer to the truth, and that there was no incentive to distort their beliefs. Participants had the option to observe the formulas but did not see it directly unless they selected that they want to see them.

Dictator Game. We implement an incentivized multiple price list as follows. 5 landlords were chosen to have the Becker-Degroot-Marschak (BDM) mechanism implemented for their game when they played against a tenant. Whether the tenant opponent was the landlord's own tenant or random of another landlord who completed the study was randomized separately and displayed to the landlord. 5 landlords were chosen to have the BDM mechanism implemented for their game when they played against a landlord. We imposed monotonicity and used a multiple price list to elicit the participant's switching point between ($10, $10) and ($x, $0). The maximum value of $x displayed was $20. The minimum value was $1. Participants could only record preferences in $1 increments. Only one question was displayed at once. We then conducted a random draw across the choices {($10, $10) versus ($1, $0); ($10, $10) versus ($2, $0); . . . , ($10, $10) versus ($20, $0)} with equal probability and implement the landlord's (implied) choice for that bundle. This mechanism preserves incentive compatability. We then send gift cards based on the choices to either the tenant or a random landlord in the survey.

Lottery Elicitation. In Section A.3.5, we describe a lottery elicitation. Landlords can choose to enroll another person in a lottery for $10. If the landlord chooses to enroll their tenant in the lottery, we draw with probability 0.01 whether their tenant receives the gift card. If the landlord chooses to enroll a random other landlord in the lottery, we draw with probability 0.01 whether the random landlord receives the gift card.

A.3.4.2 Tenant survey

Payments were similar to the landlord survey.

Overview and Fixed Payments. Tenants were paid $20 for completing the study. They choose either a Starbucks gift card or an Amazon gift card for all payments.

Beliefs. Belief payments were implemented with the same probabilities as the landlord survey and using a similar quasi-quadratic formula. The beliefs we incentivized were:

1. Prior beliefs about the percent chance of landlord settlement:

$$\text{belief bonus} = \max(0, 22 - 22 \times (\text{truth} - \text{response})/44)^2) + 3, \tag{A.77}$$

rounded to the nearest dollar, where the truth was 31.

2. Prior beliefs about the percent chance of landlords' being highly altruistic:

$$\text{belief bonus} = \max(0, 22 - 22 \times (\text{truth} - \text{response})/44)^2) + 3, \tag{A.78}$$

rounded to the nearest dollar, where the truth was 69.

Dictator Game. We employ the same structure as with landlords. The main departure is that some tenants were randomized into larger stakes. As with landlords, we obtain their switching point using a multiple price list and employ a BDM mechanism with equal probability across 20 possible questions.

Lottery Elicitation. Payments were exactly analogous to the landlord survey, but tenants could enroll their own landlord or a random landlord.

A.3.4.3 Additional Details about Outcomes

Tenants. We did not send payment plans to landlords whose tenants indicated that they did not know their landlords' email address. About 20 percent of the payment plan messages bounced.

A.3.5 Validation of Preference Measures

We included several additional tests to validate that behavior in Dictator Games reflects true preferences.

Attention and Confusion. We consider the following measures of tenant affect toward their own landlord:

1. *Free-Response Sentiment.* Before the DG, we ask tenants to share open-ended reflections about their landlord. We analyzed these free responses more systematically using VADER, a sentiment classifier from the natural-language processing literature (Hutto and Gilbert, 2014).[6] The classifier gives an aggregate sentiment score for each response.

2. *Likert Scale.* We asked tenants to subjectively rate their relationship with their landlord.

[6] VADER has been pre-trained on social media data and is designed to handle standard challenges with sentiment analysis like negation and slang.

3. *Simple Lottery.* We conducted a simple real-stakes lottery elicitation as a secondary measure of altruism. We ask the landlord whether they wish to enroll a random other landlord and their own tenant in a lottery for \$20 with probability 0.01. In the tenant experiment, we likewise ask whether to enroll her own landlord and a random tenant in the lottery. We tell participants that their choice will be kept private. We also tell participants that draws are made at random, so that enrolling others does not influence their own chances of winning or that of other entrants. Finally, we stress that choices will be anonymous.

   The idea behind this task is that it is free for the participant to enroll both the landlord and tenant in the lottery. It also shuts down instrumental preferences. They will do so if they have any altruism toward the other person.

   24% of tenants decline to enroll their landlord in the lottery. 18% of landlords decline to enroll their tenant. These rates are similar to the observed hostility in the DG.

   The benefit of this outcome is that it is a real-money choice that is perhaps simpler to understand than the DG.

Following Kling et al. (2007b), we combine these measures via an index that is the average of the standardized measures. Panel A of Table A.14 shows the index is highly correlated with tenant behavior in the DG toward their own landlord. A one-standard-deviation increase in the index is correlated with a 16-percentage-point decrease in hostility toward own landlords (column 1, $p = 0.00$). The correlation is similar with behavior towards random landlords (column 2) and markedly attenuated for behavior towards random tenants (column 3). The correlations are symmetric for high altruism (columns 3–6).

It is reassuring that this index of affect toward own landlords are most correlated with behavior toward own landlords and random landlords, but less so with random tenants. It implies that attitudes toward own landlords manifest most strongly in behaviors toward own landlords in the DG.

Panels B, C, and D of Table A.14 respectively disaggregate the index into the free response, Likert scale, and simple lottery.

The lottery results are especially informative because these are additional revealed outcomes, where we stressed anonymity, and which were simple to understand. Enrolling one's own landlord in the simple lottery is correlated with a 21.8 pp reduction in hostility (more than 50% of the control mean). We also implemented the simple lottery with landlords to rule out their inattention or confusion and find symmetric results as with tenants (Table A.15).

We further explore the lottery results and draw three conclusions (Figure A.14). First, the rates of landlords and tenants who reject enrolling their own tenant or landlord in the lottery are reassuringly similar to the levels of hostility in the DGs. For instance, 24% of tenants reject entering their own landlord in a lottery to win money. Second, landlords' rates of rejecting enrolling their own tenant exceed the rates at which they enroll a random landlord. Meanwhile, tenants' rates of rejecting enrolling their own landlord exceed the rates at which they reject enrolling their own tenant. These differences by opponent are similar to the DG. Third, these behaviors are highly correlated with DG behaviors.

That is, hostile participants are more likely not to enroll their own tenant or landlord. The increase in rejection among hostile participants is more concentrated against the participant's own landlord or tenant than against a random opponent.

Failing to enroll one's own landlord or tenant in the lottery is further evidence of hostility. It is costless to enroll one's landlord or tenant in the lottery. Participants with no social preferences would be indifferent between enrolling versus not enrolling. Yet landlords and tenants are not indifferent on average, as they exhibit different behaviors when enrolling tenants versus landlords. If they were indifferent, the opponent would not matter. Such behavior is more consistent with hostility that is directed toward particular opponents.

Global Preferences Survey. Another concern is that behavior in the DG will not extrapolate outside the lab. To rule this out, we show that behavior in the DG is correlated with externally validated survey measures of behavioral primitives. We use questions from the Global Preferences Survey (Falk et al., 2018, Forthcoming) to measure altruism, positive or negative reciprocity, and preferences for punishment. Figure A.13 shows that negative reciprocity is correlated with a lower tenant point toward random landlords but not toward random tenants. Preferences for punishment are also correlated with a lower indifference point.

A.3.5.1 Anonymity and Instrumental Motives

An important concern is that behavior in the DG may reflect instrumental motives (e.g., reputation). On the one hand, if landlords or tenants burn surplus because of a reputation game, that is perhaps still interpretable as inefficient. On the other hand, it is not clear that such behavior would reflect a social preference.

We try to set up the DG so that parties would be anonymous. If the party is anonymous, there is no instrumental reason to make a choice in the DG. For both parties, we stress anonymity in the consent process, in the context of data storage and data sharing. For tenants, we stress anonymity with respect to landlords at the time of the game, and added an additional confirmation check.[7] 88% get the question right. For the 12% who get it wrong, we correct the answer. For the remaining 88%, we reiterate anonymity. We did not include these extra checks or reminders about anonymity for landlords.

As further evidence that these concerns are not decisive, the above lottery measure of preferences explicitly emphasizes anonymity in both surveys. We write: "Your response will be kept private from your tenant" or "Your response will be kept private from your landlord." The shares of landlords who do not enroll their own tenant, and tenants who do not enroll their own landlord, are very similar to the observed hostility. Moreover, behaviors in the DG are highly correlated with behaviors in the lottery (see discussion of Table A.14, Table A.15, and Figure A.14 above). Even if one is skeptical of the DG evidence, the lottery data provide useful complementary evidence that cannot be explained with perceptions about anonymity.

[7] We updated the anonymity language partway through the tenant study. See discussion in Appendix A.3.8. Notably, comparing tenant responses before and after we changed anonymity language suggests behavior is highly similar, which further implies that perceptions that responses would be identified do not drive results.

We believe that the most relevant concern is that a share of landlords may express *altruism* for instrumental reasons. Recall that hostility simply means that the other party does not get a gift card. It does not take money away from them. Indeed, if the landlord wants to claim money from the tenant as rent, burning money means they cannot claim it. If true, this would mean that our main results about hostility are lower bounds.

A.3.6 Attention

General Attention Checks. In the landlord experiment, we included one general attention check: we ask participants what is their favorite color, but tell them to report that their favorite color is teal. We pre-registered in AEARCTR-0008053 that this attention check would not be dispositive because it could be too hard. 64% of landlords pass this check. In the tenant experiment, we also used a second attention check where we ask participants to report that their favorite number is 6. 92% of tenants pass the "6" attention check but 34% of tenants pass the "teal" check, giving further evidence that the "teal" check was too hard. Our primary findings do not condition on passing the general checks, following our pre-registrations. Because of well-known concerns about attention among Lucid participants, we do condition on the most stringent attention checks in the Memphis and National Samples.

We include numerous specific confirmation checks described in Section 1.4 that increase confidence that our participants are attentive for our key elicitations.

Specific Attention and Confirmation Checks.

- *Modified Dictator Game.* We randomize the order in which we elicit preferences toward tenants versus landlords. We randomize the order in which we ask about indifference in the multiple price lists (i.e., switching the order of Bundle A versus Bundle B). Because we randomize the order independently across elicitations, these randomizations reduce the likelihood that tenants simply click one button (to the left or right) in order to advance in the survey. To give further confidence that our results reflect actual preferences, we ask participants who report being very altruistic to explain why; we provide examples of their qualitative responses in the results. For tenants, we also included specific confirmation checks that ask whether the information would be shared with the landlord. Of the 1,175 tenants asked this question, 88% pass.[8] Finally, the fact that the DG responses are correlated with simpler elicitations and free-text responses further increases our confidence that they are not driven by inattention.

- *Prior beliefs.* For beliefs about their own tenant or landlord, the participant must report a probability. After the elicitation, we convert the probability to an odds (which may be more intuitive), and participants must confirm the odds corresponds to the probability they have in mind. For beliefs about the average tenant or landlord, the participant must report a number out of 100 who would engage in the elicited

[8] The question was added partway through the experiment.

behavior. We provide a visualization of 100 boxes that turn red as the number changes; we also ask participants to confirm that the number they report.

- *Posterior beliefs.* Participants must first report a *direction* in which they choose to update. After seeing the information, they are asked if their prior belief is "too high" or "too low." Then they must report a quantitative posterior that aligns with the direction they report.

A.3.7 Complete list of outcomes and pre-registration

We pre-registered a list of primary and secondary outcomes (AEA RCT Registry: AEARCTR-0008053 [Landlord], AEARCTR-0008436 [Random Sample], and AEARCTR-0008975 [Tenant]). Except where otherwise stated, the primary and secondary outcomes are all reported in the paper or appendix materials. Given the length of the paper, for brevity, we do not report heterogeneity listed in the preregistrations as secondary outcomes.

A.3.7.1 Landlord outcomes

As indicated in the pre-registration, the primary and secondary outcomes were as follows:

- Behavior in DG with landlords and tenants
- Choice of whether to enroll landlords or tenants in lottery (simple dictator game, Table A.14).
- Accuracy of prior beliefs ("objective," by which we meant beliefs about the average tenant, and beliefs about the tenant).

The secondary outcomes were:

- Belief updating, as well as the belief updates about the placebo belief
- Interaction with the program, including the following five outcomes: choice of whether to receive informational materials, choice of whether to refer tenants to the program, choice of whether to decline to sign any legal agreement, choice of whether to receive new offer for back rent, and choice of whether to notify tenants. As pre-registered, we aggregate these into an index in Table A.13.
- Heterogeneity.

A.3.7.2 Tenant outcomes

Primary Tenant Outcomes.
The categories of outcomes were:

- Behavior in DG.
- Beliefs (priors, posteriors, and belief update)

- Payment plan outcomes.

The specific outcomes for the DG and beliefs are reported in the text. The bargaining task outcomes were whether the tenant proposed a payment plan and the share of back rents that they proposed to repay. The extensive margin of payment is reported in the text. The intensive margin (share of back rents proposed to repay) is in Figure A.31.

Secondary Tenant Outcomes.

- Willingness to pay for information. We conduct a task that is the willingness to pay for information about how the landlord behaves. This is incentivized using a BDM mechanism with probability 1%. We implemented the mechanism when implementing all payments. We report the treatment effect of information on the above WTP in Figure A.30.

- Additional bargaining outcomes, described below.

- Validation of the altruism task. We report these in Appendix C.4 and Table A.14. We added these on March 27 as secondary outcomes (we did not collect them prior to March 27).

- Hypothetical indifference points for willingness to move and willingness to accept money in exchange for expunging an eviction record. These outcomes were elicited as multiple price lists. The latter outcome was only elicited among tenants that reported having an eviction filing. We pre-registered these as secondary because they are hypothetical. We also pre-registered that we would not examine the treatment effect on the second outcome. We report the treatment effect on the first outcome in Figure A.32. We elicited an indifference point but just focus on the willingness to accept $1,000 versus the outcome so that the outcomes are comparable when we use them in Section 1.6.2. More details on this elicitation are in Section B.1.

A.3.7.3 Random sample surveys

Representativeness. The national sample was designed to be nationally representative. The company could only guarantee a sample of Memphians that was representative at the age $\times$ gender level.

Pre-registration. Our primary outcomes were: DG outcomes and beliefs about the eviction process (Memphis sample only). DG outcomes are reported in A.10. Beliefs about the eviction process are in Figure A.29.

A.3.7.4 Other Deviations from Pre-registration

We give a detailed list of minor deviations from pre-registration for complete transparency. None of these deviations are consequential for the paper.

- The tenant survey launched February 15, 2022. On March 27, we updated the tenant survey to include a subexperiment that varied the DG stakes. We also collected additional altruism validation outcomes, based on feedback from the first waves. We updated the pre-registration materials at the AEA website at this time. The full pre-registration history and dates are available there.

- In the tenant survey, we also made minor changes to the registration of specific DG outcomes or the information treatment before March 27. About 59% of the sample was collected on March 27 or later, and the main DG results do not change if we limit only to that sample (see Figure A.9 or Table A.1 for the disaggregation). On April 8, we changed an item pertaining to the sample size across the information treatments, which does not change what we report.

- As one of the outcomes under the DG behavior in both the landlord and tenant pre-registrations, we write that we would create "difference in indifferences" measures for the DG. For the DGs with landlords as dictators, this would correspond to: the preference for (Own tenant minus random landlord) minus (Own tenant minus random tenant). We realized after completing the study that these differences are not informative, as "own-tenant" would simply get subtracted off both sides. We therefore do not report them for brevity. These outcomes can be easily computed from the estimates reported in Table A.16 by subtracting.

A.3.8 Survey Changes

A.3.8.1 Landlords

We made two minor small changes to the landlord survey while implementing it, based on feedback from participants or the Memphis ERAP officials. First, we reworded the question in a secondary outcome about declining to sign legal agreements (a secondary outcome for landlords) from a double negative that appeared to be confusing to participants. Second, we added a qualitative question about whether the participants truly wanted to split the DG bundle evenly.

A.3.8.2 Tenants

We made some minor changes partway through the tenant survey and one major change. The major change is that we added the stakes subexperiment about 40 days through the study, when about 40% of the sample had been collected. Changes are detailed below.

Elicitation Changes Associated with Stakes Randomization. The most important change we made to the tenant survey is that, after collecting roughly 450 observations, we paused the survey to begin randomizing stakes for the DG. Doing so required us to lightly change some of the wording. We also made an update to our RCT registration at the time of adding the stakes subexperiment to detail the changes to our experiment procedures.

Before March 27, all tenants who participated in the DG played the DG where we elicited the value $S(10)$ that made them indifferent between $(10, 0)$ and $(10, 10)$. Starting March 27, tenants were randomized into $x \in \{10, 100, 1000\}$. In order to newly randomize stakes for the survey, we needed to make the following minor changes to the DG elicitation: (i) we indicate to tenants that the funding for the DGs came from a separate research budget; (ii) we indicate that "at least one" tenant will have their choice implemented in each DG implementation; (iii) we emphasize that their responses will be kept private from their landlord.

For (i), we wanted to reassure tenants randomized into the $1,000 gift card condition that the funding for this activity was not coming from money that would otherwise be used for back rent repayments.

For (ii), we had previously told tenants that five tenants would be paid for their responses about landlords and five for their responses about tenants. Thus, this change reduced the probability of payment *only for people exposed to the new stakes*, though as noted above it does not appear to have affected behaviors. To keep the elicitation truthful for people who had previously enrolled, we paid five tenants exposed to the lowest stakes. Hence, more than one tenant was indeed paid. We needed to make this change because the budget precluded us from making ten payments in the $1,000 stakes condition.

For (iii), we newly emphasized to tenants that their choice would be kept private, to reduce separate concerns about anonymity and tenants' incentives.

Because we elicit the DG at for $(10, 10)$ both before and after March 27, we test whether behavior changes for participants at the same stakes ($10) but where preferences are elicited using different wordings. Figure A.9 disaggregates pre-randomization and post-randomization (left two ticks) and finds no important difference in behavior, which justifies pooling all the data for power in our main analysis. In particular, comparing tenant behavior at the same $10 stakes before and after March 27 tests whether changes to the anonymity language affect results.[9]

Other Tenant Survey Changes. We made several other edits to the survey when fielding it:

- Based on our experience sending landlords payment plans, we added additional language when describing the payment plans, e.g. about the tenants' rights when talking with landlords.

- Based on tenants' free responses, we added additional language emphasizing to tenants that the gift cards in the DG would not be linked to them in particular and would not count as rent, and we added a confirmation question about whether the gift card is split to own versus random landlord.

A.3.9 Balance and Attrition

In each test, we include all available characteristics that we elicited before the interventions. For instance, because we conduct the DGs before the information treatments, we include the outcomes of the DGs in our sample balance tests for the information treatments. Among landlords, we find no evidence of important imbalances for the random tenant

[9]Because we had limited opportunities to pilot, we further changed anonymity language slightly between March 27 and March 30, which affected 10% of participants. Results are similar if we exclude this group or limit only to tenants who completed after March 30 (Table A.1).

treatment (Table A.2, joint $p = 0.284$). We find evidence of imbalances for the information treatment (Table A.3, joint $p = 0.01$), driven by an imbalance in the share who has ever evicted. In robustness checks, we control for observable demographic characteristics and results are quite stable. Among tenants, we find no evidence of any imbalances across the random landlord treatment or the two information treatments (Tables A.4, A.5, A.6).

A.4 Data Appendix

A.4.1 Data Linkages

Fuzzy Merges to Eviction Data. We process first names, last names, addresses, and ZIP codes from the ERAP application or experiment. We retain only households whose applications had non-missing data for each of these. We merge only on street address and not unit number. To improve the match, we drop street suffixes, as they are not entered consistently (e.g., "road" versus "rd"). We merge these onto eviction records using Stata's `reclink`, with equal weights on first name, last name, address, and ZIP. We indicate a match if the threshold exceeds 0.9. In practice, few households have merges between 0.75 and 0.9.

We record several additional details about merging to the experiment details. We had sufficient information on about 330 of the landlord–tenant pairs to conduct a fuzzy match onto court records using tenant names and addresses. As with landlords, we match tenant behavior to eviction court records using fuzzy merges on name and address. Note that because we match on tenant's current address, the tenants with eviction filings still remained housed at the address at the time of taking survey.

Merges to ERAP Reciept. Merges to ERAP data are *not* fuzzy, and instead are based on an internal Case ID created when tenants initiate an application for ERAP. As tenant experiment participants were recruited using from ERAP application data, all of them are associated to a Case ID, which means we can track how they proceed through the ERAP receipt process.

We make several sample restrictions throughout Section 1.6.3. First, we only have accurate data on ERAP receipt time for tenants who applied after September 2021. Second, the experiment measures tenant hostility for their *current landlord* at that time, which may not be the same landlord as when they applied to the program. We limit to participants who have not moved between applying for ERAP and the survey.

A.4.2 Preparing the Data for Section 1.6.1

The tenant sample restricts to those with rental debts at the time of the survey.

Throughout, we let tildes ($\tilde{\cdot}$) represent proxies that we input into the construction of $\tilde{\theta}_i$, whereas primes ($'$) represent the raw data.

Misperceptions. For landlords, we construct $\Delta\tilde{p}_i := p_i' - 6$.

Tenant beliefs do not exactly map onto Δp_i in the model, as they do not represent beliefs about eviction repayment. However, they are beliefs about bargaining costs. This section posits that beliefs about bargaining can proxy for the inverse of beliefs about eviction, so if we find that tenants have a 1 pp misperception about bargaining relative to the truth, then they have a -1 pp misperception about judgments.[10]

For tenants, we have two measures of misperceptions. For tenants, we form

$$\Delta \tilde{p}_i := \frac{1}{2}\left(\left(p'_{1i} - 0.31\right) + \left(0.69 - \tilde{p}'_{2i}\right)\right) \tag{A.79}$$

where p'_{1i} and p'_{2i} represent beliefs about bargaining and altruism respectively. Notice that higher beliefs about landlord dropping the filing is associated with less bargaining, and vice-versa for beliefs about landlord altruism. Thus, as Δp_i grows, the chances of the tenant wanting to evict rise.

Altruism. We form $\tilde{a}_i := (a'_i - 100)/100$. We only take households whose $\tilde{a}_i$ is obtained for their own landlord or tenant. To accomodate the restrictions in Section 1.2 we winsorize $\tilde{a}_i$ at 0.95 if larger than 1.

Costs. For landlords, we ask whether they would accept an offer to repay future rent at a "discount." The landlord would have to agree not to file an eviction during the three months the future rent was paid. The discount would operate as a percent reduction below the price of rent. For instance, if the landlord agrees to an x% discount, then the government pays the landlord (1-x)% of the rent for 3 months.

We elicit the landlord's indifference point. For half the landlords, chosen randomly, we ask discount prices in: $\{0, 0.1, 0.3, 0.5\}$. For the other half, we ask prices in: $\{0.05, 0.2, 0.4, 0.6\}$. We generate their indifference point as the midpoint of the value that the landlord was willing to accept and control for which set the landlords were randomized into.

Adjusted Surplus. For both landlords and tenants, we read off their adjusted surplus by plugging in the misperceptions, altruism, and cost proxies from the above.

Sample Restrictions. In specifications where a_{ji} enters, we restrict only to landlords or tenants who play the game against their own tenant/landlord. Among tenants, we keep only those who report in the survey that they have positive rental debts. We did not elicit this question in the landlord survey, so we cannot make the restriction there.

A.5 Supplementary Empirical Analysis

A.5.1 U.S. and Random Samples for Dictator Game

We conducted the same DG in November to December 2021 in samples of random residents of Memphis ($N = 282$) and random, nationally-representative Americans ($N = 632$) using the online survey company Luc.id.[11] Each participant plays the game

[10]While tenants' misperceptions about recoupment are, in principle, bounded below by -6, p_i could also contain other probabilities like the chances of winning in court.

[11]In the random samples, we limit the sample to participants who pass the attention check. Results are similar without this sample restriction. The Memphis sample is not strictly representative, because Luc.id

twice, in random order, once against a random unnamed Memphis ERAP tenant applicant and once against a random unnamed landlord of a ERAP tenant applicant. We explain to participants that ERAP is designated for low-income households with rental debt.

Cross-sample comparisons suggest that the social preferences are stronger among the ERAP sample than among broader populations. ERAP applicants are more likely to be highly altruistic, consistent with positive selection based on relationships (Figure A.10, Panel A). Pooling the random samples, landlords are 14 percentage points more likely to be highly altruistic to their own tenant than the pooled random sample is to a random tenant ($p = 0.000$). Tenants are 6 pp more likely to be highly altruistic to their own landlord than the random sample is to a random landlord ($p = 0.002$).

Study participants are also more likely to be in hostile relationships than the random samples (Panel B). Pooling the samples, tenants are 4 pp more hostile toward their own landlord than the pooled random sample is to random landlords ($p = 0.014$, Panel A) and 11 pp more hostile than the random sample is to random tenants. Landlords are 3 pp more hostile toward their own tenant than either random sample is to random tenants in the program, but results are not generally significant (Memphis sample: difference = 4.8 pp, $p = 0.108$; pooled sample: $p = 0.324$). As landlords themselves are less hostile to random tenants than the benchmark, we view the differences among landlords as more informative.

A.5.2 Assortative Matching

We have two objectives with assortative matching. First, we want to figure out the total share of landlord–tenant relationships that feature at least one hostile party. Since 15 percent of landlords or 24 percent of tenants are hostile, if hostility were perfectly negatively correlated, then 39 percent could be hostile. If pairs were randomly matched, then 35% would be hostile ($0.35 \approx 0.15 + 0.24 - 0.15 \times 0.24$). Second, we need a measure of assortative matching for Section 1.7.

We form statistics that are informative about assortative matching. We posit a data-generating process for assortative matching. Based on this DGP, we simulate the amount of assortative matching that generates these statistics.

Direct and Imputed Matching. Recall that even though most landlords have multiple tenants, we only conduct the DG for a single tenant–landlord pair. Because response rates are only about 10%, we only have 19 direct landlord-tenant links who both played the DG against each other. However, for larger property management companies, we observe many tenants linked to a given landlord. Therefore, leveraging these links, we observe 471 tenants linked to 49 unique landlords who participate (and where both play the DG against their own landlord tenant). We consider the mean of the landlord's behaviors toward her own tenant.

This measure is imperfect because it assumes that landlords who are (not) hostile to one tenant where we observe hostility are (not) hostile to all. For instance, if we observe landlord A play the DG against tenant B, and then we observe tenant C's DG behavior

could only provide a sample in this area that is representative on age and gender.

against A, we regress C's DG behavior on A's.[12]

We regress tenant behavior on the landlord's mean indifference point across observations for that landlord (Table A.8, Panel A) and hostility (Panel B) among tenants and landlords we can match. Panel A, column 1 shows that raising the landlord's observed indifference point from 100% to 200% of the outside option is associated with a 16 cent increase in the tenant's indifference point, but it is not significant ($p = 0.16$). Moreover, raising the landlord's indifference point from 100% to 200% is associated with a 8 pp increase in the tenant being hostile toward her landlord, but is again insignificant.

Tenant Perceptions of Landlord Altruism. Section 1.5 documents that tenants are overly optimistic about landlord altruism. However, tenants may also have private information about landlords' altruism. We regress tenants' propensity to be highly altruistic toward own landlord on beliefs about the chance that their own landlord would be highly altruistic:

$$p_i = \beta_1 \mathbb{1}(\text{highly altruistic toward own } L)_i + \beta_0 + \varepsilon_i. \tag{A.80}$$

Our idea is that β_0 reflects average optimism. The coefficient β_1, meanwhile, represents private information as well as particular optimism among tenants who are themselves highly altruistic. If this optimism completely reflects private information, then it suggests an important role for assortative matching. Rates of high altruism among landlords, per tenants' perceptions, are 24 pp (75%) higher among tenants who are themselves highly altruistic. This measure yields a valid estimate of assortative matching if tenant beliefs are not *differentially* biased with respect to their own DG behavior.

Measuring Assortative Matching. We simulate the assortative matching in the data that would generate the above two observed statistics. In particular, within each of the landlords and tenant samples, we rank by altruism, breaking ties at random. We resample landlords (at random) so that we have the same number of landlords as tenants. We choose randomly choose a share of tenants q to be matched with landlords perfectly assortatively. That is, among these tenants, we rank the ith most altruistic tenant with the ith most altruistic landlord. We then regress: (i) the tenant's indifference point on the landlord's indifference point; (ii) the landlord's high altruism on the tenant's high altruism. The idea of this exercise is to infer how much assortative matching would generate the two statistics above.

Figure A.7 shows the share of matched relationships with at least one hostile party versus q, the share who are matched perfectly assortatively. We show the mean over 25 simulations for each q. Based on either measure, between 33–36% of relationships feature at least one hostile party and $q \in [0.175, 0.375]$.

In Section 1.7, we assume assortative matching by splitting the difference, positing that 27.5% of landlords and tenants are matched assortatively (i.e., $q = 0.275$).

[12]Unlike in the body, for this analysis, we all observations within a single large landlord or company. We conduct fuzzy matches on company name, landlord email domain, and phone number. Using tenant reports of their phones, emails, or landlord names, we form "connected sets" that link landlords. For instance, if one tenant reports landlord name A and phone number X, and another tenant reports landlord name B and phone number X, we infer that landlords A and B are the same. Thus, if we observe one landlord observation, we can potentially get a measure of hostility for many tenants.

A.5.3 IV Estimates: Beliefs

We use measures of landlord priors, interacted with their information-treatment assignment, as instruments for the magnitude of the landlord's belief update, similar to Haaland et al. (2023) and Bursztyn et al. (2020).[13] Our specification is

$$y_i = \beta \text{Update}_i + \boldsymbol{X}_i \boldsymbol{\delta} + \varepsilon_i, \tag{A.81}$$

where we instrument for the landlord's belief update (Update_i) with the following instruments, all interacted with information-treatment assignment: (i) the wedge between the landlord's beliefs about her own tenant and the truth for the average tenant, (ii) the wedge between the landlord's beliefs about the average tenant and the truth, and (iii) an indicator variable for receiving the information treatment. We include prior beliefs in $\boldsymbol{X}_i$, so that the instruments compare outcomes induced by changes in beliefs among people with the same prior beliefs. These controls address the concern that the first two instruments are inherently correlated with prior beliefs (Fuster and Zafar, 2022), but our results are little affected by controls.

We find a large elasticity of requesting materials with respect to changes in beliefs. A 10-percentage-point reduction in beliefs about one's own tenant results in a 6.0-percentage-point increase in propensity to request ERAP materials (Figure A.8B, $p = 0.037$). This finding is least driven by the variation induced by the second instrument, which uses landlords' priors about the average tenant. We also use a basic specification that simply instruments for the belief update with the treatment and obtain a slightly larger estimate.

Placebo Test. We collect data on other landlord beliefs to implement a placebo test. During our landlord study, Shelby County courts were open and processing evictions, but with a substantial backlog. Landlords' beliefs about court delays were therefore relevant to their assessment of costs and benefits of eviction. We inform landlords about how many evictions were filed in Shelby County courts between April 1 and June 30, 2021. We then elicit prior beliefs about the number of monetary evictions that were granted in that time. Following the information treatment, we ask treated landlords whether they wish to update their beliefs about the number of money evictions granted. Just 14 percent of landlords update their placebo belief, and the distributions of prior and posterior beliefs are quite similar (Figure A.18). When we control for the belief update in the IV exercise, results are almost identical.

A.5.4 Interaction Between Altruism and Misperceptions

The interaction between altruism and misperceptions is important for interpreting our results. First, there may be multiple constraints to bargaining: if relationships are severely hostile among those with misperceptions, then correcting beliefs may be insufficient to achieve efficient bargaining. Second, misperceptions and hostility could be self-reinforcing.

[13] In Figure 1.3, the orange bars illustrate that individuals who update their beliefs have substantially larger treatment effects than people who do not. Moreover, people with below-median prior beliefs have smaller treatment effects than people with above-median prior beliefs (blue bars).

Correlation Between Beliefs and Altruism. Tenants' beliefs about their own landlord's behavior in the DG are correlated with tenants' behavior in the DG toward their own landlord (Figure A.27A). They are still correlated with beliefs about the average landlord's behavior, though the slope falls by about half (Panel B). Just as in the DG, tenants appear to project their relationship with their own landlord onto beliefs about landlords as a class. Beliefs are also correlated with whether the tenant believes their own landlord will file an eviction and then settle (Figure A.27C), and the slope for average landlords is insignificant (Panel D). On the other hand, landlords tend to be more generous to tenants whom they believe would not abscond (Figure A.24A). Landlords' beliefs about average tenants are uncorrelated with behavior in the DG toward their own tenant (Figure A.24B).

Information Treatment Effects and Altruism. We examine the *interaction* between altruistic relationships and information.

Tenants. We split the tenant sample by whether the tenant had a very altruistic relationship with her landlord, which we define as preferring the bundle (\$x, \$x) to (\$2x, \$0). We find that the effect of the altruism treatment increases in magnitude, although not significantly so (Figure A.28A). However, the effect of the bargaining information treatment increases significantly (Panel B). In particular, providing information to people who were over-pessimistic about whether their landlord would bargain in court makes them less likely to bargain today. In both cases, the difference in updating behavior between people whose priors lie above and below the truth is significant, futher suggesting that beliefs about these facts mediate actions.

Why might the relationship affect the efficacy of information? One hypothesis is that damaged relationships affect information processing. We study whether participants with damagned relationships are less likely to update beliefs when presented with new information. Recall that before eliciting the quantitative belief update, we ask participants if their reported prior is too high or too low. We form binary measures of belief updating, $\mathbb{1}(\text{update})$, based on whether the participant chooses to update her belief. We regress:

$$\begin{aligned}\mathbb{1}(\text{update})_i = {} & \beta_1 \mathbb{1}\left(\text{Hostile}\right)_i + \beta_2 \mathbb{1}\left(\text{AltruismInfo}\right)_i + \beta_3 \mathbb{1}\left(\text{BargainingInfo}\right)_i \\ & + \beta_4 \left(\mathbb{1}\left(\text{Hostile}\right)_i \times \mathbb{1}\left(\text{AltruismInfo}\right)_i\right) + \beta_5 \left(\mathbb{1}\left(\text{Hostile}\right)_i \times \mathbb{1}\left(\text{BargainingInfo}\right)_i\right) + \varepsilon_i,\end{aligned} \tag{A.82}$$

where β_4 and β_5 represent the marginal effect of providing information to hostile tenants. When we study hostile tenants, we indeed find that they update beliefs less (Table A.20B), though we do not find the same result for landlords (Table A.20)A.

A.5.5 311 calls

Overview and Hypotheses. Are hostile tenants and landlords in properties that are more likely to violate city codes? We study whether hostile landlords or tenants are more likely to be involved in 311 calls about code violations. While 311 calls do not indicate whether a code violation actually occurred, they are nevertheless a proxy for animus between tenants and landlords. Calling 311 on one's landlord is a costly action that can cause the landlord to be fined or forced to make repairs.

How might we expect hostility to be linked to 311 calls? Hostile landlords may accrue 311 calls because they do not make repairs, since they have animus toward their tenants. Alternatively, they may become hostile because the tenant called 311. Similarly, tenants whose landlords violate code may be hostile, or hostile tenants may be more likely to rent from landlords who violate code. We cannot distinguish these hypotheses.

Data. We use publicly available 311 calls from the City of Memphis, which contain geocoded address information. We link them to study participants' addresses. We keep only 311 calls pertaining to code enforcement, drain maintenance, grounds maintenance, or street maintenance. We can only link to Memphis 311 calls and not calls in Shelby County, since we do not have 311 calls for the outlying areas.

311 calls are not linked to individuals. Therefore, we conduct fuzzy matches to the 311 records based on address only. For landlords, we cannot be sure whether they rented the property at the time of the 311 calls, but we keep only calls from 2020–October 2023. For tenants, we link to 311 calls at their current address.

Results. Table A.28 shows that highly hostile tenants are more likely to be linked to 311 calls. Tenants who are highly hostile (have $S(x) < \frac{x}{10}$) to either landlord opponent or their own landlord opponent have more 311 calls (Columns 3–4). These effects are large in magnitude: highly hostile tenants have more than 10 pp higher rates of code enforcement complaints (25% of those who are not highly hostile). We detect no such results for hostile landlords (Table A.28), although point estimates in some specifications are large (16 pp, Column 2). One hypothesis is that code enforcement problems drive tenant hostility (or vice-versa), but hostile landlords do not use code violations to discipline tenants.

A.5.6 Emergency Rental and Utilities Assistance Program Evaluation

This section conducts a policy evaluation of the Memphis/Shelby County Emergency Rental and Utilities Assistance Program (ERAP), with the goal of examining whether emergency rental assistance stops evictions.

ERAP Sample. We use administrative data from Memphis/Shelby County's ERAP records (Section 1.3). Our sample consists of households whose ERAP case was created after September 1, 2021 and who were paid by the time the program concluded in December 31, 2022. We use timestamps of changes to the household record to infer how the household progresses from creating an application, submitting an application, and receiving payment. Using personally identifiable information on the application, we conduct fuzzy merges on name and address to public evictions records, scraped from public records by the Legal Services Corporation and shared with us (Appendix B.1). Our merge strategy will not detect evictions if the eviction record only lists an occupant who does not appear on the ERAP application.

In Memphis, tenants may apply to the local ERAP, or landlords may apply on the tenant's behalf. Back rents are repaid to landlords, unless landlords decline or do not respond to ERAP, in which case tenants may receive a direct payment.[14]

[14] Landlords can decline payments as they can be subject to legal stipulations, such as right to random inspections of the property or an agreement not to evict the tenant within a certain period of time.

A share of paid households also received representation from an attorney who could encourage landlords to accept payments and impose eviction forbearance periods. To focus on the most externally valid portion of the sample, and obtain the treatment effect of rental assistance payment alone, we drop households who reached a legal settlement.

Several features of ERAP's program affect the interpretation of the payment treatment effects. First, ERAP payments could also include one–two months of rent for future months. These payments were intended to cover back rents accrued during the processing period. The exact amounts of additional months of rent varied by month. Second, ERAP could also pay utility bills directly to the utilities providers. Third, if the landlord declined ERAP, the program could make direct payments to the tenant. We do not currently have complete information about to whom the payments were made.

Tenant Characteristics. Tenants in the ERAP administrative data are highly similar to those in the experimental sample, by dint of how the samples were constructed (Table 1.1). Households paid in the legal program are similar to those in the non-legal program (Column 5 versus Column 6). After this restriction, there are about 4,800 paid tenants in the sample. The legal sample is more than half of all paid households in our data. The reason so many households entered the legal program is that the legal arm also paid bulk settlements with many tenants to large landlords.

Outcomes. We focus on two stages of the eviction process: (i) the eviction *filing*, which is the formal legal petition filed by the landlord that initiates a court eviction hearing; and (ii) an eviction *judgment*, which is a formal eviction (Section 1.3).

A.5.6.1 Additional Institutional Details

Legal Program. As Table 1.1 shows, more than half the paid tenants appear in the legal services program. This program enrolled two types of tenants. First, some tenants whom the program perceived to be at risk of eviction were granted legal assistance. Second, some tenants were granted bulk settlements with landlords.

Legal program participants were subject to explicit legal contracts that forbade eviction for a 45-day period. Their payment was also expedited. The lawyers could also encourage landlords into accepting the terms of the legal agreements. For this reason, we want to exclude this sample for external validity.

Excluding the legal sample poses several challenges to the empirical analysis. To begin with, by excluding at-risk tenants, we may be dropping the tenants who are likely marginal to eviction. This concern is likely valid to some extent. Three points mitigate how worried one should be. First, many tenants do obtain eviction filings and judgments in the pre-period. Second, many tenants in the legal services arm were *not* at risk, because they were granted bulk settlements together with other tenants. Third, many tenants were at risk but missed by the legal services arm. Tenants were flagged for the legal services arm in several ways: if they listed that they had an eviction notice on their application; if they were found to have an eviction record by exact-matching based on name; or if they asked for help from their screener. Clearly, many tenants did not make it to the legal services arm even if eligible, as we observe hundreds of filings in the pre period.

A second, more subtle concern is mean reversion, which we address explicitly in the analysis.

Terms of ERAP Receipt. The bundle in an ERAP payment had different valuations depending on the period. Early in the program, ERAP only paid back rents or utilities assistance. Beginning in July 2021, some bulk settlements incorporated three months of future-rent payments. In April 2022, regular (non-legal) ERAP payments also included two months of future rent, as well as late fees for all other payments, including non-legal payments. As we include calendar time and cohort fixed effects, these changes should not materially affect our results.

A.5.6.2 Event Study

Formal Specification. We now leverage quasi-random variation in ERAP *payment timing* among households who *applied* at the same time.[15]

We estimate:

$$\begin{aligned} y_{it} = \gamma_r + \delta_c + \alpha_t + \sum_s \beta_s \left(\mathbb{1}\left(\text{event period} = s\right)_{i,s(it)} \times \text{After}_{it} \right) \\ + \lambda \text{After}_{it} + \sum_{s<s'} \sigma_s \mathbb{1}\left(\text{event period} = s\right)_{i,s(it)} + \varepsilon_{it} \end{aligned} \tag{A.83}$$

for household i who was paid in 14-day period r and who applied in 14-day period c, and where t indexes 14-day calendar period. We include fixed effects for payment-period cohort γ_r, application-period cohort δ_c, and calendar period α_t. Event-time s is defined relative to the period of ERAP payment. The outcome y_{it} is an indicator for whether a household i receives an eviction filing or judgment in period t. The coefficients of interest, β_s, represent standard event-study coefficients. We present event-study estimates at the two-month level.

The indicator After_{it} is an indicator that turns on once the tenant has applied to ERAP, but potentially before she is paid. The variables $\sum_s \sigma_s \mathbb{1}\left(\text{event period} = s\right)_{i,s(it)}$ allow for periods prior to application to be correlated with eviction risk. We explain below why these controls add to the credibility of the design. Because of collinearity, we can only identify σ_s in periods $s < s'$ where s' is the event period of payment.

We cluster standard errors by household.

Identification Assumption and Tests. Our identifying assumption is that a given payment period cohort's judgment and filing rates would have trended in parallel with other cohorts if the cohort were not paid. Our approach permits standard pre-trends tests of our parallel trends assumption.

The program had explicit scope to expedite payments for two reasons. First, payments made through the legal program could be expedited. As noted above, we drop people

[15] ERAP speed being associated with altruism does not necessarily complicate the analysis. As in any difference-in-differences design, our empirical strategy permits level differences in ERAP speed that owe to individual covariates. Receipt may not be driven by time-varying changes in potential outcomes (eviction risk).

who successfully went through this program. In our primary specification, for power, we keep people who were flagged for the legal program but not ultimately paid by it.

Second, the program could expedite rent payments for households who apply with utility shutoff notices. We keep these households in our primary specification, since payments from applying with a utility shutoff would not be driven by a short-term change in filing risk *after* applying. We show results if we drop these households.

A sufficient (but not necessary) condition that ensures the parallel-trends assumption is if the timing of ERAP receipt, conditional on application date, is orthogonal to filing or judgment rates. Our approach is motivated by institutional context which suggests that there is a substantial exogenous component to the timing of ERAP payment among applicants. For instance, there is substantial variation in average time-to-payment across the screeners employed in the Memphis ERAP (Figure A.3B), though we are underpowered to use exclusively this variation in a investigator-IV design. In support of our assumption, we find that applications with above- versus below-median payment times are balanced on applicant demographic characteristics like sex, race, household size, income, and monthly rent (Table A.22). We do find that households with higher overdue rents received payments faster.

Level differences in demographics are not on-face concerning, since our specification is dynamic. To further investigate this point, we present an anticipation test. We regress judgment and filing rates prior to payment on payment speed and find no evidence of a correlation (Table A.23), once we control for calendar-date fixed effects. This table provides evidence against a remaining identification concern that landlords or tenants coordinate to return materials to ERAP in response to short-term spikes in eviction risk.

Interacting the indicator After_{it} with event-study coefficients in Equation (A.83) ensures that pre-trends are only estimated from tenants who have already applied for the program. Thus, the event-study coefficients are only identified from idiosyncratic payment-time variation once the tenant applied. This interaction reduces the risk of contamination from endogenous forces that could cause tenants to apply for the program in the first place (Ashenfelter, 1978).

The additional coefficients $\sum_{s<s'} \sigma_s \mathbb{1}\left(\text{event period} = s\right)_{i,s(it)}$ are useful but not necessary for the design. In particular, excluding these coefficients imposes that event time, indexed with respect to payment, is uncorrelated with eviction risk prior to application ($\sigma_s = 0$ for all s). This econometric restriction is justified if payment timing is quasi-random and can assist with power. Intuitively, if a household has not yet even applied, then their eviction risk should not be affected by whether it is 2 months or 4 months until payment. We call specifications where we omit the $\sum_{s<s'} \sigma_s \mathbb{1}\left(\text{event period} = s\right)_{i,s(it)}$ coefficients "semi-parametric," and the complete specification in Equation (A.83) "non-parametric."

Because we limit the sample to people paid 6 months before the end of our eviction data, our panel is balanced in the post-period. However, the pre-period event study coefficients are not identified from a balanced panel. That imbalance is required if we use only the quasi-random payment-time variation, as households will then by definition not be observed for the same number of pre-periods after applying.

A.5.6.3 Alternative Design (Non-Paid Households)

We call the above design, which leverages only payment timing among households who are paid, the "primary design." We additionally use an "alternative design" with an explicit untreated comparison group: households who apply to ERAP but never get paid, because they are lost in the screening process or do not finish the eligibility-certification process. There are more than 6,000 such households. This strategy relies on familiar difference-in-differences identifying logic: absent payment, trends for never-paid households would have been parallel to paid households in the post period.

The comparison group provides more power. On the other hand, the variation is less clean, as people who do not get paid may have unobservably different time-varying eviction risk.

We collapse to the week of payment by week of application level, which makes the source of variation explicit and eases estimation. For households that are never paid, we collapse only to the application period. That is, we obtain the mean outcome by calendar period $\times$ application period $\times$ payment period, where a period is two weeks, and the payment period is 0 if never paid.

The alternative design is similar to Equation (A.83):

$$y_{rct} = \gamma_r \times \delta_c + \alpha_t + \sum_s \beta_s \left(\mathbb{1}\left(\text{event period} = s\right)_{rcs} \times \text{After}_{rct}\right) + \lambda \text{After}_{rct} + \sum_{s<s'} \sigma_s \mathbb{1}\left(\text{event period} = s\right)_{rcs} + \varepsilon_{rct}, \tag{A.84}$$

where γ_r is 0 for all r if not paid and y_{rct} is the mean outcome at the $r \times c \times t$ level. Event dates s are 1 only if the household is paid. Equation (A.84) constitutes a "standard" event-study that compares paid households to household who applied in the same calendar week but were not paid. Another conceptual difference between Equation (A.84) and Equation (A.83) is that we interact the application by payment indicators. This interaction is natural once we think of cohorts as payment by application periods, and the interaction eliminates a modest pretrend in the filings specification.

In practice, to address concerns about negative weights, we use heterogeneity-robust estimators for Equation (A.84).

We focus on the "stacked" estimator, as in Cengiz et al. (2019b). For each payment period r, we form a "dataset" $d(r)$ that is all the unpaid households and the households paid in week r, collapsed as described above. We stack each dataset $d(r) \in \mathcal{D}$ and estimate a stacked version of Equation (A.84):

$$y_{rctd} = \gamma_{rd} \times \delta_{cd} + \alpha_{td} + \sum_s \beta_s \left(\mathbb{1}\left(\text{event period} = s\right)_{rcs} \times \text{After}_{rct}\right) + \sum_d \lambda_d \text{After}_{rctd} + \sum_{s<s'} \sigma_s \mathbb{1}\left(\text{event period} = s\right)_{rcs} + \varepsilon_{rctd}, \tag{A.85}$$

where outcomes are collapsed as described above.

This equation augments the standard event study to include dataset-specific time and cohort fixed effects. We two-way cluster this specification at the dataset by 14-day payment period level, since the data have been collapsed, and we weight by the number of underlying observations in each dataset. As Cengiz et al. (2019b) argue, the specification is robust to concerns about negative-weights because it compares each cohort of paid households to unpaid ("clean") controls.

Second, we use the Sun and Abraham (2021) estimator for Equation (A.84), where we cluster by 14-day payment period, employ the same collapse procedure, and weight by number of underlying observations.

Finally, we present the Two-Way Fixed Effect (TWFE) specification as a benchmark. For comparability, we estimate this specification by employing the same collapse procedure, clustering by 14-day payment period, and weighting by the number of underlying observations.

We limit the unpaid sample to the first three months ($3 \times$ 4-week periods) after applying. The reason is that we were concerned about linking evictions to households more than a few months after applying, if they did not get paid. The paid sample is confirmed to live at the address at least some point in the intervening time between payment and application, so this drop is not necessarily differential across samples.

A.5.6.4 Coefficient Interpretation and Rescaling

The event-study coefficients β_s characterize changes in filing and judgment rates after payment, with the units of percentage points per two-month event period. We make several adjustments to facilitate interpretability. First, the event periods are at the two-month level (for power), but the data are at the two-week level (which allows finer calendar-time fixed effects). To ease interpretation, we multiply each event-period coefficient and standard error by four, as there are four two-week observations per household per event period. Then the coefficient represents the effect in that two-month event period. We also report the average per-period effect over the 6-month post period. To obtain the cumulative effect, simply multiply the event period estimates by three.[16]

A.5.6.5 Results

Figure A.22 displays estimates of our event-study specification (Equation A.83). Panel A shows that ERAP payments have, at best, a small effect in the first two-month period which then attenuates. We present three specifications: a semi-parametric specification which imposes $\sigma_s = 0$ for all s (see above), a full non-parametric specification (Equation A.83, corresponding to the dashed line and pooled estimate reported on the figure), and a version which excludes households who applied with eviction notices or shutoffs and may have been eligible for expedited filings.[17] We see no evidence of pre-trends in any specification, and worst-case specifications for ERAP (the restricted sample) yield positive post-period effects on judgments. The estimate reported on the figure, which averages the

[16]Results are similar if we estimate event-period coefficients at the two-week level and average them to the two-month level. This specification further requires omitting the period-of-payment fixed effects, due to multicollinearity.

[17]Note that eviction *notices* are not the same as *filings* (Section 1.3). Not all households with filings listed that they had a notice on their application, and it was harder for them to be expedited if they did not do so.

post-period coefficients and subtracts from the averaged pre-period coefficients, suggests a small and non-significant effect on judgments in the average two-month period after payment.

ERAP payments cause a sharp reduction in eviction filings for two months before plateauing (Panel B). The semi-parametric version of Equation (A.83) shows no pre-trends. The full nonparametric version of Equation (A.83) shows a discernible pre-trend in filings. We eliminate this pre-trend if we exclude households with eviction shutoffs or notices, who can be flagged for expedited payment.

The alternative design yields similar results (Figures A.22C and D). Relative to the primary design, the alternative design yields a larger effect on filings and a smaller (more positive) effect on judgments. Notably, even if the program deferred filings, that it had a null or positive effect on judgments implies that the filings it stopped were less likely to yield judgments in the first place.

Further consistent with this point, we examine "non-suits," or explicit withdrawals from the court system, using the primary design (Figure A.26). We find that the program yielded a sharp increase in non-suits. The null effect on judgments implies that these non-suits came from households that would have had informal arrangements outside the courts.

Reweighted Estimates. One concern about the above approaches is that people who received filings in the interim between application and payment could be pushed into the legal program. This force pushes toward small effects, as the riskiest households have selected out. Moreover, as receiving filings is unusual, if receiving a filing in the pre-period makes one less likely to get a filing in the post-period (because not in the sample), there is a risk of regression toward the mean.

We adjust the primary strategy using propensity-score reweighting (Figure A.23). We regress an indicator for appearing in the legal program (and being dropped from the main sample) on an indicator for (i) obtaining a filing between application and payment, interacted with (ii) an indicator having an eviction notice at application.[18] We weight the primary strategy by $1/(1-p)$, where p is the propensity from this regression.[19]

Reweighting generates a persistent negative effect on filings and a more negative effect on judgments in some specifications.

Interpretation and Fiscal Costs of Eviction Prevention. Table A.24 aggregates estimates across empirical strategies. We present the non-parametric specification in the primary design, the reweighted estimate, and the stacked version of the alternative designs. The effects across research designs are consistent with a null or small negative effect on judgments (and positive in some specifications), and a moderate negative effect on filings.

We present estimates that are directly interpretable as percentage point effects. The reweighted estimates paint the most favorable picture of ERAP. For instance, the second column suggests that ERAP stopped 32 eviction judgments (s.e.: 22) and 87 eviction

[18] We include the second indicator because people with notices at application were supposed to be sent to the legal program directly.

[19] Intuitively, suppose one third of people who develop filings are put into the legal program. We upweight the remaining two thirds who are not in the legal program by $1/(1-1/3) = 3/2$, thus "filling in" the selected observations.

filings (s.e.: 32) per 1,000 paid households in the six-month period following payment (primary design, reweighted estimates). Effects are much larger in the first period. If the later periods had the same effects as in the first-period effects, then ERAP would have stopped about 160 filings per 1,000 paid households (Row 3, Column 2). Effects are smaller in the other designs.

Rows 4–9 of Table A.24 interpret these treatment effects. Rows 4 and 5 present the "maximum simulated effect" on judgments or filings. These show the treatment effects on simulated data, where we replace all treated units in the post period with having zero filings or judgments. Dividing the judgments or filings estimates in Rows 1 and 2 by the estimates in Row 4 give one way of benchmarking how large the effects are. The best-case effect in Column 2 is moderate — about 38% of the maximal effect — but also implies that the remaining 62% accepted ERAP funds and pursued evictions anyway. Effects in other columns are much smaller.

From our administrative data, we estimate the average ERAP payment amount is approximately \$5,400. We compute average prevention costs by dividing this payment amount by point estimates of the treatment effects. The fiscal costs of preventing judgments via ERAP are very high, since the point estimate is small (Row 6), whereas the costs of preventing filings are smaller but still meaningful (Row 7). In Rows 8–9, we present the minimum fiscal cost consistent with the lower bound of the 95-percent confidence intervals in Rows 1–2. Across specifications, the lower-bound cost is about \$35,000 per filing and \$70,000 per judgment. As discussed in the body, the fiscal cost is high in part because evictions are rare among this sample.

A.5.6.6 Interpretation and Discussion

ERAP payment defers filings for at least two months, and in some specifications, has a modest persistent effect. How important is stopping filings alone?

Filings are costly, but less costly than judgments. To Shelby County landlords, it costs \$127.50 directly to file, and there additional legal costs. Tenants also face costs of filings: eviction filings are matters of public record and landlords often investigate the eviction history of potential tenants. Filings also may trigger informal moves that tenants may want to avoid.

However, filings are likely less costly than judgments. Collinson et al. (2024) present estimates of the effect of judgments relative to filings, as they use a judge-IV design among households with filings. They find moderate negative effects on credit scores, employment outcomes, and hospital admissions.

To provide some evidence on the costs of filings and judgments, we elicited tenant survey participants' hypothetical *(i)* willingness to accept cash in exchange for moving from their unit, and, if they reported previously being evicted, *(ii)* willingness to accept cash in exchange for erasing their eviction from their record (see Appendix C.4 for details on these elicitations). Tenants value avoiding an eviction filing at or more than their subjective moving cost. 75 percent of surveyed tenants would decline \$1,000 to avoid a move. 83 percent of surveyed tenants who reported having an eviction would decline \$1,000 in cash to expunge their eviction record.

Appendix B

Appendix to Chapter 2

B.1 Data and Institutional Context

B.1.1 SNAP Sample Construction

We build off the sample in Ganong and Liebman (2018), and adapt their public-use code and data associated with the published paper. We extend the sample to 2016. Our main outcome (the number of people enrolled in SNAP, for different income groups) uses the USDA's Quality Control (QC) data from 1996–2016. The QC data provides information on the household's income as a fraction of the FPL. We use the QC data (together with its household weights) to obtain counts of the number of people in a given state-year that enroll in SNAP who are within some income band (as a fraction of the FPL).

In our welfare exercise and in some supplemental analyses, we are interested in SNAP take-up *rates*. For these, we treat the QC data as the numerator in the take-up rate, and form the denominator from the CPS, which contains the count of people within a household income band in each state and year.

Our data on state-level SNAP policies, including the income eligibility threshold and other policies (e.g., outreach spending), come from the USDA's SNAP Policy Database (2019).

The QC data include individuals in the household who are not in the SNAP unit. As in Ganong and Liebman (2018), we include these individuals as taking up SNAP. Many of these individuals are relatives of the individuals in the SNAP unit and may, in practice, have their consumption subsidized by SNAP. Results are very similar if we limit only to individuals in the SNAP unit.

B.1.2 Broad Based Categorical Eligibility

We provide more information about the BBCE provision that permits states to expand SNAP eligibility.

Broad Based Categorical Eligibility permits states to expand eligibility using Temporary Assistance for Needy Families (TANF) or State Maintenance of Effort (MOE) budgets.

States cannot expand eligibility beyond 200% of the FPL.

There are two concerns about other effects of the BBCE that could affect our analysis of inframarginal effects. In practice, states are legally required to fund small auxiliary services (e.g., telephone hotlines) using TANF/MOE funds in order to grant eligibility to more people in SNAP. Congressional Research Service (2019) writes:

> "As of July 2019, 42 jurisdictions have implemented what the U.S. Department of Agriculture (USDA) has called "broad-based" categorical eligibility. These jurisdictions generally make all households with incomes below a state-determined income threshold eligible for SNAP. States do this by providing households with a low-cost TANF-funded benefit or service such as a brochure or referral to a telephone hotline. There are varying income eligibility thresholds within states that convey "broad-based" categorical eligibility, though no state may have a gross income limit above 200% of the federal poverty guidelines."

The first concern, which we address in Section 2.2.6, is that this policy requires that SNAP administrators must notify households that they are eligible. In practice, the policy discussion around BBCE centers around the eligibility expansion, and the notification of receipt may not be much different than typical state efforts to notify recipients, especially for households below 115% of the FPL. The core of our robustness tests uses states that are treated with BBCE but do not expand eligibility. We find no evidence take-up increases in these states.

A secondary concern is that BBCE expansions sometimes waive asset rules. We also address this concern in Section 2.2.6.

B.1.3 Components of SNAP Policy Index

We use the SNAP policy index defined in Ganong and Liebman (2018), but without the BBCE. It is the average of dummies for each of seven policies. Six policies are directly from the SNAP Policy Database (2019). These are defined to be 1 if at least some parts of the state use the policy:

- At least one household vehicle is exempted from the asset test.
- Households with at least one recipient of Supplemental Security Income can use a simplified application for SNAP.
- Households can recertify with a telephone interview instead of a face-to-face interview.
- Households can apply to SNAP online.
- The state has fewer requirements for reporting changes in household earnings.
- There are call centers in the state for households to ask questions about SNAP, and in some places, recertify.

The final policy is a dummy if fewer than 20% of households have a certification period of 3 months or less, indicating that only a low share of SNAP households in the state must recertify at frequent intervals.

The index averages all seven policies except for when information about vehicle exemptions is unavailable; in this case, we average the remaining six.

In cases in which the index varies throughout the year, we use the minimum of the index in that year.

B.1.4 Experiment Sample Construction

We document several data cleaning decisions.

- A small number of participants had missing information about their household size or composition. We assume people with missing information were single, non-married, with no children (so had a household size of 1).

- A small number of participants had missing income. We assume they were in the bottom income bin and therefore had an income of $7,500.

- We top-coded household size at 6 because the most number of children that participants could report was 4.

- Incomes were top-coded at $250,000. We assume these participants had incomes of $300,000.

- Fewer than five participants took the experiment multiple times, and we drop them.

- **Attention checks.** The attention checks are the following. First, before treatment, we tell people: *"In this survey, we will ask you about your beliefs and attitudes about the Supplemental Nutrition Assistance Program (SNAP), also known as food stamps."* After eliciting the preferred charity (the incentive), we ask: *"What does SNAP stand for?"*. There are four multiple choice responses: *"Sufficiently Noisy Animal Parties"; "Supplementary Names Artful Program"; "Supplemental Nutrition Assistance Program"; "Salty Noodles And Pasta."* We drop the 106 participants who answer the acronym question incorrectly. Second, we drop the 145 participants who report that either 0 or 100% of people in the U.S. are eligible for SNAP.

- **Below 130% FPL Sample.** To form the "inframarginal" sample of experiment respondents, we predicted the relevant 2020 poverty threshold for each respondent using (1) the midpoint of their household income bin and (2) their household size, constructed via their marital status and number of kids. Anyone who reported a household income bin with a midpoint below 1.3 $\times$ the result is included in the sample of respondents under 130% FPL. This may have excluded some respondents from the inframarginal sample if they were also living with or supporting parents or elders.

B.1.5 Figure 2.1 Details

We collected income eligibility rules and take-up rates from various sources for a subset of U.S. social programs. To the extent possible, all values are from 2016. The set of programs was determined by the following process: We began by limiting to programs with FY 2016 budgets over $5 billion. We eliminated tax credits. Then we eliminated the following programs for specific reasons. We eliminated Section 8 Housing because the notion of participation is difficult to define where there are long wait lists and barriers to take-up are very high (often requiring moving). We eliminated Old Age Assistance and Social Security because income-based means tests are not meaningful for a population that often does not work and lives in households with other earners. Finally, we eliminated Pell Grants because eligibility is not based on a specific income threshold.

- CHIP
  - Eligibility data are from Brooks et al. (2016), Table 1, which gives income thresholds for children's eligibility to receive Medicaid or CHIP benefits, assuming a family of 3. In some states, the income threshold varies for different subgroups. The figure uses a population-weighted average of all the states' highest income thresholds.
  - The take-up rate is from Appendix Exhibit 1 of Haley et al. (2018), also as referenced by The Kaiser Family Foundation (KFF).
- EITC
  - Eligibility is calculated using the IRS.gov EITC maximum allowable AGI for a family of three.
  - The take-up rate is from the IRS.gov "About EITC" webpage (Internal Revenue Service, 2020), estimated by the Census Bureau using the CPS.
- Head Start
  - Eligibility is generally 100% of the FPL (HHS).
  - The take-up rate was calculated as follows:
    1. Participation rates are 35% (Child Trends, 2018), calculated using the total number of children enrolled in Head Start divided by the total number of children in poverty (ages 3-5).
    2. However, Head Start is oversubscribed. We use details from the Head Start Impact Study (U.S. Department of Health and Human Services, 2010): this study found that 85% of Head Start centers were oversubscribed. Within oversubscribed Head Start centers, the study randomized 60% of applicants into acceptance, while the remaining 40% were wait listed. In some centers, not all applicants were included in the randomization; in others, there were not enough applicants to attain this ratio in the randomization. We assume that take-up is 35*(1)*(15%) + 35*(10/6)*(85%). That is, the take-up rate is 35% among the 15% of centers which were not oversubscribed and 35*(10/6) in the oversubscribed centers, on average.

- Medicaid (parents only)
  - Eligibility data are from Brooks et al. (2016), Table 5, which gives income thresholds for parents' eligibility to receive Medicaid or CHIP benefits, assuming a family of 3. In some states, the income threshold varies for different subgroups. The figure uses a population-weighted average of all the states' highest income thresholds for parents.
  - The take-up rate is from Appendix Exhibit 2 of Haley et al. (2018), as referenced by KFF.
- NSLP (National School Lunch Program)
  - Eligibility for free lunch is 130% FPL in most districts; eligibility for reduced-price lunch is 185% FPL in most districts.
  - The take-up rate is calculated as follows:
    1. First, we take the total number of students eligible for free or reduced-price lunch in the 2015-2016 school year, according to Table 204.10 in National Center for Education Statistics (2017). This is around 26 million.
    2. We take the average number of free and reduced-price meals served daily in 2016, provided by the USDA Food and Nutrition Service: around 22 million (United States Department of Agriculture Food and Nutrition Service, 2020a).
    3. The take-up rate is 22 / 26
- SNAP
  - Eligibility data use a population-weighted average of states' eligibility thresholds.
  - The take-up rate is from Cunnyngham (2019), which gives estimates of 2016 take-up rates.
- TANF (Temporary Assistance for Needy Families)
  - Eligibility data are from Giannarelli et al. (2017), which provides, for all states, the income cutoff in dollars for TANF initial eligibility for a family of three. These cutoffs were converted to percent of the 2016 Federal Poverty Level for a family of three. The final eligibility level is the population-weighted average of these.
  - The take-up rate estimate comes from Giannarelli (2019).
- WIC (The Special Supplemental Nutrition Program for Women, Infants, and Children)
  - Eligibility is capped at 185% of the FPL.
  - The take-up rate is an estimate from the USDA FNS (United States Department of Agriculture Food and Nutrition Service, 2020b).

B.1.6 Figure 2.2 Details

Using JSTOR and EBSCO, a research assistant collected all *AER* and *QJE* papers that met one of 33 search terms according to the search engine.[1] The search terms were: "welfare program," "social insurance," "social program," "social assistance," "social welfare," "social benefit," "income threshold," "participation threshold," "means-testing threshold," "means-tested program," "means-tested welfare," "means-tested benefit," "means-tested subsidy," "income means testing," "eligibility rule," "eligibility threshold," "eligibility criteria," "eligibility criterion," "eligibility requirement," "woodwork effect," "program eligibility," "program benefit," "program subsidy," "program duration," "optimal program," "optimal provision," "benefit schedule," "program schedule," "public insurance," "program take-up," "incomplete take-up," "welfare take-up," "benefit take-up."

We limit the sampling frame to the 2010–2018 *AER* and 2010–2019 *QJE*. Appendix B.1.6 provides the search terms. On the authors' websites, we also provide a spreadsheet of all the papers, their inclusion criteria, and how we classified them. We also provide a list of judgment calls involved in this exercise and our rationale for our decision. We exclude the papers and proceedings but include comments. We exclude the 2019 AER because it was not available on JSTOR or EBSCO. We then read the abstract and/or introduction of each of the 278 papers that met at least one of the 33 search terms. We determine whether a paper was about a social welfare program.

We impose the following additional criteria when categorizing papers.

- We exclude papers that are principally about optimal income or capital taxation.
- We exclude transfers that are not intended to alleviate poverty (e.g., the effects of giving people computers).
- We exclude papers about credit market restrictions only, such as papers about mortgage deductions. We do include papers about consumer bankruptcy.
- We exclude papers about search and matching in labor markets if they do not have a substantial social insurance angle (e.g., UI).
- Because of the important theoretical connection between optimal social insurance and welfare design, we include papers that are about private insurance markets (including health insurance), as long as they have a significant angle about optimal policy.
- We define "program eligibility" as rules that determine whether a person has access to a social program. We do not consider eligibility to include access to different plan choices within a health program; our decision to exclude these papers is conservative, since they would only estimate a treatment effect using eligibility but not use optimal eligibility as an instrument.

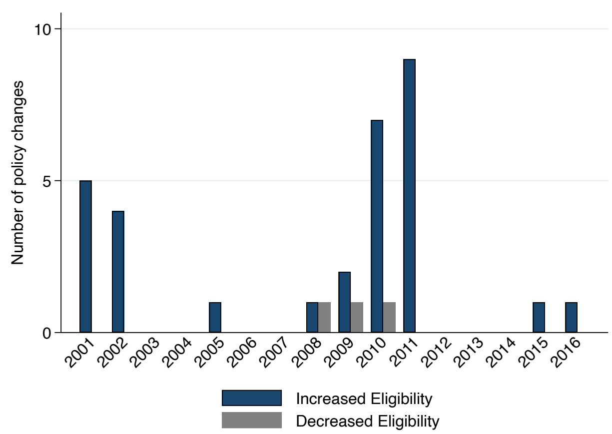
[1] The research assistant also searched the downloaded PDFs to see which search terms were most often met. Two of the papers that the search engines specified met the search terms did not actually include the search terms in the downloaded PDF, perhaps due to a bug in the search engine. Neither paper was deemed to be about social welfare programs so this issue does not substantively affect the conclusions.

B.2 Empirics Appendix

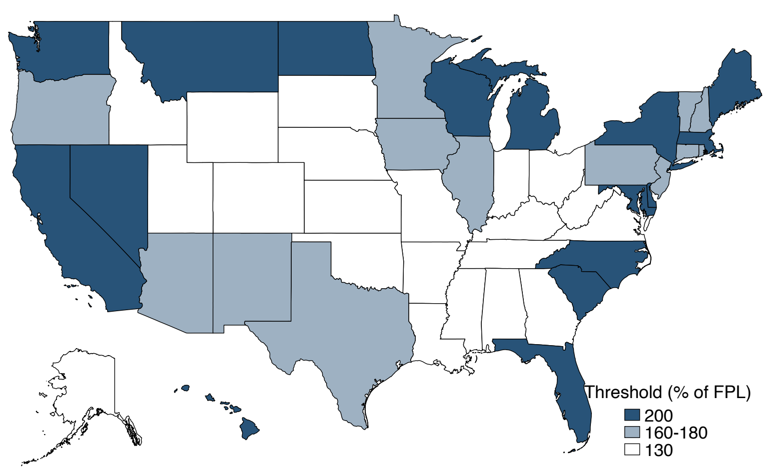
B.2.1 Additional Figures

Figure B.1: BBCE implementation background

(a) Rollout of Eligibility Changes Per Year



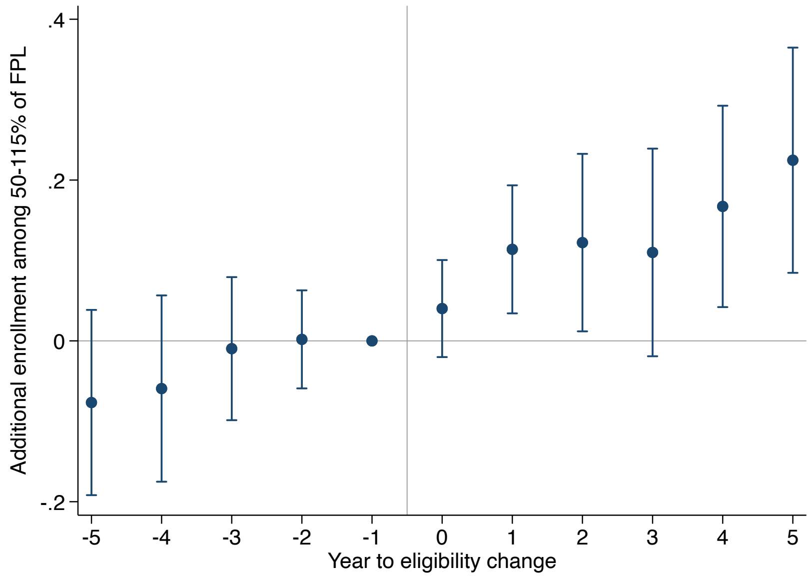
(b) Map of States that Implement Eligibility Expansions



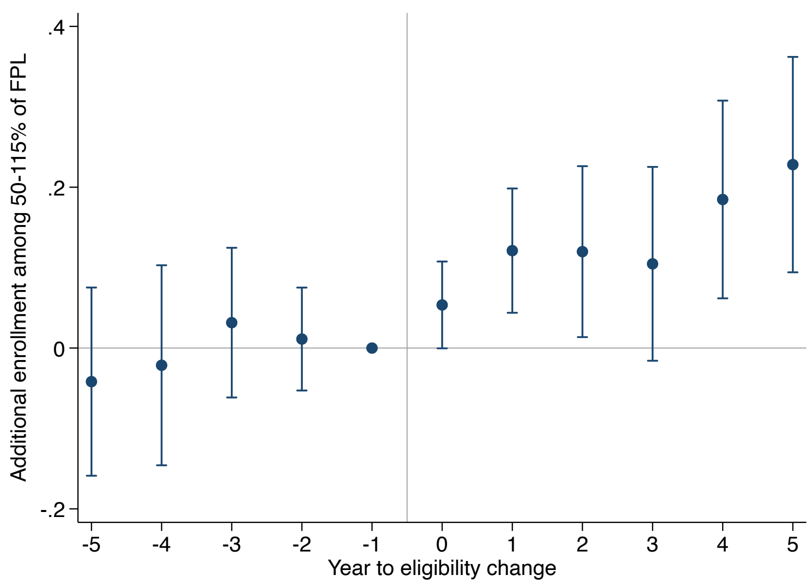
Panel A presents the number of states in each year that increased (blue bars) or decreased (gray bars) eligibility to the Supplemental Nutrition Assistance Program. Four states are counted twice, because they exhibit multiple changes. Panel B presents the maximum gross income eligibility threshold in a state from 1996–2016. The color coding refers to the maximum gross income eligibility threshold as a percent of the FPL; e.g., states colored in dark blue have maximum eligibility threshold of 200%. In two states that increase and then reduce the eligibility threshold, we present the largest eligibility threshold in the data. Source: SNAP Policy Database.

Figure B.2: Event Study of Changes to Eligibility Threshold: Without Controls

(a) Sample: 50 to 115% of FPL, No Controls

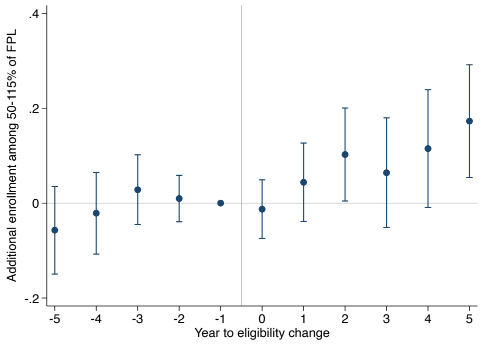


(b) Sample: 50 to 115% of FPL, Only Denominator Control

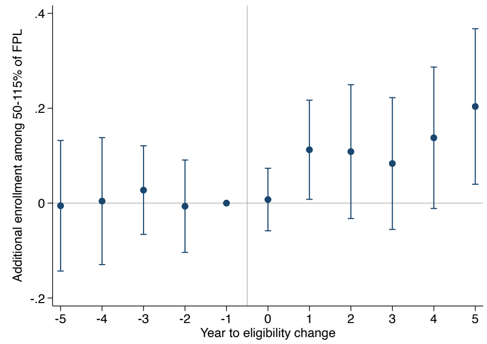


This figure is similar to Figure 2.4, but Panel A presents the specification with no controls beyond state and year fixed effects. Panel B presents the specification with state and year fixed effects, only controlling for the log of the total number of people between 50 and 115% of the FPL (from the CPS).

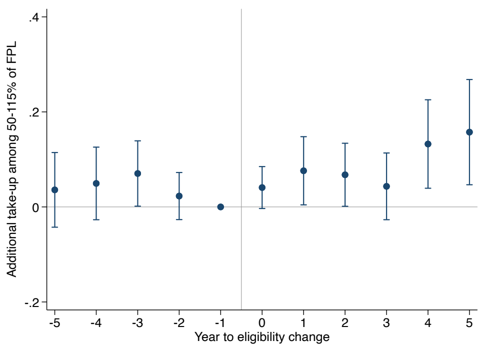
Figure B.3: Extra robustness checks



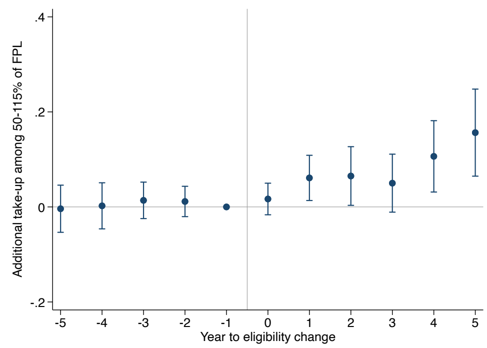
(a) Monthly data



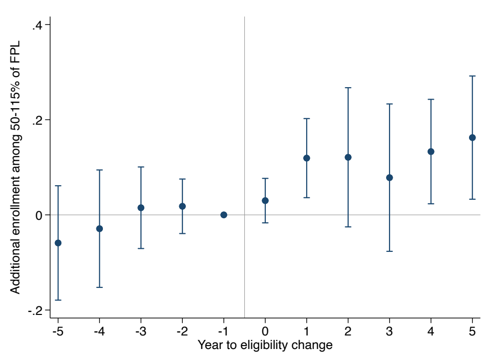
(b) Only Households with Dependents



(c) Take-up



(d) Take-up, controlling for Ln(CPS count in 50-115%)

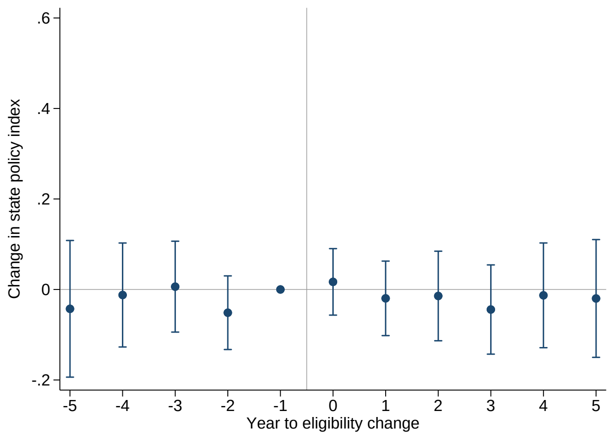


(e) Enrollment, weighted by state-year population

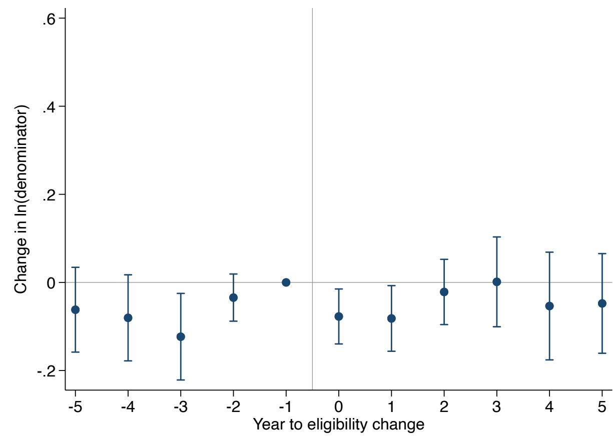
Panel A presents the results of estimating Equation (2.2) with monthly data instead of annual data. Panel B includes only SNAP recipients with any dependents—households that will not be affected by ABAWD work waivers. Panels C and D use the take-up share instead of the log of enrollment as the regressand, where the numerator in the take-up share comes from the USDA Quality Control data and the denominator uses the CPS. Panel C has no controls for state-year CPS population, while Panel D controls for the log of count of individuals in the CPS with household income in 50-115% FPL. Panel E uses the main specification and weights by population size in each state-year.

Figure B.4: Balance Tests

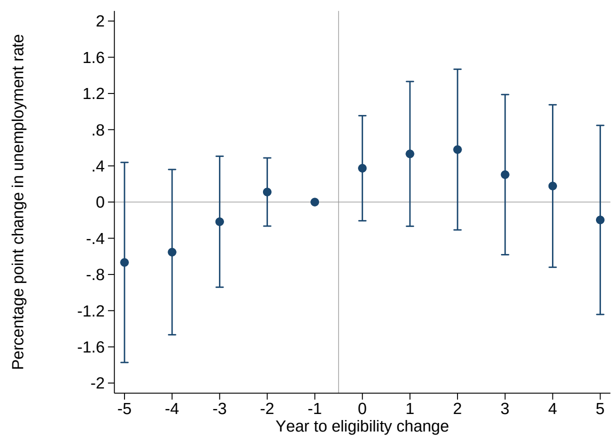
(a) SNAP Policy Index



(b) CPS Population Count: 0 to 130% of FPL

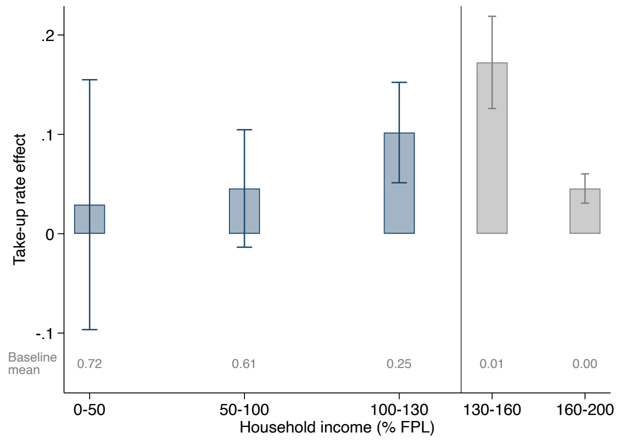


(c) Unemployment Rate



This figure presents placebo event studies with the main specification from Equation (2.1) but replacing the outcome with the main control variables. The event time is indexed around changes to state eligibility thresholds. Panel A uses the "Ganong-Liebman" index of SNAP policies, which are found in the USDA's SNAP Policy Database, as the outcome. Panel B uses the (ln of) the number of people in a state earning below 130% FPL (from the CPS) as the outcome. Panel C uses the state unemployment rate as the outcome. Standard errors are robust to heteroskedasticity and clustered by state.

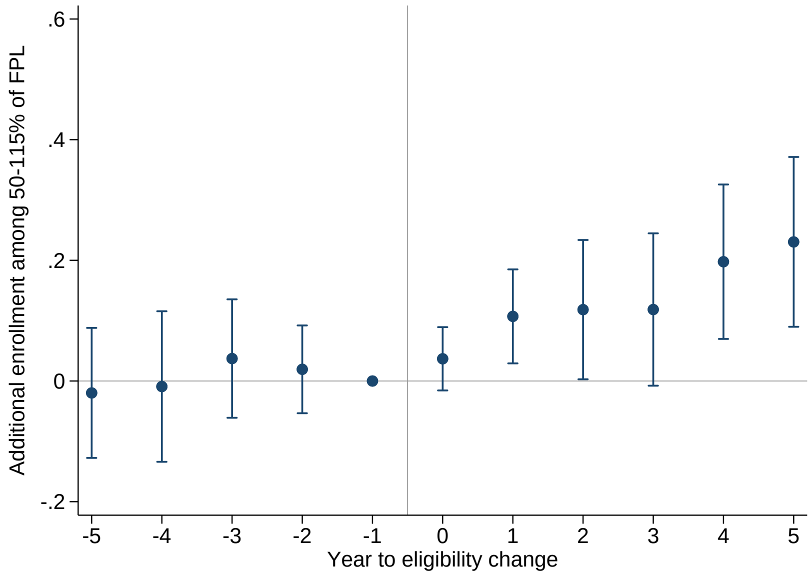
Figure B.5: Effect on Take-Up Rates by Income Group



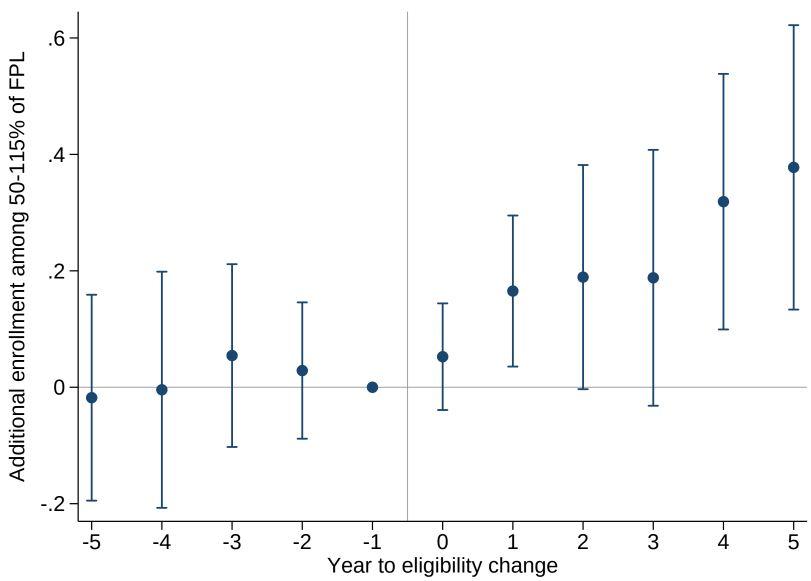
This figure presents estimates of Equation (2.2) using take-up rates as the outcome variable. The bars show the effect of the eligibility threshold on SNAP take-up by income group, and the whiskers show the 95% confidence intervals. While the regression specification is the same for all bars (with only the reference group changing), they are colored blue and gray to distinguish the effects on the inframarginal population versus the effects on the newly eligible population. Take-up rates are calculated using the enrollment counts from the USDA Quality Control (QC) data in the numerator and total counts of individuals within the income group from the Current Population Survey (CPS) in the denominator. Standard errors are robust to heteroskedasticity and clustered by state.

Figure B.6: Two-Way Fixed Effects Robustness

(a) Stacked Estimator (Cengiz et al., 2019a)



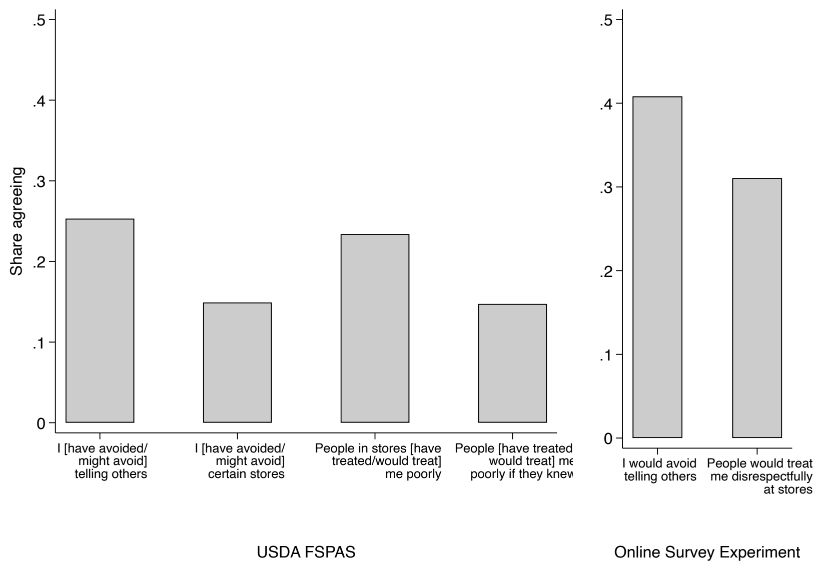
(b) Sun and Abraham (2021) Estimator



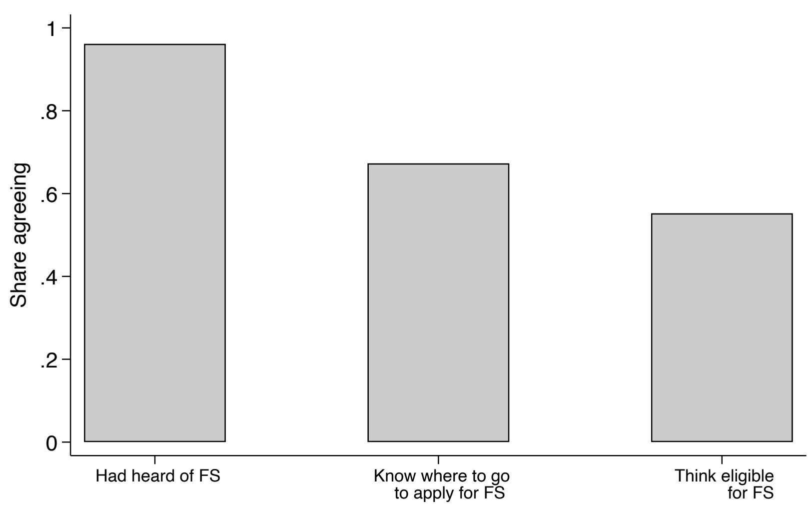
This figure presents heterogeneity-robust event study estimates using the 50–115% sample. Panel A presents the "stacked estimator" developed in Cengiz et al. (2019a). For each treated state, we form a dataset keeping just one treated state and all never-treated states. We then stack all datasets and estimate a version of Equation 2.1, controlling for dataset-state fixed effects. We employ two-way clustering by dataset and state. Panel B presents the results from the estimator in Sun and Abraham (2021), using never-treated states as a comparison group.

Figure B.7: FSPAS Descriptives

(a) Stigma



(b) Information



This figure shows the share of respondents (among approved applicants and eligible nonparticipants) agreeing with different statements presented in the USDA Food Stamp Program Access Study about the stigma around SNAP (in Panel A) and their access to information about SNAP (Panel B). In Panel A, we compare results to those from the online experiment, limited to respondents earning under 130% FPL.

B.2.2 Additional Tables

Table B.1: Estimates of the Inframarginal Effect in Alternate Samples

	(1) Main estimate	(2) Extra controls	(3) Waivers, lag unemp.	(4) Excludes recession	(5) Weighted	(6) Avg of coefficients	(7) All data
Panel A. 0–115% FPL							
Income limit (% FPL) / 100	0.064	0.067	0.052	0.064	0.063	0.054	0.072
	(0.057)	(0.056)	(0.055)	(0.060)	(0.073)	(0.065)	(0.048)
Panel B. Any dependents							
Income limit (% FPL) / 100	0.105*	0.112*	0.096*	0.114*	0.123*	0.104*	0.133***
	(0.057)	(0.060)	(0.056)	(0.059)	(0.071)	(0.062)	(0.048)
Observations	705	705	680	628	705	705	1071
N states	45	45	45	45	45	45	51

This table presents Table 2.1 with different samples, using the specification in Equation (2.2). See notes to Table 2.1 for details. Panel A uses the sample of people at 0–115% of the Federal Poverty Line (FPL). Panel B presents estimates for the sample of households with dependents, who are not subject to ABAWDs rules, in households earning 50–115% FPL. The outcome is SNAP enrollment as estimated from the USDA Quality Control data. Standard errors are robust to heteroskedasticity and clustered by state. *, **, and *** indicate $p < 0.1$, 0.05, and 0.01, respectively.

Table B.2: Pre-Policy State Characteristics

Panel A. By Ever Changed Threshold			
	No	Yes	*p*-value
Share of state pop enrolled	0.08	0.07	0.29
Unemployment rate	3.93	3.88	0.85
Average family income in state	51.87	57.42	0.01
Ganong-Liebman Index	0.06	0.10	0.09
Outreach spending	1.64	16.92	0.09
Observations	30	21	
Panel B. By New Eligibility Threshold			
	$<$ 200% FPL	200% FPL	*p*-value
Share of state pop enrolled	0.09	0.09	0.79
Unemployment rate	6.04	5.89	0.86
Average family income in state	73.57	64.79	0.08
Ganong-Liebman Index	0.52	0.43	0.48
Outreach spending	54.57	48.82	0.85
Observations	13	17	

In Panel A, we compare states which did and did not ever change their SNAP eligibility threshold in their pre-policy characteristics, as measured in the year 2000 (before any states implemented policy changes). In Panel B, we limit the sample to states which did increase their eligibility threshold and compare those which raised it to 200% FPL to those which raised it to a value below 200% FPL, where the pre-policy characteristics are measured two years before their policy change. The first row shows the share of the state population enrolled in SNAP in the given year. The second row shows the state unemployment rate. The third row shows the average family income (from the CPS). The fourth shows the Ganong-Liebman Index, excluding the BBCE indicator. The final row shows spending on SNAP outreach in the state, where the value is winsorized.

Table B.3: USDA FSPAS Characteristics

	(1) Info types	(2) Stigma types
Enrolled	0.38	0.43
Female	0.76	0.78
White	0.52	0.65
Has kids	0.47	0.53
Age	43.77	39.31
Observations	953	575

The table shows summary statistics for respondents categorized as "information-only" and "stigma-only" in the USDA Food Stamp Program Access Study (approved applicants and eligible nonparticipant samples only).

B.2.3 Measurement Error Robustness

We study whether measurement error in reported income in the Current Population Survey (CPS) could explain our main results. Figure B.4B shows that the count of people in the CPS earning below 130% FPL does not change discretely around the time of the policy implementation. The figure for people earning 50–115% FPL looks very similar. Especially given that we control for the denominator, it is implausible that state populations grow fast enough only in treated state-years, beginning exactly at the time of the eligibility increase, that this measurement error could explain our event study results. Any threat to identification requires that the mismeasured portion of the denominator grows in a way that is correlated with treatment, beginning precisely at the date of treatment.

To formalize this point, we obtain the following bound on the magnitude of measurement error in the denominator required to explain our results. In state-years with an eligibility threshold above 130% of the FPL, we simulate systematic measurement error in the denominator using an "inflated" denominator that we define as:

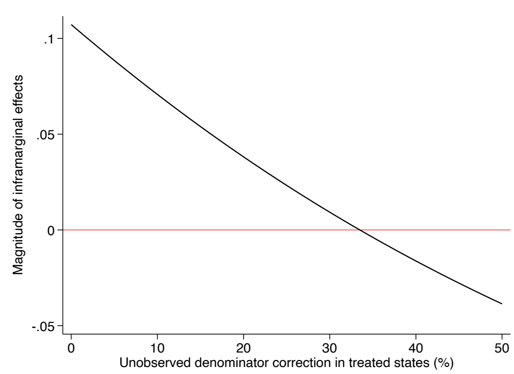
$$\text{simulated denominator} := \text{observed denominator} \times \text{inflation factor},$$

where the inflation factor represents the magnitude of simulated measurement error. For instance, an inflation factor of 1.05 represents the case where we replace the treated state-years' denominators as being 5% larger than what we observe in the CPS.

We then estimate Equation (2.2) with the simulated denominator in treated state-years. We find that the inframarginal effect vanishes only if the denominator in treated state-years is inflated by more than 30% (Figure B.8). Put another way, only when we add an additional 30% of the population to the denominator (and impose that this measurement error only exists in treated state-years) can we eliminate the inframarginal effect. As a benchmark, we note that the average state population between 50 to 115% of the FPL (i.e., the denominator) grew by 26% between 2001 and 2016. Thus the measurement error required to reverse our result would need to be larger than the entire observed population growth in the sample period. It is implausible that *only* treated states are subject to measurement error that is this extreme.

Altogether, while our denominator obtained from the CPS may be subject to some measurement error, it would have to be systematically correlated with treatment to an implausible degree in order to explain our results.

Figure B.8: Simulated Measurement Error



This figure presents a bound on the amount of measurement error in the denominator that would be required to reverse our results. In states where the eligibility threshold exceeds 130% of the FPL, we inflate the observed population between 50–115% of the FPL by the factor on the x-axis. We then present the estimate of the inframarginal effect from Equation (2.2), estimated using the simulated denominator. Only if the population is inflated by 30% can we reverse the inframarginal effect.

B.3 Online Experiment Appendix

B.3.1 Auxiliary Experiment

Table B.5 shows the auxiliary experiment is balanced between treatment and control. The results of this second experiment are mixed (Table B.9). We find no evidence for an effect of a belief correction exercise on first-order beliefs. We find a *positive* effect of the belief correction on second-order beliefs: for people whose priors were below the truth, correcting beliefs *raises* the stigma they report (point estimate: 0.069, SE: 0.041, $p = 0.091$).

We note that the treatment effect is positive for people whose beliefs are corrected down (point estimate: 0.018, SE: 0.035). This point estimate is consistent with the results from the high-state treatment. Alternatively, it may suggest that any belief correction may simply cause participants to report more stigma, e.g. because they do not like being corrected after receiving an initial hint. We also present effects with demographic controls (Table B.10), which are similar. In this case, the positive effect on second-order stigma for correcting beliefs upward is very slightly attenuated.

We are more cautious about interpreting the results from auxiliary experiment for the following reasons. First, people who are shown multiple pieces of information might simply end up confused, which could attenuate or undo its effects. Because we did not elicit beliefs after being shown the belief correction, we do not have a way of checking how the correction actually shifted posteriors. The inconclusive results suggest that providing the second piece of information might have had an unintended consequence of causing participants to tune out the second piece of information, perhaps because it was perceived as contradicting the first piece of information.

Second, the auxiliary belief correction only operates on people *after* they have been shown a hint. As a result, because it is cross-randomized, it affects the group of people that do or do not comply with the high or low treatment. The staggered nature of the design complicates this interpretation: people who have low prior beliefs after treatment are a selected group, since they have been exposed to a hint that causes them to update.[2]

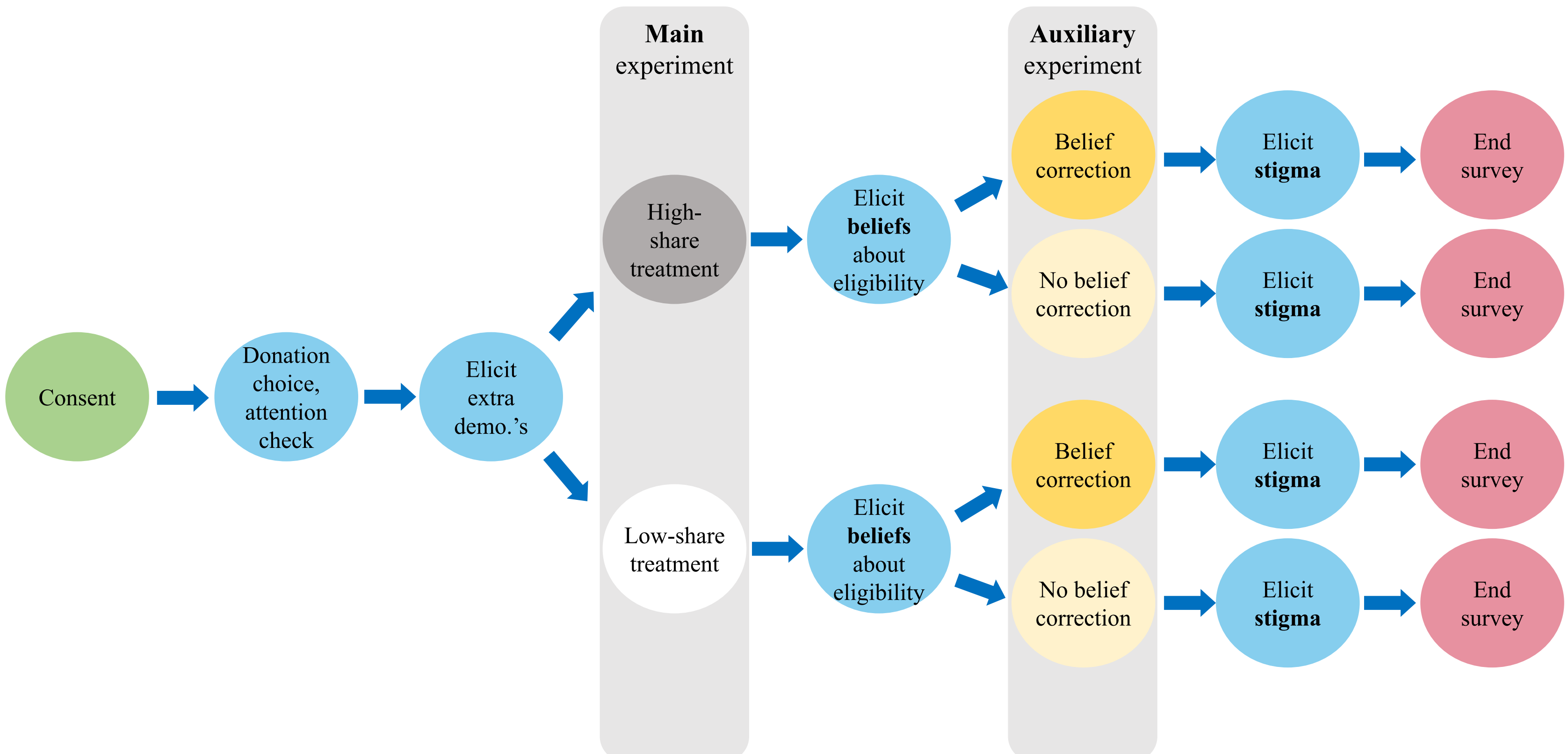
Third, the belief-correction treatment, when paired with the high-share treatment, affects people's beliefs about the *distribution* of eligibility thresholds across states. If stigma is linked to people's beliefs about the distribution of eligibility thresholds, it is not clear how the combination of experiments affects stigma.

Altogether, the main experiment provides a somewhat cleaner test of the null hypothesis that stigma plays no role in woodwork effects. Nevertheless, the inconclusive results from the auxiliary experiment lead us to interpret the experiment with some caution.

[2] Consistent with this point, the positive treatment effect on second-order stigma from correcting beliefs upward attenuates once we add demographic controls (Table B.10).

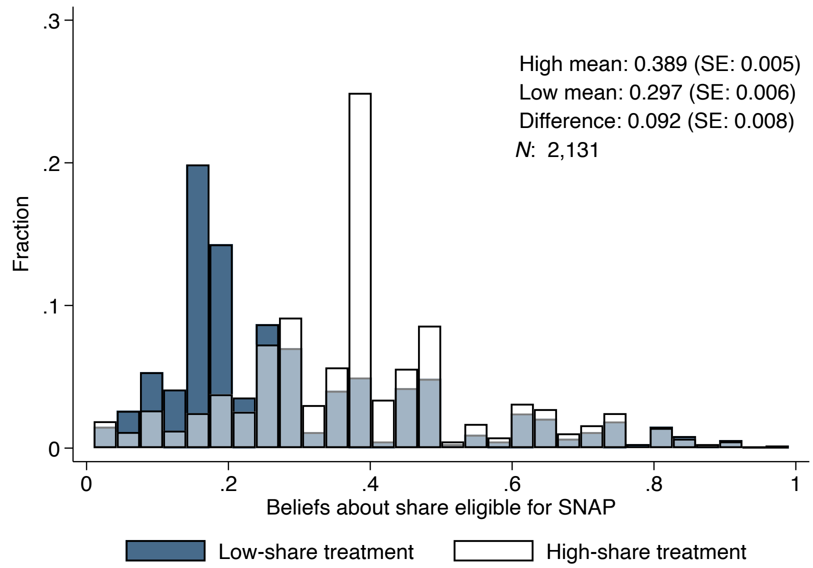
B.3.2 Additional Figures

Figure B.9: Visual Depiction of Experiment Design



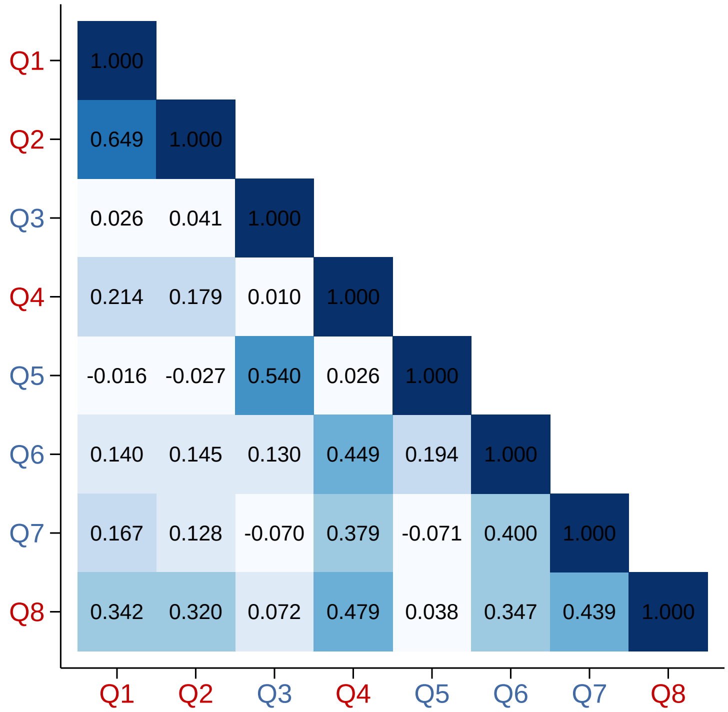
This figure presents the experiment design. The donation choice was to one of four charities (used to incentivize belief elicitation). We elicited several demographics (in addition to those provided by Lucid).

Figure B.10: Effect of High-Share Treatment on Beliefs about Eligibility



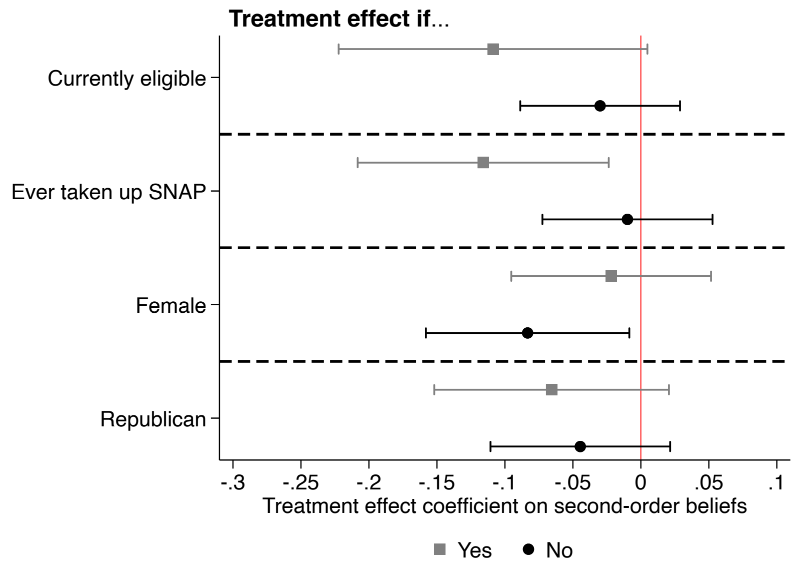
This figure presents the distribution of beliefs from the online experiment, split by treatment group, about the fraction of people who are eligible for SNAP. The y-axis shows the share of people within each treatment group who report a given fraction are eligible for SNAP. The blue bars show the values for the low-share treatment. The white bars show the values for the high-share treatment. The light blue shaded area shows the overlap.

Figure B.11: Correlations Between Stigma Questions



This figure presents correlations between the stigma questions in the order they were elicited. Section 2.3 provides the question texts. We classify questions 1, 2, 4, and 8 (labeled in red) as first-order stigma. We classify questions 3, 5, 6, and 7 (labeled in blue) as second-order stigma.

Figure B.12: Treatment Effect Heterogeneity



This figure presents treatment effects and 95% confidence intervals of the high-share treatment on the second-order stigma index (Equation (2.4)), split by demographic group.

B.3.3 Additional Tables

Table B.4: Experiment Sample Composition and Balance for High vs. Low Treatment

	CPS Sample	Full Sample			Below 130% FPL		
		Low-share	High-share	*p*-value	Low-share	High-share	*p*-value
Female	0.517	0.531	0.522	0.658	0.647	0.597	0.248
White	0.776	0.727	0.737	0.623	0.707	0.684	0.583
Hispanic	0.165	0.109	0.112	0.824	0.120	0.160	0.203
At least some college	0.611	0.778	0.772	0.737	0.606	0.612	0.894
Age	47.714	45.679	46.145	0.526	45.036	45.042	0.997
Any Children	0.254	0.537	0.531	0.790	0.618	0.597	0.619
Single	0.291	0.366	0.368	0.927	0.418	0.441	0.594
Household Size	2.296	2.519	2.517	0.973	2.687	2.692	0.970
Democrat	-	0.541	0.517	0.275	0.522	0.490	0.476
On Food Stamps (Currently or Ever)	-	0.383	0.392	0.648	0.627	0.624	0.946
Household Income (000's)	-	59.007	59.941	0.680	15.331	13.431	0.021
Census regions							
Northeast	0.175	0.208	0.191	0.308	0.169	0.095	0.014
Midwest	0.207	0.190	0.198	0.617	0.189	0.209	0.565
South	0.379	0.344	0.372	0.176	0.369	0.441	0.100
West	0.238	0.259	0.240	0.311	0.273	0.255	0.639
Joint F-test *p*-value				0.941			0.018
Observations				2131			512

Income uses the midpoint of a set of bins and is top-coded at $250,000. Household size is top-coded at 6. The CPS sample uses the 2019 NBER MORGs.

Table B.5: Online Experiment: Randomization Balance for Belief Correction

	CPS Sample	Full Sample			Below 130% FPL		
		No Correction	Belief Correction	p-value	No Correction	Belief Correction	p-value
Female	0.517	0.515	0.537	0.304	0.615	0.626	0.798
White	0.776	0.742	0.722	0.312	0.704	0.687	0.665
Hispanic	0.165	0.109	0.113	0.753	0.130	0.151	0.488
At least some college	0.611	0.783	0.767	0.378	0.615	0.604	0.788
Age	47.714	45.770	46.048	0.705	45.725	44.400	0.391
Any Children	0.254	0.528	0.540	0.559	0.623	0.592	0.473
Single	0.291	0.366	0.368	0.906	0.401	0.457	0.203
Household Size	2.296	2.530	2.507	0.724	2.757	2.626	0.350
Democrat	-	0.525	0.533	0.709	0.490	0.521	0.486
On Food Stamps (Currently or Ever)	-	0.377	0.398	0.329	0.615	0.634	0.665
Household Income (000's)	-	61.526	57.476	0.073	14.787	13.952	0.311
Census regions							
Northeast	0.175	0.199	0.200	0.965	0.162	0.102	0.044
Midwest	0.207	0.208	0.180	0.112	0.215	0.185	0.402
South	0.379	0.361	0.354	0.749	0.413	0.400	0.766
West	0.238	0.232	0.265	0.077	0.211	0.313	0.008
Joint F-test p-value				0.611			0.498
Observations				2131			512

Income uses the midpoint of a set of bins and is top-coded at $250,000. Household size is top-coded at 6. The CPS sample uses the 2019 NBER MORGs.

Table B.6: Online Experiment: Attrition Balance

	Total N	High-share treatment		Beliefs correction	
		All	<= 130% FPL	All	<= 130% FPL
1. Any attrition or drops	**567**	**0.009**	**0.018**	$-$**0.001**	**0.005**
		(**0.016**)	(**0.033**)	(**0.015**)	(**0.033**)
2. Bad priors	237	0.002	0.008	-0.000	0.003
		(0.011)	(0.024)	(0.011)	(0.025)
3. Attrited before share treatment	49	0.004	-0.004		
		(0.005)	(0.010)		
4. Attrited at or after treatment	126	0.006	0.007	-0.001	0.013
		(0.008)	(0.019)	(0.008)	(0.019)
5. Omitted any stigma answers	107	0.002	0.005	-0.006	-0.013
		(0.008)	(0.016)	(0.008)	(0.017)
6. Inattentive	106	0.000	0.021	0.005	0.003
		(0.008)	(0.013)	(0.008)	(0.014)
Observations		2,698	689	2,698	689

This table shows that attrition and drops were balanced across treatment and control. Each row tests for balance between treatment and control on a different dummy outcome. The first column gives the total number of respondents who were dropped for the reason indicated by the row. Note that respondents could be dropped for multiple reasons. The next two columns show balance for the main experiment, where respondents were provided a random hint about the share of Americans eligible for SNAP. The last two columns show balance for the secondary experiment, where respondents beliefs were corrected with the true share. Row 1's outcome is a dummy for attriting or being dropped from the sample. Row 2's outcome is a dummy for providing prior beliefs about the share of Americans eligible for SNAP that were below 1% orabove 99%, or skipping this question entirely. Row 3's outcome is a dummy for dropping out of the survey before the treatment screen. The second two columns of this row are empty because individuals who attrited before the treatment screen were not randomized into treatment or control for the beliefs correction. Row 4's outcome is a dummy for attriting at or after the share treatment screen. Row 5's outcome is a dummy for not answering any of the stigma questions. Row 6's outcome is a dummy for failing an attention check. *, **, and *** indicate $p < 0.1$, 0.05, and 0.01, respectively.

Table B.7: Online Experiment: High-Share Effect on Reported Stigma, without Demographic Controls

	Overall	Subindices	
		First-Order	Second-Order
Under 130% FPL			
High-share treatment	-0.032	0.046	-0.109*
	(0.050)	(0.065)	(0.058)
p-value	0.530	0.485	0.061
Observations	512	512	512
Full Sample			
High-share treatment	-0.013	0.025	-0.050*
	(0.024)	(0.031)	(0.027)
p-value	0.598	0.421	0.061
Observations	2,131	2,131	2,131

The table shows the effect of the "high-share" hint on individuals' level of agreement to statements measuring stigma around food stamps and welfare for individuals under 130% FPL (top panel) and the full sample (bottom panel) (Equation (2.4)). The estimates are identical to Figure 2.5. *, **, and *** indicate $p < 0.1$, 0.05, and 0.01, respectively.

Table B.8: Online Experiment: High-Share Effect on Reported Stigma, with Demographic Controls

	Overall	Subindices	
		First-Order	Second-Order
Under 130% FPL			
High-share treatment	-0.023	0.049	-0.096*
	(0.050)	(0.064)	(0.058)
p-value	0.640	0.448	0.099
Observations	512	512	512
Full Sample			
High-share treatment	-0.016	0.016	-0.048*
	(0.023)	(0.029)	(0.026)
p-value	0.489	0.580	0.072
Observations	2,131	2,131	2,131

The table shows the effect of the "high-share" hint on individuals' level of agreement to statements measuring stigma around food stamps and welfare for individuals under 130% FPL (top panel) and the full sample (bottom panel) (Equation (2.4)). It is identical to Table B.7 and Figure 2.5 except we include demographic controls for: an age quadratic, income, political party, gender, region, household size, marital status, having children, being on or ever having been on food stamps, and education and race/ethnicity fixed effects. *, **, and *** indicate $p < 0.1$, 0.05, and 0.01, respectively.

Table B.9: Online Experiment: Belief Correction, No Demographic Controls

	Overall	Subindices	
		First-Order	Second-Order
Panel A. Priors $<$ Truth			
Beliefs Correction Treatment	0.044	0.020	0.069*
	(0.036)	(0.048)	(0.041)
Observations	868	868	868
p-value	0.218	0.680	0.091
Panel B. Priors $\geq$ Truth			
Beliefs Correction Treatment	0.008	-0.002	0.018
	(0.031)	(0.041)	(0.035)
Observations	1,263	1,263	1,263
p-value	0.800	0.964	0.615

This table shows results from the second experiment embedded in our online survey, where respondents were informed of the true share of Americans eligible for SNAP after previously being asked to report their beliefs (and given a hint, which is the primary experiment discussed in the text). It presents treatment effect estimates from Equation (2.4). Panel A restricts the sample to those who initially underestimated the eligibility share, so that the treatment should have led them to revise upwards. Panel B restricts the sample to those who initially overestimated the eligibility share, so that the treatment should have decreased their beliefs. *, **, and *** indicate $p < 0.1$, 0.05, and 0.01, respectively.

Table B.10: Online Experiment: Belief Correction, With Demographic Controls

	Overall	Subindices	
		First-Order	Second-Order
Panel A. Priors < Truth			
Beliefs Correction Treatment	0.036	0.006	0.066
	(0.035)	(0.045)	(0.040)
Observations	868	868	868
p-value	0.301	0.900	0.103
Panel B. Priors ≥ Truth			
Beliefs Correction Treatment	0.032	0.034	0.030
	(0.030)	(0.038)	(0.035)
Observations	1,263	1,263	1,263
p-value	0.290	0.375	0.389

This table shows results from the second experiment embedded in our online survey, where respondents were informed of the true share of Americans eligible for SNAP after previously being asked to report their beliefs (and given a hint, which is the primary experiment discussed in the text). It presents treatment effect estimates from Equation (2.4). Panel A restricts the sample to those who initially underestimated the eligibility share, so that the treatment should have led them to revise upwards. Panel B restricts the sample to those who initially overestimated the eligibility share, so that the treatment should have decreased their beliefs. This table is identical to Table B.9 except we additionally include demographic controls for: an age quadratic, income, political party, gender, region, household size, marital status, having children, being on or ever having been on food stamps, and education and race/ethnicity fixed effects. *, **, and *** indicate $p < 0.1$, 0.05, and 0.01, respectively.

Table B.11: Online Experiment: Treatment Effect by Belief-Correction Randomization

	(1) First-order index	(2) Second-order index	(3) First-order index	(4) Second-order index
High-share treatment	0.048	-0.019	0.003	-0.081**
	(0.044)	(0.037)	(0.044)	(0.038)
Observations	1050	1050	1081	1081
Sample	Not shown truth	Not shown truth	Shown truth	Shown truth

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table presents treatment effects on first- and second-order stigma from Equation (2.4) the sample by whether the sample's beliefs were not truthfully corrected (Columns 1 and 2) or were truthfully corrected (Columns 3 and 4). *, **, and *** indicate $p < 0.1$, 0.05, and 0.01, respectively.

Table B.12: Online Experiment: Association Between Take-Up and Stigma

	(1)	(2)	(3)
	On SNAP (currently or ever)	On SNAP (currently or ever)	On SNAP (currently or ever)
First-order index	-0.133***		-0.145***
	(0.014)		(0.015)
Second-order index		-0.012	0.044**
		(0.018)	(0.018)
Constant	0.391***	0.388***	0.391***
	(0.010)	(0.011)	(0.010)
Observations	2131	2131	2131

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table presents associations between first- and second-order stigma and participants' reports about taking up SNAP (now or in the past). We elicit the take-up questions before treatment. *, **, and *** indicate $p < 0.1$, 0.05, and 0.01, respectively.

B.4 Mechanisms Appendix

This appendix provides information about the measurement and estimation in Section 2.3.2 and 2.5.2.

B.4.1 FSPAS Data

- We use surveys of (1) eligible nonparticipants and (2) successful SNAP applicants from the FSPAS, a study conducted by the USDA in the year 2000. The USDA considered someone an eligible nonparticipant if their household income was beneath 130% FPL and they were not currently enrolled in SNAP. There are 421 successful SNAP applicants and 1,323 eligible nonparticipants.
- **Stigma.** Respondents were considered affected by stigma if they answered "yes" to (agreed with) at least one of the following questions (statements).
  - If they'd ever been enrolled in SNAP:
    - Have you ever avoided telling people you got food stamps?
    - Did you ever go out of your way to shop at a store where no one knew you?
    - Have you ever been treated disrespectfully when using food stamps in a store?
    - Were you ever treated disrespectfully when you told people that you received food stamps?
  - If they'd never been enrolled in SNAP:
    - "If I got food stamps, I might go out of my way so people would not find out."
    - "I might not shop in certain stores because I don't want people there to know I use food stamps."
    - "People in stores would treat me disrespectfully when I use food stamps."
    - "People would treat me disrespectfully if they found out that I got food stamps."
- **Information.** Respondents were considered affected by information barriers if they (a) were in the eligible nonparticipant sample and (b) said "no" to any of the following questions:
  - Had you heard of food stamps or the Food Stamp Program before today's interview?
  - Do you know where you would have to go to apply for food stamps or other assistance?
  - Do you think you may be eligible to receive food stamp benefits?

- **Survey weights.** Each survey in the FSPAS is weighted to be representative of the population the respondents were sampled from. When we combine participants and eligible nonparticipants, we adjust these weights according to the share of Americans who participated in SNAP conditional on being eligible in the year 2000 (estimated in the QC data to be 40%).

B.4.2 Estimation procedures

We seek to estimate the coefficients β^s and β^i from the equation:

$$\left(\frac{\partial \ln(\text{N enrolled})}{\partial m}\right)_d = \frac{1}{p}\left(\left(\frac{\partial c}{\partial m}\right)_d s_d\right) \times \beta^s + \frac{1}{p}(1-s)_d \times \beta^i + \epsilon_d, \tag{B.1}$$

noting that the $\frac{\partial \ln \text{Number Eligible}}{\partial m}_d$ term in Equation (2.13) vanishes because our objective is to study woodwork effects and we assume (and test in Section 2.2) that the eligibility changes do not coincide with other changes to woodwork status.

We estimate this equation at the demographic-cell level d. For each demographic cell, we need to estimate three inputs: $\frac{\partial c}{\partial m}_d$, s_d, and $\frac{\partial \ln(\text{N enrolled})}{\partial m}_d$. We use the following process:

1. **Estimating $\frac{\partial \ln(\text{N enrolled})}{\partial m}_d$.** We compute estimates of the state-level population within each demographic cell d. For each demographic cell d, we then estimate $\frac{\partial \ln(\text{N enrolled})}{\partial m}_d$ using the following equation:

$$\ln(\text{N enrolled})_{s,t,d} = \eta_d \ln(\text{share eligible})_{s,t,d} + X'_{s,t,d}\phi + \delta_{s,d} + \gamma_{t,d} + \varepsilon_{s,t,d}, \tag{B.2}$$

instrumenting for $\gamma = \frac{\partial \ln(\text{N enrolled})}{\partial m}_d$ with the BBCE eligibility rate in each state-year. This a demographic-cell level version of Equation (2.3), estimated with a different outcome variable. The coefficient $\hat{\eta}_d$ corresponds to the desired parameter.

2. **Estimating $\left(\frac{\partial c}{\partial m}\right)_d$.** We assume the change in stigma costs is proportionate to the change in the second-order stigma index measured from the online experiment. Because the experiment is small at the demographic cell level, we estimate one regression:

$$y_i = \beta\mathbb{1}(\text{high})_i + \gamma\mathbb{1}(\text{truth})_i + \mathbf{X}_i\boldsymbol{\delta} + \left(\mathbf{X}_i\mathbb{1}(\text{high})_i\right)\boldsymbol{\lambda} + \varepsilon_i. \tag{B.3}$$

Equation (B.3) linearly interacts the coefficient for several demographic groups (contained in $\mathbf{X}_i$) to obtain cell-level estimates of $\left(\frac{\partial c}{\partial m}\right)_d$ by summing the relevant entries of $\boldsymbol{\lambda}$ with β. To be concrete, X contains indicators for: female, white, age bins, income groups, and household size. To obtain at the demographic *cell* level, we sum the relevant coefficients for each cell. This approach is less flexible but more precise than fully saturating the model.

3. **Estimating s_d.** We use the FSPAS. We estimate the share of individuals affected by stigma (as the mean of the indicator variable) within each demographic cell, weighted using the FSPAS survey weights described above.

B.4.3 Additional details

Demographics. We focus on the following demographic variables: female/non-female, white/non-white, age group (0–18, 19–30, 31–65, 66+), household size (1, 2, 3, or 4), and income decile (grouping deciles 40–70 and 70–100). To construct demographic *cells*, we fully interact each variable. For instance, "single white women ages 19–30 in income decile 10" is an example of a demographic cell.

To focus on the group that is most affected by woodwork effects, our estimates of $\frac{\partial \ln(\text{N enrolled})}{\partial m}_d$ use the population between 50–115% FPL. We cannot precisely limit to this group in the experiment, but we limit that to less than 130% of the FPL.

Bootstrap. We employ a bootstrap to estimate standard errors. When bootstrapping Equation (B.2), we compute a *Bayesian Bootstrap* with weights drawn from Exponential(1). We use a Bayesian Bootstrap for (B.2) because otherwise smaller demographic cells were not drawn in some bootstraps. We use a standard bootstrap for Equations (B.3) and when estimating the share of individuals affected by stigma.

Boostrap bias correction and hypothesis testing. Bootstrap estimates in Table 2.5 are bootstrap-bias corrected using the following standard procedure. Consider any parameter θ that we bootstrap. Let $\hat{\theta}$ denote the estimate from the data. Let $\hat{\theta}^b$ denote the estimate from bootstrap b. Denote the mean estimate of θ across B bootstraps as $\bar{\hat{\theta}} := B^{-1} \sum_b \hat{\theta}^b$. The bias-adjusted coefficient we present is: $2\hat{\theta} - \bar{\hat{\theta}}$. We compute a standard error by taking the sample standard deviation of bootstrap coefficients. We compute p values by testing the bias-corrected coefficient against the normal distribution.

Precision weighting. The regression used for Figure 2.6 and Table 2.5 use previously estimated demographic subgroup effects. Because we estimate these effects with noise, the dispersion in the effects — and thus in the data used to estimate Equation B.1 — will be larger than the true variation. Moreover, effects estimated in small cells will be estimated less precisely than effects estimated in larger cells. To adjust for this, in the binned scatterplots, we weight by the inverse of the product of the variance of the estimates; i.e., we give more weight to cells that are more precisely estimated. In Table 2.5, we weight by the inverse of the product of the variance of the coefficients estimated from the listed datasets (columns 1 and 2) and show robustness to the variance of the coefficient from the estimated dataset (i.e., not the product) (columns 3–5).

B.5 Welfare Analysis Appendix

B.5.1 Structural Analysis

To solve for globally optimal solutions (under different parameter values) to the social planner's problem, we require assumptions about non-local behavior. Here, we provide details of these assumptions.

Welfare weights. We assume λ_θ is linear in m and satisfies the value of λ_θ/λ_m obtained from the inverse optimum exercise. We assume that $\lambda_0 = 1$. These two assumptions pin down a unique linear welfare weight schedule of inverse optimum welfare weights.

Linear take-up probabilities. We assume the take-up probability is linear in m. We assume a representative part-stigma part-information agent who obeys: $p^i(m) = p^i_0 + \frac{\partial p}{\partial m} m + s\frac{\partial p}{\partial B} B$.

Using the values for the elasticities η_m and η_B, we obtain the slope $\frac{\partial p}{\partial m} = \eta_m \frac{p^*}{m^*}$, which then gives η^{i*}_m using that $\eta^i_m = \eta_m(1-s)^{-1}$ by Assumption 2.2. We obtain $\frac{\partial p}{\partial B} = \eta_B \frac{p^*}{B^*}$.

Obtaining optimal m **and** B**.** From the planner's budget and take-up probability, we obtain the average equilibrium SNAP benefit B^*. We invert Equation (2.8), which, together with the linearity assumptions above, delivers a unique value of optimal m^{opt} and B^{opt}. Intuitively, this approach obtains the values of m and B that satisfy the planner's optimality conditions we derived in Section 2.4. We solve this problem numerically using Matlab.

B.5.2 Marginal Value of Public Funds Approach

As a related alternative, we consider the policy of expanding eligibility within the context of its Marginal Value of Public Funds (MVPF) (Hendren, 2016; Hendren and Sprung-Keyser, 2020). This approach lets us relax the assumption that today's policy constitutes the naïve solution to Equation (2.8). It also permits us to probe other assumptions about agents' behavior and utility.

In this framework, the planner considers the ratio of benefits (willingness to pay for the policy) to the net cost to the government. Because our focus is on the redistributive nature of the policy, we ultimately consider welfare-weighted MVPFs, i.e., welfare impacts (per dollar of government expenditure). Note that equating the welfare-weighted MVPF of raising the means test to that of raising the benefit size recovers our main optimality condition.

We analyze the size of the bias in the welfare-weighted MVPF, which we define as:

$$\text{bias} := 100 \times \frac{\bar{\lambda}^n \text{MVPF}^n - \bar{\lambda}^w \text{MVPF}^w}{\bar{\lambda}^w \text{MVPF}^w}, \tag{B.4}$$

where MVPF^n is the MVPF when the eligibility threshold does not affect inframarginal recipients and MVPF^w is the MVPF when it does. $\bar{\lambda}^n$ and $\bar{\lambda}^w$ correspond to the average welfare weights of the beneficiaries of the policy in each case (denoted $\bar{\eta}$ in Hendren and Sprung-Keyser (2020)). Let Δ be the size of the eligibility threshold increase; for instance,

$\Delta = 0.01$ when we study the welfare effect of letting 1 pp more people become eligible.

Derivation. The naïve welfare impact per dollar of government expenditure of an eligibility increase is: $\text{WI}^n := \bar{\lambda}^n \frac{\Delta p_m (1-\gamma s)\text{WTP}}{\Delta p_m \kappa_m} = \bar{\lambda}^n \frac{(1-\gamma s)\text{WTP}}{\kappa_m}$, with $\bar{\lambda}^n = \mu_m \lambda_m$, where μ_m is the marginal utility of income for newly eligible population. The denominator κ_m is the entire fiscal cost per person of providing food stamps to the next share Δ of the population (including the fiscal externality).[3] Assume the WTP to take up the benefit is the same for all θ.

Let WTP^s be the willingness to pay for the reduction in stigma costs from increasing the eligibility threshold to the next share Δ of the population. Let $\tilde{\alpha}$ satisfy $\text{WTP}^s := \tilde{\alpha}\text{WTP}$, where $\tilde{\alpha} < 1$, and μ_B and μ_0 are the marginal utilities of income of previously and newly enrolled, respectively.

Stigma agents who newly take up due to the inframarginal effect are just indifferent due to the Envelope Theorem.[4] Stigma agents who previously took up the benefit have a positive willingness to pay for the reduction in stigma costs. We assume that information agents who newly take up due to the inframarginal effect have full willingness to pay for the benefit. Regardless of type, individuals who are newly eligible for the program under an eligibility expansion see first-order utility gains; information agents again gain the full WTP, and stigma agents are willing to pay $(1-\gamma)$WTP.[5] Suppose there is a share s of stigma agents.

The sophisticated welfare impact per dollar of government expenditure is:

$$\text{WI}^w := \bar{\lambda}^w \frac{\overbrace{\Delta p_m (1 - s\gamma)\text{WTP}}^{\text{Mechanical effect for marginal types}} + \overbrace{\int_0^m \text{WTP}(1-s)w^i d\theta}^{\text{Inframarginal effect for info agents}} + \overbrace{\int_0^m \tilde{\alpha} p_{\text{avg}} s \text{WTP} d\theta}^{\text{Reduction in stigma costs}}}{\underbrace{\int_0^m \kappa_\theta w d\theta}_{\text{Cost of inframarginal effect}} + \underbrace{\Delta p_m \kappa_m}_{\text{Cost of expansion to marginal types}}} \tag{B.5}$$

with

$$\bar{\lambda}^w = \frac{\Delta \lambda_m \mu_m p_m (1 - s\gamma)\text{WTP} + \int_0^m \lambda_\theta \mu_\theta \text{WTP}(1-s)w^i d\theta + \int_0^m \mu_\theta^E \lambda_\theta \tilde{\alpha} p_{\text{avg}} s \text{WTP} d\theta}{\Delta p_m (1 - s\gamma)\text{WTP} + \int_0^m \text{WTP}(1-s)w^i d\theta + \int_0^m \tilde{\alpha} p_{\text{avg}} s \text{WTP} d\theta}$$

[3]We continue to assume that labor supply is fixed and abstract from bunching. To relax this assumption, one could assume the newly eligible are willing to pay only some fraction βWTP and follow this through. This would correspond to "bunchers" having lower WTP for the higher eligibility threshold, since they are already eligible via a distortion in their labor supply. However, note that this would also correspond to a lower κ_m.

[4]We assume that there are not utility gains to individuals who are not decision-makers (e.g., the children of SNAP recipients). Otherwise, while newly-enrolled, inframarginal stigma agents have no first-order welfare gains, there would be utility gains from their children.

[5]This is analogous to the γ in the model in Section 2.4; stigma agents face costs which erode some fraction of their WTP.

where μ_θ is the marginal utility of income for person θ who is not previously enrolled on the program, μ_θ^E is the marginal utility of income for person θ who is an 'always-taker,' t_{avg} is the take-up rate for inframarginal types prior to the eligibility expansion (which we assume is constant across all types), and κ_θ is the total fiscal cost of an additional 1 pp inframarginal take-up of type θ (including the fiscal externality). w is the proportion increase in inframarginal take-up (i.e., the inframarginal effect), and is a weighted average of the effect among information agents and the effect among stigma agents $w = (1-s)w^i + sw^s$.

Assume that all individuals who would newly enroll in the program have marginal utility of income $\mu_\theta = \mu_0$ before the policy change, and individuals who would were previously enrolled have marginal utility of income $\mu_\theta^E = \mu_B$. Define α as $\tilde{\alpha} * \frac{\mu_0}{\mu_B}$.

Noting that the bias can be written as $100 \times \frac{WI^n - WI^w}{WI^w}$, algebra gives:

$$\text{bias} = 100 \times \left(\frac{p_m + \frac{m}{\Delta} w \frac{\kappa_{\text{avg}}}{\kappa_m}}{p_m + \frac{m}{(1-s\gamma)\Delta} \frac{\lambda_{\text{avg}}}{\lambda_m} \left((1-s)w^i + s p_{\text{avg}} \alpha \right)} - 1 \right), \tag{B.6}$$

where $\kappa_{\text{avg}} := \frac{\int_0^m \kappa_\theta d\theta}{m}$, by analogy to λ_{avg}; p_m (p_{avg}) is the take-up rate for those newly eligible (previously eligible); w is the percentage point increase in the take-up rate for information-types (i.e., the inframarginal effect); κ_m (κ_{avg}) is the total fiscal cost of an additional 1 pp of take-up, including fiscal externalities, for those newly eligible (previously eligible); γ represents participation (stigma) costs as a share of WTP to take up the benefit; and α corresponds to the *reduction* in costs when the eligibility threshold rises, as a share of total WTP for the policy.[6]

Note also that, if $s = 0$, $w > 0$, and $\frac{\lambda_{\text{avg}}}{\lambda_m} > \frac{\kappa_{\text{avg}}}{\kappa_m}$, then bias < 0. Intuitively, as long as the planner's valuation of inframarginal types exceeds their fiscal cost, inframarginal effects raise the welfare impact of an eligibility increase. In the stigma case, the planner also values the reduction in costs to inframarginal types.

With this approach, we relax several assumptions imposed in Section 2.4. While our model in Section 2.4 defines the gains to inframarginal stigma agents in relation to the size of the inframarginal effects, this exercise decouples them (since α and w enter separately). This provides the flexibility to incorporate welfare gains from decreases in stigma costs even in the absence of an effect of stigma on take-up. Moreover, this expression permits inframarginal program participants to have different costs from participants who are newly eligible.[7] This is at the expense of additional assumptions (e.g., on the size of α), but it is easier to see robustness to those assumptions. We also emphasize the role of the

[6]The advantage of focusing on the proportion bias in the welfare impact is that the expression does not require an estimate of willingness to pay or separate estimates of the costs κ_{avg} and κ_m. The magnitudes of these costs are difficult to estimate, because SNAP involves many fiscal externalities that plausibly vary by type. This exercise permits us to conduct welfare analysis with only the *ratio* ($\kappa_{\text{avg}}/\kappa_m$).

[7]Note that while inframarginal types tend to have higher benefits, the higher benefit may yield a *reduced* fiscal externality because people with higher SNAP benefits receive better educations or are less likely to be incarcerated; Bailey et al. (2020) show that these benefits reduce the denominator of the MVPF for a benefit increase.

welfare weights here: the policy evaluation is as much about its incidence as it is about its utility gains and fiscal costs.

Parameters. We estimate w using instrumental variables, as in Table 2.4, where we instrument for the share of a state's population that is eligible using the eligibility threshold; here, however, we regress the take-up rate on the share eligible, instead of estimating an elasticity. The result is $w = 0.0028$: take-up increases by 0.28 percentage points for every 1 percentage point increase in the share eligible. From Section 2.3, we assume that $w^s = 0$ and $s = 0.4$, such that $w^i = 0.0047$. We continue to assume today's take-up rate, $p_{avg} = p_m = 0.53$, and eligibility threshold $m = 0.27$. We assume $\gamma = 0.5$, analogous to the calibration used elsewhere in the paper. Finally, we use $\frac{\lambda_{\text{avg}}}{\lambda_m}$ derived from inverse-optimum weights, although we note that these employ an assumption MVPFs usually relax — that current policy is optimal under a certain model. However, our results are robust to a range of values for $\frac{\lambda_{\text{avg}}}{\lambda_m}$.

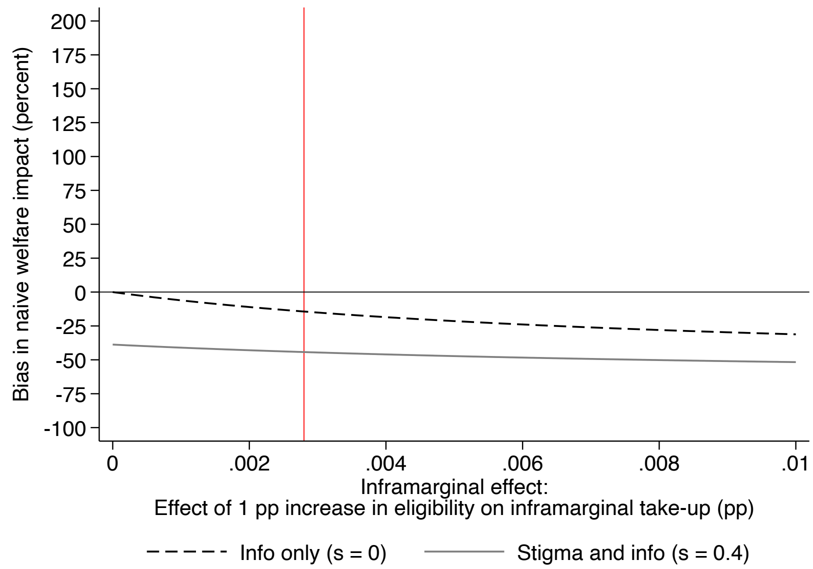
Results. We evaluate the welfare impact per dollar of government expenditure of expanding eligibility by 1 pp, i.e. we set $\Delta = 0.01$. To be conservative, we assume that the willingness to pay for a reduction in stigma costs is small, so we set $\alpha = 0.02$.

We find that, if $\frac{\kappa_{\text{avg}}}{\kappa_m} = 1$, the naïve MVPF can be about 20% below the sophisticated MVPF for the information-only case, with even larger results in the information and stigma case (Figure B.13A). However, the planner may overvalue the welfare impact for larger values of $\frac{\kappa_{\text{avg}}}{\kappa_m}$, say $\frac{\kappa_{\text{avg}}}{\kappa_m} = 10$ (Figure B.13B). This is because with $\frac{\kappa_{\text{avg}}}{\kappa_m} \gg 1$, the cost of new participants who are costly may exceed their value to the planner. Hence the naïve planner sets the eligibility threshold *too high*.[8]

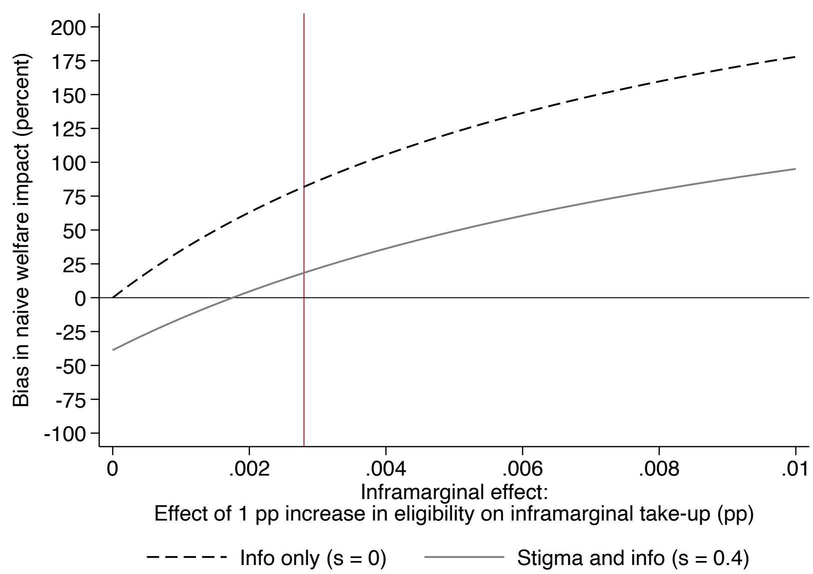
[8]Note that if $\frac{\kappa_{\text{avg}}}{\kappa_m} \gg 1$, the MVPF bias is negative for the stigma and information case ($s = 0.5$) and positive for the information-only case ($s = 0$). Here, unlike in the model, the normative conclusion that the planner may wish to raise the eligibility threshold can be stronger if there is stigma.

Figure B.13: Welfare Bias of an Eligibility Expansion using MVPF Framework

(a) $\frac{\kappa_{\text{avg}}}{\kappa_m} = 1$



(b) $\frac{\kappa_{\text{avg}}}{\kappa_m} = 10$

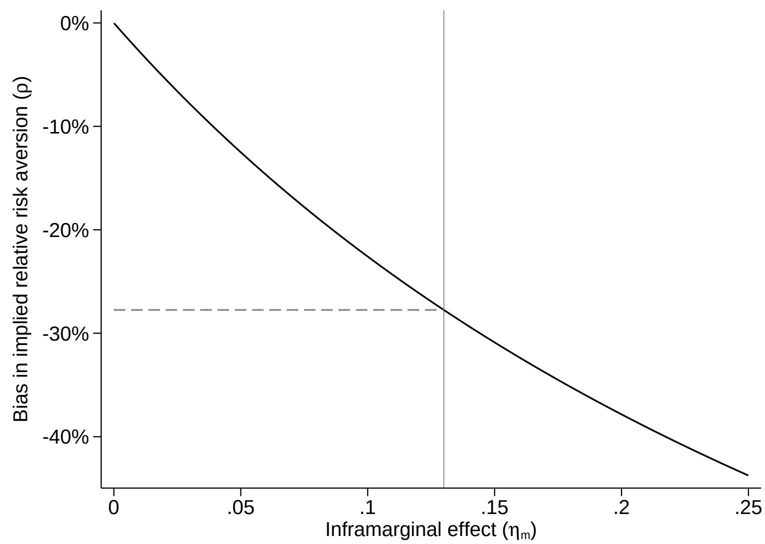


This figure shows the percent bias in the welfare impact per dollar of government expenditure (Equation (B.6)) for an inframarginal to marginal cost ratio of 1 (Panel A) and 10 (Panel B). The vertical red line plots our preferred estimate of the inframarginal effect w in terms of take-up.

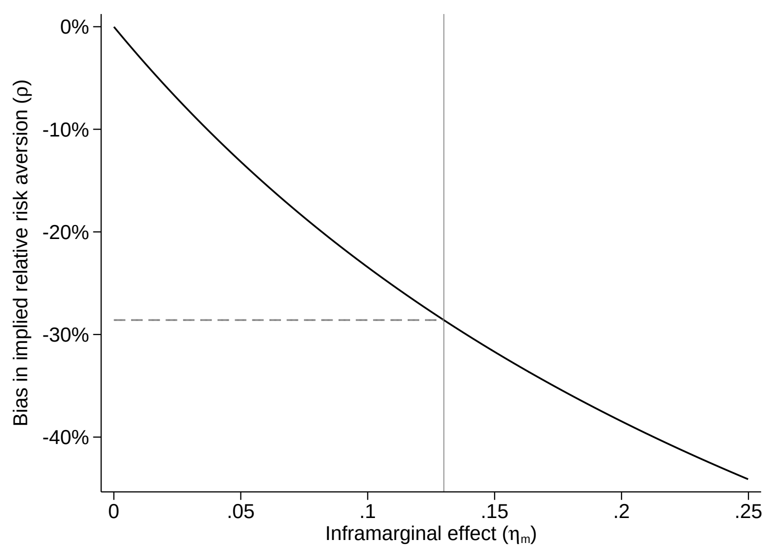
B.5.3 Robustness

Figure B.14: Robustness: Naïve Planner's Biased Risk Aversion

(a) True $\rho = 2$



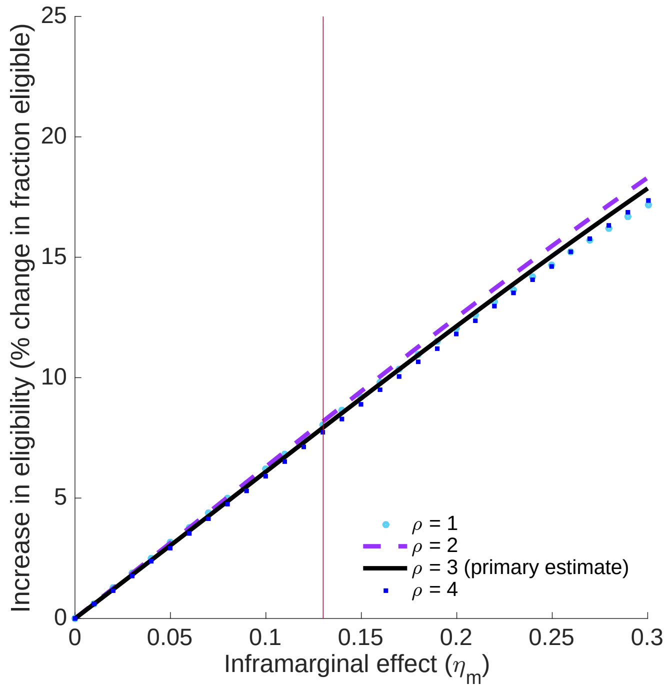
(b) $\eta_B = 0.3$



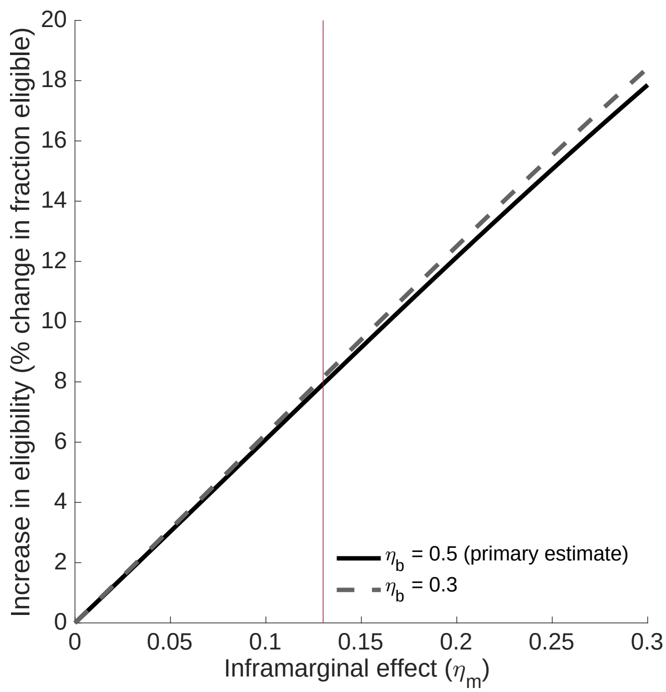
This figure shows the percent bias between the planner's "as-if" risk aversion ($\tilde{\rho}$) and the ground-truth risk aversion (ρ) (black line). It is identical to Figure 2.7A except it sets $\rho = 2$ (Panel A) or $\eta_B = 0.3$ (Panel B). Negative numbers indicate that the planner is behaving as if people are less risk averse than they really are. Panel A plots the bias as a function of the inframarginal effect; the vertical gray line plots the empirical inframarginal effect presented in Table 2.6.

Figure B.15: Numerical Simulations: Robustness

(a) Varying the Coefficient of Relative Risk Aversion (ρ)

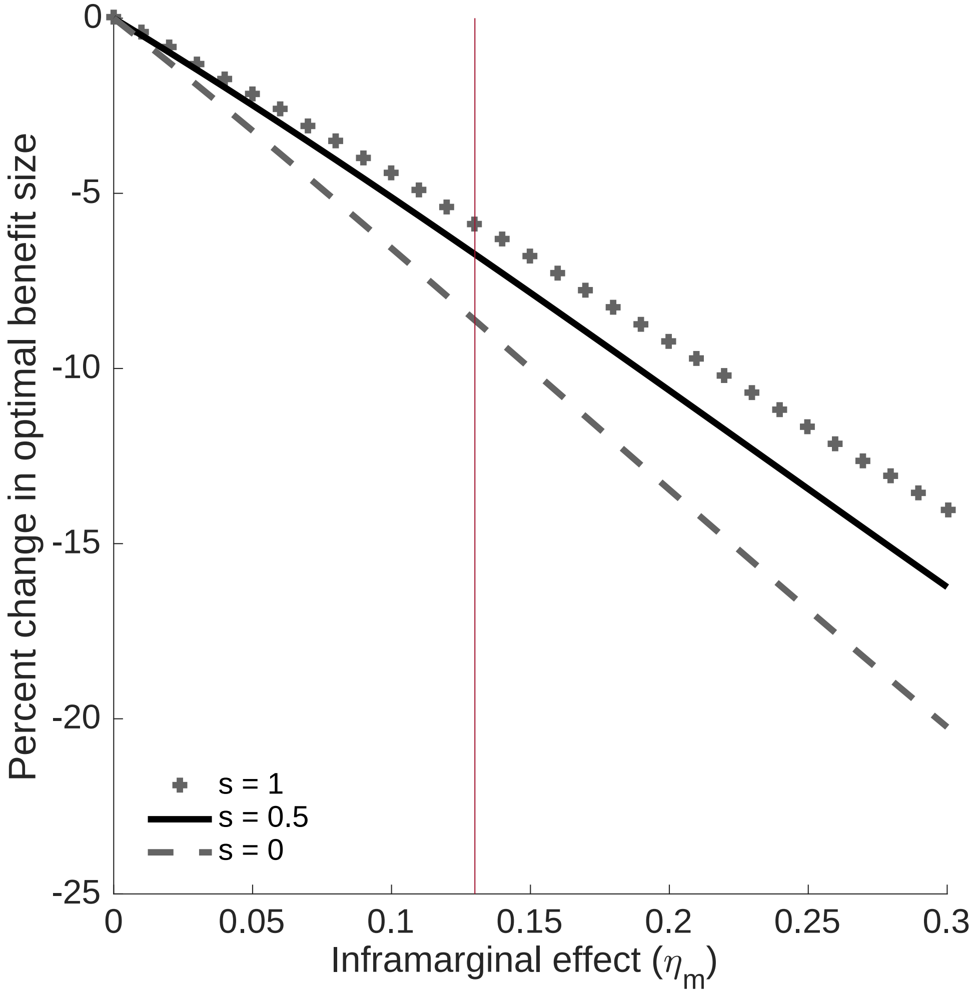


(b) Varying Take-Up Elasticity with Respect to Benefit Size (η_B)



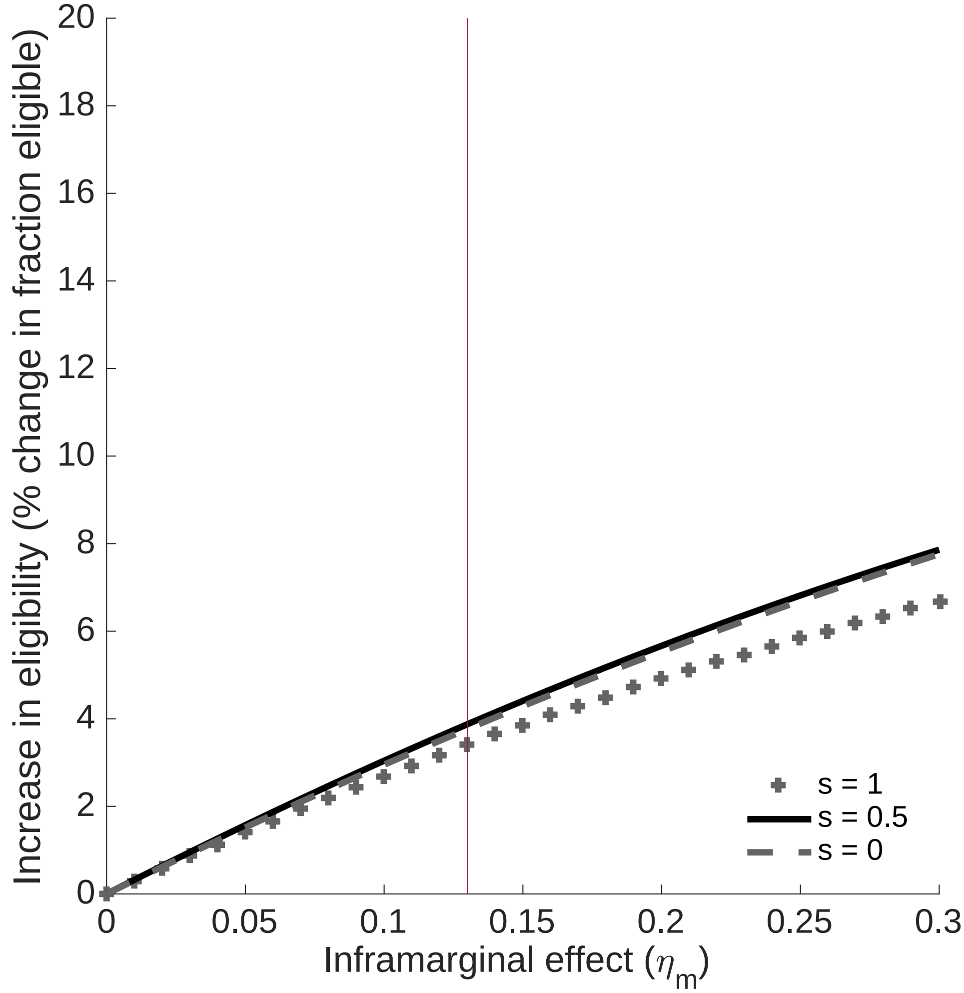
This figure shows the results from our numerical simulation exercise, which uses the optimality condition in Equation (2.8). It presents changes in the percent of people who are eligible if the planner acknowledges inframarginal effects. It shows robustness to different ρ (Panel A) and take-up elasticities with respect to the benefit size (η_B). Auxiliary parameters are set according to the values in Table 2.6.

Figure B.16: Numerical Simulations: Optimal Benefit Size



This figure shows the results from our numerical simulation exercise, which uses the optimality condition in Equation (2.8). It presents the percent change in the optimal benefit size if the planner acknowledges inframarginal effects. Auxiliary parameters are set according to the values in Table 2.6.

Figure B.17: Numerical Simulations: Robustness to Quadratic Utility



This figure shows the results from our numerical simulation exercise, which uses the optimality condition in Equation (2.8). It presents the change in the percent of people who are eligible if the planner acknowledges inframarginal effects. It is identical to Figure 2.8A except the simulations impose quadratic utility with $\rho = 3$ at equilibrium, using Equation (2.1) with $\eta_m = 0$ to infer the welfare weights.

B.6 Proofs

B.6.1 Proofs of Propositions 2.1 and 2.2.

Proof. Note that Proposition 2.1 is a special case of Proposition 2.2. We therefore prove Proposition 2.2 only.

The planner's problem is:

$$\max_{B,m} \left[sp^s(B,m) \left(\int_0^m \lambda_\theta u(B) d\theta - \int_0^m \int_{c \leq u(B)} \lambda_\theta c h(c|c < u(B), m) dc d\theta \right) \right. \tag{B.7}$$
$$\left. + (1-s) \left(\int_0^m \lambda_\theta u(B) p^i(m) d\theta \right) \right]$$

subject to

$$(1-s)p^i(m) \int_0^m B d\theta + sp^s(B,m) \int_0^m B d\theta \leq T \tag{B.8}$$
$$m \in [0,1] \tag{B.9}$$

We inspect interior solutions using the Karush-Kuhn-Tucker conditions where the constraint $m \in [0,1]$ is slack. We consider cases in which such an interior solution exists; there are possible corner solutions where $m = 1$ (i.e., the program is universal). Proposition 2.1 and 2.2 give necessary conditions for local optimality. To obtain that the statement in the proposition is sufficient for a global maximum, it is sufficient to additionally impose that the maximand is concave and the constraint is convex.

The first-order condition for B is:

$$s \left(\frac{\partial p^s}{\partial B} \lambda_{\text{avg}} m u(B) + p^s(B,m) u'(B) \lambda_{\text{avg}} m - \left(\frac{\partial}{\partial B} \int_0^m \int_{c \leq u(B)} c \lambda_\theta h(c|m) dc d\theta \right) \right)$$
$$+ (1-s) p^i(m) u'(B) \lambda_{\text{avg}} m = \sigma \left((1-s) p^i(m) m + s \frac{\partial p^s}{\partial B} B m + s p^s m \right), \tag{B.10}$$

where σ denotes the Lagrange multiplier, and we note that

$$\int_0^m \int_{c \leq u(B)} \lambda_\theta h(c|c < u(B), m) dc d\theta = \frac{1}{H(u(B)|m)} \int_0^m \int_{c \leq u(B)} \lambda_\theta h(c|m) dc d\theta. \tag{B.11}$$

Leibniz's rule gives that:

$$\frac{\partial}{\partial B}\int_0^m \int_{c\le u(B)} c\lambda_\theta h(c|m)dcd\theta = u(B)\lambda_{\text{avg}} m h(u(B)|m)u'(B) \tag{B.12}$$

$$= \lambda_{\text{avg}} m u(B)\frac{\partial p^s}{\partial B}. \tag{B.13}$$

We collect terms to obtain:

$$\lambda_{\text{avg}} m(sp^s + (1-s)p^i)u'(B) = \sigma\left(s\frac{\partial p^s}{\partial B}Bm + sp^s m + (1-s)p^i m\right). \tag{B.14}$$

We divide by $p^s m$ and rearrange, recalling that $\eta_B = \frac{\partial p^s}{\partial B}\frac{B}{p^s}$:

$$\frac{sp^s + (1-s)p^i}{p^s}u'(B)\lambda_{\text{avg}} = \sigma\left(s\eta_B + \frac{sp^s + (1-s)p^i}{p^s}\right). \tag{B.15}$$

Next we take the first-order condition with respect to m and use the shorthand $E := E[c|c < u(B)]$ to be succinct:

$$\begin{aligned} &s\frac{\partial p^s}{\partial m}\left(\lambda_{\text{avg}} m u(B) - \lambda_{\text{avg}} m E\right) + sp^s\left(\lambda_m u(B) - \lambda_m E - \lambda_{\text{avg}} m\frac{\partial E}{\partial m}\right) \\ &+ (1-s)\left(\frac{\partial p^i}{\partial m}\lambda_{\text{avg}} m u(B) + p^i\lambda_m u(B)\right) = \sigma\left((1-s)\left(\frac{\partial p^i}{\partial m}Bm + p^i B\right) + s\left(\frac{\partial p^s}{\partial m}Bm + p^s B\right)\right). \end{aligned} \tag{B.16}$$

Noting that $\frac{\partial E}{\partial m} = \frac{\partial \gamma}{\partial m}u(B)$, we collect terms to obtain:

$$\begin{aligned} &u(B)\lambda_{\text{avg}}\left((1-\gamma)s\frac{\partial p^s}{\partial m}m + (1-s)\frac{\partial p^i}{\partial m}m - sp^s\frac{\partial \gamma}{\partial m}m\right) + \lambda_m u(B)(1-\gamma)sp^s \\ &+ \lambda_m u(B)p^i(1-s) = \sigma\left((1-s)\left(\frac{\partial p^i}{\partial m}Bm + p^i B\right) + s\left(\frac{\partial p^s}{\partial m}Bm + p^s B\right)\right). \end{aligned} \tag{B.17}$$

We divide by p^s and p^i to get:

$$\begin{aligned} &u(B)\lambda_{\text{avg}}\left(\frac{(1-\gamma)s\eta_m^s}{p^i} + \frac{(1-s)\eta_m^i}{p^s} - \frac{s\frac{\partial\gamma}{\partial m}m}{p^i}\right) \\ &+ s\frac{\lambda_m u(B)(1-\gamma)}{p^i} + \frac{\lambda_m u(B)(1-s)}{p^s} = B\sigma\left((1-s)\frac{\eta_m^i + 1}{p^s} + s\frac{\eta_m^s + 1}{p^i}\right). \end{aligned} \tag{B.18}$$

Then, substituting for σ from Equation (B.15) and rearranging gives:

$$\frac{u(B)\left(\frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i}+\frac{1-s}{p^s}\eta_m^i+\frac{1-s}{p^s}\right)+\frac{s}{p^i}[(1-\gamma)\eta_m^s-m\frac{\partial\gamma}{\partial m}]\right)}{u'(B)}=\frac{B\left[\frac{(1-s)}{p^s}\left(\eta_m^i+1\right)+\frac{s}{p^i}\left(\eta_m^s+1\right)\right]}{\left(\frac{p^s}{sp^s+(1-s)p^i}s\eta_B+1\right)}. \tag{B.19}$$

At this point, rearranging terms and using a Taylor Expansion (see Section B.6.2) produces the result in Equation 2.10.

To produce Equation 2.8, we invoke the following lemma:

Lemma B.1. $\eta_m(1-\gamma)-m\frac{\partial\gamma}{\partial m}=\frac{\frac{m}{p(B,m)}\int_0^{u(B)}\frac{\partial H(c|m)}{\partial m}dc}{u(B)}$.

The proof is below. Now we can rewrite the expression as:

$$\frac{\frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i}+\frac{1-s}{p^s}\right)u(B)+\frac{1-s}{p^s}\eta_m^i u(B)+\frac{s}{p^i}\frac{m}{p^s}\int_0^{u(B)}\frac{\partial H(c|m)}{\partial m}dc}{u'(B)}=\frac{B\left[\frac{(1-s)}{p^s}\left(\eta_m^i+1\right)+\frac{s}{p^i}\left(\eta_m^s+1\right)\right]}{\left(\frac{p^s}{sp^s+(1-s)p^i}s\eta_B+1\right)} \tag{B.20}$$

An important special case is where $p^i=p^s$ and $\frac{\partial p^i}{\partial m}=\frac{\partial p^s}{\partial m}$. Then, multiplying both the numerator and denominator by $p:=p^i=p^s$, and noting that $\eta_m^i=\eta_m^s$, we get:

$$\frac{u(B)}{Bu'(B)}=\frac{\eta_m+1}{(s\eta_B+1)\left(\left(\eta_m+\frac{\lambda_m}{\lambda_{\text{avg}}}\right)(1-s\gamma)-s\frac{\partial\gamma}{\partial m}m\right)}. \tag{B.21}$$

□

B.6.1.1 Proof of Lemma B.1

Proof. Multiplying both sides by $u(B)$ and recalling that $\frac{\partial E}{\partial m}(c|c<u(B),m)=\frac{\partial\gamma}{\partial m}u(B)$, it suffices to show that $\eta_m u(B)(1-\gamma)-m\frac{\partial E[c|c\leq u(B),m]}{\partial m}=\frac{m}{p(B,m)}\int_0^{u(B)}\frac{\partial H(c|m)}{\partial m}dc$. Below, we show that $m\frac{\partial E[c|c\leq u(B),m]}{\partial m}=\eta_m(u(B)-E[c|c\leq u(B),m])-\frac{m}{p(B,m)}\int_0^{u(B)}\frac{\partial H(c|m)}{\partial m}dc$, which completes the proof.

$$m\frac{\partial E[c|c\leq u(B),m]}{\partial m}=m\int_0^{u(B)}c\frac{\partial h(c|m,c<u(B))}{\partial m}dc \tag{B.22}$$

$$= m \int_0^{u(B)} c \frac{\partial}{\partial m} \left(\frac{h(c|m)}{H(u(B)|m)} \right) dc \tag{B.23}$$

$$= m \int_0^{u(B)} c \left(\frac{\frac{\partial h(c|m)}{\partial m} H(u(B)|m) - h(c|m) \frac{\partial H(u(B)|m)}{\partial m}}{H(u(B)|m)^2} \right) dc \tag{B.24}$$

$$= m \left(\frac{1}{H(u(B)|m)} \int_0^{u(B)} c \frac{\partial h(c|m)}{\partial m} dc - \frac{1}{H(u(B)|m)^2} \int_0^{u(B)} ch(c|m) \frac{\partial p^s}{\partial m} dc \right), \tag{B.25}$$

where p^s is the take-up rate among stigma agents ($p^s = H(u(B)|m)$, the share of stigma agents with stigma costs below the utility benefits of take-up). We apply integration by parts to the first integral. We apply that $\frac{\int_0^{u(B)} chdc}{H(u(B)|m)} = E[c|c < u(B), m]$ to the second integral. Suppressing arguments of h and H to be concise, this yields:

$$m \left(\frac{1}{H} \int_0^{u(B)} c \frac{\partial h}{\partial m} dc - \frac{1}{H^2} \int_0^{u(B)} ch \frac{\partial p^s}{\partial m} dc \right) \tag{B.26}$$

$$= m \left(\frac{1}{H} \left(u(B) \frac{\partial H(u(B)|m)}{\partial m} - \int_0^{u(B)} \frac{\partial H(c|m)}{\partial m} dc \right) - \frac{\frac{\partial p^s}{\partial m}}{p^s} E[c|c < u(B), m] \right) \tag{B.27}$$

$$= \eta_m^s \left(u(B) - E[c|c < u(B), m] \right) - \frac{m}{p^s(B, m)} \int_0^{u(B)} \frac{\partial H(c|m)}{\partial m} dc, \tag{B.28}$$

recalling that $\frac{\partial p^s}{\partial m} \frac{m}{H(u(B)|m)} = \eta_m^s$. □

B.6.2 Proof of Taylor Expansion (Equation (2.8)).

Proof. Throughout the paper, we use the second-order Taylor approximation:

$$u(0) = 0 \approx u(B) - u'(B)B + \frac{u''(B)B^2}{2}, \tag{B.29}$$

which gives

$$u(B) \approx u'(B)B - \frac{u''(B)B^2}{2}. \tag{B.30}$$

We then obtain

$$\frac{u(B)/B}{u'(B)} \approx 1 + \frac{\rho}{2}. \tag{B.31}$$

Note that $u'(B) = \frac{\partial}{\partial B} (u(B) - c)$, so ρ represents the coefficient of relative risk aversion for people who would take up the program if informed. □

B.6.3 Lemma B.2 and Proof

Subsequent proofs invoke the following lemma:

Lemma B.2. *If* $\rho \geq 1$, $\frac{\partial}{\partial B}\left(\frac{u(B)/B}{u'(B)}\right) > 0$.

Proof. The quotient rule gives

$$\frac{\partial}{\partial B}\left(\frac{u(B)/B}{u'(B)}\right) > 0 \tag{B.32}$$

iff

$$(u'(B))^2 B - u(B)u'(B) - Bu''(B)u(B) > 0. \tag{B.33}$$

Dividing by $u'(B)$ (which is always greater than 0), we conclude that the left-hand side is always positive as long as

$$u'(B)B + u(B)(\rho - 1) > 0, \tag{B.34}$$

which completes the proof. □

B.7 Theory Extensions

B.7.1 Endogenous labor supply

Section 2.4 develops a proposition that gives that B and m satisfy

$$\frac{u(B)}{u'(B)B} = \frac{1+\eta_m}{\lambda_m/\lambda_{\text{avg}} + \eta_m}, \tag{B.35}$$

if $\eta_B = 0$ and $s = 0$.

We show how this expression can be microfounded in a more elaborate environment with endogenous labor supply. We focus on this parsimonious expression, nested by the more general case, for simplicity; this analysis captures many of the relevant insights.

Model environment. There is a continuum of types $\theta \sim F$, where F has support Θ. People earn labor income y from hours worked h, depending on their type θ. Let labor income $y = \theta h(\theta)$; we use this parametric form for simplicity, but the model can easily be generalized. People with labor income below r (the "eligibility threshold") earn a benefit B. People have utility $\tilde{v}(h, B, \theta)$ over labor supply and the benefit amount.[9] This utility

[9] We can think of utility over the benefit as the indirect utility of the agent's inner problem of allocating the benefit to consumption of various goods.

function induces an indirect utility function v over labor supply, the benefit amount, and the eligibility threshold:

$$v(h^*(B,r,\theta),B,r,\theta) = \max_h \begin{cases} \tilde{v}(h,B,\theta) & \theta h(\theta) \leq r \\ \tilde{v}(h,0,\theta) & \theta h(\theta) > r. \end{cases} \tag{B.36}$$

The Envelope Theorem gives the following intermediate results, which we will invoke later:

$$\frac{dv}{dB}(h^*(B,r,\theta),B,r,\theta) = \frac{\partial v}{\partial B} \tag{B.37}$$

$$\frac{dv}{dr}(h^*(B,r,\theta),B,r,\theta) = 0 \text{ if } h^*(B,r,\theta) \neq r/\theta \tag{B.38}$$

Equation (B.38) states that if the benefit constraint does not bind, there is no value to the agent to relaxing the constraint. Intuitively, for people who are very poor or very rich, adjustments to the eligibility threshold have no effect on behavior. However, the existence of a lump-sum benefit and discrete eligibility threshold can induce bunching at the threshold. A small change in eligibility will have first-order effects on utility for bunchers.

Take-up probabilities. Agents are aware of the program with probability $p(r)$ and get $v(h^*(B,r,\theta),B,r)$ if they take up. Otherwise they optimize as if the program does not exist, do not take up the program, and get $v(h^*(0,r,\theta);0,r,\theta)$ (the "outside option"). Moreover, this outside option does not depend on r: $v(h^*(0,r,\theta);0,r,\theta) = v(h^*(0,\theta);0,\theta)$ for all r.

Planner's problem. We begin with a technical assumption. Assume that income $\theta h^*(B,r,\theta)$ is weakly increasing in θ: higher types always earn weakly more labor income even though the existence of the benefit distorts labor supply. This assumption amounts to a standard single-crossing condition: even if the tax system affects labor supply or causes bunching, it will not cause high types to earn strictly less income than low types (or vice-versa).

This assumption yields a threshold type $\tilde{\theta}(B,r)$ such that all $\theta \leq \tilde{\theta}$ will choose a labor supply that is low enough that they will be eligible for the benefit. All types $\theta > \tilde{\theta}$ are not eligible.

Next, we assume that the planner has a budget T which depends on the amount of money raised through taxes on labor income. Assume the income tax schedule is exogenous, but make no other restrictions on this schedule. In that case, we can parameterize T as depending on B and r alone: $T(B,r)$.[10]

Altogether, the planner's problem is:

$$\max_{r,B}\int_0^{\tilde{\theta}(B,r)}\lambda_\theta p(r)v(h^*(B,r,\theta);B,r,\theta)f(\theta)d\theta+\int_0^{\tilde{\theta}(B,r)}\lambda_\theta(1-p(r))v(h^*(0,r,\theta);0,\theta)f(\theta)d\theta+$$
$$\int_{\tilde{\theta}(B,r)}^{\infty}\lambda_\theta v(h^*(0,r,\theta);0,\theta)f(\theta)d\theta \tag{B.39}$$

subject to

$$\int_0^{\tilde{\theta}(B,r)}p(r)Bf(\theta)d\theta\leq T(B,r). \tag{B.40}$$

Noting that $\int_0^\infty \lambda_\theta v(h^*(0);0,\theta)f(\theta)d\theta$ is a constant, we can re-write the planner's problem as:

$$\max_{r,B}\int_0^{\tilde{\theta}(B,r)}\lambda_\theta p(r)\left(v(h^*(B,r,\theta);B,r,\theta)-v(h^*(0,r,\theta);0,\theta)\right)f(\theta)d\theta \tag{B.41}$$

subject to

$$\int_0^{\tilde{\theta}(B,r)}p(r)Bf(\theta)d\theta\leq T(B,r). \tag{B.42}$$

Then, let $V(h^*(B,r,\theta);B,r,\theta):=v(h^*(B,r,\theta);B,r,\theta)-v(h^*(0,r,\theta);0,r,\theta)$ be the net utility gain from taking up the program. Note that for types $\theta>\tilde{\theta}$, $V=0$: these types choose labor supply that renders them ineligible for the benefit. For other types, $\theta\leq\tilde{\theta}$, $V>0$ assuming they earn positive utility from the benefit.

Solving for the optimum. Letting σ represent the Lagrange multiplier, take the first-order condition with respect to r:

$$\frac{\partial\tilde{\theta}}{\partial r}\left(\lambda_{\tilde{\theta}}p(r)V(h^*(B,r,\tilde{\theta});B,r,\tilde{\theta})f(\tilde{\theta})\right)$$

[10]Formally, let

$$I(B,r):=\{(\theta h^*(B,r,\theta),\theta h^*(0,r,\theta),\theta):\theta\in\Theta\}.$$

Here, I is the set of triples of: (i) labor incomes chosen if a given type θ receives the benefit, (ii) labor incomes chosen if the type θ does not receive the benefit, and (iii) the type θ, which then yields a density $f(\theta)$ and a labor supply $h^*(B,r,\theta)$. These values uniquely determine the taxes raised for a generic tax schedule that only depends on labor income, even if there is incomplete take-up of the benefit, assuming the planner knows $p(r)$. This notation shows that we can write $T(B,r)=T(I(B,r))$. Intuitively, holding F fixed, any (B,r) pair induces a distribution of labor incomes chosen across types.

$$+\int_0^{\tilde{\theta}} \lambda_\theta \left(\frac{dp}{dr} V(h^*(B,r,\theta);B,r,\theta) + p\left(\underbrace{\frac{dV}{dr} \cdot \mathbb{1}(\theta = \tilde{\theta})}_{\text{As } \frac{dV}{dr}=0 \text{ otherwise, by Equation (B.38)}} \right)\right) f(\theta)d\theta$$

$$-\sigma\left(\frac{\partial\tilde{\theta}}{\partial r}(pBf(\tilde{\theta})) + \frac{dp}{dr}BF(\tilde{\theta}) - \frac{dT}{dr}\right) = 0. \tag{B.43}$$

Take the first-order condition with respect to B:

$$\frac{\partial\tilde{\theta}}{\partial B}\left(\lambda_{\tilde{\theta}} pV(h^*(B,r,\tilde{\theta});B,r,\tilde{\theta})f(\tilde{\theta})\right) + \int_0^{\tilde{\theta}} \lambda_\theta p(r) \underbrace{\frac{\partial V}{\partial B}}_{=\frac{dV}{dB}, \text{ by Equation (B.37)}} f(\theta)d\theta -$$

$$\sigma\left(\frac{\partial\tilde{\theta}}{\partial B}p(r)Bf(\tilde{\theta}) + p(r)F(\tilde{\theta}) - \frac{dT}{dB}\right) = 0. \tag{B.44}$$

Solving for σ, we obtain:

$$\sigma = -\frac{\frac{\partial\tilde{\theta}}{\partial B}\lambda_{\tilde{\theta}} p(r)V(h^*(B,r,\tilde{\theta});B,r,\theta)f(\tilde{\theta}) + \int_0^{\tilde{\theta}}\lambda_\theta p(r)\frac{\partial V}{\partial B}f(\theta)d\theta}{\frac{\partial\tilde{\theta}}{\partial B}p(r)Bf(\tilde{\theta}) + p(r)F(\tilde{\theta}) - \frac{dT}{dB}} \tag{B.45}$$

Plugging into Equation (B.43) yields:

$$\overbrace{\frac{\partial\tilde{\theta}}{\partial r}\left(\lambda_{\tilde{\theta}} p(r)V(h^*(B,r,\tilde{\theta});B,r,\tilde{\theta})f(\tilde{\theta})\right)}^{\text{Value of } r\uparrow \text{ to otherwise ineligible people}} \tag{B.46}$$

$$+\int_0^{\tilde{\theta}}\lambda_\theta\left(\underbrace{\frac{dp}{dr}V(h^*(B,r,\theta);B,r,\theta)}_{\text{Effect of } r\uparrow \text{ on take-up of inframarginals}} + \overbrace{p(r)\left(\frac{dV}{\partial r}\cdot\mathbb{1}(\theta=\tilde{\theta})\right)}^{\text{Value of } r\uparrow \text{ to bunchers}}\right) f(\theta)d\theta$$

$$-\left(\underbrace{\frac{\partial\tilde{\theta}}{\partial B}\lambda_{\tilde{\theta}} p(r)V(h^*(B,r,\theta);B,r,\theta)f(\tilde{\theta})}_{\text{Value of } B\uparrow \text{ to bunchers}} + \overbrace{\int_0^{\tilde{\theta}}\lambda_\theta p(r)\frac{\partial V}{\partial B}f(\theta)d\theta}^{\text{Value of } B\uparrow \text{ to inframarginals}}\right)$$

$$
\times \frac{\underbrace{\frac{\partial \tilde{\theta}}{\partial r}(pBf(\tilde{\theta}))}_{\text{Mechanical cost of } r \uparrow} + \overbrace{\frac{dp}{dr}BF(\tilde{\theta})}^{\text{Indirect cost of } r \uparrow \text{ from changes in take-up}} - \underbrace{\frac{dT}{dr}}_{\text{Indirect cost of } r \uparrow \text{ from labor supply}}}{\underbrace{\frac{\partial \tilde{\theta}}{\partial B}p(r)Bf(\tilde{\theta})}_{\text{Indirect cost of } B \uparrow \text{ from changes in take-up}} + \overbrace{p(r)F(\tilde{\theta})}^{\text{Mechanical cost of } B \uparrow} - \underbrace{\frac{dT}{dB}}_{\text{Indirect cost of } B \uparrow \text{ from labor supply}}} = 0. \tag{B.47}
$$

Discussion. Equation (B.47), while involved, captures the following intuitions. At an optimum, the planner equates the following trade-offs.

- Raising r has benefits. First, it brings in more people to the program who were previously ineligible. Second, it has the value of raising take-up among inframarginal types. Third, it has a direct effect on welfare for people who bunch at the eligibility threshold, who can then adjust their labor supply (which was not necessarily at an optimum).

- Raising B has benefits. First, it brings value by affecting bunching. Second, it also has value to inframarginal types who take up the program, because it is a transfer.

- Raising r has costs. First, there is a mechanical cost of bringing more people into the program because more people are eligible. Second, there is an indirect cost of raising take-up. Third, there is an indirect cost of changing people's labor supply, which then affects the income taxes collected.

- Raising B has costs. First, there is a mechanical cost of raising the transfer to people who take up the program. Second, there is an indirect cost of bringing more people into the program via changes in labor supply. Third, there is an indirect cost of changing people's labor supply, which then affects the income taxes collected.

B.7.1.1 Simplifications

In this subsection, we show how this more general solution nests the solution in the paper.

First, we apply a change of units. Instead of considering raising the eligibility threshold by one dollar of labor income, we raise the eligibility threshold by one quantile of the population that is eligible. Let m represent the share who is eligible for the benefit: $m := F(\tilde{\theta})$.

Use the chain rule to observe that:

$$\frac{\partial p}{\partial r} = \frac{\partial p}{\partial F(\tilde{\theta})}\frac{\partial F(\tilde{\theta})}{\partial \tilde{\theta}}\frac{\partial \tilde{\theta}}{\partial r} = \frac{\partial p}{\partial m} f(\tilde{\theta})\frac{\partial \tilde{\theta}}{\partial r}. \tag{B.48}$$

Next, we invoke the following assumption:

Assumption S1: No Bunching. Assume that $h^*(\cdot) = \bar{h}(\theta)$ for all B, r, i.e. that the amount of labor supply chosen depends only on one's type.

This assumption has three implications. First, $\int_0^{\tilde{\theta}} \mathbb{1}(\theta = \tilde{\theta}) f(\theta) = 0$, assuming there are no atoms in the type distribution. Second, $\frac{\partial \tilde{\theta}}{\partial B} = 0$. Third, since labor supply is constant for all θ and the budget T only depends on r and B via h, $\frac{dT}{dr} = \frac{dT}{dB} = 0$.

As a result, employing the No Bunching assumption and dividing by $f(\tilde{\theta})\frac{\partial \tilde{\theta}}{\partial r}$ gives:

$$\begin{aligned} &\lambda_{\tilde{\theta}} p(r) V(h^*(B,r,\tilde{\theta}); B, r, \tilde{\theta}) + \int_0^{\tilde{\theta}} \lambda_\theta \frac{\partial p}{\partial m} V(h^*(B,r,\theta), B, r, \theta) f(\theta) d\theta \\ &= \left(\int_0^{\tilde{\theta}} \lambda_\theta \frac{\partial V}{\partial B} f(\theta) d\theta \right) \left(pB + \frac{\partial p}{\partial m} B F(\tilde{\theta}) \right) \frac{1}{F(\tilde{\theta}))}. \end{aligned} \tag{B.49}$$

Finally, under the No Bunching assumption, observe that for any fixed (B, r) pair, there exists $\kappa(\theta; B, r)$ such that

$$V(h(B,r,\theta); B, r, \theta) = \kappa(\theta; B) u(B)$$

for some function $\kappa(\theta)$. Put otherwise, because labor supply is fixed at $\overline{h}(\theta)$, r has no effect on utility independently of the benefit B and the type θ. Moreover, fixing B, we can always rescale utility for each type by multiplying by a real number $\kappa(\theta; B)$.

We assume that the function $\kappa(\theta)$ holds locally for all B in a neighborhood of the solution, for all types that are eligible for the benefit:

Assumption S2: Multiplicative Separability. Suppose $V(B, r, \theta) = \kappa(\theta) u(B)$ for all B in a neighborhood of B^* and for all $\theta \leq \tilde{\theta}$.

The Multiplicative Separability assumption states that utility gains from the benefit can be multiplicatively rescaled by the schedule $\kappa(\theta)$. Note that this assumption always holds if utility is homogeneous across types and all types have the same outside option;

in that case, $\kappa(\theta) = 1$ for all θ. In the body of the paper, we start directly from that more demanding homogeneity assumption.

For other utility functions, the assumption holds as long as slight changes to the benefit around the optimum do not change the relative differences in the net utility that the different types experience from receiving the benefit. These relative differences are parameterized by the $\kappa(\theta)$ schedule, which must be invariant around the optimum. This assumption fails if, e.g., high types' marginal utility from receiving B diminishes at a faster rate than low types' marginal utility even in a neighborhood around the optimum.

The Multiplicative Separability assumption permits us to rescale differences in net utility with a (re-written) λ_θ welfare weight schedule.

Define $\tilde{\lambda}_\theta := \lambda_\theta \kappa(\theta)$. Moreover, let

$$\tilde{\lambda}_{\text{avg}} = \frac{\int_0^{\tilde{\theta}} \lambda_\theta \kappa(\theta) f(\theta) d\theta}{F(\tilde{\theta})}.$$

Intuitively, these $\tilde{\lambda}_\theta$ weights capture both: (i) the differences in the planner's value for one util given to each type (parameterized via the λ_θ weights), and (ii) the differences in utility each type experiences when given B in benefits (parameterized via the $\kappa(\theta)$ schedule).

Then, working from Equation (B.49), applying the Multiplicative Separability assumption, dividing by p and using that $F(\tilde{\theta}) = m$, we obtain:

$$\frac{u(B)}{u'(B)B} = \frac{1 + \eta_m}{\frac{\tilde{\lambda}_m}{\tilde{\lambda}_{\text{avg}}} + \eta_m} \tag{B.50}$$

for $\eta_m := \frac{\partial p}{\partial m} \frac{m}{p}$, which is the equation we target.

B.7.2 Additional Discussion of Equation (2.6)

We begin the additional discussion stating the following lemma, proven in Appendix B.6.

Lemma. *If* $\rho \geq 1$, $\frac{\partial}{\partial B} \left(\frac{u(B)/B}{u'(B)} \right) > 0$.

Lemma B.2 follows from elementary properties of concavity. It establishes that the LHS of Proposition 2.1, the ratio of the average utility to the marginal utility, is increasing in $u(B)$. Henceforth we assume $\rho \geq 1$. It is useful because it allows us to determine how the planner adjusts B and m if the LHS and RHS are not equated.

To build additional intuition for Equation (2.6), we consider two sub-cases of Case 1 (without inframarginal effects), i.e. where $\eta_m = 0$.

Case 1a: no stigma, complete take-up. Assume there are no costs ($\gamma = 0$) and there is perfect take-up ($\eta_B = 0$). Rearranging Proposition 2.1 and applying Proposition B.2 gives that at an optimum,

$$1 + \frac{1}{2}\rho \approx \frac{\lambda_{\text{avg}}}{\lambda_m}.$$

The LHS of this expression is the welfare gain from transferring an additional dollar to inframarginal types. The RHS of this expression (which always weakly exceeds 1, since $\lambda_{\text{avg}} \geq \lambda_m$ as λ_θ is decreasing in θ) is the welfare-weight-adjusted cost of taking a dollar away from type λ_m to transfer to inframarginal types. A small increase in m gives $u(B)$ (valued at $u(B)/B$ per dollar) in benefits to people who have λ_m in welfare weight. A small increase in B gives $u'(B)$ to people who have λ_{avg} in welfare weights. Proposition 2.1 establishes that at an optimum, the planner is indifferent between: (i) relaxing the eligibility criterion by increasing m (and transferring to new people, but reducing the benefit to the inframarginal types), and (ii) transferring a bit more by increasing B (giving $u'(B)$ to people with weights λ_{avg}). This tradeoff is at the core of many public discussions of social welfare programs.

Case 1b: incorporating costs. Now consider the case where benefit size affects take-up probability ($\eta_B > 0$) because there are costs ($\gamma > 0$). Observe that $\eta_B > 0$ tends to reduce the RHS. Intuitively, if $\eta_B > 0$, the planner must consider that raising the benefit for inframarginal types will boost take-up. People who newly take up the benefit are just indifferent to doing so, by an envelope condition, but they have a fiscal externality. For large η_B, the planner raises B. However, if γ is large, that serves as a force against raising B: large γ implies that most of the additional gain from take-up is soaked up by costs.

B.7.3 Discussion of Assumption 2.1

B.7.3.1 Necessary Condition for Proposition 2.3

Assumption 2.1 states that the change in the eligibility threshold reduces the average stigma costs among the fraction of people who take up the program. Assumption 2.1 is difficult to validate empirically without granular information on the treatment effect of changing the eligibility threshold on people's perceived stigma cost at every part of the stigma cost distribution.

First we show that this assumption is sufficient but not necessary. Equation (B.63) from the proof of Proposition 2.3 gives that the necessary and sufficient condition is:

$$\left(\frac{1-s}{p^s}\left(\eta_m^i+1\right)+\frac{s}{p^i}\left(\eta_m^s+1\right)\right)\left(\frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i}+\frac{1-s}{p^s}\right)\right)$$
$$<\left(\frac{1-s}{p^s}+\frac{s}{p^i}\right)\left(\frac{(1-\gamma)s}{p^i}\eta_m^s+\frac{1-s}{p^s}\eta_m^i-\frac{sm}{p^i}\frac{\partial\gamma}{\partial m}+\frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i}+\frac{1-s}{p^s}\right)\right). \tag{B.51}$$

As long as Equation (B.51) holds, it is true that for all Ξ, $m^w > m^n$. Put another way, Equation (B.51) is a necessary condition that encodes the combination of Assumption 1 and either condition (i) or condition (ii) in Proposition 2.3. Thus, Equation (B.51) is weaker than Assumption 2.1 and condition (i) or condition (ii).

Equation (B.51) encodes the observation that as $s \to 0$, Proposition 2.3 always holds, because the necessary condition then reduces to:

$$\left(\eta_m^i+1\right)\left(\frac{\lambda_m}{\lambda_{\text{avg}}}\right)<\left(\eta_m+\frac{\lambda_m}{\lambda_{\text{avg}}}\right), \tag{B.52}$$

which always holds since $\frac{\lambda_m}{\lambda_{\text{avg}}} \leq 1$. Intuitively, because information-only types capture the full benefit of the program, the fully naïve planner undervalues the social value of raising m more with more information-only types. As a result, she can tolerate a larger violation of $\frac{\partial\gamma}{\partial m} > 0$.

We also note that for various configurations of $\lambda_m/\lambda_{\text{avg}}$, γ, and $\frac{\partial\gamma}{\partial m}$, as well as the other parameters, the necessary condition may hold. For instance, as $\gamma \to 0$, the necessary condition always holds. Intuitively, as stigma agents become more like information agents, we no longer need a separate condition governing the behavior of η_m^i and η_m^s.

B.7.3.2 Discussion of Assumption 2.1

How could Assumption 2.1 fail? Suppose there are no information-only types ($s = 1$). Suppose moving the eligibility threshold reduces costs for people who are just indifferent to taking up the program (i.e., for whom $c \approx u(B)$). Suppose it has no effect on people for whom $c < u(B)$. Then, changing the eligibility threshold will first-order stochastically reduce the cost distribution. However, $\frac{\partial\gamma}{\partial m}$ will perhaps counterintuitively *rise*. Intuitively, the average cost conditional on taking up the program will feature a larger density at $c \approx u(B)$.

For a concrete example of the assumption failing, suppose $c \sim H_{\text{pre}} = N(1, \sigma)$ for known σ. Suppose $u(B) = 2$. Suppose raising m changes moves all costs larger than 2 to be at 2:

$$H_{\text{post}} = \begin{cases} N(1, \sigma) & c \leq 2 \\ c = 2 & \text{otherwise} \end{cases} \tag{B.53}$$

where we denote this truncated distribution by H_{post}. Changing the eligibility threshold induces a first-order stochastic reduction in the cost distribution. It raises the share of people in the population who take up the program. However, it also raises the average cost conditional on taking up the program. The average cost before raising m is $E[c|c \leq 2, H_{\text{pre}}] \approx 0.71$, whereas the average cost after raising m is $E[c|c \leq 2, H_{\text{post}}] \approx 0.91$.

However, violations of Assumption 2.1 are unlikely in practice. To see why, note that the counter-example above requires a large change in the cost distribution *only* for draws of the cost distribution that are about as large as $u(B)$. If raising m also affects the draws of the cost distribution for $c < u(B)$, that serves as a force pushing $\frac{\partial \gamma}{\partial m}$ downward.[11]

Second, Equation (B.51) shows that the necessary and sufficient condition for Proposition 2.3 to fail is much weaker than $\frac{\partial \gamma}{\partial m} \leq 0$.

Third, the specific counter-example changed the *shape* of the cost distribution. H_{pre} is normal; H_{post} is a truncated normal. We develop propositions showing that for the normal and exponential cost distributions, any any change in the (unconditional) mean costs that maintains the distributional family from which the costs are drawn will feature $\frac{\partial \gamma}{\partial m} < 0$.

Proposition B.1. *Let $c \sim N(\mu(m), \sigma)$ with $\mu'(m) < 0$. Then $\frac{\partial \gamma}{\partial m} < 0$.*

We prove Proposition B.1 in Appendix B.6. A change in m reduces the mean (unconditional) cost but the cost distribution remains normal. Then the change in ratio of costs to benefits, conditional on taking up the program, will shrink in m; i.e., Assumption 2.1 holds. We develop a similar proposition if costs are exponentially distributed:

Proposition B.2. *Let $c \sim Exp(\theta(m))$, where $1/\theta$ is the mean of the exponential distribution Exp and $\theta'(m) > 0$. Then $\frac{\partial \gamma}{\partial m} < 0$.*

Note that $\theta'(m) > 0$ implies the average unconditional cost $1/\theta$ falls in m, so Proposition B.2 is qualitatively similar to Proposition B.1.

[11]Note that we suppose all types receive draws from the same cost distribution. Thus, the violation of Assumption 2.1 is *not* that changing m only affects costs for $\theta = m$ at the marginal of eligibility. Rather, Assumption 2.1 is only likely to fail if changing m affects people for whom $c \approx u(B)$, i.e. they are indifferent to signing up (regardless of their income).

B.7.4 Formal Statement of Proposition 2.3.

Fix a vector of parameters $\Xi = (p^i, p^s, \lambda_\theta, s, \gamma, \eta_B, u(\cdot))$. Notice that, for any Ξ, any given η^i_m, η^s_m, and function $\frac{\partial \gamma}{\partial m}(B, m)$ induce a pair $(B^*(\eta^i_m, \eta^s_m, \frac{\partial \gamma}{\partial m}), m^*(\eta^i_m, \eta^s_m, \frac{\partial \gamma}{\partial m}))$ that satisfies Equation (2.8).

We call $(B^n, m^n) := (B^*(0,0,0), m^*(0,0,0))$ the *naïve* choice of (B, m): this is the choice of eligibility threshold and benefit size if (i) the planner neglects inframarginal effects arising from either agent, and (ii) does not realize that the eligibility threshold affects stigma. Call $(B^w, m^w) := (B^*(\eta^i_m, \eta^s_m, \frac{\partial \gamma}{\partial m}), m^*(\eta^i_m, \eta^s_m, \frac{\partial \gamma}{\partial m}))$ the *sophisticated* choice of (B, m).

We make several assumptions to rule out edge cases. First we assume, (i) $\lambda_m / \lambda_{\text{avg}} < 1$ at the naïve solution. This implies there exists some point up to the naïve planner's choice of m at which the welfare weight schedule is strictly decereasing. We require (ii) $\gamma > 0$, that stigma costs are positive for stigma agents (if they exist). We also require (iii) $\rho \geq 1$.

Finally, as discussed in the body and this appendix, we impose that (iv) Assumption 2.1 holds.

Under these assumptions we can show the following:

Proposition B.3 (Formal statement of Proposition 2.3)**.** *If $\eta^i_m > 0$ or $\eta^s_m > 0$, then $m^w > m^n$ for all Ξ as long as $\eta^s_m \leq \eta^i_m$ (condition (i)). Moreover, there exists $\varepsilon > 0$ such that $m^w > m^n$ for all Ξ as long as $\eta^s_m \in [\eta^i_m, \eta^i_m + \varepsilon]$ (condition (ii)).*

The proof is in Appendix B.7.5. This version of Proposition 2.3 is slightly more general than the version stated in the body. As in the body, one hypotheses that delivers the sharp policy implication is if stigma agents are less elastic than information agents (condition (i)). However, we also add a second condition: each vector Ξ yields an interval $\eta^s_m \in [\eta^i_m, \eta^i_m + \varepsilon]$ for $\varepsilon > 0$ in which the proposition still holds (condition (ii)). The utility of having condition (ii) as an alternative is that then the statement holds for some $\eta^s_m > \eta^i_m$ for all parameterizations. These conditions are also sufficient but not necessary.

B.7.5 Proofs in Extensions

B.7.5.1 Proof of Proposition B.1

Proof. It suffices to prove that $\frac{\partial}{\partial m}\left(E[c|c \leq u(B), \mu(m)]\right) < 0$. First, let $\chi(Z) := \phi(Z)/\Phi(Z)$ for normal PDF ϕ and normal CDF Φ. Equation (3) in Sampford (1953) gives that

$$0 < \frac{\partial}{\partial Z}\left(\frac{\phi(Z)}{1-\Phi(Z)}\right) < 1 \tag{B.54}$$

for all Z. Thus

$$-1 < \frac{\partial \chi}{\partial Z} < 0 \tag{B.55}$$

since the normal PDF is even and $1 - \Phi(Z) = \Phi(-Z)$.

The usual properties of the normal distribution give:

$$E[c|c \leq u(B), \mu(m)] = \mu(m) - \sigma \frac{\phi(Z(m))}{\Phi(Z(m))} \tag{B.56}$$

for $Z(m) := (u(B) - \mu(m))/\sigma$.

The chain rule gives

$$\frac{\partial}{\partial m}\left(\frac{\phi(Z(m))}{\Phi(Z(m))}\right) = -\frac{\partial \chi}{\partial Z}\frac{\mu'(m)}{\sigma}. \tag{B.57}$$

Then evaluating Equation (B.56) at the bounds in Equation (B.55) gives

$$\mu'(m) < \frac{\partial}{\partial m}\left(E[c|c < u(B), \mu(m)]\right) < 0. \tag{B.58}$$

□

B.7.5.2 Proof of Proposition B.2

Proof. It suffices to prove that $\frac{\partial}{\partial m}\left(E[c|c < u(B), \mu(m)]\right) < 0$. The mean of the truncated exponential distribution is:

$$\mu(\theta(m)) = \frac{1}{\theta} - u(B)\left(\exp(\theta u(B)) - 1\right)^{-1} \tag{B.59}$$

for $u(B) > 0$. This function is monotonically decreasing for all $u(B)$ (Al-Athari, 2008). □

B.7.6 Proof of Proposition 2.3/Proposition B.3

Proof. We want to show that the naïve planner would raise the eligibility threshold m and lower the benefit size B. First, we note that for a given (B, m) pair, the budget constraint ensures that raising B requires lowering m, and raising m requires lowering B. Thus, it is sufficient to argue that the naïve planner sets B too high. Consider the following rearrangement of Equation (B.19):

$$\frac{u(B)}{Bu'(B)} = \frac{\frac{(1-s)}{p^s}\left(\eta_m^i + 1\right) + \frac{s}{p^i}\left(\eta_m^s + 1\right)}{\left(\frac{p^s}{sp^s+(1-s)p^i} s\eta_B + 1\right)\left(\frac{(1-\gamma)s}{p^i}\eta_m^s + \frac{1-s}{p^s}\eta_m^i - \frac{sm}{p^i}\frac{\partial\gamma}{\partial m} + \frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i} + \frac{1-s}{p^s}\right)\right)}. \tag{B.60}$$

Using Lemma B.2, we note that the LHS of Equation (B.60) is increasing in B. Thus, noting that the naïve planner solves Equation (B.60) for B^n and the sophisticated planner solves the equation for B^s, we want to show:

$$\frac{u(B^s)}{B^s u'(B^s)} - \frac{u(B^n)}{B^n u'(B^n)} < 0, \tag{B.61}$$

and substituting Equation (B.60), we have that we want to show:

$$\begin{aligned}&\frac{\frac{(1-s)}{p^s}\left(\eta_m^i + 1\right) + \frac{s}{p^i}\left(\eta_m^s + 1\right)}{\left(\frac{p^s}{sp^s+(1-s)p^i} s\eta_B + 1\right)\left(\frac{(1-\gamma)s}{p^i}\eta_m^s + \frac{1-s}{p^s}\eta_m^i - \frac{sm}{p^i}\frac{\partial\gamma}{\partial m} + \frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i} + \frac{1-s}{p^s}\right)\right)} \\ &- \frac{\frac{(1-s)}{p^s}\left(0 + 1\right) + \frac{s}{p^i}\left(0 + 1\right)}{\left(\frac{p^s}{sp^s+(1-s)p^i} s\eta_B + 1\right)\left(\frac{(1-\gamma)s}{p^i}0 + \frac{1-s}{p^s}0 - \frac{sm}{p^i}0 + \frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i} + \frac{1-s}{p^s}\right)\right)} < 0.\end{aligned} \tag{B.62}$$

Observe that

$$\left(\frac{p^s}{sp^s + (1-s)p^i} s\eta_B + 1\right) > 0.$$

Cross-multiplying, it is then sufficient to show:

$$\begin{aligned}&\left(\frac{1-s}{p^s}\left(\eta_m^i + 1\right) + \frac{s}{p^i}\left(\eta_m^s + 1\right)\right)\left(\frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i} + \frac{1-s}{p^s}\right)\right) \\ &< \left(\frac{1-s}{p^s} + \frac{s}{p^i}\right)\left(\frac{(1-\gamma)s}{p^i}\eta_m^s + \frac{1-s}{p^s}\eta_m^i - \frac{sm}{p^i}\frac{\partial\gamma}{\partial m} + \frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i} + \frac{1-s}{p^s}\right)\right).\end{aligned} \tag{B.63}$$

Under Assumption 2.1, $-\frac{sm}{p^i}\frac{\partial\gamma}{\partial m} \geq 0$, so it is sufficient to show that:

$$\begin{aligned}
&\left(\frac{1-s}{p^s}\left(\eta_m^i+1\right)+\frac{s}{p^i}\left(\eta_m^s+1\right)\right)\left(\frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i}+\frac{1-s}{p^s}\right)\right) \\
&< \left(\frac{1-s}{p^s}+\frac{s}{p^i}\right)\left(\frac{(1-\gamma)s}{p^i}\eta_m^s+\frac{1-s}{p^s}\eta_m^i+\frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i}+\frac{1-s}{p^s}\right)\right). \qquad \text{(B.64)}
\end{aligned}$$

Rearranging gives that this condition is equivalent to:

$$\left(\frac{1-s}{p^s}\eta_m^i+\frac{s}{p^i}\eta_m^s\right)\left(\frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i}+\frac{1-s}{p^s}\right)\right) < \left(\frac{1-s}{p^s}+\frac{s}{p^i}\right)\left(\frac{(1-\gamma)s}{p^i}\eta_m^s+\frac{1-s}{p^s}\eta_m^i\right) \tag{B.65}$$

$$\iff \frac{\frac{1-s}{p^s}\eta_m^i+\frac{s}{p^i}\eta_m^s}{\frac{(1-\gamma)s}{p^i}\eta_m^s+\frac{1-s}{p^s}\eta_m^i} < \frac{\frac{1-s}{p^s}+\frac{s}{p^i}}{\frac{\lambda_m}{\lambda_{\text{avg}}}\left(\frac{(1-\gamma)s}{p^i}+\frac{1-s}{p^s}\right)}. \tag{B.66}$$

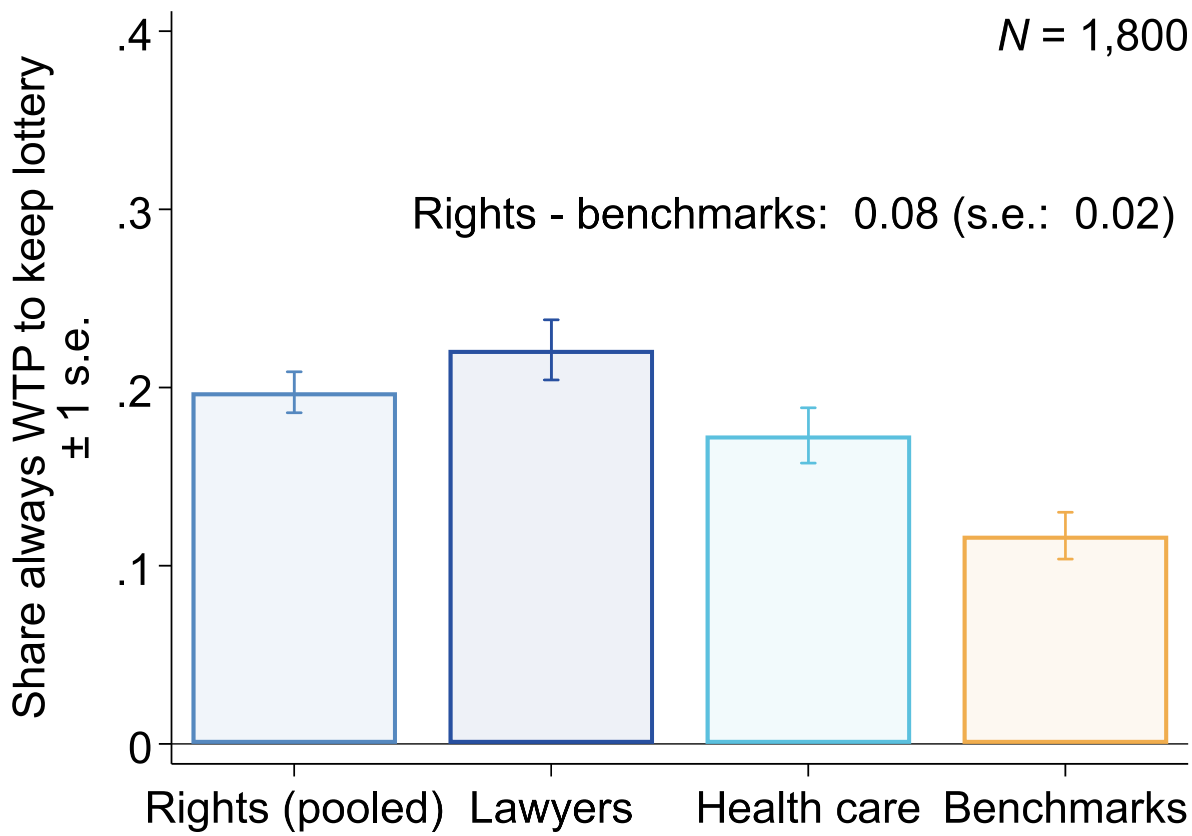
The statement holds strictly if $\eta_m^s = \eta_m^i$ as long as $\lambda_m < \lambda_{\text{avg}}$, by factoring the LHS and canceling. Moreover, holding η_m^s fixed, the LHS is strictly decreasing in η_m^i. Thus, if the statement holds for $\eta_m^s = \eta_m^i$, it also holds for $\eta_m^i < \eta_m^s$. This shows that the desired statement holds under condition (i). To argue that the desired statement holds under condition (ii), notice that the LHS is strictly increasing in η_m^s, holding η_m^i fixed. As a result, there exists $\tilde{\eta}_m^s > \eta_m^i$ such that the statement holds with equality. Since the LHS is increasing in η_m^s, the statement holds strictly for $\eta_m^s < \tilde{\eta}_m^s$. Thus, there exists an interval $\eta_m^s \in [\eta_m^i, \eta_m^i+\varepsilon]$ for $\varepsilon > 0$ such that Equation (B.66) holds, which completes the proof. □

Appendix C

Appendix to Chapter 3

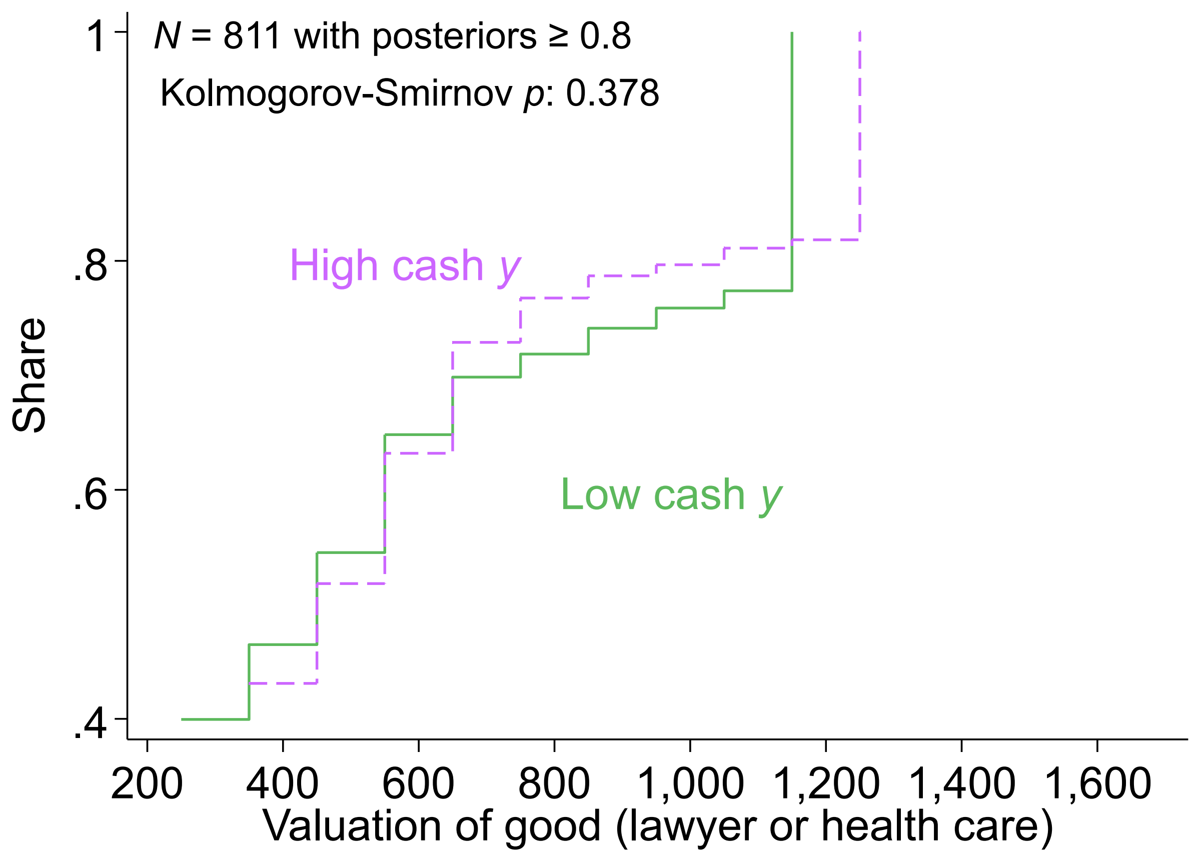
C.1 Additional Figures

Figure C.1: Share Always Willing to Pay to Keep Lottery



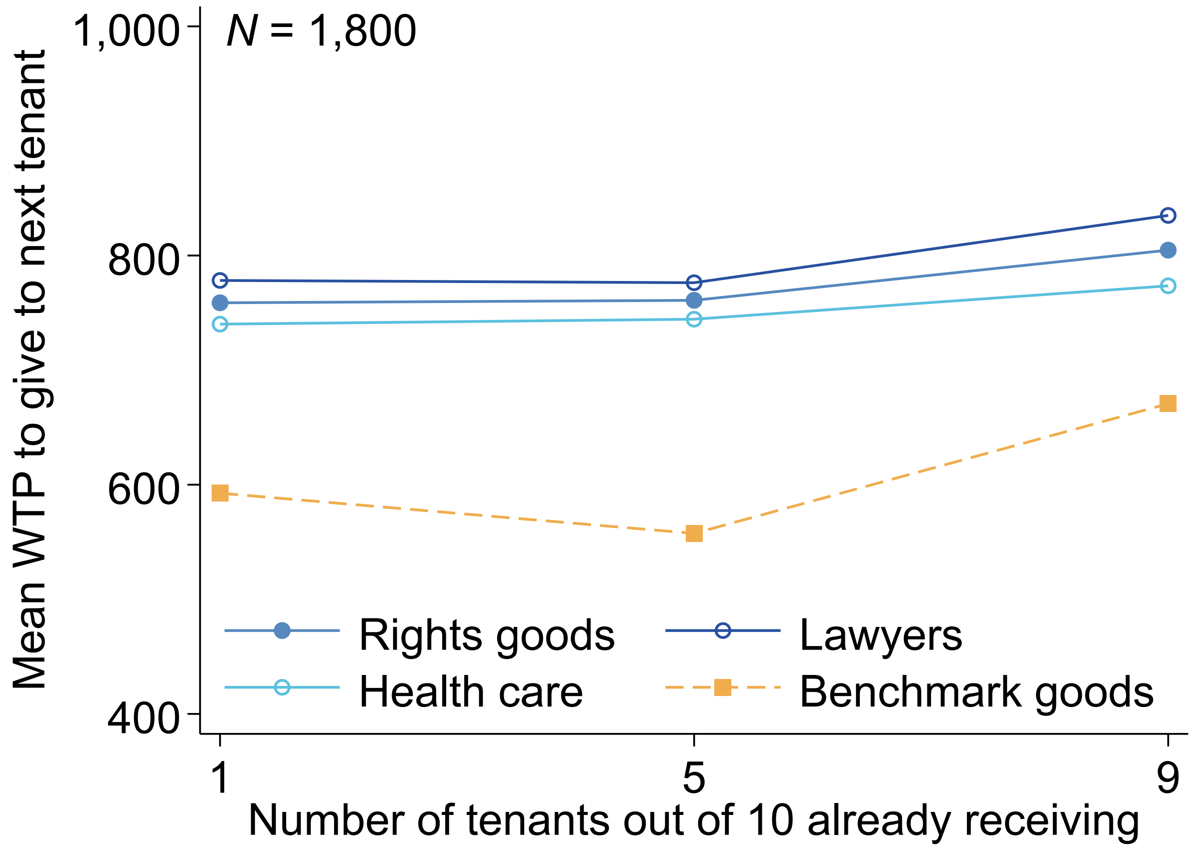
Note: This figure shows participants' decisions about keeping or rerunning a lottery in order to save money for future programs in Experiment 1 (Section 3.4.1). The y-axis is the share of participants who are willing to pay the maximum value we elicit to keep the lottery. The far left blue bar shows both rights goods pooled, while the next two bars show results for lawyers and health care disaggregated.

Figure C.2: WTP for Dignity of Choice by Cash Value



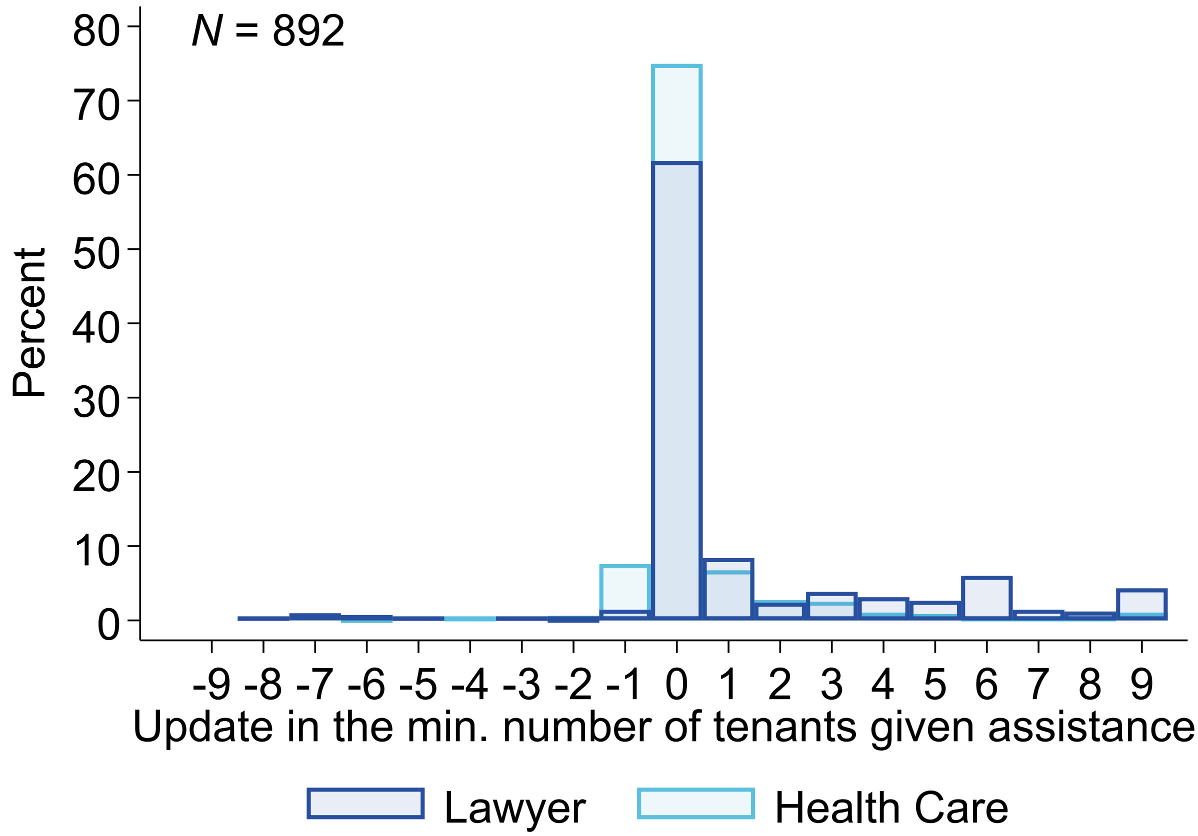
Note: This figure plots the distributions of participants' valuation of the good by the randomized cash value seen in Experiment 2. The high cash $y = \$300$ (dashed pink line) and the low cash $y = \$200$ (solid green line). The sample is restricted to those who did the experiments with rights goods and with posteriors about the percent of tenants expected to choose cash over the good $\geq 80\%$.

Figure C.3: Egalitarianism



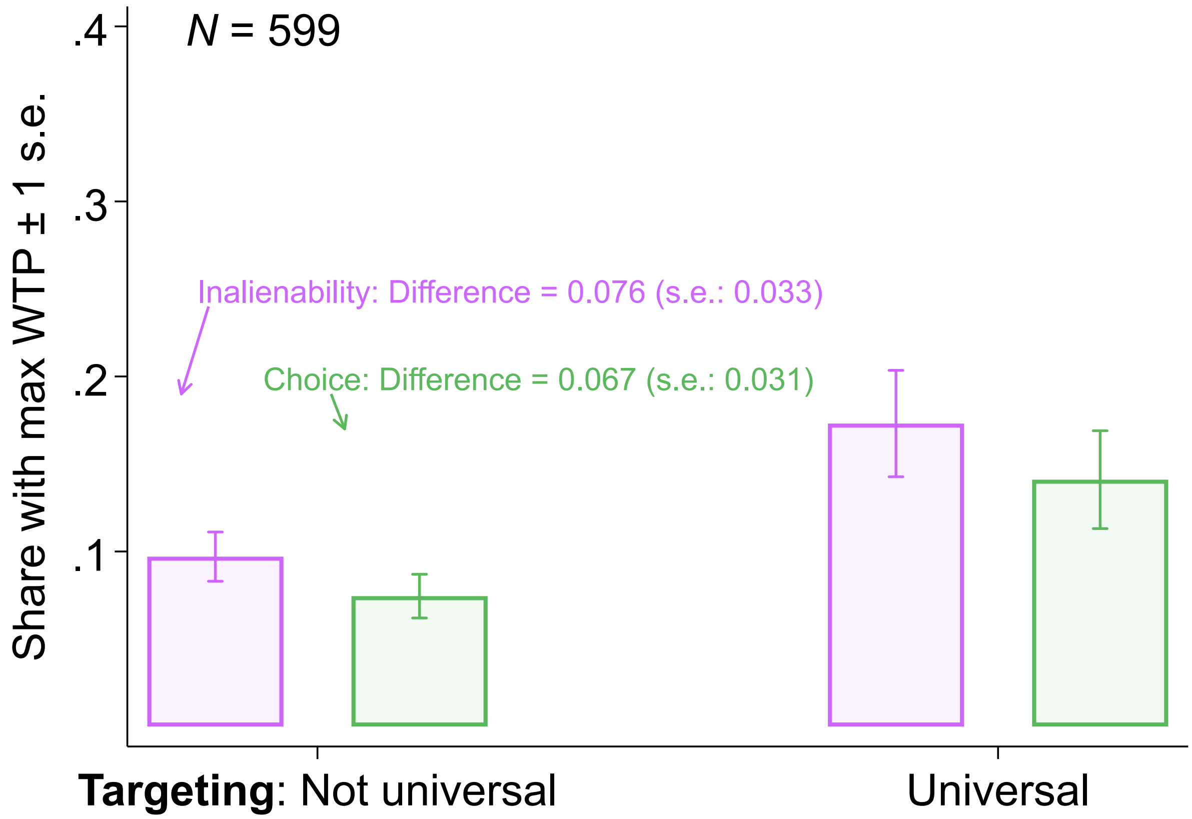
Note: This figure plots mean WTP to give the good to the next tenant in Experiment 3 (Section 3.4.3). Along the x-axis is the number of tenants out of 10 already assigned to receive the good. Rights goods are presented both pooled and disaggregated by lawyers health care.

Figure C.4: Anti-Targeting: Updates After Information Treatment



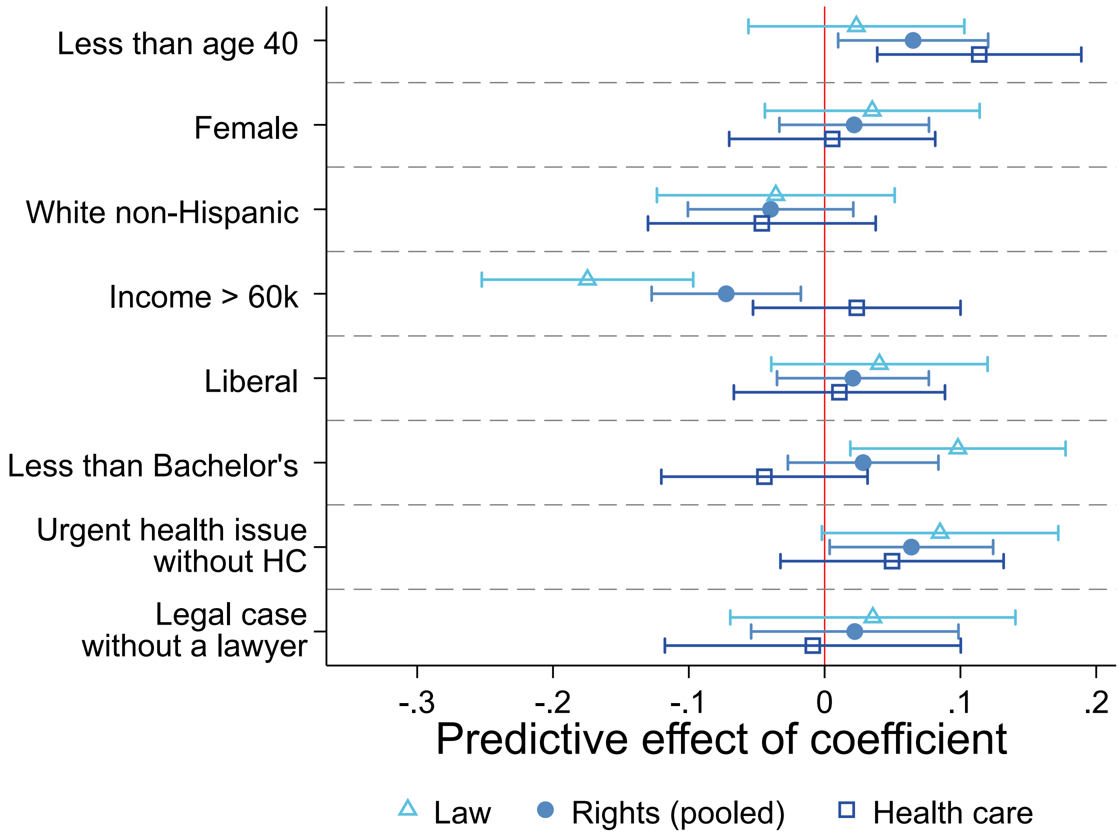
Note: This figure shows the distribution of updates in the minimum number of tenants given assistance (Figure 3.4) after being shown either a high or low information treatment (Appendix C.4.7). Only participants in lawyer or health care treatments saw information and were offered the chance to update. We restrict the sample to participants who saw rights goods, and further to those we can classify as welfarist or non-welfarist based on their prior beliefs. Zero indicates the participant declined to update. Positive numbers indicate the participant chose to update in the direction of information shown, relative to their priors.

Figure C.5: Correlations Between Universalism and Features of Rights for Benchmark Goods



Note: This figure replicates Figure 3.5 for the sample of those who saw benchmark goods (YMCA membership or bus pass). It shows the share of participants with the maximum possible WTP in the inalienability (pink) and dignity of choice (green) experiments. Shares are split by whether participants distributed the good universally in Experiment 4. The

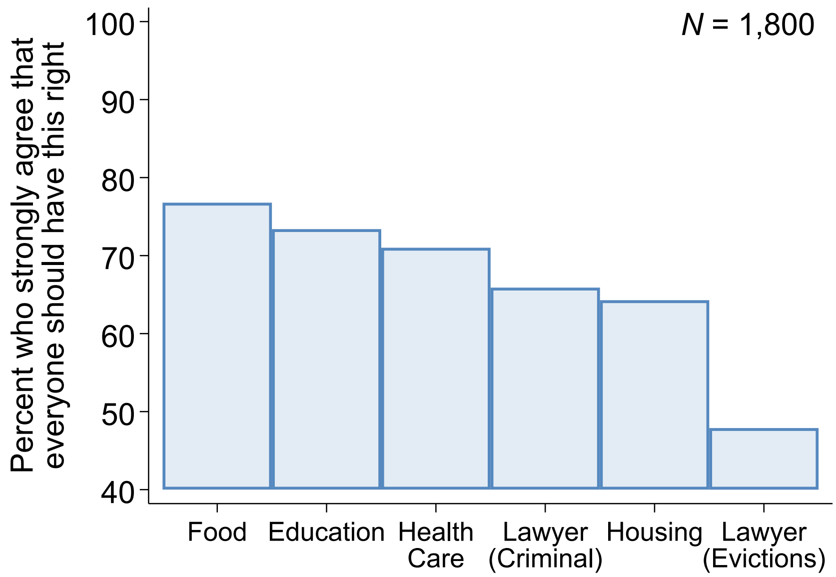
Figure C.6: Demographics and Rights Preferences



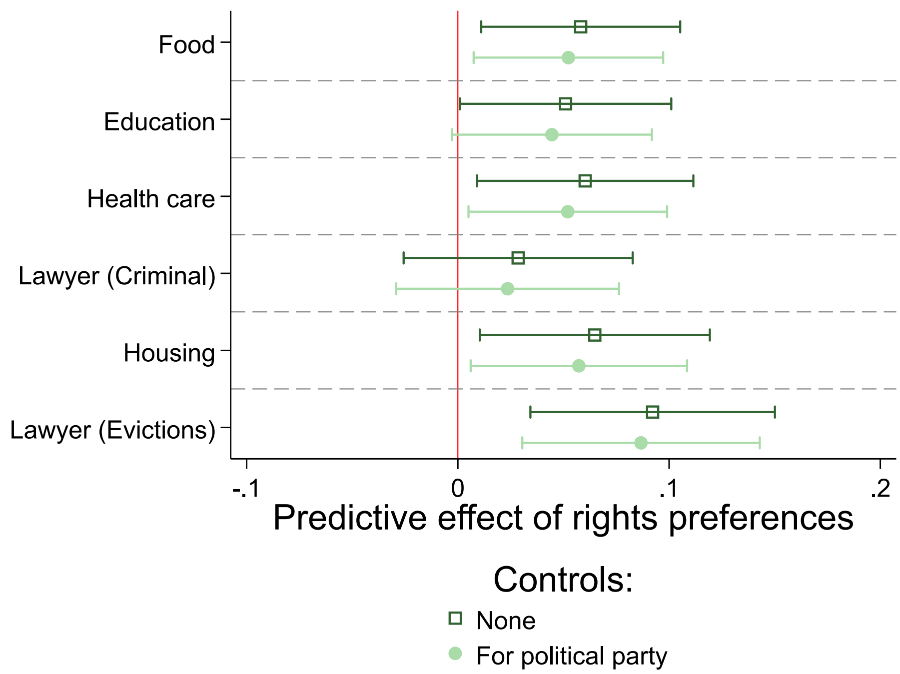
Note: This figure shows estimates of rights preferences for different demographics. Here, rights preferences are defined as positive WTP in both Experiment 1 and Experiment 2. We present estimates for rights good (pooled) as well as for lawyers and health care. *Urgent health issue without HC* captures responses to the question "Have you ever needed urgent medical care that you did not seek because of the cost?" and *Legal case without a lawyer* captures responses to "Have you ever been involved in a legal matter without a lawyer representing you?" Estimates shown with ± 1.96 standard errors.

Figure C.7: Rights to Other In-Kind Goods

(a) Strong Agreement with Rights to Other In-Kind Goods



(b) Rights Preferences and Rights to Other In-Kind Goods



Note: This figures plots results from asking Spectators if they agree there is a right to several types of in-kind goods. Panel A plots the percent who "strongly agree" that each good should be a right. Panel B shows the predictive effect of having positive WTP in both Experiments 1 and 2 on strongly agreeing each good should a right. Estimates shown with ±1.96 standard errors.

C.2 Additional Tables

Table C.1: Balance

	Rights goods			Benchmark goods			p-values		
	(1) Pooled	(2) Lawyers	(3) Health care	(4) Pooled	(5) YMCA	(6) Bus pass	(7) 1=4	(8) 2=4	(9) 3=4
White non-Hispanic	0.71	0.71	0.70	0.72	0.72	0.72	0.72	0.85	0.71
	[0.46]	[0.45]	[0.46]	[0.45]	[0.45]	[0.45]			
Liberal	0.59	0.57	0.61	0.54	0.53	0.55	0.03	0.23	0.01
	[0.49]	[0.50]	[0.49]	[0.50]	[0.50]	[0.50]			
Income $>$ 60k	0.47	0.49	0.45	0.50	0.53	0.48	0.15	0.59	0.05
	[0.50]	[0.50]	[0.50]	[0.50]	[0.50]	[0.50]			
Less than Bachelor's	0.45	0.45	0.45	0.45	0.40	0.49	0.96	0.79	0.76
	[0.50]	[0.50]	[0.50]	[0.50]	[0.49]	[0.50]			
Legal case without a lawyer	0.16	0.17	0.14	0.16	0.15	0.16	0.72	0.25	0.55
	[0.36]	[0.38]	[0.35]	[0.36]	[0.36]	[0.37]			
Urgent health issue without HC	0.31	0.30	0.32	0.32	0.30	0.33	0.43	0.20	0.96
	[0.46]	[0.46]	[0.47]	[0.47]	[0.46]	[0.47]			
Female	0.50	0.51	0.50	0.52	0.53	0.51	0.42	0.67	0.31
	[0.50]	[0.50]	[0.50]	[0.50]	[0.50]	[0.50]			
Less than age 40	0.57	0.55	0.58	0.56	0.54	0.57	0.97	0.90	0.82
	[0.50]	[0.50]	[0.49]	[0.50]	[0.50]	[0.50]			
F-statistic							0.996	0.510	1.586
p-value							0.437	0.849	0.124
Observations	1,201	606	595	599	299	300	1,800	1,205	1,194

Note: This table expands on Table 3.1. Columns (1)-(6) show demographic characteristics for different goods, pooled and separately. Columns (7)-(9) show the *p*-values of the differences between pooled benchmarks and each rights goods column. The *F*-statistic is from a joint test of significance for the listed demographic variables. Brackets show standard deviations.

Table C.2: Attrition

	Number of participants			
	(1) Lawyers	(2) Health care	(3) YMCA	(4) Bus pass
Started	627	623	309	316
Egalitarianism⋆	615	605	306	310
Passed at least 1 of 2 attenion checks⋆	614	604	303	309
Inalienability⋆	614	603	306	309
Dignity of Choice	612	600	302	306
Anti-targeting	610	599	302	306
Passed at least 2 of 3 attenion checks	607	596	300	300
Good valuation	607	596	300	300
Demographics & political preferences	606	595	299	300

Note: This table shows the attrition in our survey. Rows show the number of Spectators participating, for each good, at each stage of our survey flow. Rows marked with ⋆ were presented in a randomized order.

Table C.3: Inalienability Robustness: Controlling for Valuation of Good

	(1) Non-zero WTP (= 1)	(2) Max WTP (= 1)	(3) WTP (s.d.)	(4) WTP (s.d.)	(5) WTP (s.d.)
Rights good (= 1)	0.065	0.064	0.248		
	(0.027)	(0.019)	(0.058)		
	[0.017]	[0.001]	[0.000]		
Lawyers (= 1)				0.255	
				(0.077)	
				[0.001]	
Health care (= 1)					0.235
					(0.064)
					[0.000]
WTP for good FE	✓	✓	✓	✓	✓
Raw mean (benchmarks)			58.4	58.4	58.4
Raw s.d. (benchmarks)			69.9	69.9	69.9
Mean (benchmarks)	0.456	0.117	-0.000	-0.000	-0.000
Observations	1,800	1,800	1,800	1,205	1,194

Note: This table replicates Table 3.2 with added fixed effects for valuation of the good. Parentheses show robust standard errors. Brackets show p-values.

Table C.4: Balance Among High Posterior Participants

	(1) Rights-Benchmark goods
White non-Hispanic	-0.01
	(0.03)
Liberal	0.05
	(0.03)
Income $>$ 60k	-0.03
	(0.03)
Less than Bachelor's	0.02
	(0.03)
Legal case without a lawyer	0.01
	(0.04)
Urgent health issue without HC	-0.03
	(0.03)
Female	-0.01
	(0.03)
Less than age 40	0.02
	(0.03)
F-statistic	0.750
Observations	1,198

Note: This table shows balance for respondents with high posteriors ($\geq$ 90%) about the percent of tenants expected to pick cash over the good. The *F*-statistic is from a joint test of significance for the listed demographic variables. Parentheses show robust standard errors.

Table C.5: Dignity of Choice Robustness: Lawyers and Health Care

				Posterior $\geq$ 0.9			Posterior = 1		
	(1) Non-zero WTP (= 1)	(2) Max WTP (= 1)	(3) WTP (s.d.)	(4) Non-zero WTP (= 1)	(5) Max WTP (= 1)	(6) WTP (s.d.)	(7) Non-zero WTP (= 1)	(8) Max WTP (= 1)	(9) WTP (s.d.)
Panel A. Lawyers									
Lawyers (= 1)	0.250	0.168	0.642	0.263	0.180	0.677	0.131	0.237	0.752
	(0.029)	(0.024)	(0.072)	(0.037)	(0.029)	(0.089)	(0.060)	(0.050)	(0.157)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.029]	[0.000]	[0.000]
Posterior	-0.351	-0.278	-1.085						
	(0.055)	(0.059)	(0.170)						
	[0.000]	[0.000]	[0.000]						
Panel B. Health Care									
Health Care (= 1)	0.099	0.079	0.349	0.086	0.106	0.384	0.032	0.095	0.354
	(0.029)	(0.020)	(0.067)	(0.036)	(0.024)	(0.080)	(0.054)	(0.037)	(0.124)
	[0.001]	[0.000]	[0.000]	[0.016]	[0.000]	[0.000]	[0.547]	[0.011]	[0.004]
Posterior	-0.517	-0.166	-0.932						
	(0.065)	(0.059)	(0.179)						
	[0.000]	[0.005]	[0.000]						
Raw mean (benchmarks)			223.0			189.3			213.0
Raw s.d. (benchmarks)			277.4			247.5			268.3
Mean (benchmarks)	0.432	0.092	0.000	0.382	0.066	-0.121	0.438	0.087	-0.036
Observations: lawyers	606	606	606	273	273	273	102	102	102
Observations: health care	595	595	595	327	327	327	149	149	149
Observations: benchmarks	599	599	599	458	458	458	208	208	208

Note: This table replicates Table 3.3 separately for lawyers (Panel A) and health care (Panel B). Parentheses show robust standard errors. Brackets show *p*-values.

Table C.6: Dignity of Choice Robustness: High Posteriors

	Posterior = 1				Posterior ≥ 0.9
	(1) Non-zero WTP (= 1)	(2) Max WTP (= 1)	(3) WTP (s.d.)	(4) Entitled to Choice (= 1)	(5) Entitled to Choice (= 1)
Rights good (= 1)	0.077	0.155	0.547	0.115	0.132
	(0.048)	(0.034)	(0.110)	(0.045)	(0.033)
	[0.111]	[0.000]	[0.000]	[0.011]	[0.000]
Raw mean (benchmarks)			213.0		
Raw s.d. (benchmarks)			268.3		
Mean (benchmarks)	0.438	0.087	-0.036	0.308	0.276
DDML	✓	✓	✓		
Observations	459	459	459	459	804

Note: This table shows robustness checks for dignity of choice using the sub-sample of respondents who believe 100 out of 100 tenants would choose cash over the good. Columns (1)-(3) replicate columns (7)-(9) of Table 3.3 with double/de-biased machine learning (Chernozhukov et al., 2018). The model selects from the demographic controls reported in Table 3.1: race, income, education, gender, age, political beliefs, having experienced a legal case without a lawyer, and having had an urgent health issue without access to health care. The outcome in column (4) is a dummy that indicates a respondent selected "All tenants should be entitled to the choice of a [good]" as one reason for their decisions about giving the tenant a choice. Parentheses show robust standard errors. Brackets show p-values.

Table C.7: Tests of Egalitarianism

	(1) Non-zero WTP (= 1)	(2) Max WTP (= 1)	(3) WTP (s.d.)	(4) Non-zero WTP (= 1)	(5) Max WTP (= 1)	(6) WTP (s.d.)
Rights good=1	0.082	0.193	0.562	0.074	0.144	0.483
	(0.020)	(0.033)	(0.066)	(0.019)	(0.034)	(0.068)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
$\mathbb{1}(z = 9)$	0.043	0.124	0.291	0.055	0.122	0.307
	(0.023)	(0.038)	(0.079)	(0.023)	(0.038)	(0.079)
	[0.068]	[0.001]	[0.000]	[0.017]	[0.001]	[0.000]
Rights good=1 $\times$ $\mathbb{1}(z = 9)$	-0.046	0.011	-0.154	-0.055	0.011	-0.170
	(0.026)	(0.047)	(0.090)	(0.025)	(0.047)	(0.090)
	[0.075]	[0.809]	[0.087]	[0.027]	[0.815]	[0.058]
WTP for good FE				✓	✓	✓
Raw mean (benchmarks, $\mathbb{1}(z < 9)$)			575.3			575.3
Raw s.d. (benchmarks, $\mathbb{1}(z < 9)$)			328.4			328.4
Mean (benchmarks, $\mathbb{1}(z < 9)$)	0.888	0.258	0.000	0.888	0.258	0.000
Observations	1,800	1,800	1,800	1,800	1,800	1,800

Note: This table reports estimates from the differences-in-differences specification from Equation 3.12. $\mathbb{1}(z < 9)$ indicates the Spectator saw 9 out of 10 tenants had already received the good, rather than 1 or 5 tenants. Outcomes are the extensive margin (*Non-zero WTP*), an indicator for having the maximum WTP that we elicit (*Max WTP*), and the intensive margin (*WTP*). WTP is reported in terms of standard deviations relative to benchmark goods. Columns (4)-(6) add fixed effects for WTP for the good directly. Parentheses show robust standard errors. Brackets show p-values.

Table C.8: Robustness: Multiple Hypothesis Testing Corrections

	Inalienability	Dignity of Choice	Egalitarianism
	(1) WTP (s.d.)	(2) WTP (s.d.)	(3) WTP (s.d.)
Rights good	0.289 (0.053) [0.000]	0.518 (0.067) [0.000]	0.562 (0.066) [0.000]
Rights good & $\mathbb{1}(z = 9)$			-0.154 (0.090) [0.087]
Romano-Wolf p-value	< 0.001	< 0.001	0.096
Posterior ≥ 0.9		✓	
Observations	1,800	1,058	1,800

Note: This table shows robustness to multiple hypothesis testing corrections for inalienability, dignity of choice, and egalitarianism with Romano-Wolf p-values (Clarke et al., 2020) with 1,000 iterations. Parentheses show robust standard errors. Brackets show p-values.

Table C.9: Anti-Targeting Robustness: Controlling for Valuation of Good

	(1) Universal (= 1)	(2) No. Tenants	(3) Universal (= 1)	(4) No. Tenants	(5) Universal (= 1)	(6) No. Tenants
Rights good (= 1)	0.044	0.458				
	(0.024)	(0.187)				
	[0.067]	[0.015]				
Lawyers (= 1)			0.116	1.007		
			(0.034)	(0.252)		
			[0.001]	[0.000]		
Health care (= 1)					0.013	0.215
					(0.026)	(0.204)
					[0.621]	[0.293]
WTP for good FE	✓	✓	✓	✓	✓	✓
Mean (benchmarks)	0.260	4.855	0.260	4.855	0.260	4.855
Observations	1,800	1,800	1,205	1,205	1,194	1,194

Note: This table replicates Table 3.4 with added fixed effects for valuation of the good. Parentheses show robust standard errors. Brackets show p-values.

Table C.10: Correlations with Universalism Robustness: Controlling for Valuation of Good

	Dep. Var.: Universal (= 1)			
	(1)	(2)	(3)	(4)
Panel A. Rights goods (pooled)				
WTP for inalienability (s.d.)	0.022			0.008
	(0.012)			(0.012)
	[0.058]			[0.490]
WTP for dignity of choice (s.d.)		0.053		0.051
		(0.011)		(0.011)
		[0.000]		[0.000]
Mean WTP (s.d.)			0.062	
			(0.014)	
			[0.000]	
Panel B. Lawyers				
WTP for inalienability (s.d.)	0.022			-0.001
	(0.016)			(0.017)
	[0.184]			[0.958]
WTP for dignity of choice (s.d.)		0.072		0.072
		(0.016)		(0.017)
		[0.000]		[0.000]
Mean WTP (s.d.)			0.074	
			(0.020)	
			[0.000]	
Panel C. Health care				
WTP for inalienability (s.d.)	0.017			0.010
	(0.017)			(0.017)
	[0.314]			[0.554]
WTP for dignity of choice (s.d.)		0.033		0.032
		(0.016)		(0.016)
		[0.032]		[0.047]
Mean WTP (s.d.)			0.044	
			(0.021)	
			[0.035]	
WTP for good FE	✓	✓	✓	✓
Mean	0.428	0.428	0.428	0.428
Observations	1,201	1,201	1,201	1,201

Note: This table replicates Table 3.5 with added fixed effects for valuation of the good. Parentheses show robust standard errors. Brackets show p-values.

Table C.11: Tests of Universalism and Welfarism

	(1) Any rights prefs	(2) Both rights prefs	(3) Both rights prefs (non-welfarist)	(4) Welfarist
Panel A. All rights goods				
Universal (=1)	0.068	0.026	0.088	-0.190
	(0.028)	(0.034)	(0.032)	(0.030)
	[0.014]	[0.447]	[0.006]	[0.000]
Constant	0.765	0.354	0.233	0.376
	(0.018)	(0.020)	(0.017)	(0.020)
	[0.000]	[0.000]	[0.000]	[0.000]
Observations	892	892	892	892
Panel B. Posteriors $\geq$ 90				
Universal (=1)	0.081	0.064	0.131	-0.229
	(0.039)	(0.042)	(0.038)	(0.039)
	[0.037]	[0.126]	[0.000]	[0.000]
Constant	0.699	0.266	0.146	0.412
	(0.025)	(0.024)	(0.019)	(0.027)
	[0.000]	[0.000]	[0.000]	[0.000]
Observations	533	533	533	533
Panel C. Lawyers				
Universal (=1)	0.049	0.047	0.133	-0.255
	(0.037)	(0.051)	(0.048)	(0.046)
	[0.187]	[0.352]	[0.005]	[0.000]
Constant	0.812	0.410	0.244	0.474
	(0.024)	(0.030)	(0.026)	(0.031)
	[0.000]	[0.000]	[0.000]	[0.000]
Observations	417	417	417	417
Panel D. Health care				
Universal (=1)	0.079	-0.003	0.042	-0.140
	(0.041)	(0.045)	(0.043)	(0.039)
	[0.050]	[0.943]	[0.325]	[0.000]
Constant	0.726	0.308	0.224	0.296
	(0.025)	(0.026)	(0.023)	(0.026)
	[0.000]	[0.000]	[0.000]	[0.000]
Observations	475	475	475	475

Note: This table offers formal tests for Figure 3.6. Each panel shows estimates of an indicator for universal provision in Experiment 4 on four outcomes. The outcome in column (1) in an indicator for having positive WTP in either Experiment 1 or 2, in column (2) is an indicator for positive WTP in *both* experiments, and in Column (3) is an indicator for positive WTP in *both* experiments and non-welfarism. Finally, the outcome in column (4) is an indicator for welfarist preferences. The sample is Spectators who saw rights goods and who can be classified as welfarist or non-welfarist based on their prior beliefs. Panel A pools both rights goods, Panel B restricts the sample in Panel A to those with posteriors $\geq$ 90%, Panel C shows results for lawyers, and Panel D shows results for health care. Parentheses show robust standard errors. Brackets show p-values.

Table C.12: Robustness: Benchmark Incentives and Features of Rights

	Inalienability				Dignity of Choice				Anti-targeting			
	(1) WTP (s.d.)	(2) WTP (s.d.)	(3) WTP (s.d.)	(4) WTP (s.d.)	(5) WTP (s.d.)	(6) WTP (s.d.)	(7) WTP (s.d.)	(8) WTP (s.d.)	(9) No. Tenants	(10) No. Tenants	(11) No. Tenants	(12) No. Tenants
Unincentivized (= 1)	-0.002				-0.052				-0.422			
	(0.096)				(0.095)				(0.326)			
	[0.980]				[0.586]				[0.195]			
Rights good (= 1)		0.289	0.288	0.290		0.518	0.506	0.557		1.419	1.312	1.734
		(0.053)	(0.058)	(0.090)		(0.067)	(0.071)	(0.096)		(0.176)	(0.195)	(0.299)
		[0.000]	[0.000]	[0.001]		[0.000]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
Benchmarks Only	✓				✓				✓			
Incentivized Benchmarks	✓	✓	✓		✓	✓	✓		✓	✓	✓	
Unincentivized Benchmarks	✓	✓		✓	✓	✓		✓	✓	✓		✓
Observations	599	1,800	1,648	1,353	458	1,058	952	706	599	1,800	1,648	1,353

Note: This table shows the results of incentivization of benchmark goods on Spectators' decisions in Experiments 1, 2 and 4. Column (2) replicates column (3) of Table 3.2, column (6) replicates column (6) of Table 3.3, and column (10) replicates column (2) of Table 3.4, including all benchmark good observations. Columns (3), (7), and (11) restrict benchmark observations to only those who were incentivized, while column (4), (8), and (12) restrict benchmark observations to only those who were unincentivized. Columns (5)-(8) restrict to those with posteriors ≥ 90. Among those who saw incentivized goods (lawyers, YMCA memberships, bus passes), half randomly had their WTP for the good incentivized while the other half did not. Parentheses show robust standard errors. Brackets show p-values.

Table C.13: Robustness: Incentives and Valuation of Good

	Dep. Var.: WTP for good	
	(1) Lawyers	(2) Benchmarks
Unincentivized	-11.448	-7.180
	(26.478)	(28.301)
	[0.666]	[0.800]
Intercept	771.382	341.391
	(18.891)	(20.087)
	[0.000]	[0.000]
Observations	606	303

Note: This table shows the results of differential incentivization on Spectators' WTP for the good. Column (1) presents the effect of incentivization for lawyers and column (2) presents the same for benchmark goods. Among those who saw incentivized goods (lawyers, YMCA memberships, bus passes), half randomly had their WTP for the good incentivized while the other half did not. Parentheses show robust standard errors. Brackets show *p*-values.

Table C.14: Robustness: Double/debiased Machine Learning

	Inalienability		Dignity of Choice		Anti-targeting	
	(1) Non-zero WTP (= 1)	(2) WTP (s.d.)	(3) Non-zero WTP (= 1)	(4) WTP (s.d.)	(5) Universal (= 1)	(6) No. Tenants
Rights good (= 1)	0.073	0.290	0.167	0.512	0.177	1.492
	(0.025)	(0.054)	(0.031)	(0.067)	(0.023)	(0.174)
	[0.004]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Posterior ≥ 0.9			✓	✓		
Observations	1,800	1,800	1,058	1,058	1,800	1,800

Note: This table shows our main results using double/debiased machine learning for inalienability, dignity of choice, and anti-targeting (Chernozhukov et al., 2018). The model selects from demographic controls. Eligible demographic controls are those reported in Table 3.1: race, income, education, gender, age, political beliefs, having experienced a legal case without a lawyer, and having had an urgent health issue without access to health care. Column (1) modifies column (3) of Table 3.2, column (2) modifies column (6) of Table 3.3, and column (3) modifies column (2) of Table 3.4. Parentheses show robust standard errors. Brackets show p-values.

C.3 Framework Appendix

C.3.1 Proof of Proposition 3.1

Proof. Consider an allocation $\tilde{Y} \equiv (\tilde{y}_1, \ldots \tilde{y}_J)$ that maximizes welfarist social welfare for a given set of welfare weights, utility functions, and incomes. Thus $\tilde{Y}$ solves

$$\max \sum_{j=1}^{J} \gamma_j u(x_j, y_j), \text{ such that } \sum_j y_j \leq m. \tag{C.1}$$

Put $\Delta u(y_j = k) \equiv u(x_j, y_k) - u(x_j, y_{k-1})$, which is the difference in private utility generated from providing the kth good to individual j. Put $\Delta v(y_j = k) \equiv \gamma_j \Delta u(y_j = k)$, which the difference in welfarist social welfare from providing the kth good to individual j.

Now consider the $Y^* \equiv (y_1^*, \ldots y_J^*)$ which solves

$$\max \sum_{j=1}^{J} \gamma_j \left(u(x_j, y_j) + \eta_y \phi(y_j, r_{ij}) \right), \text{ such that } \sum_j y_j \leq m, \tag{C.2}$$

where $\phi(\cdot)$ is as in Equation (3.5).

As in the text, let $\tilde{\mathcal{J}} \equiv \{j \leq J : \tilde{y}_j > 0\}$ and $\mathcal{J}^* \equiv \{j \leq J : y_j^* > 0\}$. We want to show that $\tilde{\mathcal{J}} \subseteq \mathcal{J}^*$.

As is standard, we assume $\eta > 0, \lambda > 1$, and that $u(\cdot)$ is weakly concave in its inputs (thus, $\Delta u(y_j = 1) \geq \Delta u(y_j = k)$ for all $k > 0$). We also assume $r_j = 1$ for all j. Comparing the maximand in Equation (C.2) to Equation (C.1), the marginal welfare generated by the first y provided to any individual $j \notin \tilde{\mathcal{J}}$ increased by $\eta\lambda\Delta v(y_j = 1)$ and the marginal welfare generated by providing $k > 1$ to any individual $j \in \tilde{\mathcal{J}}$ increased by $\eta\Delta v(y_j = k)$.

There are two possible cases. First, $\tilde{Y}$ is already "flat," that is, $\max_{j \leq J}(\tilde{y}_j) = 1$. we show Y^* must also be flat. Second, $\tilde{Y}$ is not flat. Then, either (i) there exists a reallocation such that $\tilde{\mathcal{J}} \subsetneq \mathcal{J}^*$, or (ii) there is no reallocation and $\tilde{\mathcal{J}} = \mathcal{J}^*$.

Case 1. Consider reallocating $y = 1$ from j' to j, for $j', j \in \tilde{\mathcal{J}}$. Notice that the relative welfare value of providing a kth y to any individual $j \in \tilde{\mathcal{J}}$ relative to a first y to j' has decreased by $\eta \left(\lambda\Delta v(y_{j'} = 1) - \Delta v(y_j = k) \right)$, comparing the maximand in Equation (C.2) to Equation (C.1). Thus, if $\tilde{Y}$ is flat, then Y^* is also flat.

Case 2. Comparing the maximand in Equation (C.2) to Equation (C.1), the marginal welfare generated by providing the first good to each j' for whom $\tilde{y}_{j'} = 0$ has increased by $\eta\lambda\Delta v(y_{j'} = 1)$. Let $\underline{j} \equiv \arg\min_j \{\Delta v(\tilde{y}_j = k)\}$ for $k > 0$. That is, $\underline{j}$ is the individual to whom providing the marginal unit of y gives the least amount of social welfare, yet received at least one unit of y under $\tilde{Y}$. The marginal social welfare generated by providing the last good to $\underline{j}$ has increased by only $\eta\Delta v(y_{\underline{j}} = \tilde{y}_{\underline{j}})$. Thus, the welfare gain from flattening provision by allocating toward any $j' \notin \tilde{\mathcal{J}}$ has increased by $\eta \left(\lambda\Delta v(y_{j'} = 1) - \Delta v(y_{\underline{j}} = \tilde{y}_{\underline{j}}) \right)$. If that increase is sufficiently large relative to the initial gap in marginal social welfare, then Y^* will involve reallocating toward individuals in the loss domain and $\mathcal{J}^*$ will be strictly flatter than $\tilde{\mathcal{J}}$. Otherwise, the allocation does not change and is weakly flatter. $\square$

C.4 Experiment Details

C.4.1 Survey Recruitment

We recruited participants on Prolific. We advertised for an 18-minute survey entitled "Research Study," with $6 compensation. We restricted potential participants to fluent English speakers in the U.S.

C.4.2 Incentives

Rights Good Incentivization and Framing. We incentivize choices for rights goods in two ways. First, for Spectators who see lawyers, we use the strategy method. Second, we inform Spectators seeing both lawyers and health care that the nonprofit will be informed about participants' choices, which may impact their policies. Specifically, we inform Spectators doing experiments with lawyers:

> "Sometimes we will ask you what type of assistance to provide to tenants facing eviction. Please take these questions very seriously. Some participants will be randomly chosen to have their answers made in real life. **If you are chosen, your answers here will have significant impacts on the lives of real people**, so please take your time and respond truthfully."

We implement choices via a nonprofit partner in Memphis that works with tenants. Health care is not incentivized because it was not possible to implement. Instead, we tell these Spectators:

> "We will present you with a series of hypothetical scenarios and ask you what type of assistance to provide to tenants facing eviction. Please take these questions very seriously. A Memphis nonprofit who helps tenants facing eviction will be informed of participants' opinions on resource allocation."

Spectators then see contextual information about either lawyers[1] or health care[2], and then do a comprehension check about that fact. We then introduce the types of assistance the nonprofit provides: either lawyers *or* health care depending on which good they see, bus passes, YMCA memberships, and cash.

[1]We tell these Spectators: "Evictions often end up in court. In court, **tenants usually do not have lawyers**, because they are usually low-income and cannot afford them. **Landlords usually do have lawyers."** The government guarantees attorneys to anyone charged with a crime, but usually does not provide lawyers in civil settings, including eviction cases."

[2]We tell these Spectators: "Most low-income households in the United States say that they or a family member in their household delayed or went without some type of medical or dental care in the past year because they had difficulty affording the cost."

Benchmark Incentivization. Participants doing experiments with YMCA memberships or bus passes are randomized into seeing lawyers or health care as an alternate assistance option that the nonprofit provides. Based on which framing they see, their incentives follow that of lawyers or health care. Half of Spectators who see the lawyers framing are randomized into hypothetical good valuation.

Incentivization Tests.

First, we test the effect of incentivization on WTP in our 3 main experiments among benchmark Spectators. There is no effect of incentivization on outcomes in inalienability, dignity of choice, or anti-targeting experiments (Table C.12). Moreover, our main results looks similar using either only incentivized or unincentivized Spectators as the control group.

Our second test of incentivization is within incentivized participants. Half of Spectators seeing lawyers or benchmark goods are disincentivized in their choices about their WTP for the good. We inform selected participants:

> "At the beginning of the survey, we told you some choices will be randomly selected to be implemented for real tenants. In this section, all choices will be purely hypothetical. However, your choices are still important and the nonprofit will be informed of the results."

This is the last incentive-eligible section of the survey. We compare Spectators' valuations and find no difference by incentivization (Table C.13).

Belief Elicitations. We further incentivize both belief elicitations that are in the survey. At the start of the survey, we inform only incentivized participants:

> "Sometimes we will ask you to predict what choices tenants have made or the impacts programs affecting tenants have had. Please take these questions very seriously. Some participants will be randomly chosen to be paid bonuses if the answers they give are close enough to the truth. **If you are chosen, your answers could increase your participation bonus**, so please take your time and respond truthfully."

First, in Experiment 2, we elicit prior beliefs and posterior beliefs after information about the percent of tenants expected to choose $y in cash over the good. All incentivized participants (seeing lawyers, or seeing YMCA or bus passes with lawyers framing) have this belief incentivized. Second, in Experiment 4, we elicit prior beliefs and posterior beliefs after information about the efficacy of lawyers or health care. Here, incentivization is only for Spectators who see experiments about lawyers. In both modules, participants are informed: "Choose your responses carefully. You can earn bonuses for correct answers! You may request more details if you are curious about how the payment works." Interested participants saw that they would be enrolled in a lottery with a 10% chance of winning; for one up their upcoming predictions, they are compensated $1 in Spring 2024 if their answer is within 4 percentage points of the correct answer. We were not able to issue bonus payments at the time of survey because the accuracy of correct answers is determined by future events. In practice, all bonuses for selected participants are issued based on accuracy of prior beliefs, before participants were shown information.

C.4.3 Attention Checks

We include three attention checks. The first attention check asks participants to select a specific multiple choice option. The second two checks provide a list of cities and their populations, and ask participants to rank them from most to least populous.

Participants exit the survey if they fail two attention checks. 12 participants began our survey and were dropped because of this restriction (Table C.2). 7% of our final sample of 1,800 participants failed only one attention check. 41 failed the first check, 52 failed the second check, and 36 failed the third check.

We have several additional comprehension checks throughout the survey, which cannot be the basis for dropping participants per Prolific policy. Participants are always informed of the correct answer after completing a comprehension check.

C.4.4 Experiment 1 Details

To elicit participants' WTP for inalienability, we set up a scenario where 2 eligible, comparably needy tenants are entered in a lottery for one available good. We explain that: "After the lottery takes place, but before the tenants are informed of the outcome, the nonprofit reserves the ability to rerun the lottery in some cases. Sometimes prices change and money can be saved by rerunning the lottery and assigning the [good] to whoever wins the second lottery." This sets up the choice between either (1) leaving the lottery results as they are or (2) taking the good away from the original winner, giving it to whoever wins the second lottery, and saving some amount of money for future programs. We then reinforce this set-up by asking confirmation questions emphasizing, first, money is saved when the lottery is rerun, and second, the tenants will not know the original allocation if the lottery is rerun.

We first ask Spectators if they would prefer to keep the lottery results or rerun the lottery and save \$20. We repeatedly ask Spectators the same question in increments of \$20, until either they elect to rerun the lottery or they prefer to keep the results over rerunning and saving \$200.

C.4.5 Experiment 2 Details

Information Treatment. We begin by eliciting beliefs about the percent of tenants expected to choose cash y over the good. The cash value is randomized $y \in \{\$200, \$300\}$. We ask participants to guess how many tenants, among 100 tenants who apply for assistance, would choose $\$y$ over the good. We then truthfully inform participants: "Researchers who work with the nonprofit asked 10 tenants whether they would choose a [good] or [$\$y$] in cash. All of them chose to receive [$\$y$] over a [good]." We ensured the truthfulness of this information treatment using additional Prolific experiments, where we screened for tenants and asked about their preferences for all combinations of goods and cash y values. At least 10 tenants for each combination preferred cash. After sharing this information, we elicit posterior beliefs.

Elicitation. Following the information treatment and posterior elicitation, we proceed with the elicitation of WTP for dignity of choice. We begin by informing tenants that the good typically costs \$350, and that the current budget of the nonprofit allocates y in cash to the tenant and saves the rest for future programs. We first ask Spectators if they would prefer to give the tenant \$$y$ and save \$100, or give the tenant the choice between \$$y$ and the good. We repeatedly ask Spectators the same question in increments of \$100, until either they elect to give \$$y$ and save or they prefer to give the tenant the choice over giving \$$y$ and saving \$900.

C.4.6 Experiment 2: Selection on Gains Details

We test for selection on gains by randomizing the value of the bundle $\$y \in \{200, 300\}$ and presenting Spectators' implied valuation of the good (Figure C.2). Intuitively, randomizing \$$y$ identifies Spectators' supply curve for providing choice. If the distribution of the valuation of g does not vary with \$$y$, holding beliefs fixed, then that argues against substantial perceived selection on gains.

To see this, observe that:

$$E[\text{choice}|\$y = 200]_i := \int_{200}^{\infty} a(u(g|g \succ \$y))_i f(g|g \succ \$y)_i dy = C_i + 200 \tag{C.3}$$

$$E[\text{choice}|\$y = 300]_i := \int_{300}^{\infty} a(u(g|g \succ \$y))_i f(g|g \succ \$y)_i dy = C_i + 300. \tag{C.4}$$

In these expressions, $a(u(g|r \succ \$y))_i$ refers to i's allocative utility f from giving the recipient utility u from choosing g over y. The probability distribution function $f(\cdot)$ embeds the chance that i would choose g over y.

The key idea of our test is that fixing $f(\cdot)$,

$$E[\text{choice}|\$y = 300]_i - E[\text{choice}|\$y = 200]_i \equiv E[a(u(r|r \succ 300)) - a(u(r|r \succ 200))]_i. \tag{C.5}$$

This expression says that, holding constant beliefs about the probability of choosing g, the difference in the expected value of choice is equivalent to the difference in selection on gains.

We observe $f(\cdot)$ directly. Thus, we can consider the distribution of $C_i + \$y$ for beliefs within a small range of $f(\cdot)$.

We find no reason to be concerned about selection on gains (Figure C.2). Kolmogorov-Smirnov tests do not permit us to reject equality, and indeed, the distribution of valuations for high \$$y$ lies *below* that for low \$$y$, which suggests selection on *losses*. Such selection would only amplify our results.

C.4.7 Experiment 4 Details

Information Treatment. Following the targeting elicitation (described in detail in Section 3.5.1, we randomly provide Spectators doing experiments with rights goods either a high or low information treatment about the instrumental benefit of the good. For lawyers, we elicit priors in two parts: first asking how many out of 100 tenants without a lawyer receive an eviction judgement, and second asking the same for tenants *with* lawyers. We then tell Spectators in the [high/low] treatments: “Researchers studied a program that is providing lawyers to tenants facing eviction in Memphis. Among 100 of the tenants, having a lawyer led to a [80/20]% reduction in eviction rates. About 55% of tenants who did not receive a lawyer from the program were evicted in court, but only about [15/45]% of tenants who did receive one were.” These estimates are based on an ongoing RCT of providing lawyers to tenants facing eviction in Memphis, TN (Caspi and Rafkin, 2023).

For health care, we ask Spectators how many of 100 tenants without health care vouchers will have improved health outcomes 1 year later. We tell Spectators positive and null results from the Oregon Health Insurance Experiment (Baicker et al., 2013; Allen et al., 2013). In the high information treatment, Spectators see: “Researchers studied Medicaid expansion in Oregon and found that among people who newly gained access to Medicaid, rates of depression fell by 9 percentage points and increased the likelihood of self-reporting health as good, very good, or excellent by 13 percentage points.” In the low information treatment, Spectators see: “Researchers studied Medicaid expansion in Oregon and found that among people who newly gained access to Medicaid, it did not have a significant effect on measured blood pressure or cholesterol.”

Identifying Welfarists. Following the information, we give Spectators the opportunity to revise their targeting choices. We use participants’ decision whether or not to revise their allocation when information conflicts with beliefs to identify welfarists. For lawyers, elicited priors are directly comparable with the information provided and we classify each participant as seeing information (80% or 20%) above or below their priors. We analogously classify health care participants, comparing their priors about improved health outcomes one year later to an 80% effect for the “high” information and 20% for the “low” information. Therefore, welfarists are those who chose to revise their choices and either: initially distribute universally and see information below priors, initially distribute only to the poorest tenant and see information above priors, or initially distribute to the $R_i \in \{2, \ldots, 9\}$ poorest tenants. We cannot classify those who initially distribute universally and see information above priors or initially distribute only to the poorest tenant and see information below priors as either welfarist or non-welfarist.

C.4.8 Direct WTP Elicitation Details

To address the potential for anchoring, we randomize the initial choice participants see when eliciting their WTP directly for their assigned good. We ask if they would prefer to give one tenant the good or cash, randomizing the initial cash value from {$300, $500, $700}. We ask the same question in increments of $100 until we identify their indifference point.

Bibliography

Abbink, Klaus and Abdolkarim Sadrieh, "The pleasure of being nasty," *Economics letters*, 2009, *105* (3), 306–308.

_ **and Benedikt Herrmann**, "The moral costs of nastiness," *Economic inquiry*, 2011, *49* (2), 631–633.

Abramson, Boaz, "The Welfare Effects of Eviction and Homelessness Policies," 2021.

Abreu, Dilip and Faruk Gul, "Bargaining and Reputation," *Econometrica*, January 2000, *68* (1), 85–117.

Acemoglu, Daron, Simon Johnson, and James A Robinson, "Institutions as a fundamental cause of long-run growth," *Handbook of economic growth*, 2005, *1*, 385–472.

ACLU, "No Eviction Without Representation," Technical Report, American Civil Liberties Union 2022.

Aiken, Claudia, Isabel Harner, Vincent Reina, Andrew Aurand, and Rebecca Yae, "Emergency Rental Assistance (ERA) During the Pandemic: Implications for the Design of Permanent ERA Programs," Technical Report, Housing Initiative at Penn and National Low Income Housing Coalition, Philadelphia March 2022.

Aizer, Anna, "Low Take-Up in Medicaid: Does Outreach Matter and for Whom?," *American Economic Review*, 2003, *93* (2).

_ **and Jeffrey Grogger**, "Parental Medicaid Expansions and Health Insurance Coverage," Working Paper 9907, National Bureau of Economic Research 2003.

Akers, Joshua and Eric Seymour, "Instrumental Exploitation: Predatory Property Relations at City's End," *Geoforum*, May 2018, *91*, 127–140.

Al-Athari, Faris Muslim, "Estimation of the Mean of Truncated Exponential Distribution," *Journal of Mathematics and Statistics*, 2008, *4* (4), 284–288.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, and Julia Tobias, "Targeting the poor: evidence from a field experiment in Indonesia," *American Economic Review*, 2012, *102* (4), 1206–1240.

Allcott, Hunt, "Social norms and energy conservation," *Journal of public Economics*, 2011, *95* (9-10), 1082–1095.

_ , Daniel Cohen, William Morrison, and Dmitry Taubinsky, "When do" Nudges" Increase Welfare?," Technical Report, National Bureau of Economic Research 2022.

Allen, Heidi, Katherine Baicker, Mira Bernstein, Amy Finkelstein, Jonathan Gruber, Joseph P. Newhouse, Eric Schneider, Song, Sarah Taubman, Bill Wright, and Alan Zaslavsky, "The Oregon Health Insurance Experiment in the United States | The Abdul Latif Jameel Poverty Action Lab," 2013.

Almond, Douglas, Hillary W. Hoynes, and Diane Whitmore Schanzenbach, "Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes," *Review of Economics and Statistics*, 2011, *93* (2), 387–403.

Ambuehl, Sandro, B Douglas Bernheim, and Axel Ockenfels, "What motivates paternalism? An experimental study," *American economic review*, 2021, *111* (3), 787–830.

Andersen, Steffen, Cristian Badarinza, Lu Liu, Julie Marx, and Tarun Ramadorai, "Reference Dependence in the Housing Market," *American Economic Review*, October 2022, *112* (10), 3398–3440.

_ , Seda Ertaç, Uri Gneezy, Moshe Hoffman, and John A List, "Stakes Matter in Ultimatum Games," *American Economic Review*, December 2011, *101* (7), 3427–3439.

Andreoni, James, "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *Economic Journal*, 1990, *100* (401), 464–477.

_ , "An Experimental Test of the Public-Goods Crowding-Out Hypothesis," *American Economic Review*, 1993, *83* (5), 1317–1327.

_ **and John Miller**, "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica*, March 2002, *70* (2), 737–753.

_ , Deniz Aydin, Blake Barton, B. Douglas Bernheim, and Jeffrey Naecker, "When Fair Isn't Fair: Understanding Choice Reversals Involving Social Preferences," *Journal of Political Economy*, 2020, *128* (5), 1673–1711.

Angrist, Joshua D and Jörn-Steffen Pischke, *Mostly harmless econometrics: An empiricist's companion*, Princeton university press, 2009.

Armona, Luis, Andreas Fuster, and Basit Zafar, "Home price expectations and behaviour: Evidence from a randomized information experiment," *The Review of Economic Studies*, 2019, *86* (4), 1371–1410.

Ashenfelter, Orley, "Estimating the Effect of Training Programs on Earnings," *The Review of Economics and Statistics*, February 1978, *60* (1), 47.

Ashraf, Nava and Oriana Bandiera, "Social incentives in organizations," *Annual Review of Economics*, 2018, *10*, 439–463.

Atasoy, Sibel, "The End of the Paper Era in the Food Stamp Program: The Impact of Electronic Benefits on Program Participation," 2009.

Aubry, Tim, Rebecca Cherner, John Ecker, Jonathan Jetté, Jennifer Rae, Stephanie Yamin, John Sylvestre, Jimmy Bourque, and Nancy McWilliams, "Perceptions of Private Market Landlords Who Rent to Tenants of a Housing First Program," *American Journal of Community Psychology*, June 2015, *55* (3-4), 292–303.

Auray, Stéphane and David L. Fuller, "Eligibility, Experience Rating, and Unemployment Insurance Take-up," *Quantitative Economics*, 2020, *11* (3), 1059–1107.

Ausubel, Lawrence M, Peter Cramton, and Raymond J Deneckere, "Bargaining with incomplete information," *Handbook of game theory with economic applications*, 2002, *3*, 1897–1945.

Backus, Matthew, Thomas Blake, Brad Larsen, and Steven Tadelis, "Sequential Bargaining in the Field: Evidence from Millions of Online Bargaining Interactions," *Quarterly Journal of Economics*, 2020, *135* (3), 1319–1361.

Baicker, Katherine, Sarah Taubman, Heidi Allen, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Eric Schneider, Bill Wright, Alan Zaslavsky, and Amy Finkelstein, "The Oregon experiment—effects of Medicaid on clinical outcomes," *New England Journal of Medicine*, 2013, *368* (18), 1713–1722.

Bailey, Martha J., Hillary W. Hoynes, Maya Rossin-Slater, and Reed Walker, "Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program," National Bureau of Economic Research Working Paper 26942, Cambridge, MA April 2020.

Baily, Martin, "Some Aspects of Optimal Unemployment Insurance," *Journal of Public Economics*, 1978, *10* (3), 379–402.

Balzarini, John and Melody L Boyd, "Working with them: Small-scale landlord strategies for avoiding evictions," *Housing Policy Debate*, 2021, *31* (3-5), 425–445.

Bartfeld, Judith, Craig Gundersen, Timothy M. Smeeding, and James P. Ziliak, eds, *SNAP Matters: How Food Stamps Affect Health and Well-Being*, Stanford University Press, 2016.

Bartlett, Susan, Nancy Burstein, and William Hamilton, "Food Stamp Program Access Study: Final Report," Technical Report, USDA Economic Research Service, https://www.ers.usda.gov/webdocs/publications/43390/30283_efan03013-3_002.pdf?v=0 2004.

Bartling, Björn, Alexander W Cappelen, Henning Hermes, and Bertil Tungodden, "Free to fail? Paternalistic preferences in the United States," *NHH Dept. of Economics Discussion Paper*, 2023, (09).

_ , Ernst Fehr, and Holger Herz, "The intrinsic value of decision rights," *Econometrica*, 2014, *82* (6), 2005–2039.

Bartling, Björn, Ernst Fehr, and Holger Herz, "The Intrinsic Value of Decision Rights," *Econometrica*, 2014, *82* (6), 2005–2039.

Becker, Gary S, "A Theory of Social Interactions," *Journal of Political Economy*, 1974, *82* (6), 1063–1093.

Bell, Deborah Hodges, "Providing security of tenure for residential tenants: Good faith as a limitation on the landlord's right to terminate," *Ga. L. Rev.*, 1984, *19*, 483.

Belloni, A., V. Chernozhukov, and C. Hansen, "Inference on Treatment Effects after Selection among High-Dimensional Controls," *The Review of Economic Studies*, April 2014, *81* (2), 608–650.

Bénabou, Roland and Jean Tirole, "Identity, morals, and taboos: Beliefs as assets," *The Quarterly Journal of Economics*, 2011, *126* (2), 805–855.

Bernheim, B Douglas, "The good, the bad, and the ugly: A unified approach to behavioral welfare economics1," *Journal of Benefit-Cost Analysis*, 2016, *7* (1), 12–68.

_ and Antonio Rangel, "Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics," *The Quarterly Journal of Economics*, 2009, *124* (1), 51–104.

Bernheim, B. Douglas and Dmitry Taubinsky, "Behavioral Public Economics," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 1, Elsevier, 2018, pp. 381–516.

Bernheim, B Douglas and Dmitry Taubinsky, "Behavioral public economics," *Handbook of behavioral economics: Applications and Foundations 1*, 2018, *1*, 381–516.

Bertrand, Marianne, Erzo F. P. Luttmer, and Sendhil Mullainathan, “Network Effects and Welfare Cultures,” *Quarterly Journal of Economics*, August 2000.

Bezdek, Barbara, “Silence in the Court: Participation and Subordination of Poor Tenants’ Voices in the Legal Process,” *Hofstra Lawy Review*, 1992, *20* (3), 533–608.

Bhargava, Saurabh and Dayanand Manoli, “Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment,” *American Economic Review*, November 2015, *105* (11), 1–42.

Blouin, Arthur, “Culture and contracts: The historical legacy of forced labour,” *The Economic Journal*, 2022, *132* (641), 89–105.

Bó, Pedro Dal, “Experimental evidence on the workings of democratic institutions,” *Institutions, property rights, and economic growth: The legacy of Douglass North*, 2014, *266.*

_ , Andrew Foster, and Louis Putterman, “Institutions and behavior: Experimental evidence on the effects of democracy,” *American Economic Review*, 2010, *100* (5), 2205–2229.

Bobadilla-Suarez, Sebastian, Cass R Sunstein, and Tali Sharot, “The intrinsic value of choice: The propensity to under-delegate in the face of potential gains and losses,” *Journal of risk and uncertainty*, 2017, *54*, 187–202.

Bottan, Nicholas L. and Ricardo Perez-Truglia, “Betting on the House: Subjective Expectations and Market Choices,” Technical Report 27412, National Bureau of Economic Research, Cambridge, MA 2021.

Bronchetti, Erin T., Garret Christensen, and Hilary W. Hoynes, “Local Food Prices, SNAP Purchasing Power, and Child Health,” *Journal of Health Economics*, 2019, *68.*

Brooks, Tricia, Sean Miskell, Samantha Artiga, Elizabeth Cornachione, and Alexandra Gates, “Medicaid and CHIP Eligibility, Enrollment, Renewal, and Cost-Sharing Policies as of January 2016: Findings from a 50-State Survey,” Technical Report, The Henry J. Kaiser Family Foundation 2016.

Bruins, Marianne, James A. Duffy, Michael P. Keane, and Anthony A. Smith, “Generalized Indirect Inference for Discrete Choice Models,” *Journal of Econometrics*, July 2018, *205* (1), 177–203.

Burbidge, John B., Lonnie Magee, and A. Leslie Robb, “Alternative Transformations to Handle Extreme Values of the Dependent Variable,” *Journal of the American Statistical Association*, 1988, *83* (401), 123–127.

Bursztyn, Leonardo, Alessandra González, and David Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, 2020, *119* (10), 2997–3029.

_ , _ , and _ , "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, Forthcoming.

_ and Robert Jensen, "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure," *Annual Review of Economics*, 2017, *9* (1), 131–53.

Callaway, Brantley and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," 2020.

Camerer, Colin, *Behavioral Game Theory: Experiments in Strategic Interaction* The Roundtable Series in Behavioral Economics, New York, N.Y. : Princeton, N.J: Russell Sage Foundation ; Princeton University Press, 2003.

Cappelen, Alexander W, James Konow, Erik Ø Sørensen, and Bertil Tungodden, "Just luck: An experimental study of risk-taking and fairness," *American Economic Review*, 2013, *103* (4), 1398–1413.

_ , Ranveig Falch, and Bertil Tungodden, "Fair and unfair income inequality," *Handbook of Labor, Human Resources and Population Economics*, 2020, pp. 1–25.

Caspi, Aviv and Charlie Rafkin, "Legal Assistance for Evictions: Impacts and Demand," 2023.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, "The Effect of Minimum Wages on Low-Wage Jobs," *Quarterly Journal of Economics*, August 2019, *134* (3), 1405–1454.

_ , _ , _ , and _ , "The effect of minimum wages on low-wage jobs," *The Quarterly Journal of Economics*, 2019, *134* (3), 1405–1454.

Chandrasekhar, Arun G., Benjamin Golub, and He Yang, "Signaling, Shame, and Silence," Technical Report 25169, National Bureau of Economic Research 2019.

Charité, Jimmy, Raymond Fisman, Ilyana Kuziemko, and Kewei Zhang, "Reference points and redistributive preferences: Experimental evidence," *Journal of Public Economics*, 2022, *216*, 104761.

Charness, Gary and Matthew Rabin, "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 2002, *117* (3), 817–869.

__ , **Uri Gneezy, and Vlastimil Rasocha**, "Experimental methods: Eliciting beliefs," *Journal of Economic Behavior & Organization*, 2021, *189*, 234–256.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins, "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, 2018, *21* (1), C1–C68.

Chetty, Raj, "A General Formula for the Optimal Level of Social Insurance," *Journal of Public Economics*, 2006, *90* (10-11), 1879–1901.

__ , "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy*, April 2008, *116* (2), 173–234.

__ , "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 2009, *1* (1), 451–487.

__ , "Behavioral economics and public policy: A pragmatic perspective," *American Economic Review*, 2015, *105* (5), 1–33.

__ **and Amy Finkelstein**, "Social Insurance: Connecting Theory to Data," in "Handbook of Public Economics," Vol. 5, Elsevier, 2013, pp. 111–193.

Child Trends, "Key facts about Head Start enrollment," 2018.

Chisholm, Elinor, Philipa Howden-Chapman, and Geoff Fougere, "Tenants' Responses to Substandard Housing: Hidden and Invisible Power and the Failure of Rental Housing Regulation," *Housing, Theory and Society*, March 2020, *37* (2), 139–161.

Chopra, Felix, Christopher Roth, and Johannes Wohlfart, "Home Price Expectations and Spending: Evidence from a Field Experiment," *Available at SSRN 4452588*, 2023.

Clarke, Damian, Joseph P Romano, and Michael Wolf, "The Romano–Wolf multiple-hypothesis correction in Stata," *The Stata Journal*, 2020, *20* (4), 812–843.

Cohen, Elior, "Housing the Homeless: The Effect of Housing Assistance on Recidivism to Homelessness, Economic, and Social Outcomes," Technical Report, mimeo 2020.

Collinson, Robert, Anthony Defusco, Ben Keys, Humphries John Eric, David Phillips, Vincent Reina, Patrick Turner, and Winnie van Dijk, "Emergency Assistance Grants and Household Stability: Evidence from COVID Assistance Lotteries," Technical Report 2023.

__ , **John Eric Humphries, Nicholas Mader, Davin Reed, Daniel Tannenbaum, and Winnie van Dijk**, "Eviction and Poverty in American Cities," *Quarterly Journal of Economics*, 2024.

Congressional Research Service, "The Supplemental Nutrition Assistance Program (SNAP): Categorical Eligibility," Technical Report R42054 October 2019.

Cossyleon, Jennifer E., Philip ME Garboden, and Stefanie DeLuca, "Recruiting Opportunity Landlords: Lessons from Landlords in Maryland," Technical Report, Poverty & Race Research Action Council 2020.

Cunnyngham, Karen, "Reaching Those in Need: Estimates of State Supplemental Nutrition Assistance Program Participation Rates in 2016q," Technical Report, United States Department of Agriculture 2019.

Currie, Janet, "U.S. Food and Nutrition Programs," in Robert A. Moffitt, ed., *Means-Tested Transfer Programs in the United States*, Chicago: University of Chicago Press, 2003.

_ , "The Take Up of Social Benefits," Technical Report 10488, National Bureau of Economic Research, Cambridge, MA May 2004.

_ **and Firouz Gahvari**, "Transfers in cash and in-kind: Theory meets the data," *Journal of economic literature*, 2008, *46* (2), 333–383.

Currie, Janet M. and Jeff Grogger, "Explaining Recent Declines in Food Stamp Program Participation," *Brookings-Wharton Papers on Urban Affairs*, 2001, *2001* (1), 203–244.

Dahl, Gordon B., Katrine V. Løken, and Magne Mogstad, "Peer Effects in Program Participation," *American Economic Review*, 2014, *104* (7), 2049–2074.

Dannenberg, Astrid and Carlo Gallier, "The choice of institutions to solve cooperation problems: a survey of experimental research," *Experimental Economics*, 2020, *23* (3), 716–749.

Danz, David, Lise Vesterlund, and Alistair J Wilson, "Belief elicitation and behavioral incentive compatibility," *American Economic Review*, 2022, *112* (9), 2851–2883.

Daponte, Beth Osborne, Seth Sanders, and Lowell Taylor, "Why Do Low-Income Households Not Use Food Stamps? Evidence from an Experiment," *Journal of Human Resources*, 1999, *34* (3), 612–618.

Dávila, Eduardo, "Using elasticities to derive optimal bankruptcy exemptions," *The Review of Economic Studies*, 2020, *87* (2), 870–913.

DellaVigna, Stefano, "Chapter 7 - Structural Behavioral Economics," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 1 2018, pp. 613–723.

_ **, John A. List, and Ulrike Malmendier**, "Testing for altruism and social pressure in charitable giving," *Quarterly Journal of Economics*, 2012, *127* (1), 1–56.

DeLuca, Stefanie and Eva Rosen, "Housing Insecurity Among the Poor Today," *Annual Review of Sociology*, 2022, *48.*

Desmond, Matthew, *Evicted: Poverty and Profit in the American City*, Crown, 2016.

— **and Tracey Shollenberger**, "Forced displacement from rental housing: Prevalence and neighborhood consequences," *Demography*, 2015, *52* (5), 1751–1772.

Diamond, Peter and Eytan Sheshenski, "Economic Aspects of Optimal Disability Benefits," *Journal of Public Economics*, May 1995, *57* (1), 1–23.

East, Chloe N., "Immigrants' Labor Supply Response to Food Stamp Access," *Labour Economics*, 2018, *51* (202-226).

Eck, Chase S., "The Effect of Electronic Benefit Transfer on the Marginal Propensity to Consume Food out of SNAP," 2018, p. 43.

Engel, Christoph, "Dictator Games: A Meta Study," *Experimental Economics*, November 2011, *14* (4), 583–610.

Enke, Benjamin and Thomas Graeber, "Cognitive uncertainty," Technical Report, National Bureau of Economic Research 2019.

Eslami, Esa, "Trends in Supplemental Nutrition Assistance Program Participation Rates: Fiscal Year 2010 to Fiscal Year 2013," Technical Report, United States Department of Agriculture, Washington, D.C. August 2015.

Ewens, Michael, Bryan Tomlin, and Liang Choon Wang, "Statistical Discrimination or Prejudice? A Large Sample Field Experiment," *The Review of Economics and Statistics*, March 2014, *96* (1), 119–134.

Fairweather, Daryl, Matthew E Kahn, Robert D Metcalfe, and Sebastian Sandoval-Olascoaga, "The Impact of Climate Risk Disclosure on Housing Search and Buying Dynamics: Evidence from a Nationwide Field Experiment with Redfin," 2023.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global Evidence on Economic Preferences," *The Quarterly Journal of Economics*, November 2018, *133* (4), 1645–1692.

— , — , — , David Huffman, and Uwe Sunde, "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences.," *Management Science*, Forthcoming.

Federal Register, "Revision of Categorical Eligibility in the Supplemental Nutrition Assistance Program (SNAP)," July 2019, *84* (142), 35570–35581.

Fehr, Ernst and Klaus M Schmidt, "A theory of fairness, competition, and cooperation," *The quarterly journal of economics*, 1999, *114* (3), 817–868.

Fetter, Daniel K. and Lee M. Lockwood, "Government Old-Age Support and Labor Supply: Evidence from the Old Age Assistance Program," *American Economic Review*, August 2018, *108* (8), 2174–2211.

Fetzer, Thiemo, Srinjoy Sen, and Pedro CL Souza, "Housing insecurity, homelessness and populism: Evidence from the UK," 2020.

Finkelstein, Amy and Matthew Notowidigdo, "Take-up and Targeting: Experimental Evidence from SNAP," *Quarterly Journal of Economics*, 2019, *134* (3).

Fisman, Raymond, Shachar Kariv, and Daniel Markovits, "Individual Preferences for Giving," *American Economic Review*, December 2007, *97* (5), 1858–1876.

Frean, Molly, Jonathan Gruber, and Benjamin D. Sommers, "Premium Subsidies, the Mandate, and Medicaid Expansion: Coverage Effects of the Affordable Care Act," *Journal of Health Economics*, 2017, *53*, 72–86.

Freyberger, Joachim and Bradley Larsen, "How Well Does Bargaining Work in Consumer Markets? A Robust Bounds Approach," Technical Report, National Bureau of Economic Research 2021.

Friedberg, Leora and Steven Stern, "Marriage, divorce, and asymmetric information," *International Economic Review*, 2014, *55* (4), 1155–1199.

Friedman, David D, "Does altruism produce efficient outcomes? Marshall versus Kaldor," *The Journal of Legal Studies*, 1988, *17* (1), 1–13.

Friedrichsen, Jana, Tobias König, and Renke Schmacker, "Social Image Concerns and Welfare Take-Up," *Journal of Public Economics*, December 2018, *168*, 174–192.

Fuster, Andreas and Basit Zafar, "Survey Experiments on Economic Expectations," Technical Report w29750, National Bureau of Economic Research, Cambridge, MA February 2022.

Ganong, Peter and Jeffrey B. Liebman, "The Decline, Rebound, and Further Rise in SNAP Enrollment: Disentangling Business Cycle Fluctuations and Policy Changes," *American Economic Journal: Economic Policy*, November 2018, *10* (4), 153–176.

Garboden, Philip ME and Eva Rosen, "Serial filing: How landlords use the threat of eviction," *City & Community*, 2019, *18* (2), 638–661.

Gasparini, Leonardo C and Santiago M Pinto, "Equality of opportunity and optimal cash and in-kind policies," *Journal of Public Economics*, 2006, *90* (1-2), 143–169.

Geddes, Eilidh and Nicole Holz, "Rational Eviction: How Landlords Use Evictions in Response to Rent Control," *Available at SSRN 4131396*, 2022.

Giannarelli, Linda, "What Was the TANF Participation Rate in 2016?," Technical Report, Urban Institute 2019.

_ , Christine Heffernan, Sarah Minton, Megan Thompson, and Kathryn Stevens, "Welfare Rules Databook: State TANF Policies as of July 2016," Technical Report, Urban Institute 2017.

Gilderbloom, John I., "Social Factors Affecting Landlords In The Determination of Rent," *Urban Life*, 1985, *14* (2), 155–179.

Golosov, Mikhail and Aleh Tsyvinski, "Designing Optimal Disability Insurance: A Case for Asset Testing," *Journal of Political Economy*, 2005, *114* (2), 257–279.

Graetz, Nick, Carl Gershenson, Peter Hepburn, Sonya R Porter, Danielle H Sandler, and Matthew Desmond, "A comprehensive demographic profile of the US evicted population," *Proceedings of the National Academy of Sciences*, 2023, *120* (41), e2305860120.

Gromis, Ashley and Matthew Desmond, "Estimating the prevalence of eviction in the United States," *Cityscape*, 2021, *23* (2), 279–290.

_ , Ian Fellows, James R Hendrickson, Lavar Edmonds, Lillian Leung, Adam Porton, and Matthew Desmond, "Estimating eviction prevalence across the United States," *Proceedings of the National Academy of Sciences*, 2022, *119* (21), e2116169119.

Gruber, Jonathan, "The Consumption Smoothing Benefits of Unemployment Insurance," *American Economic Review*, 1997, *87* (1), 192–205.

Haag, Matthew, "A Landlord Says Her Tenants Are Terrorizing Her. She Can't Evict Them.," *The New York Times*, July 2021.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart, "Designing Information Provision Experiments," *Journal of Economic Literature*, 2023.

Hainmueller, Jens, "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies," *Political Analysis*, 2012, *20* (1), 25–46.

Haley, Jennifer M., Genevieve M. Kenney, Robin Wang, Victoria Lynch, and Matthew Buettgens, "Medicaid/CHIP Participation Reached 83.7 Percent Among Eligible Children in 2016," *Health Affairs*, 2018, *37* (8), 1194–1199.

Hanna, Rema and Benjamin A. Olken, "Universal Basic Incomes versus Targeted Transfers: Anti-Poverty Programs in Developing Countries," *Journal of Economic Perspectives*, November 2018, *32* (4), 201–226.

Harris, Timothy F., "Do SNAP Work Requirements Work?," *Economic Inquiry*, 2021, *59*, 72–94.

Hastings, Justine and Jesse M. Shapiro, "How are SNAP Benefits Spent? Evidence from a Retail Panel," *American Economic Review*, 2018, *108* (12), 3493–3540.

_ , Ryan Kessler, and Jesse M. Shapiro, "The Effect of SNAP on the Composition of Purchased Foods: Evidence and Implications," *American Economic Journal: Economic Policy*, Forthcoming.

Hausman, Daniel M and Michael S McPherson, "Taking ethics seriously: economics and contemporary moral philosophy," *Journal of economic literature*, 1993, *31* (2), 671–731.

Heckman, James J. and Jeffrey A. Smith, "The Determinants of Participation in a Social Program: Evidence from a Prototypical Job Training Program," *Journal of Labor Economics*, 2004, *22* (2), 243–98.

Hendren, Nathaniel, "The Policy Elasticity," *Tax Policy and the Economy*, 2016, *30.*

_ and Ben Sprung-Keyser, "A Unified Welfare Analysis of Government Policies," *Quarterly Journal of Economics*, 2020, *135* (3), 1209–1318.

_ , Camille Landais, and Johannes Spinnewijn, "Choice in Insurance Markets: A Pigouvian Approach to Social Insurance Design," *Annual Review of Economics*, Forthcoming.

Hjort, Jonas, "Ethnic Divisions and Production in Firms*," *The Quarterly Journal of Economics*, November 2014, *129* (4), 1899–1946.

Holmes, Stephen and Cass R Sunstein, *The cost of rights: why liberty depends on taxes*, WW Norton & Company, 2000.

Homonoff, Tatiana and Jason Somerville, "Program Recertification Costs: Evidence from SNAP," *American Economic Journal: Economic Policy*, November 2021, *13* (4), 271–298.

Hoy, Michael and Emmanuel Jimenez, "Squatters' rights and urban development: an economic perspective," *Economica*, 1991, pp. 79–92.

Hoynes, Hilary and Diane Whitmore Schanzenbach, "Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program," *American Economic Journal: Applied Economics*, 2009, *1* (4), 109–139.

— **and** — , "Work incentives and the Food Stamp Program," *Journal of Public Economics*, 2012, *96*, 151–162.

— **,** — **, and Douglas Almond**, "Long-Run Impacts of Childhood Access to the Safety Net," *American Economic Review*, 2016, *106* (4), 903–34.

Hutto, C. J. and Eric Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," *Proceedings of the International AAAI Conference on Web and Social Media*, 2014, *2* (1), 216–225.

Internal Revenue Service, "About EITC," 2020.

Kaplow, Louis, "Optimal income taxation," Technical Report, National Bureau of Economic Research 2022.

— **and Steven Shavell**, "Any non-welfarist method of policy assessment violates the Pareto principle," *Journal of Political Economy*, 2001, *109* (2), 281–286.

Katz, Michael B., *In the Shadow of the Poorhouse: A Social History of Welfare in America*, Basic Books, Inc., 1986.

Keniston, Daniel, Bradley J. Larsen, Shengwu Li, J.J. Prescott, Bernardo S. Silveira, and Chuan Yu, "Fairness in Incomplete Information Bargaining: Theory and Widespread Evidence from the Field," Technical Report 2022.

Klerman, Jacob Alex and Caroline Danielson, "The Transformation of the Supplemental Nutrition Assistance Program," *Journal of Policy Analysis and Management*, September 2011, *30* (4), 863–888.

Kleven, Henrik, "Sufficient Statistics Revisited," *Annual Review of Economics*, 2021, *13* (1), 515–538.

Kleven, Henrik Jacobsen and Wojciech Kopczuk, "Transfer Program Complexity and the Take-Up of Social Benefits," *American Economic Journal: Economic Policy*, February 2011, *3* (1), 54–90.

Kling, Jeffrey R, Jeffrey B Leibman, and Lawrence F Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

— **, Jeffrey B Liebman, and Lawrence F Katz**, "Experimental analysis of neighborhood effects," *Econometrica*, 2007, *75* (1), 83–119.

Kőszegi, Botond and Matthew Rabin, "A model of reference-dependent preferences," *The Quarterly Journal of Economics*, 2006, *121* (4), 1133–1165.

Kranton, Rachel, Matthew Pease, Seth Sanders, and Scott Huettel, "Groupy and non-groupy behavior: Deconstructing bias in social preferences," *Work. Pap., Duke Univ., Durham NC*, 2016.

Kreider, Brent, John V. Pepper, Craig Gundersen, and Dean Jolliffe, "Identifying the Effects of SNAP (Food Stamps) on Child Health Outcomes When Participation Is Endogenous and Misreported," *Journal of the American Statistical Association*, 2012, *107* (499), 958–975.

Kroft, Kory, "Takeup, Social Multipliers and Optimal Social Insurance," *Journal of Public Economics*, April 2008, *92* (3-4), 722–737.

Krueger, Alan B and Bruce D Meyer, "Labor Supply Effects of Social Insurance," in Alan J. Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, Vol. 4, Elsevier, 2002, pp. 2327–2392.

Larsen, Bradley J., "The Efficiency of Real-World Bargaining: Evidence from Wholesale Used-Auto Auctions," *Review of Economic Studies*, 2021, *88* (2), 851–882.

Ledyard, John O., "Is There a Problem with Public Goods Provision," in "The Handbook of Experimental Economics" 1995, pp. 111–194.

Lenk, Alexandr, "Do People Respect Other's Freedom of Choice Out of Principle? An Experimental Investigation."

Leos-Urbel, Jacob, Amy Ellen Schwartz, Meryle Weinstein, and Sean Corcoran, "Not Just for Poor Kids: The Impact of Universal Free School Breakfast on Meal Participation and Student Outcomes," *Economics of Education Review*, October 2013, *36*, 88–107.

Levine, David K., "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics*, 1998, *1* (3), 593–622.

Levitt, Steven D. and John A. List, "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?," *The Journal of Economic Perspectives*, 2007, *21* (2), 153–174.

Levitt, Steven D and John A List, "What do laboratory experiments measuring social preferences reveal about the real world?," *Journal of Economic perspectives*, 2007, *21* (2), 153–174.

Lindbeck, Assar, Sten Nyberg, and Jorgen Weibull, "Social Norms and Economic Incentives in the Welfare State," *The Quarterly Journal of Economics*, February 1999, *114* (1), 1–35.

Liscow, Zachary and Abigail Pershing, "Why is so much redistribution in-kind and not in cash? Evidence from a survey experiment," *National Tax Journal*, 2022, *75* (2), 313–354.

List, John A, "On the interpretation of giving in dictator games," *Journal of Political economy*, 2007, *115* (3), 482–493.

_ , Matthias Rodemeier, Sutanuka Roy, and Gregory K Sun, "Judging nudging: Understanding the welfare effects of nudges versus taxes," Technical Report, National Bureau of Economic Research 2023.

Low, Hamish and Luigi Pistaferri, "Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off," *American Economic Review*, October 2015, *105* (10), 2986–3029.

Lowes, Sara, "Ethnographic and field data in historical economics," in "The Handbook of Historical Economics," Elsevier, 2021, pp. 147–177.

Luttmer, Erzo FP and Monica Singhal, "Tax morale," *Journal of economic perspectives*, 2014, *28* (4), 149–168.

Mabli, James, Thomas Godfrey, Nancy Wemmerus, Joshua Leftin, and Stephen Tordella, "Determinants of Supplemental Nutrition Assistance Program Participation from 2008 to 2012," Technical Report, United States Department of Agriculture Food and Nutrition Service 2014.

Manchester, Colleen Flaherty and Kevin J. Mumford, "How Costly Is Welfare Stigma? Separating Psychological Costs from Time Costs in Food Assistance Programs," 2012, p. 44.

Marcus, Michelle and Katherine G. Yewell, "The Effect of Free School Meals on Household Food Purchases: Evidence from the Community Eligibility Provision," Technical Report 2021.

Merlo, Antonio and Xun Tang, "Bargaining with optimism: Identification and estimation of a model of medical malpractice litigation," *International Economic Review*, 2019, *60* (3), 1029–1061.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan, "Household Surveys in Crisis," *Journal of Economic Perspectives*, November 2015, *29* (4), 199–226.

Mialon, Hugo M and Paul H Rubin, "The economics of the Bill of Rights," *American Law and Economics Review*, 2008, *10* (1), 1–60.

Moffitt, Robert, "An Economic Model of Welfare Stigma," *American Economic Review*, December 1983, *73* (5), 1023–1035.

Mullainathan, Sendhil and Eldar Shafir, *Scarcity: Why having too little means so much*, Macmillan, 2013.

Myerson, Roger B and Mark A Satterthwaite, "Efficient mechanisms for bilateral trading," *Journal of economic theory*, 1983, *29* (2), 265–281.

National Center for Education Statistics, "Digest of Education Statistics," Technical Report, U.S. Department of Education 2017.

Nichols, Albert L. and Richard J. Zeckhauser, "Targeting Transfers through Restrictions on Recipients," *American Economic Review Papers and Proceedings*, 1982, *72* (2), 372–377.

North, Douglass C, "Institutions," *Journal of economic perspectives*, 1991, *5* (1), 97–112.

Obama, Barack, "Remarks by the President on the Affordable Care Act," *The White House*, 2013.

Ortiz, Miguel, "Hate, Feat, and Intergroup Conflict: Experimental Evidence from Nigeria," 2023.

Palmer, Caroline, David C Phillips, and James X Sullivan, "Does emergency financial assistance reduce crime?," *Journal of Public Economics*, 2019, *169*, 34–51.

Pollak, Robert, "Family Bargaining with Altruism," Technical Report w30499, National Bureau of Economic Research, Cambridge, MA September 2022.

Polman, Evan, "Self–other decision making and loss aversion," *Organizational Behavior and Human Decision Processes*, 2012, *119* (2), 141–150.

Rabin, Matthew, "Incorporating fairness into game theory and economics," *The American economic review*, 1993, pp. 1281–1302.

—, "Moral preferences, moral constraints, and self-serving biases," 1995.

Ramos-Toro, Diego, "Social Exclusion and Social Preferences: Evidence from Colombia's Leper Colony," *Unpublished manuscript*, 2022.

Ratcliffe, Caroline, Signe-Mary McKernan, and Kenneth Finegold, "Effects of Food Stamp and TANF Policies on Food Stamp Receipt," *Social Service Review*, June 2008, *82* (2), 291–334.

—, —, Laura Wheaton, Emma Kalish, Catherine Ruggles, Sara Armstrong, and Christina Oberlin, "Asset Limits, SNAP Participation, and Financial Stability," Technical Report, Urban Institute, Washington, D.C. June 2016.

Rawls, John, "A theory of justice," *Cambridge (Mass.)*, 1971.

Romano, Joseph P and Michael Wolf, "Stepwise multiple testing as formalized data snooping," *Econometrica*, 2005, *73* (4), 1237–1282.

Ruggles, Stephen, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek, "IPUMS USA: Version 10.0," www.ipums.org 2020.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Danika Brockman, Grace Cooper, Stephanie Richards, and Megan Schouweiler, "IPUMS USA: Version 13.0 [dataset]," 2023.

Sacarny, Adam, Katherine Baicker, and Amy Finkelstein, "Out of the Woodwork: Enrollment Spillovers in the Oregon Health Insurance Experiment," *American Economic Journal: Economic Policy*, August 2022, *14* (3), 273–295.

Sadka, Joyce, Enrique Seira, and Christopher Woodruff, "Information and Bargaining through Agents: Experimental Evidence from Mexico's Labor Courts," Technical Report 2020.

Saez, Emmanuel and Stefanie Stantcheva, "Generalized social marginal welfare weights for optimal tax theory," *American Economic Review*, 2016, *106* (01), 24–45.

Sampford, M. R., "Some Inequalities on Mill's Ratio and Related Functions," *The Annals of Mathematical Statistics*, March 1953, *24* (1), 130–132.

Sen, Amartya, *Commodities and Capabilities*, North-Holland, 1985.

Shildrick, Tracy and Robert MacDonald, "Poverty talk: how people experiencing poverty deny their poverty and why they blame 'the poor'," *The Sociological Review*, 2013, *61* (2), 285–303.

Sommers, Benjamin D. and Arnold M. Epstein, "Why States Are So Miffed about Medicaid — Economics, Politics, and the "Woodwork Effect"," *New England Journal of Medicine*, July 2011, *365* (2), 100–102.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

Tobin, James, "On limiting the domain of inequality," *The Journal of Law and Economics*, 1970, *13* (2), 263–277.

Train, Kenneth E, *Discrete choice methods with simulation*, Cambridge university press, 2009.

United States Department of Agriculture Food and Nutrition Service, "SNAP Quality Control Data," https://snapqcdata.net/ 2019.

_ , "National School Lunch Program: Participation and Lunches Served," 2020.

_ , "WIC 2017 Eligibility and Coverage Rates," 2020.

U.S. Department of Agriculture Economic Research Service, "SNAP Policy Data Sets," https://www.ers.usda.gov/data-products/snap-policy-data-sets/ August 2019.

U.S. Department of Health and Human Services, "Head Start Impact Study. Final Report.," Technical Report, Administration for Children and Families 2010.

Vasserman, Shoshana and Muhamet Yildiz, "Pretrial negotiations under optimism," *The RAND Journal of Economics*, 2019, *50* (2), 359–390.

Vaughn, Ted, "The Landlord-Tenant Relation in a Low-Income Area," *Social Forces*, 1964, *16* (2), 208–218.

Wood, Thomas and Ethan Porter, "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence," *Political Behavior*, March 2019, *41* (1), 135–163.

Yildiz, Muhamet, "Bargaining with Optimism," *Annual Review of Economics*, September 2011, *3* (1), 451–478.