Obviously Strategy-Proof Task Allocation*

Quitzé Valenzuela-Stookey† Jason Baron‡ Richard Lombardo§

August 23, 2026

Preliminary and incomplete

Abstract

We study strategically simple mechanisms for multi-dimensional screening when agents have linear preferences and the right to demand a status-quo allocation. We show that static obviously strategy-proof (OSP) mechanisms take a simple form in which agents choose from a diverse menu of narrowly defined veto rights. We call these *polarized-ray mechanisms*. We extend the results to general extensive-form games. We then characterize the convex structure of polarized-ray mechanisms and develop tools for optimizing over this set. Finally, we examine the potential of this simple class of mechanisms as a practical market-design tool and provide quantitative evidence of their potential in two empirical applications.

*We are grateful for helpful comments from Federico Echenique, Francesco Fabbri and Fabio Maccheroni; and to seminar participants at UC Berkeley and Theros Conference 2026.

†Department of Economics, University of California, Berkeley. Email: quitze@berkeley.edu.

‡Department of Economics, Duke University and NBER. Email: jason.baron@duke.edu.

§Department of Economics, Harvard University. Email: richardlombardo@fas.harvard.edu.

Multi-dimensional screening problems, in which heterogeneous objects must be allocated among agents with unknown multi-dimensional preferences, arise frequently in organizations. Such problems are notoriously difficult to analyze in generality due to the complexity of agents' incentive constraints – a problem that is most acute when analyzing mechanisms under a permissive solution concept, such as Bayes-Nash equilibrium (Rochet, 1987; Lahr and Niemeyer, 2024). Applying a more demanding solution concept, such as strategic dominance, is one avenue along which progress on multi-dimensional screening can be made (e.g. Gale and Shapley (1962); Shapley and Scarf (1974)). Not only is the reduction of strategic complexity in such mechanisms normatively appealing; from a positive perspective they may provide a better model of the types of mechanisms used in practice.

In this paper we study obviously strategy-proof mechanisms (Li, 2017) for an important class of multi-dimensional task-allocation problems. Informally, a mechanism is obviously strategy-proof (OSP) if the worst outcome any agent could obtain if they follow the strategy prescribed by the mechanism is better than the best possible outcome from any deviation (Li, 2017). OSP is a strengthening of dominant-strategy incentive compatibility, and there is extensive experimental evidence that OSP implementation is more robust to agents' levels of strategic sophistication (Li, 2017; Breitmoser and Schweighofer-Kodritsch, 2022).

The setting that we study is as follows. A designer has a continuum of tasks (or objects) to allocate among a set of $J < \infty$ agents, where each task is of one of $I < \infty$ types.[1] Agents have preferences over bundles of tasks, which are their private information. Each task must be assigned to one agent. Each agent has an outside option assignment of tasks, denoted by $\sigma_j \in \mathbb{R}^I_{\geq 0}$, which they have the right to demand.[2] Thus, agents' preferences shape the set of implementable allocations, and it may benefit the designer to use a mechanism that responds to agents' preferences (we discuss possible designer objectives below). Examples of such problems include the assignment of Child Protective Services investigators to cases (Baron et al., 2026), prosecutors to cases (Agan et al., 2023), tax collectors to households

[1] Equivalently, we can interpret the model as one with a finite set of tasks, rather than a continuum, by allowing for random allocations. Under this interpretation, the measure of task type i assigned to agent j represents the expected number of task i that j receives.

[2] For example, an agent's outside option could represent the terms of their current contract. More abstractly, the outside option could reflect the designer's notion of a fair workload: the designer wishes to assign tasks efficiently, without unfairly burdening any individual agent. See, for example, Baron et al. (2026).

(Bergeron et al., 2024), workers to tasks (Acemoglu and Restrepo, 2018; Valenzuela-Stookey, 2025), and refugees to locations (Bansak et al., 2018; Ahani et al., 2021; Delacrétaz et al., 2023). In general, these are multi-dimensional screening problems without transfers, in which the designer is constrained by the need to ensure participation in the face of agents' unknown preferences.[3]

Throughout, we maintain the assumption that agents' preferences over bundles of tasks are linear. This is an important simplifying assumption, which allows us to make significant progress on characterizing OSP outcomes. Clearly, this assumption is a better fit in some applications than in others, but we feel that the gains in tractability compensate for the loss of generality. Moreover, the mechanisms we propose retain many of their desirable properties even if agents' preferences are not separable across tasks (see Section 2.2).

Our first main result (Theorem 1) is a complete characterization of the set of static (i.e. simultaneous-move) OSP mechanisms. Static mechanisms may be appealing in settings where speed is a priority and repeated communication with agents is costly. To understand the characterization, it is helpful to introduce a class of *veto games*, defined as follows: each agent $j \in \mathcal{J} := \{1, \dots, J\}$ is offered a menu of *veto sets* $A_j \subset 2^{\mathbb{R}^I_{\geq 0}}$, and is asked to select a set from their menu. The final allocation is then determined by a *selection rule* $\Gamma : \prod_{j \in \mathcal{J}} A_j \to \mathcal{Z}$, where $\mathcal{Z}$ is the set of market-clearing allocations of tasks to agents (i.e. such that each task is assigned to exactly one agent). Let $\Gamma_j((a_j)_{j \in \mathcal{J}})$ denote j's marginal assignment under allocation $\Gamma((a_j)_{j \in \mathcal{J}})$. The defining property of a veto game is that $\Gamma_j((a_j)_{j \in \mathcal{J}})$ must lie in the set a_j for every agent j and any profile of selections $(a_{j'})_{j' \in \mathcal{J}}$. In other words, by selecting $a_j \in A_j$, j vetoes all allocations in which their own assignment is outside of a_j.

Theorem 1 shows that under a richness condition on the space of preferences, every joint distribution of types and allocation that can be implemented with a static OSP mechanism can also be implemented with a veto game.[4] More importantly, it suffices to restrict attention

[3]The primary focus here is on problems in which monetary transfers between the designer and the agents are either infeasible (as in school choice or organ exchange (Roth, 2007)) or beyond the scope of the particular mechanism-design question. For example, a firm may seek to improve the allocation of tasks to workers whose wages have already been determined by a collective bargaining agreement. That said, the model can be extended to include transfers, as discussed in Section 1.1.

[4]There is one additional potential feature of OSP mechanisms, related to messages that can be used to coordinate the allocation of agent j with the type of agent k when k receives their status-quo σ_k. These

to the following narrow class of veto games: either an agent is given no choice and A_j is the singleton set $\left\{\{x \in \mathbb{R}^I_{\geq 0} : x \leq \sigma_j\}\right\}$, or each action in A_j is a ray with an endpoint at the status-quo assignment.[5] Economically, a ray is a fixed-composition adjustment right: the agent chooses the direction in which their assignment may be changed, while the designer determines how far to move in that direction. In addition, the set of rays that constitute an agent's menu must satisfy a "polarization" condition, which constrains the degree of similarity between any two rays. We refer to these as *polarized-ray* mechanisms. In such mechanisms, each veto set is simple (a one-dimensional ray) but the menu offered to each agent is diverse, in that the rays must be polarized. Moreover, the menu consists of at most $I + 1$ sets, which helps to rationalize the use of relatively simple mechanisms in practice.

It is well-known that obvious dominance is sensitive to the extensive form of the game; the standard revelation principle of Myerson (1981) does not apply. Thus we next relax the restriction to simultaneous-move games. Theorem 2 characterizes the set of OSP mechanisms in which each agent moves at most once, which turns out to be equivalent to the set of all mechanisms satisfying a stronger notion of obvious dominance, the *one-step dominance* notion of Pycia and Troyan (2023b). Essentially, these mechanisms reduce to sequentially offering each agent the option to choose from an action set satisfying the conditions from a static polarized-ray mechanism.

Return now to static mechanisms. Any polarized-ray mechanism, in addition to being OSP, satisfies a number of desirable properties. The characterization implies that all agent j needs to know about the mechanism is their own menu of veto sets; additional details about the designer's selection rule or the veto sets of other agents are irrelevant for j's choice problem. Thus, given the agents' menus of veto sets, the designer is free to choose any selection rule, and in doing so optimally can guarantee a (weak) improvement over the status-quo allocation. This is true regardless of the designer's objective, the distribution over agents' preferences, or even the form that these preferences take; the mechanism is robust to model misspecification in these dimensions. Theorem 1 thus reveals a conceptually interesting modular structure of static OSP mechanisms: agents' incentives are entirely

messages introduce additional notation, and are not central to the characterization, so we defer their discussion.

[5]That is, for each action $a_j \in A_j$ there exists $\kappa \in \mathbb{R}^I_{\geq 0}$ such that $a_j = \{\sigma_j + y\kappa : y \in \mathbb{R}_{\geq 0}\}$.

determined by a layer of veto rights, on top of which the designer is free to pursue their objective.

From a normative perspective, these observations suggest that any polarized-ray mechanism (with a suitable selection rule) is a reasonable candidate to apply in practice. We next consider the designer's problem of choosing veto rights. We assume that when agent j is assigned to a task of type i they generate a social cost that is known to the designer, and potentially heterogeneous across tasks and agents, and the designer's objective is to minimize the total social cost of the allocation. Our main interest is in selecting among static OSP mechanisms. However, before studying this problem we first take a step back and answer a more basic question: under what conditions is it optimal to use a choice-based mechanism, i.e. one that responds non-trivially to agents' preferences? Theorem 3 provides necessary and sufficient conditions for choice-based mechanisms to be optimal when the designer can choose among all Bayesian incentive-compatible (BIC) mechanisms, including those that are neither static nor OSP. In fact, it turns out that whenever there is a general BIC mechanism that improves upon the status quo, there is a static OSP mechanism that can do so. This suggests that static OSP mechanisms preserve at least some flexibility for the designer.

We then consider the problem of finding optimal mechanisms. While the ability to restrict attention to polarized-ray mechanisms significantly reduces the dimensionality of the feasible space, this remains a challenging problem to which we are unable to provide a definitive solution. We are, however, able to make some progress by simplifying the problem in two ways. First, we characterize the convex structure of the set of polarized-ray menus: we show how this set can be partitioned into a finite collection of cones, and characterize the extremal generators of these cones. By rewriting the optimization problem so that the choice variable is a distribution over these extremal generators, we make the constraint to polarized-ray mechanisms implicit, which simplifies the program. Second, we study a "large market" relaxation of the optimal design problem. In this relaxation, the dual program has a convenient formulation, which can facilitate the search for an optimal mechanism. The solution to the large-market problem is asymptotically optimal in finite markets as the number of agents grows.

Our results suggest that polarized-ray mechanisms are a promising place to begin in

searching for mechanisms that can be used in applications. To illustrate their potential, we apply our framework to two high-stakes public-sector assignment problems: the allocation of Child Protective Services (CPS) investigations to caseworkers and the allocation of misdemeanor cases to prosecutors. Using administrative data and quasi-random assignment designs, we estimate agent-specific performance across task types, identify candidate static OSP mechanisms using the large-market optimization procedure, and evaluate the performance of these mechanisms relative to the status-quo assignment systems. In both settings, the mechanisms generate meaningful improvements in the planner's objective while respecting agents' status-quo assignment rights. Moreover, despite their simplicity, the mechanisms recover a substantial share of the gains achievable under a full-information benchmark in which the planner directly observes agents' preferences. These results suggest that polarized-ray mechanisms may indeed be useful in practice.

Related literature

Our main contribution is the characterization of obvious strategy-proofness in an important class of multi-dimensional screening problems.[6] In such problems, extensive work has demonstrated both the difficulty of characterizing incentive compatibility (Rochet, 1987; Bikhchandani et al., 2006; Lahr and Niemeyer, 2024) and the difficulty of solving for optimal mechanisms (Armstrong, 1996; Rochet and Choné, 1998; Thanassoulis, 2004; Manelli and Vincent, 2006; Pavlov, 2011; Daskalakis et al., 2014, 2015; Hart and Nisan, 2019) within broad classes of mechanisms. Our results show that such mechanisms reduce to granting agents diverse menus of narrowly defined veto rights. We then study the question of how veto rights should be designed. In this regard, we contribute to what Budish (2012) calls the "mechanism-design" approach to multi-dimensional screening, i.e. maximizing an objective subject to incentive and feasibility constraints.[7]

Obvious strategy-proofness, introduced by Li (2017), has been widely studied due to

[6] In particular, the problem we study here is one of multi-dimensional screening with type-dependent participation constraints and no transfers.

[7] This is in contrast to the "good properties" (or axiomatic) approach, in which the designer specifies a set of desirable properties (e.g. stability, Pareto efficiency) and identifies a mechanism that satisfies them. Deferred acceptance (Gale and Shapley, 1962), top trading cycles (Shapley and Scarf, 1974), and market-based mechanisms (Varian, 1973; Hylland and Zeckhauser, 1979; Budish, 2011) are perhaps best understood in this way.

its normative appeal and has been identified as a promising criterion for guiding mechanism design in multi-dimensional settings (Palacios-Huerta et al., 2024). Studies of OSP in specific problems or classes of mechanisms include Feldman et al. (2014); Dütting et al. (2016); Golowich and Li (2021); Mandal and Roy (2022). However, OSP has not been applied to the type of task-allocation problem studied here.

The most closely related work is that of Pycia and Troyan (2023b), who characterize OSP in general environments and introduce the notion of one-step dominance.[8] Their characterizations cover multi-good allocation without transfers. The major differences between our results and the characterization in Pycia and Troyan (2023b) are (i) the restriction to linear preferences, which allows us to place significantly more structure on the mechanisms, and (ii) the additional constraint arising from the fact that agents have the right to demand their status-quo allocation. Thus, the mechanisms that we identify are (modulo a few technical distinctions) a subset of the millipede games characterized by Pycia and Troyan (2023b).

Finally, this paper is related to other work on similar task-allocation problems (Baron et al., 2026; Valenzuela-Stookey, 2025). The setting is closest to that of Baron et al. (2026). There, the authors study BIC mechanisms for the special case in which there are two types of tasks. In that case, it turns out that with a continuum of agents the solution is a polarized-ray mechanism, so restricting to static OSP mechanisms is without loss of optimality. Baron et al. (2026) also introduce a two-step procedure for solving a large-market version of their program, which we generalize in Section 5.2.

The remainder of the paper is organized as follows. We present the model in Section 1. In Section 2 we characterize static OSP mechanisms. In Section 3 we relax the static restriction. We then turn to the designer's problem. We first characterize when choice-based mechanisms are useful (without restricting to static or OSP mechanisms) in Section 4; we then develop a procedure for optimizing over polarized-ray mechanisms by characterizing the convex structure of this set (Section 5.1) and introducing a large-market relaxation of the program (Section 5.2). Finally, we apply this procedure to identify promising mechanisms in two empirical applications (Section 6). Proofs omitted from the main text are presented

[8]Pycia and Troyan (2023a) study OSP for a more general class of payoff structures, but this generality does not bring their setting closer to the current one, so we focus the discussion on Pycia and Troyan (2023b).

in the appendix.

1 Model

There is a continuum of tasks (or objects), each of which is of one of a finite set of types indexed by $i \in \mathcal{I} = \{1, \ldots, I\}$ (with $I \geq 2$), and a finite set of agents, indexed by $j \in \mathcal{J} = \{1, \ldots, J\}$. An allocation is a matrix $\mu \in \mathbb{R}_{\geq 0}^{\mathcal{I} \times \mathcal{J}}$ such that row i sums to n^i, where n^i is the mass of task type i that needs to be completed (each task must be assigned to one agent). Let $\mathcal{Z}$ be the set of allocations. Given an allocation μ, we denote the j^{th} column by μ_j, and refer to this as j's *assignment*.

The designer does not know the preferences of agents over tasks. Let $c_j(i) > 0$ be the private cost to agent j of completing task i, so that j's total *workload* from allocation μ is

$$\sum_{i \in \mathcal{I}} c_j(i)\mu_j(i).$$

We maintain the normalization that agents prefer lower workload.[9] Let $\mathbf{C}$ be the set of preference profiles, with typical element $\mathbf{c} = (c_j(\cdot))_{j=1}^J$. Let F_j be the distribution of j's preference type c_j. We maintain the assumption that types are independent across agents, and that F_j has full support on a hypercube $C_j \subset \mathbb{R}_{\geq 0}^I$.

The agents' outside options are defined by an allocation $\sigma \in \mathcal{Z}$, which we refer to as the status quo. Each agent has the right to demand their marginal allocation from the status quo, σ_j. This participation decision occurs at the interim stage, i.e. after each agent observes their own preferences but before they engage with the mechanism.[10] (There is a sense in which the class of static OSP mechanisms is independent of the status quo. See the discussion following Theorem 1, and Section 5.1.)

While the designer's objective is not relevant for characterizing agents' incentives (our primary focus), when we turn to the question of choosing a mechanism we assume that the

[9] The substantive assumption is $c_j(i) > 0$, i.e. all tasks have the same qualitative effect on agents' payoffs.

[10] While we only impose interim participation constraints, it turns out that static OSP mechanisms, characterized below, satisfy a stronger ex-post participation constraint that holds after all types are revealed.

designer's objective is to minimize the expected total social cost,[11] given by

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\pi_j(i)\mu_j(i),$$

for some known weights $\pi_j(i)$.[12] We can think of $\pi_j(i)$ as representing the (observable) performance of agent j on task i (where lower $\pi_j(i)$ means higher performance).

If the designer could implement any allocation in $\mathcal{Z}$ then the social-cost minimization problem would be a straightforward linear program. However, the designer here is constrained by the fact that (i) agents can demand their status-quo assignment and, (ii) the designer does not know the agents' preferences.

Remark 1. An alternative interpretation of the model is that there is a finite set of tasks, and μ_j is the expected assignment of agent j under a random allocation (random allocations are not needed with a continuum of tasks).[13] This model is equivalent from the agents' and designer's perspectives, given that all are assumed to have linear preferences. We adopt this interpretation in the applications of Section 6.

1.1 Discussion of modeling assumptions

Why OSP?

There are a number of reasons for which we find it worthwhile to focus on OSP mechanism. For one, an extensive empirical literature has shown that when players have an obviously dominant strategy they are far more likely to play it, relative to strategies that are merely dominant, especially in complex environments such as the multi-dimensional problem at hand (Li, 2017; Breitmoser and Schweighofer-Kodritsch, 2022). This suggests that OSP mechanisms are not only normatively appealing, but also more likely to be observed in practice. Theorem 3 also provides some indirect evidence that OSP is not overly restrictive in this context, as noted above.

[11] Describing the objective as minimizing social cost, as opposed to maximizing welfare, is simply a normalization.

[12] It is straightforward to extend Theorem 3 (concerning the value of choice) to any objective for the designer that is a smooth function of the expected allocation. In that case, we interpret $(\pi_j(i))_{i\in\mathcal{I},j\in\mathcal{J}}$ as the gradient of the objective. Theorem 1 has no relationship with the designer's objective.

[13] We need not even appeal to the Birkhoff-von-Neumann Theorem here, as there is no constraint on the column sums of an allocation (see Budish et al. (2013) for further discussion).

A more practical reason for focusing on static OSP mechanisms from a market-design perspective is tractability. The large theoretical literature on multi-dimensional screening has shown that weaker implementability notions (e.g. Bayesian incentive compatibility) are difficult to characterize in general. In contrast, we are able to provide a relatively simple characterization of static OSP mechanisms. Thus the restriction to OSP allows us to make progress on an otherwise intractable problem.[14] Moreover, static OSP mechanisms turn out to be robust to model misspecification regarding the distribution of agents' preferences, about which the designer may in practice have only limited information.

Transfers

As discussed above, the focus here is on settings in which transfers between the designer and agents are either infeasible or beyond the scope of the mechanism-design exercise. While there are many applications that fit this description, there are certainly others in which transfers are an important part of the designer's toolkit.[15]

The model can be extended to incorporate transfers by letting one of the "tasks", m, be payments made by the agents to the designer. This (potentially) requires two changes to the model. First, we have assumed that all tasks have the same qualitative effect on agents' payoffs. If tasks are burdensome then we can continue to use the normalization $c_j(m) > 0$. However, if tasks are "goods" then m would have the opposite effect, so we would need $c_j(m) < 0$. (We have also assumed that allocations are non-negative; we may wish to relax this restriction on m in some settings so that the designer can make payments to agents). Second, we have assumed a fixed supply of each task. In a setting with transfers, this would correspond to requiring that the mechanism generate a fixed level of revenue. Relaxing this constraint changes the space of feasible mechanisms.

In summary, under some conditions we can incorporate transfers directly into the current model (e.g. tasks are burdensome, payments must be non-negative, and revenue must equal

[14] To this point, Palacios-Huerta et al. (2024) write "We argue that the behavioral implications for future developments moving toward OSP [combinatorial allocation] formats should not be underestimated."

[15] Roth (2007) discusses reasons why transfers may be infeasible in some settings. Even in settings where transfers between the designer and agents are feasible, e.g. where the designer is a firm and agents are employees, transfers are frequently not tied to completion of specific tasks. In restricting attention to counterfactual mechanisms which fix transfers, our approach can be viewed as an instance of minimalist market design (Sönmez, 2023).

some predetermined level). More generally, the modifications needed to incorporate transfers do not appear to be large. However, we leave the analysis of these cases for future work.

2 Static OSP mechanisms

We first consider static mechanisms, in which agents move simultaneously. A static mechanism consists of actions A_j for each $j \in \mathcal{J}$, a selection rule $\Gamma : \prod_{j\in\mathcal{J}} A_j \to \mathcal{Z}$, and type-contingent strategies for the agents $s_j : C_j \to \Delta(A_j)$.[16] We let $A_{\mathcal{J}} = (A_j)_{j\in\mathcal{J}}$ and use $a_{\mathcal{J}}$ to denote an action profile $(a_j)_{j\in\mathcal{J}}$ and $s_{\mathcal{J}}$ to denote a strategy profile $(s_j)_{j\in\mathcal{J}}$. We write $\Gamma_j(a_{\mathcal{J}}) \in \mathbb{R}^I_{\geq 0}$ for j's assignment in allocation $\Gamma(a_{\mathcal{J}})$.

Given a simultaneous-move game $(A_{\mathcal{J}}, \Gamma)$, let $\mathcal{O}_j(a) := \{\Gamma_j(a, a_{-j}) : a_{-j} \in A_{-j}\}$ be the set of possible assignments for j if they take action $a \in A_j$. We say that a static mechanism *respects the agents' outside options* if for all j there exists an action $a \in A_j$ such that $\mathcal{O}_j(a) = \{\sigma_j\}$. In other words, agents have the right to demand their outside option. By definition, we restrict attention to mechanisms that respect the outside option. We can now introduce the solution concept for simultaneous-move games.

Definition (Static OSP)**.** Type-contingent strategy s_j is *obviously dominant* for j in simultaneous-move game $(A_{\mathcal{J}}, \Gamma)$ if

$$\max_{\mu_j \in O_j(s_j(c_j))} \mu_j \cdot c_j \leq \min_{\mu_j \in O_j(a)} \mu_j \cdot c_j \qquad \forall\ c_j \in C_j, \text{ and } s_j(c_j) \neq a \in A_j$$

Mechanism $(A_{\mathcal{J}}, \Gamma, s_{\mathcal{J}})$ is *obviously strategy-proof* (OSP) if s_j is obviously dominant $\forall\ j \in \mathcal{J}$.

We begin with a preliminary observation on stratgies. Given a game $\mathcal{G} = (A_{\mathcal{J}}, \Gamma)$, define the *myopically optimal choices* for agent j as

$$a_j^*(c_j) := \arg\min_{a\in A_j} \{\inf \{c_j \cdot x \ : \ x \in \mathcal{O}_j(a)\}\}$$

In other words, if agent j with preference c_j could choose their allocation freely from $\cup_{a\in A_j} \mathcal{O}_j(a)$, it would be optimal to choose an $x \in a$ for some $a \in a_j^*(c_j)$. We say that

[16] It is straightforward to allow for randomness in Γ. Indeed we can think of $\mathcal{Z}$ as a set of expected allocations (see Remark 1). As agents are risk neutral, randomization plays no role. For clarity we focus on deterministic mechanisms and interpret $\mathcal{Z}$ as deterministic allocations.

an agent's strategy is myopically optimal if they always make such choices.[17] Then we have the following simple observation.

Lemma 1. If a static mechanism is OSP then agents' strategies are myopically optimal.

Proof. If not then there exists $a \neq s_j(c_j)$ and $x \in \mathcal{O}_j(a)$ such that $c_j \cdot x < c_j \cdot x'$ for all $x' \in s_j(c_j)$. But then $s_j(c_j)$ does not obviously dominate a. □

Lemma 1 narrows the range of OSP mechanisms to those with myopically optimal strategies; it is silent on the form game should take.

Definition. A *veto game* is a simultaneous-move game with the following features

- $a \subset \mathbb{R}^I_{\geq 0}$ for each $a \in A_j$ and $j \in \mathcal{J}$.
- $\Gamma_j(a_{\mathcal{J}}) \subset a_j$ for all $a_{\mathcal{J}} \in \prod_{k \in \mathcal{J}} A_k$ and $j \in \mathcal{J}$

In other words, a veto game in which each agent selects a set of possible assignments for themselves (thus unilaterally vetoing allocations in which their assignment does not belong to this set) and the designer chooses an allocation consistent with these choices. Note that the ability of the designer to do so imposes some restrictions on the action sets. A *unilateral-veto (UV) mechanism* is a veto game, together with a profile of type-contingent strategies. A particular class of UV mechanisms plays a central role in the analysis.

Definition. In a veto game, the action set for agent j consists of *polarized rays* if

1. Each $a \in A_j$ is a non-monotone ray with endpoint σ_j, i.e. $a = \{x \in \mathbb{R}^I : (x - \sigma_j) = y\kappa(a),\ y \in \mathbb{R}_{\geq 0}\}$ for some $\kappa(a) \in \mathbb{R}^I$ such that $\kappa(a) \not\geq 0$ and $\kappa(a) \not\leq 0$,[18] and
2. For any distinct $a', a'' \in A_j$ there exists $\alpha \in [0, 1]$ such that $\alpha\kappa(a') + (1 - \alpha)\kappa(a'') \geq 0$.

Intuitively, polarization of the rays defining an agent's veto sets means that these sets move the agent's allocation in different directions. One way to see this is that a necessary (but not sufficient) condition for polarization is that no two rays can "take away" the same task, i.e. the sets of negative coordinates for any two rays must be disjoint. Figure 1 depicts an example. In other words, polarized rays are diverse menus of narrowly defined rights.

[17] Note that the set of myopically optimal choices may be empty, but Lemma 1 implies that this cannot be the case in an static OSP mechanism.

[18] The fact that these rays are not contained in $\mathbb{R}^I_{\geq 0}$ is not important, as the selection rule is restricted to non-negative allocations.

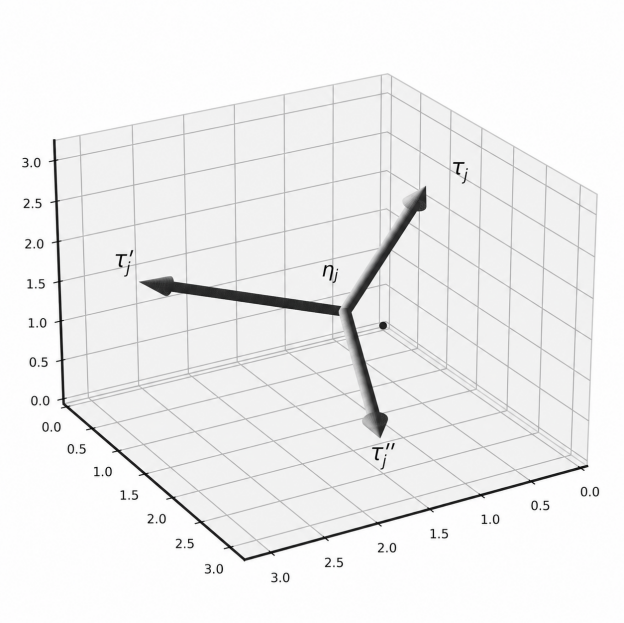


Figure 1: Polarized rays

Definition. A *polarized-ray mechanism* consists of a veto game $(A_{\mathcal{J}}, \Gamma)$ and a strategy profile $s_{\mathcal{J}}$ such that for any agent j

1. either j's action set consists of polarized rays, or j has a single action $\{x \in \mathbb{R}^I_{\geq 0} \ : \ x \leq \sigma_j\}$; and

2. j's strategy is myopically optimal, i.e. $s_j(c_j) \in a_j^*(c_j)$ for all $c_j \in C_j$.

The first condition says that there are two options for the action sets of each agent. One is that the agent makes no choice, but is guaranteed an assignment that is lower in every dimension than their status quo. We call such agents *remedial*. The other is that j chooses from a set of polarized rays.

We say that two mechanisms are *equivalent* if they induce the same joint distribution of types and allocations. In order to fully span the set of implementable joint distributions, we need a small generalization of polarized-ray mechanisms.

Definition. An *extended polarized-ray mechanism* appends a set of "opt-out messages", W_j, to j's action set: $A_j = \hat{A} \cup W_j$, where actions $\hat{A}_j$ satisfy the conditions of a polarized ray mechanism. The selection rule satisfies $\Gamma_j(a_{\mathcal{J}}) \subset a_j$ if $a_j \in \hat{A}_j$ and $\Gamma_j(a_{\mathcal{J}}) = \sigma_j$ if $a_j \in W_j$. Strategies are myopically optimal.

We refer to the actions in W_j as *withdrawal actions*, and the action set $\hat{A}_j \cup W_j$ as *extended polarized-ray actions*. The only thing an extended polarized-ray mechanism does beyond a polarized-ray mechanism is give agents multiple ways to claim their outside option. Because these different options are identical from j's perspective, the extension is only useful for correlating j's type with the assignments of other agents. Thus in general, unless the designer's preferences over one agent's assignment depend on the type of another, these additional messages have no value. However, they are needed to state the full equivalence result.

We say that j's preferences are *strictly monotone* if $C_j = \mathbb{R}^I_{>0}$ (note that this assumption is about both monotonicity and richness of preferences).

Theorem 1. Every extended polarized-ray mechanism is OSP. If preferences are strictly monotone then every static OSP mechanism is equivalent to an extended polarized-ray mechanism.

Theorem 1 reveals interesting structural features of static OSP mechanisms. For one, every action induces a one-dimensional set of possible outcomes, $\mathcal{O}_j(a)$, regardless of the number of task types (unless the agent is remedial, in which case they take no actions). Intuitively, this is because OSP requires that there be limited ambiguity about what assignment an agent will receive once they have acted. Moreover, veto sets must be sufficiently distinct (in the sense of being polarized). In other words, the menu of veto sets offered to agents must be sufficiently diverse. Polarization ensures that, for any preference type, at most one ray can offer an improvement over the status quo. The agent therefore needs only to identify whether one of the narrowly defined rights in their menu is beneficial; they need not compare several potentially attractive sets of assignments or anticipate how far the designer will move along the selected ray. As a result, Theorem 1 also helps rationalize the use of fairly simple mechanisms.

Corollary 1. Any collection of polarized rays in $\mathbb{R}^I$ has at most I members. Thus if preferences are strictly monotone, an OSP mechanism needs only I non-withdrawal actions.

Proof. Let $\kappa(a_j) \in \mathbb{R}^I$ be a vector defining the ray containing a_j, i.e. $a_j \subset \{y\kappa(a_j) + \eta_j : y \in \mathbb{R}_{\geq 0}\}$. By definition $\kappa(a'_j) \not\geq 0$ for any $a'_j \in A_j$, so it must have at least one strictly negative

coordinate. Moreover, if $\kappa(a'_j)_i < 0$ then $\kappa(a''_j)_i \geq 0$ for any $a''_j \neq a'_j$, because otherwise $y'\kappa(a'_j) + y''\kappa(a''_j) \not\geq 0$ for any $y', y'' \geq 0$ with at least one strictly positive. Thus there can be at most I such vectors. □

Note also that changing the endowment does not affect the polarization property. In other words, if the designer communicates only the *directions* of the rays, but not the endowment, then a polarized-ray mechanism remains OSP even if agents do not observe their endowment. Thus, the model can accommodate settings in which the status-quo allocation is uncertain at the time that agents submit preferences.

Another implication of Theorem 1 is that the selection rule is in a sense irrelevant for determining whether a unilateral-veto mechanism is OSP. Observe that the definition of polarized rays depends only on the action sets, not the selection rule. The only role that the selection rule plays in shaping incentives in an extended polarized-ray mechanism is in defining the myopically optimal strategies, which depend in principle on Γ through $\mathcal{O}_j$. However, in such a mechanism, identifying myopically optimal strategies is particularly easy.

Lemma 2. Let A_j consist of polarized rays, and let $\kappa_j : A_j \to \mathbb{R}^I$ be the directions defining the rays. Then for any $c_j \in \mathbb{R}^I_>$, if $c_j \cdot \kappa(a) < 0$ then $c_j \cdot \kappa(a') \geq 0$ for all $a' \neq a$.

Proof. This follows from the part of proof of Theorem 1 where it is shown that any extended polarized-ray mechanism is OSP. □

In other words, there is at most one direction that the agent can move which reduces their workload relative to the status quo. This in turn implies the selection rule does not matter for determining OSP.

Proposition 1. Let $(A_{\mathcal{J}}, \Gamma, s_J)$ be an extended polarized-ray mechanism. Then so is $(A_{\mathcal{J}}, \Gamma', s_J)$ for any feasible Γ', i.e. such that $\Gamma'_j(a_{\mathcal{J}}) \subset a_j$ for all $a_{\mathcal{J}} \in \prod_{k\in\mathcal{J}} A_k$ and $j \in \mathcal{J}$.

In other words, in extended polarized-ray mechanisms agents' incentives are shaped entirely by the action sets, rather than the designer's selection rule. In light of Proposition 1, it is without loss of optimality for the designer to use an ex-post-optimal allocation rule:

$$\Gamma(a_{\mathcal{J}}) \in \arg\min_{\mu\in\mathcal{Z}} \sum_{j\in\mathcal{J}} \pi_j \cdot \mu_j \quad s.t. \quad \mu_j \in a_j \ \ \forall\ j \in \mathcal{J}. \tag{1}$$

We call a polarized-ray mechanism with this selection rule an *ex-post-optimal polarized-ray* (EOPR) mechanism. Together, the previous observations greatly simplify the problem of choosing an OSP mechanism.

2.1 Proof of Theorem 1

Let $(A_{\mathcal{J}}, \Gamma, s_{\mathcal{J}})$ be a static OSP mechanism. Because the mechanism respects the outside option by assumption, there is an action $a_j^0 \in A_j$ be an action such that $\mathcal{O}_j(a_j^0) = \{\sigma_j\}$. We call an action $a \in A_j$ *remedial* if $x \leq \sigma_j$ for all $x \in \mathcal{O}_j(a)$, and *redundant* if $x \geq \sigma_j$ for all $x \in \mathcal{O}_j(a)$. The key result is the following observation regarding the structure of $\mathcal{O}_j$.

Proposition 2. If type-contingent strategy s_j is obviously dominant for j then exactly one of the following holds

1. There is at most one $a \in A_j$ that is remedial and has $\mathcal{O}_j \neq \{\sigma_j\}$. Every other action is redundant.

2. There is a subset of actions $\hat{A}_j \subset A_j$ such that $(\mathcal{O}_j(a))_{a \in \hat{A}_j}$ consists of a collection of polarized rays. Every action not in $\hat{A}_j$ is redundant.

Proof. We begin with some intermediate claims.

Claim 1. $\sigma_j \in \mathcal{O}_j(a)$ for all $j \in \mathcal{J}$ and $a \in A_j$.

Proof of Claim 1. If every agent $j' \neq j$ chooses $a_{j'}^0$ then market clearing requires that j be assigned σ_j. Thus it must be that $\sigma_j \in \mathcal{O}_j(a)$ for all $a \in A_j$. □

Claim 2. For any $c_j \in \mathbb{R}^I_{>0}$ it must be that $c_j \cdot x \geq c_j \cdot \sigma_j$ for all $s_j^*(c_j) \neq a \in A_j$ and $x \in \mathcal{O}_j(a)$.

Proof of Claim 2. Suppose there exists $a \neq s_j^*(c_j)$ and $x \in \mathcal{O}_j(a)$ such that $c_j \cdot x < c_j \cdot \sigma_j$. By Claim 1, $\sigma_j \in \mathcal{O}_j(s_j(c_j))$. But then σ_j is possible following action $s_j(c_j)$ while x is possible following action a, so $s_j(c_j)$ does not obviously dominate a. □

Claim 3. If $x', x'' \in \mathcal{O}_j(a)$ then there cannot exist $c_j \in \mathbb{R}^I_{\geq 0}$ such that $c_j \cdot x' < c_j \cdot \sigma_j < c_j \cdot x''$.

Proof of Claim 3. If such a c_j exists, we can take it to be in $\mathbb{R}^I_{>0}$ because the two inequalities in the claim are strict. By Claim 2, if such a c_j exists then $s_j(c_j) = a$, because $c_j \cdot x' < c_j \cdot \eta_j$. But because $x'' \in \mathcal{O}_j(a)$ and $c_j \cdot x'' > c_j \cdot \sigma_j$, action a does not obviously dominate a_j^0. □

Claim 4. For each $j \in \mathcal{J}$ and $a \in A_j$, one of the following holds

1. $x \geq \sigma_j$ for all $x \in \mathcal{O}_j(a)$.
2. $x \leq \sigma_j$ for all $x \in \mathcal{O}_j(a)$.
3. $\mathcal{O}_j(a)$ is contained in a non-monotone ray originating at σ_j, i.e. $\mathcal{O}_j(a) \subset \{x \in \mathbb{R}^I : (x - \sigma_j) = y\kappa,\ y \in \mathbb{R}_{\geq 0}\}$ for some $\kappa \in \mathbb{R}^I$ such that $\kappa \not\geq 0$ and $\kappa \not\leq 0$.

Proof of Claim 4. If $\mathcal{O}_j(a) = \{\sigma_j, x\}$ for some $x \in \mathbb{R}^I$ then one of the three conditions is automatically satisfied. Suppose instead that there exist $x' \neq x''$ in $\mathcal{O}_j(a)$ such that neither is equal to σ_j. Consider the condition from Claim 3. By Farkas' lemma, the following are equivalent

1. there does not exist $c_j \in \mathbb{R}^I_{\geq 0}$ such that $c_j \cdot x' < c_j \cdot \sigma_j$ and $c_j \cdot x'' > c_j \cdot \sigma_j$.
2. There exist $y', y'' \in \mathbb{R}_{\geq 0}$, with at least one strictly positive, such that $y'(x' - \sigma_j) \geq y''(x'' - \sigma_j)$.

So Claim 3 implies that the second condition holds. Reversing the roles of x' and x'' we also have that there exist $\hat{y}', \hat{y}'' \in \mathbb{R}_{\geq 0}$, with at least one strictly positive, such that $\hat{y}'(x' - \sigma_j) \leq \hat{y}''(x'' - \sigma_j)$.

Suppose $y' > 0$ (a symmetric argument applies if $y' = 0, y'' > 0$). Then $x' - \sigma_j \geq \frac{y''}{y'}(x'' - \sigma_j)$, so $\hat{y}''(x'' - \sigma_j) \geq \hat{y}'\frac{y''}{y'}(x'' - \sigma_j)$. This implies that one of the following holds

1. $x'' - \sigma_j \geq 0$ and $\hat{y}'' \geq \hat{y}'\frac{y''}{y'}$
2. $x'' - \sigma_j \leq 0$ and $\hat{y}'' \leq \hat{y}'\frac{y''}{y'}$.
3. $x'' - \sigma_j \not\geq 0$, $x'' - \sigma_j \not\leq 0$ and $\hat{y}'' = \hat{y}'\frac{y''}{y'}$.

Suppose $x'' - \sigma_j \geq 0$. By assumption, $x'' \neq \sigma_j$, so we must be in the first case, with $\hat{y}'' \geq \hat{y}'\frac{y''}{y'}$, so $x' - \sigma_j \geq 0$ as well. Because x' and x'' are arbitrary selections from $\mathcal{O}_j(a)$, we have that $x - \sigma_j \geq 0$ for all $x \in \mathcal{O}_j(a)$. Similarly in the second case, with $x'' - \sigma_j \leq 0$ and $\hat{y}'' \leq \hat{y}'\frac{y''}{y'}$, we have that $x - \sigma_j \leq 0$ for all $x \in \mathcal{O}_j(a)$.

Consider the third case, where $x'' - \sigma_j \not\geq 0$, $x'' - \sigma_j \not\leq 0$ and $\hat{y}'' = \hat{y}' \frac{y''}{y'}$. Note that if this holds then $y'', \hat{y}', \hat{y}''$ must be strictly positive (and we are still assuming $y' > 0$), so $\frac{\hat{y}''}{\hat{y}'} = \frac{y''}{y'}$. Then we have

$$\frac{y''}{y'}(x'' - \sigma_j) \geq (x' - \sigma_j) \quad \text{and} \quad \frac{y''}{y'}(x'' - \sigma_j) \leq (x' - \sigma_j)$$

so $\frac{y''}{y'}(x'' - \sigma_j) = (x' - \sigma_j)$. In other words, $(x'' - \sigma_j)$ and $(x' - \sigma_j)$ lie in the same ray. Then $\mathcal{O}_j(a)$ is contained in a ray originating at σ_j. □

Claim 5. If a' and a'' in A_j are not remedial or redundant then they are polarized.

Proof of Claim 5. By Claim 4, both $\mathcal{O}_j(a')$ and $\mathcal{O}_j(a'')$ are contained in rays from σ_j. Thus it suffices to show that there exists $x' \in \mathcal{O}_j(a')$, $x'' \in \mathcal{O}_j(a'')$, and $y', y'' \geq 0$, with at least one strictly positive, such that $y'(x' - \sigma_j) + y''(x'' - \sigma_j) \in \mathbb{R}^I_{\geq 0}$.

Let $x' \in \mathcal{O}_j(a')$, $x'' \in \mathcal{O}_j(a'')$ be any selections such that $x' \neq \sigma_j$, $x'' \neq \sigma_j$. Observe that there cannot exist $c_j \geq 0$ such that $c_j \cdot x' \leq c_j \cdot \sigma_j$ and $c_j \cdot x'' < c_j \cdot \sigma_j$. If this were the case then, because $x' - \sigma_j \not\geq 0$, we could find $c'_j \in \mathbb{R}^I_{>0}$ such that $c'_j \cdot x' < c'_j \cdot \sigma_j$ and $c'_j \cdot x'' < c'_j \cdot \sigma_j$, violating Claim 2. By Farkas' lemma, the following are equivalent

i. There does not exist $c_j \geq 0$ such that $c_j \cdot x' \leq c_j \cdot \sigma_j$ and $c_j \cdot x'' < c_j \cdot \sigma_j$.

ii. There exist $y' \geq 0, y'' > 0$, such that $y'(x' - \sigma_j) + y''(x'' - \sigma_j) \geq 0$.

The second condition is precisely what we wanted to show. □

Claim 6. If $a \in A_j$ is remedial and $\mathcal{O}_j(a) \neq \{\sigma_j\}$ then every $a' \in A_j$ is either remedial or redundant.

Proof of Claim 6. If a' is not remedial or redundant then it is contained in a non-monotone ray from σ_j, by Claim 4. Then there exists $c_j \in \mathbb{R}^I_{>0}$ and $x' \in \mathcal{O}_j(a')$ such that $c_j \cdot x' < c_j \cdot \sigma_j$. Then for any $x \in a_j$, $x \neq \sigma_j$ we can also choose c_j so that $c_j \cdot x < c_j \cdot \sigma_j$. But this violates Claim 2. □

We have concluded that for any $j \in \mathcal{J}$, exactly one of the following holds

1. Every $a \in A_j$ is either remedial or redundant.

2. $(\mathcal{O}_j(a))_{a \in A_j}$ consists of a collection of polarized rays and a collection of redundant sets.

Moreover, by Claim 2 there can be at most one remedial action such that $\mathcal{O}_j(a) \neq \{\sigma_j\}$. This completes the proof of Proposition 2 □

To prove the necessity direction of Theorem 1, observe that as preferences are strictly monotone, no redundant action such that $\mathcal{O}_j(a) \neq \{\sigma_j\}$ can ever be chosen (doing so is obviously dominated by choosing a_j^0), and so we obtain an equivalent mechanism by removing the redundant actions. Thus excluding the set of actions such that $\mathcal{O}_j(a) = \{\sigma_j\}$, there is a one-to-one map between the set $\mathcal{O}_j(a)$ and the action a. Thus the mechanism is an extended polarized-ray mechanism (without the extension if for each j there is only one action such that $\mathcal{O}_j(a) = \{\sigma_j\}$).

For the sufficiency direction of Theorem 1, it remains to show that any extended polarized-ray mechanism is OSP. Any j with a remedial mechanism has a trivial dominant strategy. Suppose j's mechanism consists of polarized rays, and let $\kappa(a)$ be the ray direction associated with action a. Then (as shown using Farkas' lemma in the proof of Proposition 2, Claim 5) $c_j \cdot \kappa(a) < 0$ implies $c_j \cdot \kappa(x') \geq 0$ for all $a' \neq a$. Thus choosing a is obviously dominant for type c_j.

2.2 Static OSP mechanisms for market design

Polarized-ray mechanisms are a fairly robust way to improve outcomes in practice. Consider any EOPR mechanism. Because $\sigma_j \in a_j$ for all $j \in \mathcal{J}$ and $a_j \in A_j$, the designer always has the option to select the status-quo allocation, regardless of the agents' actions. Thus the mechanism ensures that the expected social cost is no larger than under the status quo. In particular, the mechanism ensures gains in expectation regardless of the distribution over agents' preferences, and is thus robust to misspecification in this dimension. Indeed, this conclusion does not even rely on the agents' preferences being linear.[19] Any EOPR mechanism is thus a reasonable candidate for use in practice. Nonetheless, if the designer has some knowledge of the type distribution they may want to use it to try to select among EOPR mechanisms. We explore this problem below.

[19]If preferences are non-linear then the mechanism may no longer be OSP. However, the designer is still guaranteed a (weak) improvement in any equilibrium of the induced game.

3 Beyond static mechanisms

As is well-known, the standard revelation principle does not apply to OSP implementation; the extensive for matters as well. Thus we consider all extensive-form games with perfect recall.

Formally, an extensive-form game is defined in the usual way as a tuple

$$\mathcal{G} = \left(\mathcal{H}, \bar{\mathcal{H}}, P, \left(\mathcal{M}_j\right)_{j\in\mathcal{J}}, \left(\rho_h\right)_{h:P(h)=0}, \Gamma\right)$$

with the following components:

- $\mathcal{H}$ is a collection of partially ordered *histories*, i.e. sequences actions. We write $h \preceq h'$ when h precedes h', i.e. h is a subhistory of h'. The initial history $\varnothing$ is in $\mathcal{H}$, and for any non-empty history $h = (a_1, \ldots, a_k, a_{k+1})$, the subhistory $(a_1, \ldots, a_k)$ is also in $\mathcal{H}$.

- $\bar{\mathcal{H}} := \{h \in \mathcal{H} : \nexists\ a \text{ such that } (h, a) \in \mathcal{H}\}$ is the set of *terminal histories*. For every nonterminal history $h \in \mathcal{H} \setminus \bar{\mathcal{H}}$, the set of available actions is $A(h) := \{a : (h, a) \in \mathcal{H}\}$.

- $P : \mathcal{H} \setminus \bar{\mathcal{H}} \longrightarrow \mathcal{J} \cup \{0\}$ is the *player function*. If $P(h) = j \in \mathcal{J}$, agent j moves at h; if $P(h) = 0$, nature moves at h. For each $j \in \mathcal{J}$, let $\mathcal{H}_j := \{h \in \mathcal{H} \setminus \bar{\mathcal{H}} : P(h) = j\}$ be the set of histories at which j moves.

- At every history h with $P(h) = 0$, nature chooses according to a probability distribution $\rho_h \in \Delta(A(h))$ where ρ_h has full support.[20]

- $\mathcal{M}_j$ is a partition of $\mathcal{H}_j$ into *information sets*, with typical element M. If $h, h' \in M$, then $A(h) = A(h')$. We denote this common action set by $A(M)$, and denote by $M_j(h)$ the unique element of $\mathcal{M}_j$ containing h. For an information set M and a history h, write $M \preceq h$ if $h' \preceq h$ for some $h' \in M$, and $h \preceq M$ if $h \preceq h'$ for some $h' \in M$.

- $\Gamma : \bar{\mathcal{H}} \longrightarrow \mathcal{Z}$ is the outcome function. Let Γ_j be the corresponding assignment for j.

To define perfect recall, consider a history

$$h = (a_1, \ldots, a_K)$$

[20]The full-support assumption is relevant under OSP: actions in $A(h)$ that are never taken by nature could still affect a strategies obvious dominance, so dropping such actions from $A(h)$ weakly expands the implementable set.

and let $h^t = (a_1, \ldots, a_t)$ denote its subhistory of length t, with $h^0 = \varnothing$. Let

$$0 \leq t_1 < \cdots < t_L < K$$

be the indices at which agent j previously moved along h, so that $P(h^{t_\ell}) = j$. Define agent j's personal history at h by

$$R_j(h) := \left(\left(M_j(h^{t_1}), a_{t_1+1}\right), \ldots, \left(M_j(h^{t_L}), a_{t_L+1}\right)\right).$$

The game has *perfect recall* if, for every $j \in \mathcal{J}$, every $M \in \mathcal{M}_j$, and every $h, h' \in M$,

$$R_j(h) = R_j(h').$$

Thus, whenever agent j cannot distinguish two current histories, the agent has the same record of all previous information sets visited and all previous actions taken.

For each $j \in \mathcal{J}$, let

$$\mathcal{M}_j^0 := \{M \in \mathcal{M}_j : R_j(h) = \varnothing \text{ for some } h \in M\}.$$

By perfect recall, if $R_j(h) = \varnothing$ for some $h \in M$, then this holds for every $h \in M$. Thus, $\mathcal{M}_j^0$ is the collection of information sets at which agent j moves for the first time.

A strategy for agent j is a function s_j that assigns an action to every information set: $s_j(M) \in A(M)$ for every $M \in \mathcal{M}_j$. Let

$$\Sigma_j := \prod_{M \in \mathcal{M}_j} A(M) \qquad \text{and} \qquad \Sigma_{-j} := \prod_{k \in \mathcal{J} \setminus \{j\}} \Sigma_k$$

denote the corresponding strategy spaces. A type-contingent strategy for agent j is a mapping $s_j : C_j \longrightarrow \Sigma_j$ where $s_j(c_j)$ is the strategy prescribed to type c_j. We write type-contingent strategy profile as $s_{\mathcal{J}} = (s_j)_{j \in \mathcal{J}}$.

A realization of nature's moves is a complete contingent specification

$$d_0 \in D_0 := \prod_{\{h : P(h) = 0\}} A(h).$$

Given a history h, a strategy profile (s_j, s_{-j}), and a realization d_0, let $\zeta(h; s_j, s_{-j}, d_0) \in \bar{\mathcal{H}}$ denote the resulting terminal history when play begins at h.

For two strategies $s_j, s'_j \in \Sigma_j$, let $\alpha_j(s_j, s'_j)$ be the set of their possible earliest points of departure:

$$\begin{aligned}\alpha_j(s_j, s'_j) := \big\{ M \in \mathcal{M}_j : s_j(M) \neq s'_j(M), \\ \exists (s_{-j}, d_0) \in \Sigma_{-j} \times D_0 \text{ such that } M \preceq \zeta(\varnothing; s_j, s_{-j}, d_0), \\ \exists (\widehat{s}_{-j}, \widehat{d}_0) \in \Sigma_{-j} \times D_0 \text{ such that } M \preceq \zeta(\varnothing; s'_j, \widehat{s}_{-j}, \widehat{d}_0) \big\}.\end{aligned}$$

Under perfect recall, these are precisely the information sets at which the two strategies can first diverge along some path of play. Let $\alpha_j^{-1}(M, s_j) := \{s'_j \in \Sigma_j : M \in \alpha_j(s_j, s'_j)\}$.

For $M \in \mathcal{M}_j$ and $s_j \in \Sigma_j$, define the set of assignments for j that can result from following s_j starting at M by

$$\mathcal{O}_j(M, s_j) := \big\{ \Gamma_j\big(\zeta(h; s_j, s_{-j}, d_0)\big) : h \in M,\ s_{-j} \in \Sigma_{-j},\ d_0 \in D_0 \big\}.$$

Definition. A strategy $s_j \in \Sigma_j$ is *obviously dominant* for type $c_j \in C_j$ if, for every alternative strategy $s'_j \in \Sigma_j$ and every $M \in \alpha_j(s_j, s'_j)$,

$$\max_{\mu_j \in \mathcal{O}_j(M, s_j)} \mu_j \cdot c_j \ \leq \ \min_{\mu_j \in \mathcal{O}_j(M, s'_j)} \mu_j \cdot c_j.$$

The mechanism $(\mathcal{G}, s_{\mathcal{J}})$ is *obviously strategy-proof* if, for every $j \in \mathcal{J}$ and every $c_j \in C_j$, the prescribed strategy $s_j(c_j)$ is obviously dominant for type c_j.

Thus, at every possible earliest point of departure, the worst possible workload from following s_j is no larger than the best possible workload from following s'_j.

As in the static setting, we restrict attention throughout to extensive forms in which agents always have the right to demand their outside option.

Definition. The game form $\mathcal{G}$ *respects the agents' outside options* if for every $j \in \mathcal{J}$ there exists a strategy $s_j^0 \in \Sigma_j$ such that

$$g_j\big(\zeta(h; s_j, s_{-j}, d_0)\big) = \sigma_j$$

for every for every nonterminal history $h \in \mathcal{H} \setminus \bar{\mathcal{H}}$, every $s_{-j} \in \Sigma_{-j}$, and every $d_0 \in D_0$.

In other words, in any sub-game every player has a strategy that guarantees them their outside option. We call s_j^0 a *withdrawal strategy* for j.[21]

[21] This is slightly weaker than the following:

The critical class of games is a straightforward generalization of static polarized-ray games.

Definition. A *sequential extended polarized-ray game* is an extensive-form game of the following form

- In each terminal history each player moves exactly once.
- $A(M) = \hat{A}_j \cup W_j$ is an extended polarized-ray action set for every $j \in \mathcal{J}$ and $M \in \mathcal{M}_j$.
- The final assignment $\Gamma(a_{\mathcal{J}})$ satisfies $\Gamma_j(a_{\mathcal{J}}) \in a_j$ if $a_j \in \hat{A}_j$ and $\Gamma_j(a_{\mathcal{J}}) \in a_j$ if $a_j \in W_j$.

A *sequential extended polarized ray mechanism* consists of such a game, together with strategies that are myopically optimal at each information set, i.e.

$$s_j(c)(M) \in \underset{a_j \in A(M)}{\arg\min} \left\{ \inf \left\{ c_j \cdot x \ : \ x \in \{\Gamma_j\big(\zeta(h; a_j, s_{-j}, d_0)\big) : h \in M, s_{-J} \in \Sigma_{-j}, d_0 \in D_0\} \right\} \right\}$$

for all $M \in \mathcal{M}_j$.

We say that two extensive-form mechanisms are equivalent if they induce the same joint distribution of types and allocations. We call a game *single-move* if along each path of play, each agent moves at most once.

Theorem 2. Any sequential polarized-ray mechanism is OSP. If preferences are strictly monotone, any single-move OSP mechanism is equivalent to a sequential polarized-ray mechanism.[22]

Proof. Fix a single-move OSP mechanism $(\mathcal{G}, s_{\mathcal{J}})$ and an agent j. Without loss of generality, assume that $|A(M)| > 1$ for all $M \in \mathcal{M}_j$. Let $M \in \mathcal{M}_j^0$ be an information set at which j along some path of play. Observe that M is reachable under $s_j(c_j)$ for all $c_j \in C_j$, i.e. $\exists\ (s_{-j}, d_0) \in \Sigma_{-j} \times D_0$ such that $M \preceq \zeta(\varnothing; s_j(c_j), s_{-j}, d_0)$ for all $c_j \in C_j$.

Definition. The game form $\mathcal{G}$ *strongly respects the agents' outside options* if for every non-terminal history $h \in \mathcal{H} \setminus \bar{\mathcal{H}}$ and every $j \in \mathcal{J}$, either $g_j(\bar{h}) = \sigma_j$ for all $\bar{h} \succeq h$, or for any $\bar{h} \succeq h$ there exists $M \in \mathcal{M}_j$ and $a \in A(M)$ such that $h \preceq M \preceq \bar{h}$ and $g_j(\bar{h}') = \sigma_j$ for all $h' \in M$ and $\bar{h}' \succeq (h', a)$.

Imposing the stronger condition does not affect the results.

[22] The same characterization obtains if we drop the single-move restriction, but impose the stronger one-step dominance notion of Pycia and Troyan (2023b), rather than obvious dominance.

Consider a simultaneous-move game constructed as follows: The action space for each agent $k \in \mathcal{J} \setminus \{j\}$ is Σ_k, and the action space for nature is $M \times D_0$. That is, nature chooses a history in $h \in M$, in addition to the chance moves in D_0. The action space for j is $A(M)$. The selection rule is given by $\hat{\Gamma}(a_j, s'_{-j}, h, d_0) := \Gamma(\zeta(h; a_j, s'_{-j}, d_0))$ for any strategy profile $s'_{-j} \in \prod_{k \neq j} \Sigma_k$, $h \in M$, and $d_0 \in D_0$ (note that j's action in $A(M)$ is sufficient to define the outcome because it is a single-move game).

If strategy s_j is obviously dominant for type c_j of agent j in game $\mathcal{G}$ then it must also be obviously dominant for type c_j in the simultaneous-move game defined above. Then Theorem 2 follows from Theorem 1. □

4 When is choice useful?

Before we turn to the problem of choosing among OSP mechanisms, we take a step back and answer a more basic question: can sacrificing some control of the allocation to the agents help the designer achieve a lower expected social cost? We are interested in the answer to this question both when the designer is restricted to static OSP mechanisms, and when they can use any BIC mechanism. It is instructive, moreover, to also consider the value of choice under a relaxation of the status-quo constraint. Let m be a mechanism, where $m(\mathbf{c}) \in \mathcal{Z}$ is the allocation when agents report type profile $\mathbf{c}$, and write $m_j(i|\mathbf{c})$ for the measure of task i that goes to agent j. The relaxed status-quo constraint is

$$\sum_{i \in \mathcal{I}} c_j(i) \mathbb{E}_{-j}[m_j(i|c_j, c_{-j})] \leq \beta_j \sum_{i \in \mathcal{I}} c_j(i) \sigma_j(i) \quad \forall\ j \in \mathcal{J},\ c_j \in C_j,$$

where $\beta_j \geq 1$. In other words, no agent can be given an expected workload that is more than β_j times their expected workload under the status quo. The baseline case is $\beta_j = 1$, while $\beta_j > 1$ allows for slack in this constraint.

Consider the general BIC mechanism design problem, where aside from the status-quo constraint and the market-clearing condition (i.e. every task instance is assigned to exactly one agent), we impose no further restrictions on the mechanism used to allocate tasks. By the standard revelation principle, it is without loss of optimality to restrict attention to

incentive-compatible direct-revelation mechanisms. Then the designer's program is

$$\min_{m} \mathbb{E}\left[\sum_{i=1}^{I}\sum_{j=1}^{J}\pi_j(i)\ m_j(i|\mathbf{c})\right] \tag{2}$$

$$s.t. \quad \sum_{i\in\mathcal{I}} c_j(i)\mathbb{E}_{-j}[m_j(i|c_j,c_{-j})] \leq \beta_j \sum_{i\in\mathcal{I}} c_j(i)\sigma_j(i) \quad \forall\ j\in\mathcal{J},\ c_j\in C_j \tag{SQ}$$

$$\sum_{i\in\mathcal{I}} c_j(i)\ \mathbb{E}_{-j}[m_j(i|c_j,c_{-j})] \leq \sum_{i\in\mathcal{I}} c_j(i)\ \mathbb{E}_{-j}[m_j(i|c'_j,c_{-j})] \quad \forall\ j\in\mathcal{J},\ c_j,c'_j\in C_j \tag{IC}$$

$$\sum_{j\in\mathcal{J}} m_j(i|\mathbf{c}) = n^i \quad \forall\ i\in\mathcal{I},\ \mathbf{c}\in\mathbf{C} \tag{MC}$$

The designer here faces a multi-dimensional screening problem with a type-dependent participation constraint. While the solution to this problem is an open question, it turns out that we can answer the more limited question of whether or not the optimal mechanism must respond non-trivially to agents' preferences.

Say that *choice improves upon the status quo* if there is a mechanism m that satisfies the IC, SQ, and MC constraints, is not constant in the preference profile, and yields a strictly lower expected social cost than the status quo. Say that *choice is optimal* if all solutions to the designer's program in eq. (2) are non-constant mechanisms.[23]

Our first result shows that in many circumstances we should expect choice to be optimal. Say that *the status quo has full support* if $\sigma_j(i) > 0$ for all $i\in\mathcal{I}, j\in\mathcal{J}$. Say that *preferences have a common support* if $C_j = C_{j'}$ for all $j,j'\in\mathcal{J}$ (the distributions may still differ across agents). When preferences have a common support, it must be non-degenerate for choice to play any role. Say that preferences have *bounded support* if C_j is a bounded hypercube. Let $\bar{c}^i_j = \max\{c_j(i) : c_j\in C_j\}$ and $\underline{c}^i_j = \min\{c_j(i) : c_j\in C_j\}$, so $[\underline{c}^i_j, \bar{c}^i_j]$ is the domain of j's preference in dimension i. Say that *performance is identical* for all agents if $\pi_{j'} = \pi_{j''}$ for all $j',j''\in\mathcal{J}$.

Theorem 3. If preferences have a common bounded support and the status quo has full support, then the following are equivalent

i. Agents' preference distributions are non-degenerate, and performance is not identical

[23]From a design perspective, we could first solve for the optimal deterministic mechanism and set this as the status quo, in which case the only question is whether choice can improve upon the status quo. However, it is also useful to understand when choice is optimal under a given status-quo constraint.

across agents.

ii. Choice improves upon the status quo among static OSP mechanisms.

iii. Choice improves upon the status quo among all BIC mechanisms, for any $(\beta_j)_{j\in\mathcal{J}}$.

If, moreover, $\beta_j = 1$ for all j, and the support of agents' preferences has dimension I,[24] then the following is also equivalent to *i.* - *iii.*

iv. Choice is optimal among static OSP mechanisms (and a fortiori among all BIC mechanisms).[25]

Proof. Proof in Section A.2. □

Condition *i.* in Theorem 3 is clearly a necessary condition for choice to improve upon the status quo; if preference distributions are degenerate there is no role for choice, and if performance is identical there is no scope for improvement. The equivalence between *i.*, *ii.*, and *iii.* shows that under fairly mild conditions (common support of preferences and full-support status quo) choice improves upon the status quo whenever the problem is non-trivial, whether or not the designer is restricted to static OSP mechanisms. Moreover, in the baseline case where there is no slack in the SQ constraints and agents' preferences are not degenerate along any dimension, choice is also optimal whenever the problem is non-trivial. The assumption that the status quo has full support simplifies the characterization, but is not essential; it can be relaxed at the cost of additional notation.

At a high level, the conclusion that choice is generally valuable should not be surprising. However, it is worth emphasizing that Theorem 3 is not entirely straightforward because of the dual role played by the status-quo allocation: it is the benchmark against which the designer compares an alternative choice-based mechanism, and it determines the feasible set of mechanisms. Without the second role, i.e. if the feasible set did not depend on the

[24] As preferences are invariant with respect to strictly increasing affine transformations, it is always without loss of generality to normalize $\bar{c}^i_j = \underline{c}^i_j$ for some i. By saying that the support of an agent's preferences has dimension I, we mean that it is not possible to simultaneously normalize two dimensions to a constant. This is also equivalent to assuming that $\bar{c}^i_j = \underline{c}^i_j$ for at most one $i \in \mathcal{I}$. However, if this holds then it is also without loss of generality to assume that $\bar{c}^i_j > \underline{c}^i_j$ for all $i \in \mathcal{I}$.

[25] "Choice is optimal among static OSP mechanisms" means that every solution to the designer's program when they are additionally restricted to use static OSP mechanisms is a non-constant mechanism.

status quo, then there would certainly be no reason to suspect that an arbitrary degenerate mechanism could not be improved upon by choice. The question here is (slightly) more subtle because the alternative must constitute an improvement for each of the agents, as well as for the designer.

Theorem 3 establishes that there is, in general, value in using mechanisms that respond to agents' preferences. In fact, the proof shows that to get gains it suffices to use a simple static OSP mechanism that takes the form of bilateral trade: the designer specifies an allowable trade between two agents, which is implemented if both agents assent.

5 Towards optimal veto rights

We now turn to the problem of choosing among static OSP mechanisms. By Theorem 1 and Proposition 1, when preferences are strictly monotone the designer can without loss of optimality restrict attention to EOPR mechanisms. Despite the significant dimension reduction that this entails, the problem remains challenging. There are three main difficulties with optimizing over the set of EOPR mechanisms. First, the polarization constraint contains an existential qualifier, which makes it difficult to work with. Second, the problem is not separable across agents, due to the market-clearing constraint. In particular, the selection rule is a complicated function of the menus offered to each agent, determined by the linear program in (1). Third, while we assume that the designer's objective is a linear function of the expected assignment of each agent, the expected assignments are determined by the underlying polarized-ray mechanism in ways that are difficult to analyze.

While the last point is inherent to the exercise of optimizing over EOPR mechanisms, we develop techniques to deal with the first two challenges. First, we characterize the convex structure of polarized-ray mechanisms. This allows us to represent such mechanisms as distributions over a set of extreme points, thus ensuring that the polarization constraint is satisfied implicitly. Second, we introduce a "large-market" relaxation of the designer's program, which simplifies the market-clearing condition. A solution to the relaxed program can then be taken as a starting point in the search for solutions to the original program; we show that the solution to the relaxed program is approximately optimal in the original

program for large markets. It is not in general possible to solve the relaxed program analytically (unless there are only two types of tasks), and its computational complexity remains an open question. Nonetheless, the parallelization enabled by the large-market relaxation significantly reduces the computational burden of searching for approximate solutions.

5.1 The convex structure of polarized rays

In this section we study the structure of the set of polarized-ray veto-set menus. This set is not necessarily convex; indeed, it is not even obvious what it means to add two menus. However, we show that the set can be divided into a collection of subsets, such that each subset is convex (in a suitably defined vector space).

As this section concerns the menu offered to a given agent, we suppress the j subscript in the notation. To begin, observe that we can discuss whether the veto sets for a given agent are polarized rays without reference to the endowment. That is, given a menu of rays A (polarized or not) with endowment η, let $R \subset \mathbb{R}^I$ be the collection of vectors such that $\kappa \in R$ if and only if $a = \{x \in \mathbb{R}^I : x = \eta + y\kappa,\ y \in \mathbb{R}_{\geq 0}\}$ for some $a \in A\{\eta\}$ (this last condition is just a normalization, meaning that we drop the 0 vector from R). In other words, R is the collection of directions defining the rays in A. The polarization condition concerns only R, and is independent of η. If the rays in R are polarized, we call R a polarized-ray menu (in which case it can have at most I elements). Our objective is to characterize the set of such menus, which we denote by $\mathcal{P}$.

Let $\mathcal{P}^n$ be the subset of $\mathcal{P}$ in which there are exactly n rays, for $1 \leq n \leq I$. As we have already observed, in any polarized ray menu each ray must lie in a different closed orthant of $\mathbb{R}^I$ (and cannot lie in the positive or negative orthants). We use the following notation: for each sign vector $\zeta \in \{1, -1\}^I$ we define the closed orthant

$$\mathcal{O}_\zeta := \{x \in \mathbb{R}^I : x_i \zeta_i \geq 0 \ \forall\ i\}$$

and let $\mathcal{O} = \{\mathcal{O}_\zeta\}_{\zeta \in \{1,-1\}^I}$ be the set of orthants. A *class* $P \subset \mathcal{P}$ consists of the set of polarized ray menus in which the rays occupy a given set of orthants. Formally, we define a class P by a set of sign vectors $E(P) \subset \{1, -1\}^I$, such that a menu R is in P if and only if

i. For every $\zeta \in E(P)$ there is exactly one ray $\kappa \in R$ such that $\kappa \in \mathcal{O}_\zeta$, and

ii. For any $\kappa \neq \kappa' \in R$ there exists $\alpha \in (0,1)$ such that $\alpha\kappa + (1-\alpha)\kappa' \geq 0$.

Note that by construction, all menus in a given class have the same number of rays, and the collection of all classes covers $\mathcal{P}$.

We view a class $P \subset \mathcal{P}^n$ as a subset of $\mathbb{R}^{I \times E(P)}$. For a polarized-ray menu R, we abuse notation and write $R(\zeta)$ for the unique $\kappa \in R \cap \mathcal{O}_\zeta$. The convex combination $\alpha R + (1-\alpha)R'$ between menus R, R' in class P is defined in the standard way: $[\alpha R + (1-\alpha)R'](\zeta) = \alpha R(\zeta) + (1-\alpha)R'(\zeta)$ (note that this operation is well defined only if R, R' belong to the same class). Denote the closure of class P by $\bar{P}$. We want to describe the structure of $\bar{P}$ for any $1 \leq n \leq I$ and any class $P \subset \mathcal{P}^n$. We begin with a preliminary characterization.

Lemma 3. Menu R is in $\bar{P}$ if and only if

i. For every $\zeta \in E(P)$ there is at least one ray $\kappa \in R$ such that $\kappa \in \mathcal{O}_\zeta$, and

ii. For any $\kappa \neq \kappa' \in R$ there exists $\alpha \in [0,1]$ such that $\alpha\kappa + (1-\alpha)\kappa' \geq 0$.

Proof. Proof in Section A.3. □

Note the differences between the characterizations of P and $\bar{P}$, which are due to the fact that rays in $R \in \bar{P}$ can lie on the boundary of a closed orthant, whereas for $R \in P$ the rays must lie in the interior.

Both $\bar{P}$ and P are convex cones: If $R \in P$ (or $\bar{P}$) then so is αR for any $\alpha > 0$, and if $R, R' \in P$ (or $\bar{P}$) then so is $\alpha R + (1-\alpha)R'$ for any $\alpha \in [0,1]$. Moreover, $\bar{P}$ is a polyhedral cone: there exists a finite set of *extremal generators* $\{R_1, \ldots, R_n\} \subset \bar{P}$ such that $\bar{P} = \{a_1 R_1 + \cdots + a_n R_n : a_j \in \mathbb{R}_{\geq 0} \ \forall\ j\}$, and such that for all $1 \leq j \leq n$ if $R_j = R' + R''$ and $R', R'' \in \bar{P}$ then $R' = \alpha R_j, R'' = \beta R_j$ for some $\alpha, \beta \in \mathbb{R}_{\geq 0}$. Notice also that $\bar{P}$ is closed under separate scalar multiplication of each ray in a menu:

Definition. Menu R' is *equivalent* to $R \in \bar{P}$ if there exists $\beta \in \mathbb{R}_{\geq 0}^{E(P)}$ such that $R' = \{\beta(\zeta)R(\zeta)\}_{\zeta \in E(P)}$.

From Lemma 3, we can see that if $R \in \bar{P}$ and R' is equivalent to R then $R' \in \bar{P}$. In light of this observation, we have the following equivalent definition of extremal generators.

Definition. Menu $R \in \bar{P}$ is an *extremal generator* of $\bar{P}$ if there exist no $R', R'' \in \bar{P}$, with neither equivalent to R, such that $R = \alpha R' + (1-\alpha)R''$ for some $\alpha \in [0,1]$.

We show that it suffices to restrict attention to simple menus consisting only of rays that are collinear with the standard basis vectors (which we denote by e_j for $j \in I$), and such that only one ray is non-positive.

Definition. A menu R is *simple* if

i. For all $\kappa \in R$, either $\kappa = e_j$ or $\kappa = -e_j$ for some $j \in \{1, \ldots, I\}$.

ii. There is at most one $\kappa \in R$ such that $\kappa \not\geq 0$.

The set of extremal generators of $\bar{P}$ is unique only up to scalar multiplication. We therefore refer to *the* extremal generators of $\bar{P}$ as those in which each ray has length 1. The main result of this section is a characterization of the extremal generators of $\bar{P}$ for any $1 \leq n \leq I$ and any class $P \subset \mathcal{P}^n$.

Proposition 3. For any class P, the set of extremal generators of $\bar{P}$ consists of all and only the menus R such that

i. $|R| = |E(P)|$ and for every $\zeta \in E(P)$ there is at least one $\kappa \in R$ such that $\kappa \in \mathcal{O}_\zeta$

ii. R is simple.

The core of the proof of Proposition 3 is the following intermediate result.

Proposition 4. If R is an extremal generator of $\bar{P}$ then for each $\kappa \in R$ there is exactly one $1 \leq i \leq I$ such that $\kappa(i) \neq 0$.

Proof. Proof in Section A.4. □

Given Proposition 4, the proof of Proposition 3 is straightforward.

Proof of Proposition 3. It is immediate from Lemma 3 that if R satisfies conditions *i.* and *ii.* of Proposition 3 then $R \in \bar{P}$, so we can write it as $R = \{R(\zeta)\}_{\zeta \in E(P)}$. Moreover, because R is simple, $R(\zeta)$ cannot be written as a convex combination of any two $\kappa, \kappa' \in \mathcal{O}_\zeta$ unless $R(\zeta), \kappa, \kappa'$ are collinear. Thus R is an extremal generator of $\bar{P}$.

For the converse, first observe that if $R \in \bar{P}$ then it must satisfy condition $i.$ of Proposition 3, so we only need to show that any extremal generator of $\bar{P}$ is simple. Given Proposition 4, we need to show that there is at most one $\kappa \in R$ such that $\kappa \not\geq 0$. If $\kappa \not\geq 0$ and $\kappa' \not\geq 0$ for $\kappa \neq \kappa' \in R$ then $\kappa = -\mathbb{1}_i$ and $\kappa' = -\mathbb{1}_{i'}$ for some $i, i' \in \{1, \ldots, I\}$, by Proposition 4. But then $\alpha\kappa + (1-\alpha)\kappa' \not\geq 0$ for all $\alpha \in [0,1]$, so the menu is not polarized. □

As detailed in Section 5.2, the key step towards identifying (approximately) optimal polarized-ray mechanisms is to be able to solve linear programs over the expected allocations induced by polarized ray menus for each agent. For this purpose, the characterization of the extremal generators of $\bar{P}$ for each class P is extremely useful: rather than optimize over P directly, we can optimize the set of distributions supported on $\bar{P}$. This is convenient because it replaces the (complex) constraint that the rays in a menu be polarized with the (simple) constraint that distributions we consider lie in the simplex. Of course, we must still solve a high-dimensional non-linear optimization problem over each class. Thus the approach will be particularly useful when there are not too many classes, and the closure of each class does not have too many extremal generators.

With I task types, we can identify each class P with a set of sign vectors $E(P)$ such that for any coordinate $i \in \{1, \ldots, I\}$ there is at most one $\zeta \in E(P)$ such that $\zeta_i = -1$. Moreover, each ζ must have at least one positive entry, and no more than $I-1$ positive entries (by definition of polarized rays). Let $p(\zeta)$ be the number of positive entries of ζ. Then by Proposition 3 we have the following immediate observation

Lemma 4. The number of extremal generators for class P is

$$\sum_{\zeta \in E(P)} (I - p(\zeta)) \prod_{\zeta' \neq \zeta} p(\zeta') + \prod_{\zeta \in E(P)} p(\zeta)$$

where the first term is the number of extremal generators in which one ray is non-positive, and the second term is the number in which none are.

It is easy to show that the expression in Lemma 4 is increasing in $p(\zeta)$ for any ζ, so it is maximized when $p(\zeta) = I-1$ for all $\zeta \in E(P)$, which yields $|E(P)|(I-1)^{|E(P)|-1} + (I-1)^{|E(P)|}$. For example, for $I = 4$ and $I = 5$ the maximal number of extremal generators for any class is 189 and 2,304 respectively (and this number is significantly lower for many classes). However,

with $I = 10$ there are classes with $7,360,989,291$ extremal generators.

Let $C(I)$ be the number of classes as a function of the number of task types. For $h > 1$, let $c(N, h)$ be the number of classes such that $|E(P)| = h$ when there are N task types. Note that the number of classes such that $|E(P)| = 1$ is $\sum_{j=1}^{N-1} \binom{N}{j} = 2^N - 2$.

Lemma 5. $C(I) = 2^I - 2 + \sum_{h=2}^{I} c(I, h)$, and $c(I, h)$ can be written recursively as

$$c(I, h) = \frac{1}{h} \sum_{j=1}^{I-(h-1)} \binom{I}{j} c(I - j, h - 1) \tag{3}$$

for $I \geq h > 1$, where we define $c(N, 1) := \sum_{j=1}^{N} \binom{N}{j}$.[26]

Proof. Proof in Section A.5. □

When the number of task types is moderate, we can thus exploit the convex structure of polarized ray menus to significantly reduce the computational burden of identifying (approximately) optimal static OSP mechanisms.

5.2 The large-market model

In the designer's true program every task must be assigned to an agent for every realized preference profile. Suppose instead that we require only that every task be assigned in expectation, taken over the preference profiles. That is, for a mechanism m we replace the market-clearing condition from eq. (2) with

$$\mathbb{E}\left[\sum_{j \in \mathcal{J}} m_j(i|\mathbf{c})\right] = n^i \quad \forall\, i \in \mathcal{I}. \tag{4}$$

One interpretation of this relaxation is that each agent is actually a unit mass population of agents with identical performance, and F_j describes the distribution of preferences in this population. We thus refer to (4) as the large-market market-clearing (LM-MC) constraint.

Fix an endowment $\eta \in \mathcal{Z}$. Let A_j be a collection of veto sets for each j, with endowment η_j, and let $s_j : C_j \mapsto A_j$ be a myopically optimal strategy for j. Let $(\Gamma_j : A_j \to \mathbb{R}^I_{\geq 0})_{j \in \mathcal{J}}$ be a collection of *individual-level selection rules* for the designer, where we require that

[26] For example $C(2) = 3, C(3) = 13, C(4) = 50, C(6) = 875, C(8) = 21,145, C(10) = 678,568$, and $C(12) = 27,644,435$.

$\Gamma_j(a_j) \in a_j$, but do not impose that together $(\Gamma_j)_{j\in\mathcal{J}}$ satisfy ex-post market clearing. Then the mechanism in which each agent chooses from A_j according to strategy s_j, and the designer chooses from $s_j(c_j)$ according to Γ_j, satisfies the LM-MC constraint if and only if

$$\sum_{j\in\mathcal{J}} \mathbb{E}_j[\Gamma_j(s_j(c_j))] = (n_1, \ldots, n_I).$$

Conversely, if we replace the individual-level selections rules $(\Gamma_j)_{j\in\mathcal{J}}$ with a (legitimate) selection rule $\Gamma : \prod_{j\in\mathcal{J}} A_j \mapsto \mathcal{Z}$, which satisfies ex-post market clearing, then LM-MC is satisfied a fortiori. Thus the following program is a lower bound on the social cost from choosing the optimal static OSP mechanism with endowment η:

$$\min_{(A_j,\Gamma_j)_{j\in\mathcal{J}}} \quad \sum_{j=1}^{J} \pi_j \cdot \mathbb{E}\left[\Gamma_j(s_j(c_j))\right] \tag{5}$$

$$s.t. \quad \text{either } j\text{'s veto sets are polarized rays, or } j \text{ receives}$$

$$\text{two veto sets: } \{\eta_j\} \text{ and } \{x \in \mathbb{R}^I_{\geq 0} \ : \ x \leq \eta_j\}; \text{ and} \tag{OSP}$$

$$\sum_{j\in\mathcal{J}} \mathbb{E}_j[\Gamma_j(s_j(c_j))] = (n_1, \ldots, n_I). \tag{LM-MC}$$

where s_j is a myopically optimal strategy for j (which is unique up to zero-measure perturbations). We can relax this program further by allowing the designer to randomize over the veto sets and selection rule for each j. Under the population interpretation of the large-market program, randomization is equivalent to splitting each "agent" into subsets that receive different mechanisms. Therefore to economize on notation we do not explicitly model randomization in program (5), as it can be represented by redefining $\mathcal{J}$.[27] We refer to the program in (5) (allowing for randomization) as the *large-market OSP program.*

[27] Technically, this notational trick is only valid if the randomization involves rational weights, but this should not cause confusion.

5.3 Solving the large-market OSP program

It is convenient to separate out the parts of the program that depend only on $\mathbb{E}_j[\Gamma_j(s_j(c_j))]$. Define the *OSP-feasible set* for j by

$$\mathcal{F}_j := \text{cvh}\Big\{\mathbb{E}_j[\Gamma_j(s_j(c_j))] \in \mathbb{R}^I_{\geq 0} \ : \ \Gamma_j(a_j) \in a_j \ \forall a_j \in A_j \ ; \ s_j \text{ is myopically optimal; and either } A_j \text{ consists of polarized rays, or } j \text{ receives two veto sets: } \{\eta_j\} \text{ and } \{x \in \mathbb{R}^I_{\geq 0} : x \leq \eta_j\}\Big\}.$$

where cvh denotes the convex hull. In other words, $\mathcal{F}_j$ is the set of expected allocations for j that can be induced using a (random) veto protocol and selection rule satisfying all the conditions of program (5) aside from (LM-MC). Then the large-market OSP program has the same value as

$$\min_{(x_j)_{j\in\mathcal{J}}} \sum_{j\in\mathcal{J}} \pi_j \cdot x_j \tag{6}$$

$$s.t. \quad x_j \in \mathcal{F}_j \quad \forall \ j \in \mathcal{J}$$

$$\sum_{j\in\mathcal{J}} x_j(i) = n^i \quad \forall \ i \in \mathcal{I} \tag{LM-MC}$$

We refer to eq. (6) as the *outer program*. It is instructive to study the dual of this program, which can be written as

$$\sup_{\lambda\in\mathbb{R}^I} \sum_{i\in\mathcal{I}} \lambda^i n^i - \sum_{j\in\mathcal{J}} S_j(\lambda - \pi_j) \tag{7}$$

where $S_j(w) := \max\{w \cdot y \ : \ y \in \mathcal{F}_j\}$ is the support function of the convex set $\mathcal{F}_j$. Strong duality holds, meaning that the values of the outer program in eq. (6) and its dual in eq. (7) coincide.

Proposition 5. If $\mathcal{F}_j$ is bounded for all j then strong duality for the outer program holds. Moreover, if λ_* is a solution to the program in eq. (7) then there exist selections

$$x_j \in \arg\max\{(\lambda_* - \pi_j) \cdot y \ : \ y \in \mathcal{F}_j\}$$

such that $(x_j)_{j\in\mathcal{J}}$ satisfy market clearing, and these solve the program in eq. (6).

Proof. Proof in Section A.6. □

The condition that $\mathcal{F}_j$ is bounded is mild. It holds if the marginal rate of substitution

between different tasks is bounded, i.e. $\underline{c} \leq c_j(i) \leq \bar{c}$ for all i and some $0 < \underline{c} < \bar{c} < \infty$. It also holds if j's type distribution is sufficiently concentrated on such a set.

Remark 2. Baron et al. (2026) study a version of the large-market program with only two types of tasks, but allowing for any BIC mechanism. However, they show that under a regularity condition on the type distribution, the optimal mechanism is in fact a polarized-ray mechanism (with two rays). In this case, the restriction to static OSP mechanisms is without loss of optimality (in the large-market program) and the support function can be characterized analytically.

Given the support functions $(S_j)_{j\in\mathcal{J}}$, the dual program in eq. (7) is a simple I-dimensional concave maximization program. The only difficulty lies in characterizing the support functions. Writing the program defining these functions more explicitly, we have

$$\begin{aligned} S_j(w) = \max_{A_j,\Gamma_j,s_j} \quad & w \cdot \mathbb{E}_j[\Gamma_j(s_j(c_j))] \\ s.t. \quad & s_j \text{ is myopically optimal;} \\ & \Gamma_j : A_j \to \mathbb{R}^I_{\geq 0} \text{ ; and} \\ & \text{either } A_j \text{ consists of polarized rays, or } j \text{ receives} \\ & \quad \text{two veto sets: } \{\eta_j\} \text{ and } \{x \in \mathbb{R}^I_{\geq 0} : x \leq \eta_j\}. \end{aligned}$$

Ultimately this program needs to be solved computationally. The fact that the support function can be characterized in parallel for each agent is of some assistance, but the task of optimizing over the set of polarized-ray menus remains challenging. However, this problem can be simplified by exploiting the convex structure of this set, as described in Section 5.1: within each class of polarized rays (i.e. those in which the ray directions occupy a given set of orthants) we can write the program defining $S_j(w)$ as a maximization problem over the set of distributions on the extremal generators of this class, which are characterized by Proposition 3. In principle, we could then compare the solutions across all classes to obtain a global solution and define the value of $S_j(w)$. The benefit of this approach is that it imposes the challenging constraint (polarized rays) by construction. Still, there are two problems with this technique in practice. First, when I is large there are many classes (Lemma 5) and many extremal generators for each class (Lemma 4). Second, the objective is a complicated

non-linear function of the menu. Nonetheless, the extremal-generators approach may be useful for obtaining approximate solutions. We employ this approach in Section 6.

5.4 From the large market back to the small

As discussed above, any EOPR mechanism improves upon the status quo, both from the designer's perspective and from that of every agent. A veto protocol that solves the large-market program is a natural candidate. This mechanism is (unsurprisingly) approximately optimal with a large number of agents. To formalize this claim, consider a sequence of replica economies, indexed by an integer N. In the N-replica economy, there are N copies of each agent with identical performance, but with preferences drawn independently from the same distribution. We refer to these agents as "group j". There are also Nn^i copies of task i.

Fix a solution to the large-market program, which is an EOPR mechanism that we denote by m^*. For each N, we implement the mechanism m^*. If m^* involves randomization for agent j (as permitted in the large-market program) then we replicate the randomization in the N-replica economy by offering each veto protocol in the support of the randomized mechanism to a proportional fraction of the N group-j agents.[28] Let $V^*_{LM}(N)$ be the random variable corresponding to the social cost under this mechanism, and let $V^*(N)$ be the corresponding random variable for the optimal mechanism in the N-replica economy (they are random because they depend on the realized preference profile). We say that the solution to the large-market program is *asymptotically optimal as the market grows* if $\frac{1}{N}\mathbb{E}[V^*_{LM}(N)] \to \frac{1}{N}\mathbb{E}[V^*(N)]$ as $N \to \infty$.

Proposition 6. The optimal mechanism in the large-market program is asymptotically optimal in the original program as the market grows.

Proof. Proof in Section A.7. □

[28] That is, if under m^* agent j receives veto protocol A_j with probability ε, then $\lfloor \varepsilon N \rfloor$ of the group-j agents receive this veto protocol (any residual agents can be allocated to any mechanism in the support of m^*, as this will not affect the asymptotic results).

5.5 Optimization procedure

In summary, we have identified the following procedure for identifying promising static OSP mechanisms in applications:

1. *Estimate the support functions of the OSP-feasible sets for each agent.* This is done by solving linear programs over the expected assignments for j that can be induced when j chooses from a menu of polarized rays. The program is simplified by using Proposition 3 to rewrite the problem as choosing distributions over the extremal generators, but with many task types it is generally only possible to obtain approximate solutions (see discussion in Section 5.3).

2. *Solve the large-market problem to identify promising polarized-ray veto protocols.* Given an estimate of the support function for each agent, the dual is a straightforward concave maximization problem. Given a solution to the dual, the protocol outlined above (using Proposition 3 to optimize over menus of polarized rays) can be used to identify a veto protocol.

6 Empirical Applications

This section evaluates whether the class of mechanisms characterized in the theory can generate gains relative to existing assignment systems in two important public-sector settings: the assignment of CPS investigations to caseworkers and the assignment of misdemeanor cases to prosecutors. Both environments share a common organizational structure. Heterogeneous tasks are allocated across heterogeneous agents through quasi-random assignment systems designed to equalize workloads and preserve procedural fairness. These systems, however, do not account for two forms of heterogeneity emphasized in the model: agents may differ in their performance across task types, and they may have heterogeneous preferences over assignments. Appendix B.1 provides additional institutional background.

In both settings, the planner faces a constrained task-allocation problem in which the potential gains from specialization must be balanced against institutional constraints to preserve existing workloads and assignment rights. Both environments are high-stakes policy

settings. Approximately 37% of children in the United States are the subject of a CPS investigation by age 18, while misdemeanor cases account for more than 80% of criminal cases in the United States (Kim et al., 2017; Agan et al., 2023). In the CPS setting, investigators determine whether children should remain in the home or be placed into foster care. In the prosecution setting, assistant district attorneys determine whether misdemeanor cases should be advanced to prosecution. Although misdemeanor offenses often involve relatively low-level conduct, the decision to prosecute can have lasting consequences, as prosecution typically results in a criminal record even in the absence of a conviction. Improving the quality of these decisions therefore has potentially large social benefits.

The institutional environments closely match the structure of the model. In both settings, cases are allocated through rotational systems that approximate random assignment conditional on office and time variation. These systems generate a status-quo allocation that is independent of agents' preferences and approximately equalizes caseloads in expectation. At the same time, they may fail to exploit comparative advantage across agents. For example, some agents may be relatively effective at handling high-risk maltreatment investigations or criminal prosecutions while others may perform relatively better on lower-risk cases.

The participation constraint emphasized in the theory is particularly relevant in these environments. Both CPS agencies and prosecutor offices operate under substantial staffing constraints (e.g., see Casey Family Programs (2023)), and the costs of handling cases may vary both across agents and case types. For example, handling a high-risk maltreatment investigation or criminal prosecution may impose greater cognitive, emotional, or administrative burdens on some agents. Reforms that systematically assign agents to less desirable case mixes may therefore increase burnout, reduce morale, and exacerbate turnover. These effects may in turn be self-reinforcing: if some agents exit or reduce effort, the remaining staff must absorb higher workloads, further straining the system. In practice, assignment mechanisms that make a subset of agents worse off relative to the status quo are therefore unlikely to be implementable. Moreover, as in many public-sector settings, transfers are not naturally available as a design instrument. Compensation is typically determined by rigid salary schedules tied to tenure, rank, or civil-service rules rather than to specific assignments, limiting the extent to which monetary transfers can offset undesirable case allocations.

For expositional simplicity, we abstract from the dynamic features of these environments and study static assignment problems in which a fixed set of cases is allocated across agents. In practice, however, cases arrive sequentially and must be assigned without knowledge of future arrivals. Nevertheless, Baron et al. (2026) show how over sufficiently long horizons a static mechanism can be well-approximated by one that works in the dynamic setting.

6.1 Measurement, Identification, and Data

We now describe how we measure agent performance and implement the empirical counterpart to the planner's objective. Despite the institutional differences between the CPS and prosecution applications, the empirical structure is similar in both settings. Agents make binary intervention decisions on assigned cases based on assessments of downstream risk. In the CPS setting, investigators decide whether a child should remain in the home or be placed into foster care. In the prosecution setting, assistant district attorneys decide whether a misdemeanor case should be advanced to prosecution.

A natural way to organize both decisions is in terms of expected downstream outcomes absent intervention. In the CPS setting, the relevant outcome is subsequent maltreatment if a child remains in the home, while in the prosecution setting an important consideration is subsequent criminal justice involvement absent prosecution. Framed this way, both applications can be viewed as statistical classification problems in which interventions are allocated based on expected downstream risk. Appendix B.1 provides additional institutional details and further discussion of the objectives faced by decision-makers in each setting.

Modeling performance. Let k index individual cases, j index agents, and $i \in \mathcal{I}$ index case types. Let $D_{kj} \in \{0, 1\}$ denote agent j's intervention decision for case k. In the CPS application, $D_{kj} = 1$ indicates foster care placement, while in the prosecution application $D_{kj} = 1$ indicates that the case is advanced to prosecution.

Let $Y_k^* \in \{0, 1\}$ denote the relevant downstream outcome absent intervention. In the CPS setting, $Y_k^* = 1$ indicates that the child would experience subsequent maltreatment if left in the home (not placed in foster care). In the prosecution setting, $Y_k^* = 1$ indicates that the defendant would experience subsequent criminal justice involvement absent prosecution.

The joint realization of (D_{kj}, Y_k^*) generates four possible classification outcomes: false negatives, false positives, true negatives, and true positives. Define the corresponding realized indicators

$$\mathrm{FN}_{kj} = \mathbf{1}\{D_{kj} = 0, Y_k^* = 1\}, \quad \mathrm{FP}_{kj} = \mathbf{1}\{D_{kj} = 1, Y_k^* = 0\},$$

$$\mathrm{TN}_{kj} = \mathbf{1}\{D_{kj} = 0, Y_k^* = 0\}, \quad \mathrm{TP}_{kj} = \mathbf{1}\{D_{kj} = 1, Y_k^* = 1\}.$$

Exactly one of these indicators equals one for each case-agent pair. We define the expected social cost generated by assigning a type-i case to agent j as

$$\pi_j(i) = \mathbb{E}[w_{\mathrm{FN}}\mathrm{FN}_{kj} + w_{\mathrm{FP}}\mathrm{FP}_{kj} + w_{\mathrm{TN}}\mathrm{TN}_{kj} + w_{\mathrm{TP}}\mathrm{TP}_{kj} \mid g(k) = i]$$

where $g(k) \in \mathcal{I}$ denotes the type of case k and $(w_{\mathrm{FN}}, w_{\mathrm{FP}}, w_{\mathrm{TN}}, w_{\mathrm{TP}})$ are policy weights reflecting the relative social costs of each classification outcome. These weights are primitives of the planner's objective and must be specified by the designer.

Identification challenge. A central empirical challenge in both applications is that $\pi_j(i)$ is not nonparametrically identified because of the standard "selective labels" problem. Downstream outcomes are only observed under one realization of the intervention decision. For example, in the CPS setting we do not observe whether maltreatment would have occurred had a child placed into foster care instead remained in the home. Similarly, in the prosecution setting we do not observe whether a prosecuted defendant would have experienced subsequent criminal justice involvement absent prosecution.

Our approach, following Baron et al. (2026) (Lemma 1), is therefore to focus on relative performance differences across agents rather than on absolute levels of performance. Under quasi-random assignment, they show that differences in expected outcomes across agents within the same rotational assignment unit (e.g., office-by-time cells) are identified separately for each case type. Intuitively, the planner need not know the absolute social cost generated by each agent-case pair; it is sufficient to know which agents perform relatively better on different classes of cases. Because the planner's objective depends only on relative differences in $\pi_j(i)$, these comparisons are sufficient to rank agents and evaluate the welfare consequences of alternative assignment mechanisms. Throughout the empirical analysis, we therefore

interpret $\pi_j(i)$ as performance relative to a reference agent within each application. Appendix B.2 discusses identification in more detail.

Data and Estimation. For the CPS application, we use records from the Michigan Department of Health and Human Services covering the universe of child maltreatment investigations in the state between 2008 and 2016. These records include investigator identifiers, placement decisions, case characteristics, and subsequent CPS involvement. The resulting analysis sample contains 168,331 investigations assigned to 377 investigators across 45 counties, approximately 3% of which result in foster care placement.

For the prosecution application, we use records from the Suffolk County District Attorney's Office covering nonviolent misdemeanor cases between 2004 and 2018. The dataset, originally assembled by Agan et al. (2023), includes prosecutor identifiers, prosecution decisions, case characteristics, and subsequent criminal justice outcomes. The resulting analysis sample contains 31,840 cases assigned to 62 prosecutors, approximately 80% of which are advanced to prosecution.

We define the downstream outcome Y_k^* using measures of subsequent CPS or criminal justice involvement observed over fixed post-assignment horizons: six months in the CPS application, following Baron et al. (2026), and two years in the prosecution application, following Agan et al. (2023). We restrict attention to samples for which the quasi-random assignment assumption is plausible and for which downstream outcomes can be consistently measured. Additional details on sample construction are provided in Appendix B.3.

To map the empirical environments into the finite-type framework of the model, we partition cases into a finite set of task categories. While there are many possible ways to define case types—for example using allegation or offense categories, prior histories, case severity, or demographic characteristics—we focus on variation in predicted downstream risk absent intervention. In each application, we therefore estimate a predictive model using pre-assignment case characteristics to construct a measure of each case's predicted downstream risk. We then partition cases into quartiles of the predicted risk distribution, which define the task types used throughout the analysis.

This approach captures a dimension of heterogeneity that is directly relevant for both

comparative advantage and heterogeneous preferences. Higher-risk cases are likely to require greater investigative effort, scrutiny, and emotional burden, making them a natural source of both performance heterogeneity and assignment preferences across agents.

Using these task categories, we estimate agent-level performance measures separately by case type. To mitigate overfitting and estimation noise, we implement a split-sample procedure in which mechanisms are estimated in a training sample and evaluated in a held-out sample. We additionally apply empirical Bayes shrinkage to address measurement error in agent-level estimates and calibrate the social-cost parameters $(w_{\text{FN}}, w_{\text{FP}}, w_{\text{TN}}, w_{\text{TP}})$. Full estimation details are provided in Appendix B.4.

Preferences and task partitions. Agents' preferences over assignments, represented by c_j, are unobserved. In the model, these preferences are drawn from a distribution F_j, which is assumed to be known to the planner. We therefore evaluate mechanisms under calibrated preference distributions rather than estimating investigator- or prosecutor-specific preferences directly. In the CPS application, the preference distribution is informed by survey evidence on investigators' relative willingness to handle higher-risk cases (Baron et al., 2026). In the prosecution application, comparable survey evidence on prosecutors' case-type preferences is unavailable. We therefore use a stylized preference distribution with increasing expected costs across risk quartiles, reflecting the possibility that higher-risk cases generate greater workload, stress, or emotional burden. Full preference distribution details are provided in Appendix B.5.

We consider mechanisms defined over several possible task partitions. The richest specification treats the four predicted-risk quartiles as distinct task types, so that agents have quartile-specific assignment costs $c_j = (c_{j1}, \ldots, c_{j4})$. We also consider coarser partitions formed by aggregating subsets of these quartiles, yielding seven binary partitions and six three-group partitions of the four quartiles. For each coarser partition, we collapse the quartile-level social-cost estimates, status-quo allocations, and preference draws to the coarser task groups, as described in Appendix B.6. Thus, the binary and three-group mechanisms are re-optimized from the same underlying quartile-level performance and preference primitives, rather than from separately estimated lower-dimensional moments.

For each candidate partition, we solve the mechanism using training-sample, social-cost estimates and compute its training-moment small-market welfare gain. We define the optimal mechanism as the candidate with the largest training-moment, small-market reduction in social cost. We then evaluate and report welfare gains for this selected mechanism using held-out test moments. We contrast these gains with a full-information benchmark, which uses the quartile-level social-cost estimates and allows the planner to observe each agent's full four-dimensional preference type.

6.2 Simulation Results

Panel A of Table 1 reports the estimated gains from the optimal CPS assignment mechanisms relative to the status-quo rotational assignment system. The table reports changes in the planner's weighted social-cost objective—the aggregate value of $\pi_j(i)$ across assigned investigator-case pairs—as well as changes in false negatives, false positives, and overall foster care placements. The table compares three environments: the small-market mechanism derived from the finite-agent model, the corresponding large-market approximation, and a full-information benchmark in which the planner directly observes investigators' task-specific preferences. See Appendix B.7 for details on the computation of welfare changes.

The optimized mechanisms generate economically meaningful improvements relative to the status quo. The small-market mechanism reduces weighted social costs by approximately 3%, while the large-market approximation reduces costs by nearly 5%. Both mechanisms reduce false negatives and false positives, with the small-market mechanism reducing false positives by nearly 7%, while also slightly reducing overall foster care placements. The mechanisms also recover a substantial share of the gains achievable under full information. The full-information benchmark reduces weighted social costs by approximately 17%, implying that the small-market mechanism captures roughly one-fifth of the gains available if the planner directly observed investigators' preferences (which is a loose upper bound on the gains from any mechanism in the true asymmetric-information problem). Importantly, these gains arise entirely from reallocating existing cases across investigators and do not require

Table 1: Performance Gains from Optimized Assignment Mechanisms

	(1) Small-Market Mechanism	(2) Large-Market Approximation	(3) Full-Information Benchmark
Panel A. CPS Investigations			
Weighted social cost	-317.4**	-500.8***	-1,762.5***
	(156.2)	(65.3)	(546.1)
	[-3.0%]	[-4.8%]	[-16.8%]
False negatives	-229.4*	-369.0***	-1,070.0***
	(121.5)	(49.4)	(390.2)
	[-0.9%]	[-1.4%]	[-4.1%]
False positives	-268.6**	-417.4***	-1,508.0***
	(129.2)	(53.8)	(449.1)
	[-6.8%]	[-10.6%]	[-38.4%]
Placements	-39.3	-48.4**	-438.0**
	(46.5)	(22.9)	(202.6)
	[-0.8%]	[-0.9%]	[-8.5%]
Panel B. Misdemeanor Prosecution			
Weighted social cost	-589.6*	-1,010.4***	-2,297.3***
	(333.3)	(92.4)	(454.9)
False negatives	-68.8	-125.5***	-267.7***
	(57.7)	(16.3)	(76.8)
	[-5.1%]	[-9.4%]	[-20.0%]
False positives	-315.1**	-522.4***	-1,208.4***
	(144.0)	(44.6)	(219.9)
Prosecutions	-246.3**	-396.9***	-940.7***
	(115.5)	(41.0)	(207.9)
	[-1.0%]	[-1.6%]	[-3.8%]

Notes. This table reports gains from the optimized assignment mechanisms relative to the status-quo rotational assignment systems. Columns 1–3 report results for the finite-agent small-market mechanism, the corresponding large-market approximation, and a full-information benchmark in which the planner directly observes agents' task-specific preferences. Candidate mechanisms include all binary and three-group partitions of the four risk quartiles, as well as the full quartile mechanism. Estimates are evaluated using held-out test moments. Panel A reports results for the CPS application. The selected mechanism in Columns 1 and 2 is a binary partition which pools the lower two quartiles into low-type and the upper two quartiles into high-type. Percent changes relative to the status quo are reported in brackets. Because false positive outcomes are selectively observed, baseline false positive counts are not directly identified. Percent changes for false positives and weighted social costs therefore use extrapolation-based estimates of baseline false positive rates, described in Appendix B.8. Panel B reports results for the prosecution application. The selected mechanism is a three-group partition which splits cases into low-, medium-, and high-risk groups corresponding to quartiles $\{1\}$, $\{2,3\}$, and $\{4\}$. Because prosecution rates are high, extrapolation-based estimates of baseline false positive rates are not credible in this setting. Percent changes are therefore reported only for directly identified outcomes. False-positive changes are computed using the accounting identity $\Delta FP = \Delta P + \Delta FN$. Standard errors are computed using the bootstrap procedure described in Appendix B.7. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

increasing investigator workloads.[29]

Panel B reports the corresponding results for the misdemeanor prosecution application. The qualitative patterns are similar. Relative to the status-quo rotational assignment system, the optimized mechanisms reduce weighted social costs, false negatives, false positives, and overall prosecutions, with the large-market approximation again generating larger gains than the finite-market implementation.[30] As in the CPS setting, the small-market mechanism captures a substantial fraction of the gains achievable under full information. Relative to the full-information benchmark, the small-market mechanism achieves approximately one-quarter of the attainable reduction in weighted social costs.

Altogether, the two applications suggest that obviously strategy-proof unilateral-veto mechanisms can deliver meaningful efficiency gains while preserving the participation guarantees provided by the status quo.

References

D. Acemoglu and P. Restrepo. Modeling automation. In *AEA papers and proceedings*, volume 108, pages 48–53. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2018.

A. Agan, J. L. Doleac, and A. Harvey. Misdemeanor prosecution. *The Quarterly Journal of Economics*, 138(3):1453–1505, 2023.

N. Ahani, T. Andersson, A. Martinello, A. Teytelboym, and A. C. Trapp. Placement optimization in refugee resettlement. *Operations Research*, 69(5):1468–1486, 2021.

M. Armstrong. Multiproduct nonlinear pricing. *Econometrica: Journal of the Econometric Society*, pages 51–75, 1996.

D. Arnold, W. S. Dobbie, and P. Hull. Measuring racial discrimination in bail decisions. *American Economic Review*, 112(9):2992–3038, 2022.

[29]Because false positive outcomes are selectively observed, baseline false positive rates are not directly identified. Following Arnold et al. (2022) and Baron et al. (2026), we estimate baseline false positive rates using an extrapolation strategy based on investigators with very low placement rates. Intuitively, if some investigators almost never place children into foster care, the subsequent maltreatment rates among children left at home approximate the unselected baseline risk distribution. This approach is particularly plausible in the CPS setting because placement rates are low and many investigators have placement rates near zero. Appendix B.8 provides additional details.

[30]Unlike the CPS application, prosecution rates are sufficiently high that analogous extrapolation-based estimates of baseline false positive rates are not credible in this setting. We therefore report percent changes only for directly identified outcomes.

K. Bansak, J. Ferwerda, J. Hainmueller, A. Dillon, D. Hangartner, D. Lawrence, and J. Weinstein. Improving refugee integration through data-driven algorithmic assignment. *Science*, 359(6373):325–329, 2018.

E. J. Baron, J. J. Doyle Jr, N. Emanuel, P. Hull, and J. Ryan. Discrimination in multiphase systems: Evidence from child protection. *The Quarterly Journal of Economics*, 139(3): 1611–1664, 2024.

E. J. Baron, R. Lombardo, J. P. Ryan, J. Suh, and Q. Valenzuela-Stookey. Reforming rotations. Technical report, National Bureau of Economic Research, 2026.

A. Bergeron, P. Bessone, J. K. Kabeya, G. Z. Tourek, and J. L. Weigel. Supermodular bureaucrats: Evidence from randomly assigned tax collectors in the DRC. Technical report, Working paper, 2024.

S. Bikhchandani, S. Chatterji, R. Lavi, A. Mu'alem, N. Nisan, and A. Sen. Weak monotonicity characterizes deterministic dominant-strategy implementation. *Econometrica*, 74 (4):1109–1132, 2006.

Y. Breitmoser and S. Schweighofer-Kodritsch. Obviousness around the clock. *Experimental Economics*, 25(2):483–513, 2022.

E. Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6):1061–1103, 2011.

E. Budish. Matching "versus" mechanism design. *ACM SIGecom Exchanges*, 11(2):4–15, 2012.

E. Budish, Y.-K. Che, F. Kojima, and P. Milgrom. Designing random allocation mechanisms: Theory and applications. *American economic review*, 103(2):585–623, 2013.

Casey Family Programs. Turnover costs and retention strategies, 2023. URL `https://www.casey.org/turnover-costs-and-retention-strategies/`. Accessed: 2024-08-27.

D. C. Chan, M. Gentzkow, and C. Yu. Selection with Variation in Diagnostic Skill: Evidence from Radiologists. *The Quarterly Journal of Economics*, 137(2):729–783, 01 2022.

C. Daskalakis, A. Deckelbaum, and C. Tzamos. The complexity of optimal mechanism design. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 1302–1318. SIAM, 2014.

C. Daskalakis, A. Deckelbaum, and C. Tzamos. Strong duality for a multiple-good monopolist. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 449–450, 2015.

D. Delacrétaz, S. D. Kominers, and A. Teytelboym. Matching mechanisms for refugee resettlement. *American Economic Review*, 113(10):2689–2717, 2023.

P. Dütting, M. Feldman, T. Kesselheim, and B. Lucier. Posted prices, smoothness, and combinatorial prophet inequalities. *arXiv preprint arXiv:1612.03161*, 2016.

M. Feldman, N. Gravin, and B. Lucier. Combinatorial auctions via posted prices. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, pages 123–135. SIAM, 2014.

D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

L. Golowich and S. Li. On the computational properties of obviously strategy-proof mechanisms. *arXiv preprint arXiv:2101.05149*, 2021.

S. Hart and N. Nisan. Selling multiple correlated goods: Revenue maximization and menu-size complexity. *Journal of Economic Theory*, 183:991–1029, 2019.

A. Hylland and R. Zeckhauser. The efficient allocation of individuals to positions. *Journal of Political economy*, 87(2):293–314, 1979.

H. Kim, C. Wildeman, M. Jonson-Reid, and B. Drake. Lifetime prevalence of investigating child maltreatment among us children. *American Journal of Public Health*, 107(2):274–280, 2017.

P. Lahr and A. Niemeyer. Extreme points in multi-dimensional screening. arXiv:2412.00649, 2024.

S. Li. Obviously strategy-proof mechanisms. *American Economic Review*, 107(11): 3257–3287, 2017.

P. Mandal and S. Roy. Obviously strategy-proof implementation of assignment rules: A new characterization. *International Economic Review*, 63(1):261–290, 2022.

A. M. Manelli and D. R. Vincent. Bundling as an optimal selling mechanism for a multiple-good monopolist. *Journal of Economic Theory*, 127(1):1–35, 2006.

R. B. Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

I. Palacios-Huerta, D. C. Parkes, and R. Steinberg. Combinatorial auctions in practice. *Journal of Economic Literature*, 62(2):517–553, 2024.

G. Pavlov. Optimal mechanism for selling two goods. *The BE Journal of Theoretical Economics*, 11(1), 2011.

E. Putnam-Hornstein, J. Prindle, and I. Hammond. Engaging families in voluntary prevention services to reduce future child abuse and neglect: A randomized controlled trial. *Prevention Science*, 22(7):856–865, 2021.

M. Pycia and P. Troyan. Obviously strategyproof mechanisms in general environments. *Available at SSRN 5161185*, 2023a.

M. Pycia and P. Troyan. A theory of simplicity in games and mechanism design. *Econometrica*, 91(4):1495–1526, 2023b.

J.-C. Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of mathematical Economics*, 16(2):191–200, 1987.

J.-C. Rochet and P. Choné. Ironing, sweeping, and multidimensional screening. *Econometrica*, pages 783–826, 1998.

A. E. Roth. Repugnance as a constraint on markets. *Journal of Economic perspectives*, 21 (3):37–58, 2007.

L. Shapley and H. Scarf. On cores and indivisibility. *Journal of mathematical economics*, 1 (1):23–37, 1974.

T. Sönmez. Minimalist market design: A framework for economists with policy aspirations. *arXiv preprint arXiv:2401.00307*, 2023.

J. Thanassoulis. Haggling over substitutes. *Journal of Economic theory*, 117(2):217–245, 2004.

Q. Valenzuela-Stookey. Automation and task allocation under asymmetric information. Working paper, 2025.

H. R. Varian. Equity, envy, and efficiency. *Journal of Economic Theory*, 1973.

A Omitted proofs

A.1 Proof of Theorem 1

Assume that preferences are strictly monotone and we have a mechanism $(\mathcal{T}, s^*_{\mathcal{J}})$ that is OSP. Because agents' choices depend only on the essential reduction of $\mathcal{T}$, we simplify notation by assuming that $\mathcal{T} = \tilde{\mathcal{T}}$. We begin with some intermediate claims.

Claim 1. $\eta_j \in a_j$ for all $j \in \mathcal{J}$ and $a_j \in A_j$.

Proof of Claim 1. If every agent $j' \neq j$ chooses $\{\eta_{j'}\}$ then market clearing requires that j be assigned η_j. Thus it must be that $\eta_j \in a_j$ for all $a_j \in A_j$. □

Claim 2. For any $c_j \in \mathbb{R}^I_{>0}$ it must be that $c_j \cdot x \geq c_j \cdot \eta_j$ for all $a_j \in A_j \setminus \{s^*_j(c_j)\}$ and $x \in a_j$.

Proof of Claim 2. Suppose there exists $a_j \neq s^*_j(c_j)$ and $x \in a_j$ such that $c_j \cdot x < c_j \cdot \eta_j$. By Claim 1, $\eta_j \in s_j(c_j)$. But then η_j could be chosen following action $s_j(c_j)$ while x could be chosen for action a_j, so $s_j(c_j)$ does not obviously dominate a_j. □

Claim 3. If $a_j \in A_j$ and $x', x'' \in a_j$ then there cannot exist $c_j \in \mathbb{R}^I_{\geq 0}$ such that $c_j \cdot x' < c_j \cdot \eta_j$ and $c_j \cdot x'' > c_j \cdot \eta_j$.

Proof of Claim 3. If such a c_j exists, we can take it to be in $\mathbb{R}^I_{>0}$ because the two inequalities in the claim are strict. By Claim 2, if such a c_j exists then $s_j(c_j) = a_j$, because $c_j \cdot x' < c_j \cdot \eta_j$. But because $x'' \in a_j$ and $c_j \cdot x'' > c_j \cdot \eta_j$, choosing a_j does not obviously dominate choosing $\{\eta_j\}$. □

Claim 4. For each $j \in \mathcal{J}$ and $a_j \in A_j$, one of the following holds

1. $x \geq \eta_j$ for all $x \in a_j$.
2. $x \leq \eta_j$ for all $x \in a_j$.
3. a_j is contained in a non-monotone ray originating at η_j, i.e. $\tilde{a}_j \subset \{x \in \mathbb{R}^I : (x - \eta_j) = y\kappa,\ y \in \mathbb{R}_{\geq 0}\}$ for some $\kappa \in \mathbb{R}^I$ such that $\kappa \not\geq 0$ and $\kappa \not\leq 0$.

Proof of Claim 4. If $a_j = \{\eta_j, x\}$ for some $x \in \mathbb{R}^I$ then one of the three conditions is automatically satisfied. Suppose instead that there exist $x' \neq x''$ in a_j such that neither is equal to η_j. Consider the condition from Claim 3. By Farkas' lemma, the following are equivalent

1. there does not exist $c_j \in \mathbb{R}^I_{\geq 0}$ such that $c_j \cdot x' < c_j \cdot \eta_j$ and $c_j \cdot x'' > c_j \cdot \eta_j$.
2. There exist $y', y'' \in \mathbb{R}_{\geq 0}$, with at least one strictly positive, such that $y'(x' - \eta_j) \geq y''(x'' - \eta_j)$.

So Claim 3 implies that the second condition holds. Reversing the roles of x' and x'' we also have that there exist $\hat{y}', \hat{y}'' \in \mathbb{R}_{\geq 0}$, with at least one strictly positive, such that $\hat{y}'(x' - \eta_j) \leq \hat{y}''(x'' - \eta_j)$.

Suppose $y' > 0$ (a symmetric argument applies if $y' = 0, y'' > 0$). Then $x' - \eta_j \geq \frac{y''}{y'}(x'' - \eta_j)$, so $\hat{y}''(x'' - \eta_j) \geq \hat{y}'\frac{y''}{y'}(x'' - \eta_j)$. This implies that one of the following holds

1. $x'' - \eta_j \geq 0$ and $\hat{y}'' \geq \hat{y}'\frac{y''}{y'}$
2. $x'' - \eta_j \leq 0$ and $\hat{y}'' \leq \hat{y}'\frac{y''}{y'}$.
3. $x'' - \eta_j \not\geq 0$, $x'' - \eta_j \not\leq 0$ and $\hat{y}'' = \hat{y}'\frac{y''}{y'}$.

Suppose $x'' - \eta_j \geq 0$. By assumption, $x'' \neq \eta_j$, so we must be in the first case, with $\hat{y}'' \geq \hat{y}' \frac{y''}{y'}$, so $x' - \eta_j \geq 0$ as well. Because x' and x'' are arbitrary selections from a_j, we have that $x - \eta_j \geq 0$ for all $x \in a_j$. Similarly in the second case, with $x'' - \eta_j \leq 0$ and $\hat{y}'' \leq \hat{y}' \frac{y''}{y'}$, we have that $x - \eta_j \leq 0$ for all $x \in a_j$.

Consider the third case, where $x'' - \eta_j \not\geq 0$, $x'' - \eta_j \not\leq 0$ and $\hat{y}'' = \hat{y}' \frac{y''}{y'}$. Note that if this holds then $y'', \hat{y}', \hat{y}''$ must be strictly positive (and we are still assuming $y' > 0$), so $\frac{\hat{y}''}{\hat{y}'} = \frac{y''}{y'}$. Then we have

$$\frac{y''}{y'}(x'' - \eta_j) \geq (x' - \eta_j) \quad \text{and} \quad \frac{y''}{y'}(x'' - \eta_j) \leq (x' - \eta_j)$$

so $\frac{y''}{y'}(x'' - \eta_j) = (x' - \eta_j)$. In other words, $(x'' - \eta)$ and $(x' - \eta)$ lie in the same ray. Then a_j is contained in a ray originating at η_j. □

We call a set $a_j \in A_j$ *remedial* if $x \leq \eta_j$ for all $x \in a_j$, and *redundant* if $x \geq \eta_j$ for all $x \in a_j$.

Claim 5. If a_j' and a_j'' are not remedial or redundant then they are polarized.

Proof of Claim 5. By Claim 4, both a_j' and a_j'' are contained in rays from η_j. Thus it suffices to show that there exists $x' \in a_j'$, $x'' \in a_j''$, and $y', y'' \geq 0$, with at least one strictly positive, such that $y'(x' - \eta_j) + y''(x'' - \eta_j) \in \mathbb{R}^I_{\geq 0}$.

Let $x' \in a_j'$, $x'' \in a_j''$ be any selections such that $x' \neq \eta_j$, $x'' \neq \eta_j$. Observe that there cannot exist $c_j \geq 0$ such that $c_j \cdot x' \leq c_j \cdot \eta_j$ and $c_j \cdot x'' < c_j \cdot \eta_j$. If this were the case then, because $x' - \eta_j \not\geq 0$, we could find $c_j' \in \mathbb{R}^I_{>0}$ such that $c_j' \cdot x' < c_j' \cdot \eta_j$ and $c_j' \cdot x'' < c_j' \cdot \eta_j$, violating Claim 2. By Farkas' lemma, the following are equivalent

i. There does not exist $c_j \geq 0$ such that $c_j \cdot x' \leq c_j \cdot \eta_j$ and $c_j \cdot x'' < c_j \cdot \eta_j$.

ii. There exist $y' \geq 0, y'' > 0$, such that $y'(x' - \eta_j) + y''(x'' - \eta_j) \geq 0$.

The second condition is precisely what we wanted to show. □

Claim 6. If $a_j \in A_j$ is remedial and $a_j \neq \{\eta_j\}$ then a_j' is remedial or redundant for all $a_j' \in A_j$.

Proof of Claim 6. If a_j' is not remedial or redundant then it is contained in a non-monotone ray from η_j, by Claim 4. Then there exists $c_j \in \mathbb{R}^I_{>0}$ and $x' \in a_j'$ such that $c_j \cdot x' < c_j \cdot \eta_j$.

Then for any $x \in a_j$, $x \neq \eta_j$ we can also choose c_j so that $c_j \cdot x < c_j \cdot \eta_j$. But this violates Claim 2. □

Say that agent j's mechanism is *remedial* if a_j is remedial for all $a_j \in A_j \setminus \{\eta_j\}$.

We have concluded that for any $j \in \mathcal{J}$, the collection of veto sets for j takes one of the following two forms.

1. Every $a_j \in A_j$ is either remedial or redundant.

2. A_j is the union of a collection of polarized rays with a collection of redundant veto sets.

Moreover, if preferences are strictly monotone then no redundant set can ever be chosen (doing so is obviously dominated by choosing $\{\eta_j\}$), and so we obtain an equivalent mechanism by removing the redundant sets. Then the mechanism is dominated by a polarized-ray mechanism.

To conclude the proof of Theorem 1, it remains to show that any polarized-ray mechanism is OSP. Any j with a remedial mechanism has a trivial dominant strategy. Suppose j's mechanism consists of polarized rays. Then (as shown using Farkas' lemma in the proof of Claim 5) if $c_j \cdot x' < c_j \cdot \eta_j$ for some $x' \in a'_j$ then $c_j \cdot x'' \geq c_j \cdot \eta_j$ for all $x'' \in a''_j$, with $a''_j \neq a'_j$. Thus choosing a'_j is obviously dominant.

A.2 Proof of Theorem 3

To understand Theorem 3, it is useful to separate it into constituent parts. First, consider the gap between "choice improves upon the status quo" and "choice is optimal". We relate this gap to the degrees of slack in the SQ constraints, parameterized by β_j, and the range of preferences, $[\underline{c}^i_j, \bar{c}^i_j]$. Define $\hat{c}^i := \max_{j \in \mathcal{J}} \underline{c}^i_j$, so $\hat{c}^i \in [\underline{c}^i_j, \bar{c}^i_j]$ for all i, j if and only if $\cap_{j \in \mathcal{J}} C_j \neq \varnothing$.

Lemma 6. Assume $\cap_{j \in \mathcal{J}} C_j \neq \varnothing$. If an allocation $m \in \mathcal{Z}$ (i.e. a non-choice-based mechanism) satisfies (SQ) and (MC) then

$$\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \left(\bar{c}^i_j - \hat{c}^i\right) \left(m_j(i) - \beta_j \sigma_j(i)\right)^+ \leq \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \hat{c}^i \left(\beta_j - 1\right) \sigma_j(i).$$

Proof. Assume that $m \in \mathcal{Z}$ satisfies (SQ), i.e.

$$\sum_{i\in\mathcal{I}} m_j(i)c_j(i) \leq \beta_j \sum_{i\in\mathcal{I}} \sigma_j(i)c_j(i) \quad \forall\ j \in \mathcal{J},\ \forall\ c_j \in C_j. \tag{SQ}$$

Then for any $\mathbf{c} \in \mathbf{C}$ we must have

$$\begin{aligned} 0 &\geq \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \left(m_j(i) - \beta_j\sigma_j(i)\right) c_j(i) \\ &= \sum_{i\in\mathcal{I}} \hat{c}^i \sum_{j\in\mathcal{J}} \left(m_j(i) - \beta_j\sigma_j(i)\right) + \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \left(m_j(i) - \beta_j\sigma_j(i)\right) \left(c_j(i) - \hat{c}^i\right) \end{aligned} \tag{8}$$

where the inequality in the first line follows by summing the (SQ) constraint over j, and the second line is simple algebra. If m satisfies (MC) then the first term in eq. (8) is equal to

$$\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \hat{c}^i \left(1 - \beta_j\right) \sigma_j(i)$$

Then for eq. (8) to hold, it must be that

$$\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \hat{c}^i \left(\beta_j - 1\right) \sigma_j(i) \geq \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \left(m_j(i) - \beta_j\sigma_j(i)\right) \left(c_j(i) - \hat{c}^i\right)$$

for all $\mathbf{c} \in \mathbf{C}$. Under the assumption that $\cap_{j\in\mathcal{J}} C_j \neq \varnothing$ we have $\bar{c}^i_j \geq \hat{c}^i$ for all $i \in \mathcal{I}, j \in \mathcal{J}$. Then setting $c_j(i) = \hat{c}^i$ if $\left(m_j(i) - \beta_j\sigma_j(i)\right) < 0$ and $c_j(i) = \bar{c}^i_j$ otherwise, we have

$$\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \hat{c}^i \left(\beta_j - 1\right) \sigma_j(i) \geq \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \left(\bar{c}^i_j - \hat{c}^i\right) \left(m_j(i) - \beta_j\sigma_j(i)\right)^+$$

as desired. □

Corollary 2. If $\beta_j = 1$ for all $j \in \mathcal{J}$ and $\cap_{j\in\mathcal{J}} C_j$ has dimension I then the only non-choice-based mechanism satisfying (SQ) and (MC) is the status quo. In particular, choice is optimal if and only if choice improves upon the status quo.

Proof. From Lemma 6 for $\beta_j = 1$, (SQ) and (MC) imply that for any non-choice-based mechanism $m \in \mathcal{Z}$,

$$\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \left(\bar{c}^i_j - \hat{c}^i\right) \left(m_j(i) - \sigma_j(i)\right)^+ \leq 0.$$

If m satisfies (MC) and $m \neq \sigma$ then $m_j(i) - \sigma_j(i) > 0$ for some i, j. If $\cap_{j\in\mathcal{J}} C_j$ has dimension I then $\bar{c}^i_j > \hat{c}^i$ for all i, j, so the above inequality is violated. Thus the only deterministic

mechanism satisfying (SQ) and (MC) is the status quo. □

Now consider the equivalence between *i.*, *ii.*, and *iii.* in Theorem 3. Underlying this equivalence is the simple class of bilateral trading mechanisms, introduced above, which can be used to improve upon the status quo. Let j', j'' be the agents involved in the mechanism, let $\gamma \in \mathbb{R}^I$ be the trade, and assume that agent j' (agent j'') follows the strategy of trading if and only if $\gamma \cdot c_{j'} \leq 0$ ($\gamma \cdot c_{j''} \geq 0$). We begin with the following simple observation.

Lemma 7. The bilateral trading mechanism for agents j', j'' with trade γ is feasible and improves upon the status quo if and only if

i. $\sigma_{j'} + \gamma \geq 0$ and $\sigma_{j''} - \gamma \geq 0$,

ii. $(\pi_{j'} - \pi_{j''}) \cdot \gamma < 0$,

iii. $\min \{\gamma \cdot c_{j'} : c_{j'} \in C_{j'}\} < 0$,

iv. $\max \{\gamma \cdot c_{j''} : c_{j''} \in C_{j''}\} > 0$.

Proof. Trade is feasible by condition *i*. Agent j' (resp. j'') is willing to trade if and only if $\gamma \cdot c'_j \leq 0$ (resp. $\gamma \cdot c_{j''} \geq 0$). Thus conditions *iii*. and *iv*. ensure that trade occurs with positive probability (given the maintained assumption that the distribution has full support). Condition *ii*. implies that when trade does occur, it reduces the social cost. □

The real question is whether we can find a bilateral trade that satisfies these conditions. The following sufficient conditions guarantee that such a mechanism exists, and immediately imply the conclusions of Theorem 3.

Proposition 7. If there exist j', j'' such that

i. $\pi_{j'} \neq \pi_{j''}$;

ii. $C_{j'} \cap C_{j''}$ is non-empty and non-degenerate; and

iii. $\sigma_{j'}(i) > 0$ and $\sigma_{j''}(i) > 0$ for all $i \in \mathcal{I}$;

then there is a bilateral trading mechanism that improves upon the status quo.

Proof. We begin with an intermediate result.

Lemma 8. If $\pi'_j \neq \pi_{j''}$; $\sigma_{j'}(i), \sigma_{j''}(i) > 0$ for all $i \in \mathcal{I}$; and there exist points $c', c'' \in C_{j'} \cap C_{j''}$ such that $\text{cvh}(c'', 0) \cap \text{cvh}(\pi_{j'} - \pi_{j''}\ ,\ c'\ ,\ 0) = \{0\}$; then there is a bilateral trading mechanism satisfying the conditions of Lemma 7.

Proof. First, because $\sigma_{j'}(i), \sigma_{j''}(i) > 0$ for all i, if we can find γ satisfying Lemma 7 $ii. - iv.$ then we can scale it to satisfy Lemma 7 i.

Assume without loss of generality that there exists $i \in \mathcal{I}$ such that $\pi^i_{j'} - \pi^i_{j''} > 0$ (otherwise simply reverse the roles of j' and j''). Then 0 is an extreme point of $\text{cvh}(\pi_{j'} - \pi_{j''}, c', 0)$. Thus $\text{cvh}(\pi_{j'} - \pi_{j''}, c', 0) \setminus \{0\}$ and $\text{cvh}(c'', 0) \setminus \{0\}$ are disjoint, non-empty, and convex. By the separating hyperplane theorem there exists $\gamma \in \mathbb{R}^I$ such that $(\pi_{j'} - \pi_{j''}) \cdot \gamma < 0$, $\gamma \cdot c' < 0$, and $\gamma \cdot c'' > 0$. Because $\min\{\gamma \cdot c_{j'} : c_{j'} \in C_{j'}\} \leq \gamma \cdot c'$ and $\max\{\gamma \cdot c_{j''} : c_{j''} \in C_{j''}\} \geq \gamma \cdot c''$, conditions $ii. - iv.$ of Lemma 7 are satisfied. □

To prove Proposition 7, it remains to show that if $C_{j'} \cap C_{j''}$ is non-empty and non-degenerate then there exist points $c', c'' \in C_{j'} \cap C_{j''}$ satisfying the conditions of Lemma 8.

Let $\delta = \pi_{j'} - \pi_{j''}$. Because $C_{j'} \cap C_{j''}$ is a non-degenerate hypercube, there are $c^1, c^2 \in C_{j'} \cap C_{j''}$ such that c^1, c^2, δ are pairwise linearly independent. It suffices to show that either $\text{cvh}(c^1, 0) \cap \text{cvh}(\delta, c^2, 0) = \{0\}$ and/or $\text{cvh}(c^2, 0) \cap \text{cvh}(\delta, c^1, 0) = \{0\}$.

To see this, note that $\text{cvh}(c^1, 0) \cap \text{cvh}(\delta, c^2, 0) \neq \{0\}$ if and only if c^1 is in the convex cone generated by (δ, c^2). That is, there exist $\alpha, \rho \geq 0$ such that $c^1 = \alpha c^2 + \rho\delta$. Moreover, $\alpha, \rho > 0$ because c^1, c^2, δ are pairwise linearly independent. Symmetrically $\text{cvh}(c^2, 0) \cap \text{cvh}(\delta, c^1, 0) \neq \{0\}$ if and only if there exist $\hat{\alpha}, \hat{\rho} > 0$ such that $c^2 = \hat{\alpha} c^1 + \hat{\rho}\delta$. Suppose, towards a contradiction, that both are true. Then

$$\begin{aligned} c^1 &= \alpha(\hat{\alpha} c^1 + \hat{\rho}\delta) + \rho\delta \\ &= \alpha\hat{\alpha} c^1 + (\alpha\hat{\rho} + \rho)\delta \end{aligned}$$

which contradicts the linear independence of c^1 and δ. □

Proposition 7 shows that bilateral trading mechanisms work to improve upon the status quo under mild conditions. The most restrictive condition of Proposition 7 is that $\sigma_{j'}(i) > 0$

and $\sigma_{j''}(i) > 0$ for all $i \in \mathcal{I}$. However even this can be relaxed, at the cost of additional notation, by restricting attention only to the dimensions such that $\sigma_{j'}(i) > 0$ and $\sigma_{j''}(i) > 0$.

A.3 Proof of Lemma 3

Proof. Suppose $R \in \bar{P}$. Let (R_n) be a sequence in P converging to R. That R satisfies condition *i.* is immediate from the fact that each R_n satisfies condition *i.* in the definition of P. From condition *ii.* in the definition of P, we can choose $\alpha_n \in (0,1)^{E(P)\times E(P)}$ such that $\alpha_n(\zeta,\zeta')R_n(\zeta) + (1-\alpha_n(\zeta,\zeta'))R_n(\zeta') \geq 0$ for all $\zeta \neq \zeta' \in E(P)$. Then the sequence (α_n) contains a subsequence that converges to $\hat{\alpha} \in [0,1]^{E(P)\times E(P)}$. Because the set of $\alpha \in [0,1]$, $\kappa \in \mathcal{O}_\zeta$ and $\kappa' \in \mathcal{O}_{\zeta'}$ satisfying $\alpha\kappa + (1-\alpha)\kappa' \geq 0$ is closed, $\hat{\alpha}$ must satisfy condition *ii.* of the Lemma.

Conversely, suppose that R satisfies conditions *i.* and *ii.* of the Lemma. We want to show that for any $\varepsilon > 0$ there exists $R' \in P$ such that $|R' - R| < \varepsilon$. Consider the following perturbation of R: For each $\zeta \in E(P)$ and $\delta^-, \delta^+ > 0$ let $R'(\zeta)_j = R(\zeta)_j + \delta^+$ if $\zeta_j = 1$ and $R'(\zeta)_j = R(\zeta)_j - \delta^-$ if $\zeta_j = -1$. Then $R'(\zeta)$ is in the interior of $\mathcal{O}_\zeta$ for all $\zeta \in E(P)$. It remains to show that we can choose δ^-, δ^+ so that R' satisfies condition *ii.* in the definition of P. We know that for any $\zeta \neq \zeta' \in E(P)$ there exists $\alpha(\zeta,\zeta') \in [0,1]$ such that $\alpha(\zeta,\zeta')R(\zeta) + (1-\alpha(\zeta,\zeta'))R(\zeta') \geq 0$. If $\alpha(\zeta,\zeta') \in (0,1)$ then $\alpha(\zeta,\zeta')R'(\zeta) + (1-\alpha(\zeta,\zeta'))R'(\zeta') \geq 0$ if

$$\frac{\delta^-}{\delta^+} \leq \min\left\{\frac{\alpha}{(1-\alpha)}, \frac{1-\alpha}{\alpha}\right\}.$$

So for (ζ,ζ'), condition *ii.* in the definition of P is satisfied if δ^-/δ^+ is small enough. If instead $\alpha(\zeta,\zeta') = 1$ then it must be $R(\zeta) \geq 0$. Then for $\delta^-, \delta^+ > 0$ there exists $\underline{\alpha}(\delta^-,\delta^+) < 1$ such that $\alpha R'(\zeta)_j + (1-\alpha)R'(\zeta')_j \geq 0$ for all $\alpha \geq \underline{\alpha}(\delta^-,\delta^+)$ and all j such that $\zeta_j = 1$. Note that we can let $\underline{\alpha}(\delta^-,\delta^+)$ be non-decreasing in δ^- and non-increasing in δ^+. Moreover, for any j such that $\zeta_j = -1$, we have $\underline{\alpha}(\delta^-,\delta^+)R'(\zeta)_j + (1-\underline{\alpha}(\delta^-,\delta^+))R'(\zeta')_j \geq 0$ if

$$\frac{\delta^-}{\delta^+} \leq \frac{1-\underline{\alpha}(\delta^-,\delta^+)}{\underline{\alpha}(\delta^-,\delta^+)}$$

which, because $\delta^- \mapsto \underline{\alpha}(\delta^-,\delta^+)$ is non-decreasing, we can satisfy by choosing δ^- small enough. A symmetric argument applies to the remaining case where $\alpha(\zeta,\zeta') = 0$. Because there are finitely many pairs (ζ,ζ'), for any $\delta^+ > 0$ we can choose $\delta^- > 0$ sufficiently small

such that R' satisfies condition *ii.* in the definition of P (for all pairs). □

A.4 Proof of Proposition 4

Proof. By definition, $\kappa \neq 0$ for all $\kappa \in R$, so we just need to show that there do not exist dimensions $i \neq i'$ such that $\kappa(i) \neq 0$ and $\kappa(i') \neq 0$. Suppose towards a contradiction that this holds for some $\kappa^1 \in R$. Define $B_1 = \{\kappa^1\}$.

Notice that because $\kappa^1(i') \neq 0$, increasing or decreasing $\kappa^1(i)$ while fixing $\kappa^1(i')$ does not amount to a rescaling of κ^1, and so any such perturbations generate a new menu that is not equivalent to R. Thus for R to be an extreme point, there cannot exist $\varepsilon_1 > 0$ such that replacing κ^1 with either one of $\kappa^1 + \varepsilon_1 \mathbb{1}_i$ and $\kappa^1 - \varepsilon_1 \mathbb{1}_i$ yields two more polarized-ray protocols. Let α be weights that satisfy the condition for polarization of R. Because $\kappa^1(i) \neq 0$, we could find such an ε_1 unless polarization binds, i.e. unless there exists some $\kappa' \neq \kappa^1$ such that

$$\kappa^1(i)\alpha(\kappa^1, \kappa') + \kappa'(i)(1 - \alpha(\kappa^1, \kappa')) = 0 \tag{9}$$

and $\alpha(\kappa^1, \kappa') > 0$ (because otherwise reducing $\kappa^1(i)$ would not affect eq. (9)). Let $B_2 \subset R \setminus \{\kappa^1\}$ be the set of such κ'. Notice that for any $\kappa' \in B_2$ we have $\kappa'(i) \neq 0$ and $\text{sign}(\kappa'(i)) = -\text{sign}(\kappa^1(i))$ (because otherwise the equality in eq. (9) could not hold). Abusing notation, write $\text{sign}_1 = \text{sign}(\kappa^1(i))$ and $\text{sign}_2 = \text{sign}(\kappa'(i))$ such that $\kappa' \in B_2$.

There exist $\varepsilon_1 > 0$ and $\varepsilon(\kappa') > 0$ for $\kappa' \in B_2$, such that without violating eq. (9) we could replace κ^1 with $\kappa^1 - \varepsilon_1 \mathbb{1}_i$ (resp. $\kappa^1 + \varepsilon_1 \mathbb{1}_i$) and κ' with $\kappa' + \varepsilon(\kappa')\mathbb{1}_i$ (resp. $\kappa' - \varepsilon(\kappa')\mathbb{1}_i$) for all $\kappa' \in B_2$. Because $\kappa^1(i) \neq 0$ and $\kappa'(i) \neq 0$, these changes would only violate the polarization conditions if there exists some $\kappa^2 \in B_2$ and $\kappa' \neq \kappa^2$ such that

$$\kappa^2(i)\alpha(\kappa^2, \kappa') + \kappa'(i)(1 - \alpha(\kappa^2, \kappa')) = 0 \tag{10}$$

and $\alpha(\kappa^2, \kappa') > 0$. Let $B_3(\kappa^2) \subset \mathbb{R} \setminus \{\kappa^2\}$ be the set of such κ', and let $B_3 = \cup_{\kappa^2 \in B_2} B_3(\kappa^2)$. Notice that for any $\kappa' \in B_3$ we have $\kappa'(i) \neq 0$ with $-\text{sign}_2 = \text{sign}(\kappa'(i)) := \text{sign}_3$ (because otherwise the equality in eq. (10) could not hold). Therefore $B_3 \cap B_2 = \varnothing$. Suppose $B_3 = B_1 := \{\kappa^1\}$. Then there exists $\varepsilon_1 > 0$ such that without violating any polarization constraints we could replace κ' with $\kappa' - \varepsilon_1 \mathbb{1}_i$ (resp. $\kappa' + \varepsilon_1 \mathbb{1}_i$) and κ'' with $\kappa'' + \varepsilon(\kappa'')\mathbb{1}_i$ (resp. $\kappa'' - \varepsilon(\kappa'')\mathbb{1}_i$) for all $\kappa' \in B_1 = B_3 = \{\kappa^1\}$ and $\kappa'' \in B_2$. But then R would not be an

extremal generator of $\bar{P}$. Therefore it must be that $B_3 \not\subset B_1$.

Suppose that we have constructed sets $\{\kappa^1\} = B_1, \ldots B_N$ in this way, so that

1. if $\kappa' \in B_n$ and κ'' is such that $\alpha(\kappa', \kappa'') > 0$ and

$$\kappa'(i)\alpha(\kappa', \kappa'') + \kappa''(i)(1 - \alpha(\kappa', \kappa'')) = 0$$

   then $\kappa'' \in B_{n+1}$, and

2. $\text{sign}_n = -\text{sign}_{n+1}$

If $B_N \subset \cup_{n=1}^{N-1} B_n$ then there exists $\varepsilon : \cup_{n=1}^{N-1} B_n \to \mathbb{R}_+$ such that without violating any polarization constraints we could replace κ' with $\kappa' - \varepsilon(\kappa')\mathbb{1}_i$ (resp. $\kappa' + \varepsilon(\kappa')\mathbb{1}_i$) and κ'' with $\kappa'' + \varepsilon(\kappa'')\mathbb{1}_i$ (resp. $\kappa'' - \varepsilon(\kappa'')\mathbb{1}_i$) for all κ' in $\cup_{n \text{ is even}} B_n$ and $\kappa'' \in \cup_{n \text{ is odd}} B_n$. Then R would not be an extremal generator. Therefore it must be that $B_N \not\subset \cup_{n=1}^{N-1} B_n$. But because R can contain at most $I < \infty$ rays, we cannot continue this construction indefinitely. Therefore R cannot be an extremal generator. □

A.5 Proof of Lemma 5

Proof. With I task types, we can identify each class P such that $|E(P)| = h$ with a set of sign vectors $\{\zeta^1, \ldots, \zeta^h\}$ such that for any coordinate $i \in \{1, \ldots, I\}$ there is at most one $k \in \{1, \ldots, h\}$ such that $\zeta_i^k = -1$. Moreover, each ζ^k must have at least one negative entry, and no more than $I - 1$ negative entries (by definition of polarized rays).

Notice that sign vector ζ^1 can have no more than $I - (h - 1)$ negative entries (because otherwise there would be no way to give distinct negative entries to $\{\zeta^2, \ldots, \zeta^h\}$). For any $1 \leq j \leq I - (h - 1)$ there are $\binom{I}{j}$ possible specifications for ζ^1 such that ζ^1 has j negative entries. Fixing ζ^1, every ζ^k for $k > 1$ must have positive entries where ζ^1 has negative entries. Thus the problem of choosing the remaining $\{\zeta^2, \ldots, \zeta^h\}$ is equivalent to that of choosing $h - 1$ sign vectors when there are only $I - j$ task types (as long as $h - 1 > 1$, the base case of $h - 1 = 1$ is discussed below). Hence there are $\binom{I}{j} c(I - j, h - 1)$ possible ways to choose ζ^1 and $\{\zeta^2, \ldots, \zeta^h\}$ such that ζ^1 has j negative entries. However, the ordering of the sign vectors is arbitrary; a class P is determined by the set of sign vectors. As there are h possible ways to determine which sign vector is labeled "1", the $\binom{I}{j} c(I - j, h - 1)$ possible

choices of ζ^1 and $\{\zeta^2, \ldots, \zeta^h\}$ correspond to $\frac{1}{h}\binom{I}{j} c(I-j, h-1)$ distinct classes. Summing across the possible values of j yields the expression in eq. (3).

Notice that we set $c(N,1) = \sum_{j=1}^{N} \binom{N}{j}$, although the number of classes such that $|E(P)| = 1$ with N task types is $\sum_{j=1}^{N-1} \binom{N}{j} = 2^N - 2$. The difference is that in the specification of $c(N,1)$ we allow for sign vectors in which all N entries are negative. This is because once the recursion arrives at $c(N',1)$ we have $N' < I$, which means that each sign vector already has some positive entries, and we are free to assign N' negative entries to the last one. This is also why we treat the case of $h = 1$ separately in writing $C(I) = 2^I - 2 + \sum_{h=2}^{I} c(I,h)$ $\square$

A.6 Proof of Proposition 5

Proof. We first verify that the program in eq. (7) is indeed the dual of eq. (6). Letting λ^i be the multipliers on the LM-MC constraints, we have

$$\begin{aligned}
&\min_{(x_j)_{j\in\mathcal{J}}} \sup_{\lambda\in\mathbb{R}^I} \sum_{j\in\mathcal{J}} \pi_j \cdot x_j - \sum_{i\in\mathcal{I}} \lambda^i \left(\sum_{j\in\mathcal{J}} x_j(i) - n^i \right) \quad s.t. \quad x_j \in \mathcal{F}_j \quad \forall\, j \in \mathcal{J} \\
&\quad \geq \sup_{\lambda\in\mathbb{R}^I} \min_{(x_j)_{j\in\mathcal{J}}} \sum_{j\in\mathcal{J}} \pi_j \cdot x_j - \sum_{i\in\mathcal{I}} \lambda^i \left(\sum_{j\in\mathcal{J}} x_j(i) - n^i \right) \quad s.t. \quad x_j \in \mathcal{F}_j \quad \forall\, j \in \mathcal{J} \\
&\quad = \sup_{\lambda\in\mathbb{R}^I} \sum_{i\in\mathcal{I}} \lambda^i n^i - \sum_{j\in\mathcal{J}} \max_{x_j\in\mathcal{F}_j} \sum_{i\in\mathcal{I}} \left(\lambda^i - \pi_j(i)\right) x_j^i \\
&\quad = \sup_{\lambda\in\mathbb{R}^I} \sum_{i\in\mathcal{I}} \lambda^i n^i - \sum_{j\in\mathcal{J}} S_j(\lambda - \pi_j)
\end{aligned}$$

where inequality in the second line follows from weak duality, and the last line from the definition of the support function.

Because giving the single veto set $\{\eta_j\}$ to agent j is feasible, we have $\eta_j \in \mathcal{F}_j$. If η_j is in the interior of $\mathcal{F}_j$ for all j then Slater's condition is satisfied, so we are done. Otherwise, define a perturbed problem by letting

$$\mathcal{F}_j^\varepsilon = \{x \in \mathbb{R}^I_{\geq 0} \ : \exists\ y \in \mathcal{F}_j \text{ with } |x - y| \leq \varepsilon\}.$$

for $\varepsilon \geq 0$, and let S_j^ε be the support function of this set. Then η_j is in the interior of $\mathcal{F}_j^\varepsilon$ for all j, so Slater's condition is satisfied for any $\varepsilon > 0$ if we replace $\mathcal{F}_j$ with $\mathcal{F}_j^\varepsilon$. Let P_ε and D_ε be the values of the primal and dual programs at ε. Notice that $\mathcal{F}_j^\varepsilon$ is compact and convex

valued and $\varepsilon \mapsto \mathcal{F}_j^\varepsilon$ is upper- and lower-hemicontinuous on $\mathbb{R}_{\geq 0}$. Thus $\varepsilon \mapsto P_\varepsilon$ is continuous by Berge's maximum theorem, and converges to P_0 as $\varepsilon \to 0$.

Now note that $\varepsilon \mapsto S_j^\varepsilon(w)$ is increasing, so $\varepsilon \mapsto D_\varepsilon$ is decreasing. Moreover, weak duality (for a minimization problem) implies that $P_0 \geq D_0$. Then $P_\varepsilon = D_\varepsilon \leq D_0 \leq P_0$ for all $\varepsilon > 0$, and $P_\varepsilon \to P_0$ as $\varepsilon \to 0$, which implies $D_0 = P_0$. □

A.7 Proof of Proposition 6

Proof. To simplify notation, assume that the large-market mechanism does not involve randomization. It is straightforward to extend the argument to the case with randomization by sub-dividing the groups.

As the mechanism is OSP, the strategy of each agent is independent of N. Let $d_j^a(N)$ be the random variable corresponding to the set of group-j agents who choose veto set $a \in A_j$ in the N-replica economy. The mechanism allows the designer to choose different allocations for agents in $d_j^a(N)$. However, towards establishing asymptotic convergence we can assume that (perhaps sub-optimally) the designer does not do so. Define the random variable

$$\tilde{V}_{LM}^*(N) = \min_{(\mu_j^a)} \sum_{j \in \mathcal{J}} \sum_{a \in A_j} |d_j^a(N)| \pi_j \cdot \mu_j^a \quad s.t. \quad \mu_j^a \in a \ \forall \ j \in \mathcal{J}, \ a \in A_j,$$
$$\frac{1}{N} \sum_{j \in \mathcal{J}} \sum_{a \in A_j} \mu_j^a |d_j^a(N)| = (n_1, \ldots, n_I)$$

Let $V_{LM}^*(\infty)$ be the value obtained in the large-market program, which is a lower bound on $\frac{1}{N}\mathbb{E}[V^*(N)]$ for any N. Then to prove the result it suffices to show that $\frac{1}{N}\tilde{V}_{LM}^*(N) \to V_{LM}^*(\infty)$ almost surely as $N \to \infty$. Define the random variable $\rho_j^a(N) := \frac{|d_j^a(N)|}{N}$ (note that $\sum_{a \in A_j} \rho_j^a = 1$ for all j). Then we can write

$$\frac{1}{N}\tilde{V}_{LM}^*(N) = \min_{(\mu_j^a)} \sum_{j \in \mathcal{J}} \sum_{a \in A_j} \rho_j^a(N) \pi_j \cdot \mu_j^a \tag{11}$$

$$s.t. \quad \mu_j^a \in a \ \forall \ j \in \mathcal{J}, \ a \in A_j, \tag{12}$$

$$\sum_{j \in \mathcal{J}} \sum_{a \in A_j} \mu_j^a \rho_j^a(N) = (n_1, \ldots, n_I) \tag{13}$$

Abusing notation, let $\tilde{V}_{LM}^*(\rho)$ be the value of this program as a function of $\rho = (\rho_j^a)_{j \in \mathcal{J}, a \in A_j}$.

If $\rho_j^a = \mathbb{E}_j[\mathbb{1}\{s_j(c_j) = a\}]$ then $\tilde{V}_{LM}^*(\rho) = V_{LM}^*(\infty)$. By the strong law of large numbers $\rho_j^a(N) \xrightarrow{a.s.} \mathbb{E}_j[\mathbb{1}\{s_j(c_j) = a\}]$ as $N \to \infty$ for all j, a. Thus we just need to show that $\tilde{V}_{LM}^*(\rho)$ is upper-semicontinuous at $\rho_j^a = \mathbb{E}_j[\mathbb{1}\{s_j(c_j) = a\}]$.

Let $\kappa_j(a)$ be a vector defining the ray a, chosen so that the non-negativity constraint binds at $\eta_j + \kappa_j(a)$. Then any μ_j^a satisfying constraint (12) can be written as $\mu_j^a = \eta_j + \alpha_j^a \kappa_j(a)$ for some scaling factors $\alpha_j^a \in [0, 1]$. Let $(\hat{\alpha}_j^a)$ be the scaling factors that define the solution to the large-market program. Without loss of generality, we can assume that $\hat{\alpha}_j^a > 0$, since otherwise a can just be replaced by η_j. Note also that the allocation defined by $(\delta\hat{\alpha}_j^a)$ also satisfies market clearing for any $\delta \in [0, 1]$, because the resulting mechanism is just a convex combination of the large-market solution and the endowment η, and thus $(\delta\hat{\alpha}_j^a)$ is also a valid mechanism in the large-market program. Then for any $\varepsilon \geq 0$ we can choose $\delta < 1$ so that the value of $(\delta\hat{\alpha}_j^a)$ in the large-market program, denoted by V_δ^*, is no more than $V_{LM}^*(\infty) + \varepsilon$. Then to show that $\tilde{V}_{LM}^*(\rho)$ is upper-semicontinuous at $\rho_j^a = \mathbb{E}_j[\mathbb{1}\{s_j(c_j) = a\}]$ it suffices to show that for any $\varepsilon > 0$ and $\delta \in (0, 1)$ there exists $\sigma > 0$ such that $|\mathbb{E}_j[\mathbb{1}\{s_j(c_j) = a\}] - \rho_j^a| < \sigma$ implies $\tilde{V}_{LM}^*(\rho) \leq V_\delta^* + \varepsilon/2$.

As the objective is continuous in ρ, we just need to show that by taking a ρ_j^a sufficiently close to $\mathbb{E}_j[\mathbb{1}\{s_j(c_j) = a\}]$ we can ensure that there exists (α_j^a) such that the resulting mechanism satisfies the market-clearing constraint (13), and such that (α_j^a) is arbitrarily close to $(\delta\hat{\alpha}_j^a)$. As a function of (α_j^a) the market-clearing constraint is

$$\sum_{j \in \mathcal{J}} \eta_j + \sum_{j \in \mathcal{J}} \sum_{a \in A_j} \alpha_j^a \kappa_j(a) \rho_j^a = (n_1, \ldots, n_I)$$

As discussed above, $(\delta\hat{\alpha}_j^a)$ satisfies market clearing in the large-market program, meaning the previous condition is satisfied if $\alpha_j^a = \delta\hat{\alpha}_j^a$ and $\rho_j^a = \mathbb{E}_j[\mathbb{1}\{s_j(c_j) = a\}]$. Then it is also satisfied if $\rho_j^a = (1 + \gamma_j^a)\mathbb{E}_j[\mathbb{1}\{s_j(c_j) = a\}]$ for some γ_j^a if we set $\alpha_j^a = \frac{\delta}{1+\gamma_j^a}\hat{\alpha}_j^a$. Moreover, as $\delta < 1$ we have $\frac{\delta}{1+\gamma_j^a}\hat{\alpha}_j^a \leq 1$ as long as $|\gamma_j^a|$ is sufficiently small, so that (α_j^a) are valid scaling factors (i.e. the resulting mechanism does not violate the non-negativity constraint). $\square$

B Empirical Appendix

B.1 Institutional Background

This appendix provides additional institutional background for the empirical applications discussed in the paper. We begin with the CPS setting in Michigan and then describe the misdemeanor prosecution system in Suffolk County, Massachusetts.

CPS and Foster Care

The CPS and foster care systems in Michigan are broadly representative of those used in other U.S. states. The process begins when a report of suspected child abuse or neglect is made to the state's centralized hotline. Reports may come from a wide range of sources, including mandated reporters such as teachers and medical professionals. Incoming calls are routed to screeners, who determine whether the report should be "screened in" for a formal investigation. Screened-out cases involve no further CPS action, while screened-in cases are referred to the child's local CPS office for investigation.

Once a case is assigned to a local office, it is allocated to an investigator using a rotational system, under which new cases are assigned to the investigator at the top of a queue. Each county in Michigan has its own CPS office, and in some cases investigators are organized into geographically defined teams based on the child's location. As a result, assignment is effectively random within these units, which we capture using zip-code-by-year fixed effects. There are two exceptions to this rule. First, cases involving allegations of sexual abuse are typically assigned to more experienced investigators. Second, cases involving children with a recent prior investigation are often reassigned to the original investigator. Accordingly, we exclude such cases from the analysis.

After assignment, the investigator conducts an investigation, which involves interviewing the individuals involved and reviewing relevant police and medical reports, and is typically required to be completed within 30 days. The investigator then determines whether the child should be placed into foster care. Under CPS guidelines, this decision is based on the perceived risk of subsequent maltreatment if the child remains in the home. Investigators are instructed to recommend removal when the child faces imminent risk, but to maintain

the child in the home otherwise, and they retain substantial discretion in making placement decisions. If a placement is recommended, it is reviewed by a supervisor and typically approved by the court; in practice, it is rare for either supervisors or courts to overturn the investigator's recommendation. Children placed into foster care are temporarily assigned to foster families, relatives, or, much less commonly, to residential settings.

Misdemeanor Prosecution

Our second application considers the misdemeanor prosecution system in Suffolk County, Massachusetts, following Agan et al. (2023). Misdemeanor cases constitute the vast majority of criminal cases in the U.S., and for many individuals they represent the primary point of contact with the justice system. These cases often involve relatively low-level offenses, such as minor drug possession, disorderly conduct, or driving-related violations. Despite their relatively low severity, the decision to prosecute can have meaningful consequences, as it typically results in the creation of a criminal record even in the absence of a conviction.

In Suffolk County, misdemeanor cases are processed through a set of municipal and district courts, each with jurisdiction over a particular geographic area. Cases are scheduled for arraignment in the court corresponding to the location of the alleged offense. At arraignment, assistant district attorneys (ADAs) are assigned to courtrooms according to a centralized scheduling process. On a given day, each ADA is responsible for handling all cases assigned to their courtroom, and neither defendants nor prosecutors can select which cases they handle. As a result, assignment of cases to prosecutors approximates random assignment within court and time period. There are a small number of exceptions to this process. Cases involving more serious charges, such as felonies or certain violent offenses, may be subject to additional oversight or assigned to more experienced prosecutors. Accordingly, the analysis focuses on nonviolent misdemeanor cases, where the assignment process is most consistent with the quasi-random structure described above.

After a case is scheduled for arraignment, it is assigned to an ADA for screening. In the arraignment courtroom, the ADA reviews the case and decides whether to advance it to formal arraignment. If the ADA advances the case, the defendant's name and charges are entered into the official court record at the arraignment hearing, and the defendant

acquires a criminal record of the charges. If the ADA declines to advance the case, no such record is created and the case does not proceed further in the system. For cases that are prosecuted, subsequent stages of the case are handled by other prosecutors assigned through a separate process, and other court actors (e.g., judges and defense attorneys) are assigned independently of the initial ADA.

B.2 Identification Details

To connect the empirical objects to the theory, let k index individual cases, j index agents, and $i \in \mathcal{I}$ index case types. Let $g(k) \in \mathcal{I}$ denote the type of case k. Let Y_k^* denote the adverse outcome that would be realized absent intervention. In the CPS application, $D_{kj} = 1$ denotes foster care placement and $Y_k^* = 1$ denotes subsequent maltreatment if the child remains at home. In the prosecution application, $D_{kj} = 1$ denotes prosecution and $Y_k^* = 1$ denotes subsequent criminal justice involvement if the defendant is not prosecuted.[31] Given planner-chosen classification weights $(w_{\text{FP}}, w_{\text{FN}}, w_{\text{TP}}, w_{\text{TN}})$, we define the empirical counterpart of the theory primitive type-specific $\pi_j(i)$ as the expected classification loss from assigning a type-i case to agent j:

$$\pi_j(i) = \mathbb{E}\Big[w_{\text{FP}}\mathbf{1}\{D_{kj} = 1, Y_k^* = 0\} + w_{\text{FN}}\mathbf{1}\{D_{kj} = 0, Y_k^* = 1\} \\ + w_{\text{TP}}\mathbf{1}\{D_{kj} = 1, Y_k^* = 1\} + w_{\text{TN}}\mathbf{1}\{D_{kj} = 0, Y_k^* = 0\} \mid g(k) = i\Big]$$

Thus $\pi_j = (\pi_j(1), \ldots, \pi_j(I))$ is the agent-specific social-cost vector used by the mechanism, with lower values corresponding to better performance.

In both empirical applications, the relevant adverse outcome is observed only for untreated cases. This gives rise to the standard "selective labels" problem: although treatment decisions are observed for all cases, counterfactual outcomes under the alternative treatment status are not. Thus, $\pi_j(i)$ is not directly observed because Y_k^* is observed only for untreated cases. As a result, agent-specific social costs are not point-identified without

[31] For prosecuted defendants, subsequent criminal justice involvement is observed, but it reflects both underlying risk and the causal effect of prosecution itself. Because the mechanism objective is defined in terms of the outcome that would occur absent intervention, the relevant outcome in the prosecution setting is subsequent criminal justice involvement if not prosecuted. This parallels the CPS setting, where the relevant outcome is subsequent maltreatment if the child remains at home.

additional assumptions. Following Baron et al. (2026), our approach instead focuses on identifying differences in performance across agents. Under quasi-random assignment within institutional strata and an exclusion restriction – namely, that agents affect outcomes only through the treatment decision itself – differences in social costs across agents are identified even though their absolute levels are not.

Let $P_j(i) = \Pr(D_{kj} = 1 \mid g(k) = i)$ be the treatment rate and $\mathrm{FN}_j(i) = \Pr(D_{kj} = 0, Y_k^* = 1 \mid g(k) = i)$ be the false negative rate. For simplicity, suppose agents j and j' operate in the same institutional strata; we address differences in institutional strata assignment via a linear adjustment in estimation, discussed below. Under quasi-random assignment of case types to agents within institutional strata, the latent risk distribution for type-i cases does not vary across agents. Therefore, up to a type-specific constant that is common across agents,

$$\pi_j(i) = (w_{\mathrm{FP}} - w_{\mathrm{TN}})P_j(i) + (w_{\mathrm{FN}} + w_{\mathrm{FP}} - w_{\mathrm{TP}} - w_{\mathrm{TN}})\mathrm{FN}_j(i) + \kappa(i),$$

where $\kappa(i)$ does not depend on j. Consequently, for any benchmark agent j^0,

$$\pi_j(i) - \pi_{j^0}(i) = (w_{\mathrm{FP}} - w_{\mathrm{TN}})(P_j(i) - P_{j^0}(i)) + (w_{\mathrm{FN}} + w_{\mathrm{FP}} - w_{\mathrm{TP}} - w_{\mathrm{TN}})(\mathrm{FN}_j(i) - \mathrm{FN}_{j^0}(i)).$$

These relative social costs are the empirical objects that identify the performance vector used in the mechanism. To see the intuition for this result from Baron et al. (2026), note that under random assignment, all agents face the same underlying distribution of latent risk within a case type, so differences in false positive rates can be recovered from differences in treatment rates and false negative rates. Intuitively, if two agents face cases with the same underlying prevalence of adverse outcomes, then an agent who intervenes more frequently but achieves the same false negative rate must necessarily generate more false positives. This allows relative performance comparisons across agents even though latent risk is not directly observed for treated cases. As a result, pairwise welfare comparisons across agents are identified.

We define $\tilde{\pi_j}(i) = \pi_j(i) - \pi_{j^0}(i)$ as differences in social cost between agent j and an arbitrary benchmark agent j^0. Thus, $\tilde{\pi_j} = (\tilde{\pi_j}(1), \ldots, \tilde{\pi_j}(I))$ is sufficient for determining the optimal mechanism. The identification arguments extend naturally to comparisons of

assignment mechanisms. In particular, expected welfare differences across feasible reallocations depend only on the relative social costs $\tilde{\pi_j}(i)$'s. See Baron et al. (2026), Lemma 1 for the formal identification proof.

B.3 Data Sources and Analysis Samples

We use the administrative dataset introduced in Baron et al. (2026), which contains the universe of child maltreatment investigations in Michigan between January 2008 and November 2016. The data include investigator identifiers, placement decisions, and a rich set of child and investigation characteristics, including demographic information, allegation types, and prior CPS involvement.

Cases are assigned to investigators through a rotational system within offices, generating quasi-random assignment conditional on zip-code-by-year. We follow the sample construction in Baron et al. (2026) closely and impose standard restrictions to ensure quasi-random assignment and reliable measurement of outcomes. In particular, we exclude cases that are not assigned through rotation (e.g., sexual abuse cases and recent repeat investigations), and exclude observations from small rotations with fewer than four investigators. We further drop cases without at least six months of follow-up to measure subsequent outcomes and observations with missing zip code information. Finally, we drop cases assigned to an investigator who did not receive at least 50 cases from each risk quartile grouping to ensure reliable estimates of performance. Our final analysis sample consists of 168,331 investigations involving 146,801 children assigned to 377 investigators.

Following the literature, we measure subsequent maltreatment as a new CPS investigation within six months among children who are not removed from the home (e.g., Baron et al., 2024; Putnam-Hornstein et al., 2021). While re-investigations are an imperfect proxy for true maltreatment, they are commonly used in this setting and have the advantage of being outside the control of the initial investigator. We refer to this measure as "subsequent maltreatment" throughout for ease of exposition. As discussed in the main text, this outcome is unobserved for children placed into foster care, which gives rise to the selective labels problem central to our identification strategy.

Summary statistics of our analysis sample for the CPS application are reported in Table

C.1. The average child in the sample is 6.8 years old. 3.2% of investigations result in foster care placement, and among children who are not removed from the home, 16.3% are re-investigated for maltreatment within six months. These moments highlight both the relatively low frequency of placement decisions and the substantial risk of subsequent CPS contact among children who remain at home.

We also use the administrative dataset introduced and made publicly available by Agan et al. (2023), which contains the universe of nonviolent misdemeanor cases processed by the Suffolk County District Attorney's Office (SCDAO) between 2004 and 2018. The data include prosecutor identifiers, prosecution decisions, detailed case characteristics, and subsequent criminal justice outcomes. Cases are defined at the complaint level and linked across defendants using unique identifiers, allowing subsequent outcomes to be tracked over time.

Cases are assigned to prosecutors through a centralized scheduling process at arraignment, which approximates quasi-random assignment conditional on court and time. We follow the analysis sample in Agan et al. (2023), imposing standard restrictions to ensure quasi-random assignment and reliable measurement of outcomes. In particular, we restrict attention to nonviolent misdemeanor cases, excluding any case involving felony charges, violent offenses, or firearm-related charges. As in the CPS application, we additionally drop cases assigned to prosecutors who did not receive at least 50 cases from each risk quartile grouping to ensure reliable estimates of performance. Our main estimation sample consists of 31,840 cases assigned to 62 prosecutors. Summary statistics for this sample are reported in Table C.2.

Following Agan et al. (2023), we measure subsequent criminal justice involvement as the incidence of a new criminal complaint within two years following arraignment. This outcome is constructed using linked administrative records and captures subsequent contact with the criminal justice system. In this sample, 77.7% of nonviolent misdemeanor cases are advanced to prosecution. Among cases that are not prosecuted, 20.9% of defendants receive a new criminal complaint within two years.

B.4 Performance Moment Estimation Details

Risk prediction and case types. For each application, we construct an algorithmic prediction of the probability that an individual would experience an adverse subsequent outcome absent intervention. In the CPS application, this corresponds to the probability that a child would experience subsequent maltreatment if left at home. In the misdemeanor prosecution application, it corresponds to the probability that a defendant would experience subsequent criminal justice involvement if not prosecuted. Formally, we estimate $\Pr(Y_k^* = 1 \mid X_k, D_k = 0)$, where X_k includes the characteristics observed by the decision-maker at the time of assignment.

We estimate these risk predictions using gradient boosted decision trees, with hyperparameters selected via 3-fold cross-validation. In both settings, the relevant counterfactual outcome Y_k^* is observed only for untreated cases: children left at home in the CPS setting and defendants not prosecuted in the misdemeanor setting. Accordingly, the prediction models are trained exclusively on these subsamples.

The feature sets include the information available to the relevant decision-maker at the time of assignment. In the CPS setting, this includes allegation types, relationships between the alleged perpetrator and the child, prior CPS involvement, and demographic characteristics. In the prosecution setting, this includes offense characteristics, prior criminal history, and defendant demographics. In both applications, we partition observations into four case types corresponding to quartiles of the predicted risk distribution. These case types constitute the task categories used throughout the analysis.

We evaluate predictive performance on held-out samples in Figure C.1. In the CPS application, the model achieves an out-of-sample AUC of 0.633 among children left at home. In the prosecution application, the corresponding out-of-sample AUC is 0.743 among non-prosecuted defendants.

Performance estimation. We next estimate the agent-by-case-type performance moments used as inputs to the mechanism. Our estimation procedure addresses two concerns common to both applications: (*i*) overfitting and measurement error in agent-specific moments, and (*ii*) the fact that assignment is quasi-random only within institutional strata.

To avoid overstating the gains from reassignment due to overfitting, we implement a split-sample procedure. We randomly split cases within each agent into a "training" sample (50%) and an "evaluation" sample (50%). The training moments are used to construct the mechanism, while all reported welfare gains are evaluated using the held-out evaluation moments.

We first estimate agent-specific performance measures across case types. In both applications, this requires estimating treatment rates and false negative rates by agent and case type. Let Z_{kj} indicate that case k is assigned to agent j, and let $g(k) \in \{1, 2, 3, 4\}$ denote the predicted-risk quartile of case k. Let D_k denote whether case k receives the intervention—foster care placement in the CPS application and prosecution in the misdemeanor application. Let FN_k denote a false negative, equal to one when the case is untreated and the adverse outcome is realized. We estimate the following regressions separately in the training and evaluation samples:

$$D_k = \sum_j \beta_{j1}^D Z_{kj} + \sum_j \sum_{i=2}^{4} \beta_{ji}^D \mathbf{1}\{g(k) = i\} Z_{kj} + \mathbf{X}_k' \alpha^D + u_k, \tag{14}$$

$$\text{FN}_k = \sum_j \beta_{j1}^{\text{FN}} Z_{kj} + \sum_j \sum_{i=2}^{4} \beta_{ji}^{\text{FN}} \mathbf{1}\{g(k) = i\} Z_{kj} + \mathbf{X}_k' \alpha^{\text{FN}} + \epsilon_k. \tag{15}$$

Here $\mathbf{X}_k$ denotes the relevant assignment-strata fixed effects: zip-code-by-year fixed effects in the CPS application and court-by-time fixed effects in the prosecution application. Assignment is quasi-random only within institutional strata. In the CPS application, investigators are rotationally assigned within office-by-year cells, which we approximate using zip-code-by-year fixed effects. In the prosecution application, prosecutors are assigned within court-by-year-month and court-by-day-of-week cells (Agan et al., 2023). The linear adjustment then allows us to make comparisons of agent performance across strata.

The coefficient β_{j1} gives agent j's adjusted moment for the first risk quartile, while $\beta_{j1} + \beta_{ji}$ gives the corresponding adjusted moment for quartile $i \in \{2, 3, 4\}$. We then convert the estimated treatment and false-negative moments into estimates of relative social costs. Let j^0 denote the benchmark agent. For quartile 1,

$$\widehat{\tilde{\pi_j}(1)} = (w_{\text{FP}} - w_{\text{TN}})(\widehat{\beta_{j1}^D} - \widehat{\beta_{j^0 1}^D}) + (w_{\text{FN}} + w_{\text{FP}} - w_{\text{TP}} - w_{\text{TN}})(\widehat{\beta_{j1}^{\text{FN}}} - \widehat{\beta_{j^0 1}^{\text{FN}}}),$$

and for quartiles $i \in \{2, 3, 4\}$,

$$\widehat{\tilde{\pi_j(i)}} = (w_{\text{FP}} - w_{\text{TN}}) \left[(\widehat{\beta^D_{j1}} + \widehat{\beta^D_{ji}}) - (\widehat{\beta^D_{j^0 1}} + \widehat{\beta^D_{j^0 i}}) \right] + (w_{\text{FN}} + w_{\text{FP}} - w_{\text{TP}} - w_{\text{TN}}) \left[(\widehat{\beta^{\text{FN}}}_{j1} + \widehat{\beta^{\text{FN}}}_{ji}) - (\widehat{\beta^{\text{FN}}_{j^0 1}} + \widehat{\beta^{\text{FN}}_{j^0 i}}) \right].$$

Following Chan et al. (2022), we assume $w_{\text{TP}} = w_{\text{TN}} = 0$, so that welfare depends only on prediction mistakes. The relative values of w_{FN} and w_{FP} must ultimately be specified by the planner and reflect the relative social costs of false negatives and false positives. We calibrate these parameters here primarily for expositional purposes, using observed treatment and subsequent adverse outcome rates in each setting to guide benchmark values.

In the CPS application, we set $w_{\text{FP}} = 1$ and $w_{\text{FN}} = 0.25$. To motivate this choice, note that investigators place only 3% of children into foster care, while 16% of children who remain at home experience a subsequent maltreatment investigation within six months. This mismatch suggests that CPS may place greater weight on avoiding unnecessary removals than on avoiding false negatives, implying $w_{\text{FN}} < w_{\text{FP}}$. Normalizing $w_{\text{FP}} = 1$ therefore suggests values of w_{FN} in the interval $(0, 1)$, and the relative frequencies of placements and subsequent maltreatment imply a benchmark value near 0.25.

In the misdemeanor prosecution application, we follow a similar approach. Approximately 80% of cases are advanced to prosecution, while 21% of defendants who are not prosecuted receive a new criminal complaint within two years. These moments suggest that prosecutors may place substantial weight on avoiding false negatives. Normalizing $w_{\text{FP}} = 1$, the relative frequencies suggest a benchmark value of roughly 4 for w_{FN}.

Finally, because agent-by-type performance moments are noisy, we apply empirical Bayes shrinkage to the benchmark-differenced social-cost estimates $\widehat{\tilde{\pi_j(i)}}$. The resulting shrunk estimates are the performance vectors used by the mechanism.

B.5 Preference Distributions

The empirical welfare calculations require the planner to know the distribution of agents' task-specific assignment costs. However, preference data is generally unavailable. We therefore treat the preference distribution as a calibrated input to the mechanism design problem.

At the quartile level, each agent has a cost vector $c_j = (c_{j1}, c_{j2}, c_{j3}, c_{j4})$, where c_{ji} is the private cost of receiving a case in risk quartile i. The planner knows the distribution from which these costs are drawn but not an individual agent's realized cost vector. The full-information benchmark uses the same realized preference draws, but gives the planner direct knowledge of each agent's full four-dimensional type.

In the CPS application, we calibrate the preference distribution using the investigator survey evidence from Baron et al. (2026). The survey elicits investigators' marginal rates of substitution (MRS) between fourth-quartile cases and lower-risk cases. We interpret the survey MRS as evidence on the scalar relative cost

$$\theta_j = \frac{c_{j4}}{(c_{j1} + c_{j2} + c_{j3})/3}.$$

From the raw elicited marginal-rate-of-substitution (MRS) data, we first smooth this distribution using a Gaussian kernel density estimator on the observed MRS support. We then compute the hazards from the kernel-smoothed distribution and select the nondecreasing hazard sequence that minimizes the discrete sum of squared deviations in hazard rates. The CDF defined by the selected nondecreasing hazard sequence is the distribution for θ_j that we assume for investigator preferences.

To map the monotone-hazard MRS draws θ_j into a four-dimensional cost vector, we normalize $c_{j3} = 1$. If $\theta_j \leq 1$, we set $c_{j1} = c_{j2} = c_{j3} = 1$ and $c_{j4} = \theta_j$. If $\theta_j > 1$, we draw (c_{j1}, c_{j2}) uniformly from the feasible set

$$0 \leq c_{j1} \leq c_{j2} \leq 1, \qquad \theta_j \frac{c_{j1} + c_{j2} + 1}{3} \geq 1,$$

and set

$$c_{j4} = \theta_j \frac{c_{j1} + c_{j2} + 1}{3}.$$

This construction preserves the survey-implied relative cost θ_j and allows heterogeneity in the relative costs of the first three risk quartiles.

In the prosecution application, comparable survey evidence on prosecutors' case-type preferences is unavailable. We therefore use a stylized preference distribution designed to capture the idea that higher-risk cases are expected to impose greater workload, stress, or other private costs. We draw $c_{j2} \sim U[0.9, 1.2]$ and $c_{j3} \sim U[1.2, 1.5]$, set $c_{j1} = 3 - c_{j2} - c_{j3}$,

and draw $c_{j4} \sim U[1,3]$. This normalizes the average cost of the first three quartiles to one while allowing the fourth-quartile cost to vary substantially across prosecutors. Under this distribution, expected costs increase across the four risk quartiles.

B.6 Coarser Partitions and Optimal Mechanism Choice

In the main empirical exercise, we allow the planner to consider coarsenings of the task partition if they achieve greater reductions in social cost. Starting from the four predicted-risk quartiles, we consider all possible binary partitions, all possible three-group partitions, and the full quartile partition. This gives seven binary partitions, six three-group partitions, and one quartile partition.

All coarser mechanisms are constructed from the same quartile-level empirical primitives used by the quartile mechanism. We do not estimate separate binary or three-group performance regressions. Instead, when a coarser task group pools several quartiles, we average the quartile-level performance moments within that group using the relevant market's quartile shares. For example, if a binary partition pools quartiles 1 and 2 into a low-risk task, then the performance moment for that low-risk task is the weighted average of the quartile-1 and quartile-2 performance moments, with weights equal to the shares of quartile-1 and quartile-2 cases in that market. We apply the same procedure separately in each CPS county, as well as across the full sample of cases in Suffolk County.

We collapse the preference distribution in the same way. For each simulated preference draw, we first draw a four-dimensional quartile preference vector from the application-specific preference distribution described above. We then average the quartile-level preferences within each coarser task group using the same market-specific quartile shares. Thus, the binary and three-group mechanisms are not based on separate preference assumptions. They are induced by aggregating the same quartile-level preference draws using the market-specific composition of each coarser task group.

The status-quo endowment is also defined from the quartile-level market totals. Within each market, every agent is assigned the market-average number of status-quo cases in each quartile. For a coarser task group, the status-quo endowment is the sum of these market-average quartile counts over the quartiles in that group. This is the endowment used when

solving the binary and three-group mechanisms. Thus, the binary and three-group mechanisms are re-optimized from the same underlying quartile-level performance estimates, preference draws, and case composition as the quartile mechanism.

After solving a coarser mechanism, we expand its assignments back to quartiles before evaluating welfare. We do this using the same market-specific quartile shares used to construct the coarser problem. For example, if a mechanism assigns an agent ten cases from a pooled low-risk task containing quartiles 1 and 2, and quartile 1 accounts for 60 percent of that pooled group in the market, we record that assignment as six quartile-1 cases and four quartile-2 cases. All welfare changes, false negative changes, false positive changes, and intervention changes are then computed using these expanded quartile allocations and the quartile-level evaluation moments. This keeps the evaluation scale identical across binary, three-group, and quartile mechanisms.

We select the optimal partition using only training moments. For each candidate partition, we solve the mechanism using training-sample performance moments and compute its small-market social-cost change, averaged over simulated preference profiles and evaluating the social cost change using training performance moments. We define the optimal mechanism as the candidate with the largest training-moment small-market reduction in social cost.

Tables C.3 and C.4 report the training-moment, small-market criterion used to select the task partition. In the CPS application, the selected mechanism is the binary partition that pools quartiles 1 and 2 and pools quartiles 3 and 4. In the prosecution application, the selected mechanism is the three-group partition that separates quartile 1, pools quartiles 2 and 3, and separates quartile 4.

The optimal mechanism is chosen before using the evaluation moments. After selecting the optimal mechanism, we evaluate it using held-out test moments. The main tables report the selected small-market mechanism, the corresponding large-market allocation, and the full-information benchmark. The full-information benchmark is always defined at the quartile level: even when the selected mechanism is binary or three-group, the benchmark allows the planner to observe each agent's full four-dimensional quartile preference type.

B.7 Social Welfare Change

Using the quartile performance moments and preference distribution described above, we then solve for the optimal mechanism. We evaluate the performance of each mechanism by comparing its implied assignment to the status-quo assignment. Let m index a candidate mechanism and let $q^i_{jm}(b)$ denote the number of quartile-i cases assigned to agent j under mechanism m in simulated preference profile b. Let q^i_{j0} denote the corresponding status-quo assignment. In the empirical implementation, we define each agent's status-quo allocation as the average number of cases per worker of each type in the county/office, matching the feasibility constraints imposed in the mechanism.[32]

Following Baron et al. (2026), the estimated change in social cost for a given preference profile b is

$$\widehat{\Delta C}_m(b) = \sum_j \sum_{i=1}^{4} \left(q^i_{jm}(b) - q^i_{j0}\right) \widehat{\widetilde{\pi_j(i)}}.$$

Since $\widehat{\widetilde{\pi_j(i)}}$ is measured relative to an arbitrary benchmark agent, this expression identifies welfare changes rather than welfare levels. When computing welfare changes, we use evaluation sample estimates of the $\widehat{\widetilde{\pi_j(i)}}$, which are separately estimated from the training sample $\widehat{\widetilde{\pi_j(i)}}$ to fit the mechanism as described above. Negative values of $\widehat{\Delta C}^e_m$ therefore correspond to reductions in social cost, or welfare gains.

For the large-market mechanism, we set q^i_{jm} as the expected allocations from solving the large-market problem. For the finite-agent small-market mechanism, allocations $q^i_{jm}(b)$ vary across draws of preference profiles b, so we compute the welfare change separately for each simulated preference profile and report the average,

$$\widehat{\Delta C_m} = \frac{1}{B} \sum_{b=1}^{B} \widehat{\Delta C_m(b)}$$

The full-information benchmark is computed analogously, except that the planner is allowed to observe each agent's realized four-dimensional quartile preference type before assigning cases.

We compute changes in intermediate outcomes using the same assignment comparisons.

[32] Observed assignments may differ from this equal-split allocation, but under random assignment, this is the expected status-quo assignment.

Let $\widehat{P_j(i)}$ denote the estimated treatment rate for agent j and type i, and let $\widehat{\mathrm{FN}_j(i)}$ denote the corresponding false negative rate. The change in interventions is

$$\widehat{\Delta D_m(b)} = \sum_j \sum_{i=1}^{4} \left(q^i_{jm}(b) - q^i_{j0}\right) \widehat{P_j(i)},$$

where interventions correspond to foster care placements in the CPS application and prosecutions in the misdemeanor application. The change in false negatives is

$$\widehat{\Delta FN_m(b)} = \sum_j \sum_{i=1}^{4} \left(q^i_{jm}(b) - q^i_{j0}\right) \widehat{\mathrm{FN}_j(i)}.$$

While false positives are not observed, their changes can be recovered from the identity:

$$\widehat{\Delta FP_m(b)} = \widehat{\Delta D_m(b)} + \widehat{\Delta FN_m(b)}.$$

Reported percent changes divide each level change by the corresponding status-quo baseline. In the CPS application, baseline false positives and baseline social costs require the extrapolation procedure described in Appendix B.8. In the prosecution application, we do not extrapolate baseline false-positive levels as the untreated rate is too low, so percent effects are reported only for directly identified baseline outcomes.

We also report standard errors that account for sampling uncertainty in the estimated evaluation performance moments as well as uncertainty across preference profile draws. For each bootstrap draw, we perturb the raw agent-by-type evaluation performance estimates using the agent-clustered standard errors from the performance regressions, re-apply the empirical Bayes shrinkage procedure to the perturbed moments, and recompute the welfare and outcome changes using the derived mechanism allocation. For the small-market mechanism and the full-information benchmark, we simultaneously use different draws from the preference distribution across bootstrap draws, affecting $q^i_{jm}(b)$. The reported standard error is the standard deviation of the resulting bootstrap distribution.

B.8 Identification of False Positive Rates

The identification results above are sufficient for the relative welfare comparisons that are the primary focus of the paper. In the CPS application, however, we additionally estimate

false positive rates and corresponding social-cost levels via the extrapolation-based approach used in Baron et al. (2026), which follows the "identification at infinity" methods of Arnold et al. (2022). This allows us to express the estimated effects of the mechanism as percentage changes relative to the baseline.

The key challenge is that false positive rates depend on the unconditional latent maltreatment probability, $\mathbb{E}[Y_k^*]$, which is not directly observed because subsequent maltreatment outcomes are only observed for children who remain at home. Following Baron et al. (2026), we recover this quantity by extrapolating investigator-specific subsequent maltreatment rates toward a hypothetical investigator with a zero placement rate. Intuitively, selective labels vanish as placement rates approach zero because outcomes become observed for nearly all assigned children.

This approach is particularly well-suited to the CPS setting because foster care placement rates are very low in practice ($\approx 3\%$ in our sample), so the required extrapolation is limited. In contrast, approximately 80% of misdemeanor cases are prosecuted in the Suffolk County application. Implementing a comparable extrapolation strategy in that setting would therefore require extrapolating far beyond the support of the observed prosecutor treatment-rate distribution, making the resulting estimates highly sensitive to functional form assumptions. Accordingly, in the prosecution application we restrict attention to differences across mechanisms without attempting to recover baseline false positive levels. See Appendix E.6 of Baron et al. (2026) for additional details and formal derivations.

C Supplementary Tables and Figures

Table C.1: Summary Statistics for the CPS Application

	Mean
Female child	48.0%
White child	64.2%
Black child	21.8%
Prior CPS investigation	43.9%
Age at investigation	6.8
Domestic violence allegation	10.1%
Physical abuse allegation	28.8%
Physical neglect allegation	41.9%
Substance abuse allegation	16.9%
Foster care placement	3.2%
Six-month reinvestigation, among not placed	16.3%
Investigations	168,331
Children	146,801
Investigators	377
Counties	45

Notes. This table summarizes the CPS case-level analysis sample used to estimate the quartile investigator performance moments. Six-month reinvestigation is reported among investigations not resulting in foster care placement, since the outcome is observed only for children left in the home.

Table C.2: Summary Statistics for the Prosecution Application

	Mean
Male defendant	79.2%
Predicted White defendant	33.1%
Predicted Black defendant	33.1%
Predicted Hispanic defendant	27.2%
Age at filing	33.4
Misdemeanor conviction within past year	8.1%
Felony conviction within past year	4.0%
Number of counts	1.7
Motor vehicle offense	41.9%
Drug offense	14.9%
Public order offense	26.5%
Prosecuted	77.7%
Two-year criminal complaint, among not prosecuted	20.9%
Cases	31,840
Defendants	27,690
Prosecutors	62

Notes. This table summarizes the misdemeanor prosecution case-level analysis sample used to estimate the quartile prosecutor performance moments. Predicted race and ethnicity rows report average predicted probabilities.

Table C.3: CPS Task Partition Comparison

Task Partition	(1) Small Market	(2) Large Market	(3) Full Information
Binary: $\{1,2\} \mid \{3,4\}$	**-1,329.5**	**-2,160.2**	**-9,556.4**
Binary: $\{1,3\} \mid \{2,4\}$	-1,264.9	-1,763.6	-9,556.4
Three-group: $\{1,2\} \mid \{3\} \mid \{4\}$	-1,252.2	-3,107.6	-9,556.4
Binary: $\{1,4\} \mid \{2,3\}$	-1,182.9	-1,867.1	-9,556.4
Binary: $\{1,2,3\} \mid \{4\}$	-1,103.8	-1,967.2	-9,556.4
Three-group: $\{1,3\} \mid \{2\} \mid \{4\}$	-1,098.8	-3,044.7	-9,556.4
Binary: $\{1,2,4\} \mid \{3\}$	-1,081.4	-1,753.7	-9,556.4
Three-group: $\{1,4\} \mid \{2\} \mid \{3\}$	-1,049.5	-2,959.1	-9,556.4
Binary: $\{1,3,4\} \mid \{2\}$	-971.7	-1,799.2	-9,556.4
Quartile: $\{1\} \mid \{2\} \mid \{3\} \mid \{4\}$	-684.3	-4,224.5	-9,556.4
Three-group: $\{1\} \mid \{2,4\} \mid \{3\}$	-535.1	-3,010.7	-9,556.4
Three-group: $\{1\} \mid \{2,3\} \mid \{4\}$	-445.4	-3,100.5	-9,556.4
Binary: $\{1\} \mid \{2,3,4\}$	-429.7	-2,196.3	-9,556.4
Three-group: $\{1\} \mid \{2\} \mid \{3,4\}$	-361.2	-3,410.8	-9,556.4

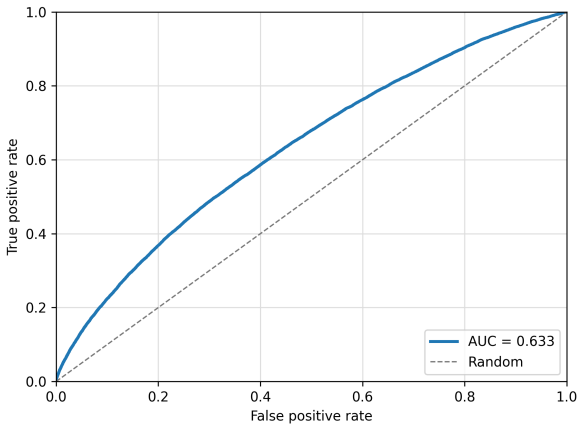
Notes. This table reports training-moment social-cost changes for each task partition considered in the CPS coarsening exercise. Partitions are written using predicted-risk quartile indices. Rows are sorted by small-market social-cost changes, with more negative values indicating larger cost reductions. Bold rows denote the partition selected by training-moment small-market performance. The full-information column reports the quartile full-information benchmark, so it does not vary across coarsening decisions within a sample.

Table C.4: Prosecution Task Partition Comparison

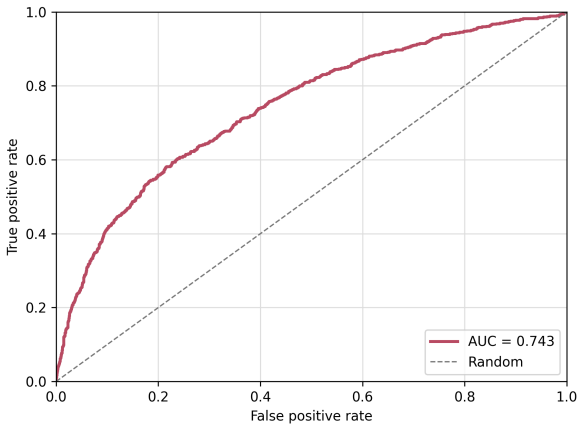
Task Partition	(1) Small Market	(2) Large Market	(3) Full Information
Three-group: $\{1\} \mid \{2,3\} \mid \{4\}$	**-1,157.8**	**-1,567.9**	**-4,659.2**
Binary: $\{1,2\} \mid \{3,4\}$	-1,129.4	-1,041.8	-4,659.2
Three-group: $\{1\} \mid \{2,4\} \mid \{3\}$	-883.7	-1,538.3	-4,659.2
Binary: $\{1\} \mid \{2,3,4\}$	-851.9	-892.9	-4,659.2
Three-group: $\{1,3\} \mid \{2\} \mid \{4\}$	-700.8	-1,271.4	-4,659.2
Quartile: $\{1\} \mid \{2\} \mid \{3\} \mid \{4\}$	-532.9	-1,943.8	-4,659.2
Binary: $\{1,3,4\} \mid \{2\}$	-483.8	-375.9	-4,659.2
Binary: $\{1,4\} \mid \{2,3\}$	-480.4	-419.3	-4,659.2
Binary: $\{1,3\} \mid \{2,4\}$	-414.6	-406.3	-4,659.2
Three-group: $\{1,2\} \mid \{3\} \mid \{4\}$	-404.6	-1,579.1	-4,659.2
Binary: $\{1,2,3\} \mid \{4\}$	-385.8	-763.7	-4,659.2
Binary: $\{1,2,4\} \mid \{3\}$	-369.7	-271.6	-4,659.2
Three-group: $\{1,4\} \mid \{2\} \mid \{3\}$	-334.9	-1,045.0	-4,659.2
Three-group: $\{1\} \mid \{2\} \mid \{3,4\}$	-74.6	-1,977.9	-4,659.2

Notes. This table reports training-moment social-cost changes for each task partition considered in the prosecution coarsening exercise. Partitions are written using predicted-risk quartile indices. Rows are sorted by small-market social-cost changes, with more negative values indicating larger cost reductions. Bold rows denote the partition selected by training-moment small-market performance. The full-information column reports the quartile full-information benchmark, so it does not vary across coarsening decisions within a sample.

Figure C.1: Algorithm Performance: ROC Curves



Panel A: CPS Sample



Panel B: Prosecution Sample

Notes. This figure summarizes the predictive performance of the algorithms $m(X_k)$ in the CPS and prosecution samples, evaluated on the held-out sample. Within each sample, the figure plots the receiver operating characteristic (ROC) curves of $m(X_k)$ as a classifier of Y_k^* and reports the AUC. The sample is limited to untreated cases.