

STRATEGIC SIMPLICITY IN MULTIDIMENSIONAL SCREENING^{*}

Quitze Valenzuela-Stookey[†] Jason Baron[‡] Richard Lombardo[§]

October 1, 2026

Abstract

We study strategically simple mechanisms for a class of multidimensional screening problems that arise frequently in applications: a designer must allocate a set of heterogeneous items among agents, while each agent has the right to demand a status-quo assignment. We first model strategic simplicity via the notion of *obvious strategy-proofness* (OSP). Static OSP mechanisms take a simple form in which agents choose from a diverse menu of narrowly defined *veto rights*: each agent chooses from a menu of one-dimensional sets in which their assignment can lie, and these sets must represent sufficiently differentiated directions of trade away from the status quo. We call these *polarized-ray mechanisms*. For general extensive-form games, the characterization of OSP mechanisms preserves much of the polarized-ray structure from the static case. We then consider a more demanding notion of *strong obvious strategy-proofness*, and show that such mechanisms essentially reduce to sequential implementation of polarized-ray mechanisms. We 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 performance in two empirical applications: the assignment of Child Protective Services investigations to caseworkers and the assignment of misdemeanor cases to prosecutors.

^{*}This draft subsumes an earlier working paper that circulated as “Unilateral Veto Mechanisms.” We are grateful to Federico Echenique, Francesco Fabbri, Fabio Maccheroni, and seminar participants at UC Berkeley and the Theros Conference 2026 for helpful comments.

[†]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.

The need to allocate heterogeneous items (e.g. tasks, objects) among agents arises frequently in organizations. Child Protective Services investigators must be assigned to maltreatment reports (Baron et al., 2026), prosecutors to cases (Agan et al., 2023), tax collectors to households (Bergeron et al., 2026), and refugees to locations (Bansak et al., 2018; Ahani et al., 2021; Delacrétaz et al., 2023). When agents’ preferences are private information and there are many items, these become problems of multidimensional screening. Such problems are notoriously challenging. Due to the complexity of incentive constraints, it is not in general possible to tractably characterize the set of implementable maps from agents’ preferences to outcomes (Rochet, 1987; Bikhchandani et al., 2006; Lahr and Niemeyer, 2024). Optimal mechanisms are thus difficult to identify both analytically and computationally (Armstrong, 1996; Rochet and Choné, 1998; Daskalakis et al., 2014). To the extent that the literature has been able to describe optimal mechanisms, it has shown that they tend to be quite complex (Manelli and Vincent, 2006; Hart and Reny, 2015; Hart and Nisan, 2019). For example, even in the simple problem of a monopolist selling two goods to a single buyer, the optimal mechanism may consist of a menu with infinitely many options (Daskalakis et al., 2013).

Nonetheless, organizations routinely confront multidimensional screening problems in practice. In doing so, they must contend not only with the theoretical challenges outlined above, but also with the difficulty that real-world agents might have understanding and participating in complex mechanisms. This has inspired a growing interest in mechanism-design analysis that incorporates simplicity considerations. Various notions of simplicity have been proposed, including explicit restrictions on the form of the mechanism (e.g. Hart and Nisan 2017), or constraints on the complexity of the choice problem faced by agents (Valenzuela-Stookey, 2023; Camara, 2025; Yang and Dworczak, 2026).

In this paper, we focus on simplicity as it pertains to agents’ ability to engage in contingent reasoning, as formalized by Li (2017) via the notion of obvious strategy-proofness (OSP). Loosely, OSP mechanisms are those in which the worst possible consequence to an agent from following their prescribed strategy is at least as good as the best possible consequence from any deviation. This is a strengthening of dominant-strategy incentive compatibility, and there is experimental evidence that OSP implementation is more robust to agents’ levels of strategic sophistication (Li, 2017; McGee and Levin, 2019). Our main ob-

jective is to characterize OSP mechanisms in a practically relevant class of multidimensional screening problems, and develop tools for optimizing over this set.

The settings that we consider are characterized by two key features: (i) a designer needs to allocate a set of “tasks” (or objects) among agents, while ensuring that all agents prefer their assignment to some default option, and (ii) agents’ preferences over possible assignments are their private information.¹ For example, a hospital needs to assign nurses to shifts, while nurses have preferences over how many night, weekend, and weekday shifts they work. The default option in this setting might arise from a contractual option that nurses have to demand a certain mix of shifts, or it could simply represent what a designer concerned about retention and burnout considers to be a “fair” schedule (Bard and Purnomo, 2005). Similar considerations arise in a number of applications, including those mentioned above.

We focus on settings without transfers.² Moreover, we assume that agents have linear preferences over their assignment.³ The designer’s objective can differ depending on the setting; they may be concerned directly with agents’ welfare, or they may have preferences over the allocation that are orthogonal to the agents’ own, and view the agents’ preferences merely as a constraint.

Our first main result is a complete characterization of the set of static (i.e., simultaneous-move) OSP mechanisms. When the set of possible preference types is sufficiently rich, every static OSP mechanism is equivalent to an *extended polarized-ray* (ePR) mechanism, and all such mechanisms are OSP (Theorem 1). In an ePR mechanism, each action available to an agent can be identified with a one-dimensional ray originating at their status-quo assignment. By choosing a ray, the agent fixes the *direction* in which their assignment may change, while leaving the designer discretion over the magnitude of the change. Each action thus operates

¹The special case of our model with only two types of tasks can be formulated as a one-dimensional screening problem. Baron et al. (2026) and Valenzuela-Stookey (2025) study versions of this problem in the one-dimensional case. Here, we are primarily interested in problems with more than two types of tasks.

²In some settings, 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.

³On one hand, the restriction to linear preferences is an important simplifying assumption, which helps us to explicitly describe the structure of OSP mechanisms. On the other hand, restricting the preference space also expands the set of OSP mechanisms: there are mechanisms that are OSP for all agents with linear preferences, but not for agents with more general preferences. See Section 1.1 for further discussion.

as a *unilateral veto*: by acting, an agent rules out all assignments for themselves except those lying along a particular ray, regardless of what other agents do. Moreover, the different rays from which an agent can choose must point in directions that are sufficiently dispersed, as formalized by a property we call *polarization*.⁴ Polarization can be interpreted as a notion of diversity for the menu of veto rights.

A consequence of polarization is that, aside from the option to demand their status quo, each agent can have at most as many actions as there are types of tasks. This implies that static OSP mechanisms also involve simple menus, which helps to rationalize the use of relatively simple mechanisms in practice. Moreover, in such mechanisms the incentive problem and the allocation problem are largely separable: whether a menu of polarized rays is OSP depends on the rights offered to agents, not on the rule the designer subsequently uses to choose among allocations consistent with those rights. Once agents have selected their rays, the designer can therefore choose the feasible allocation that best advances their own objective (Corollary 2). Moreover, because the status-quo allocation remains feasible following every profile of choices, the designer can be sure to do weakly better than under the status quo. This guarantee does not require correctly specifying the distribution of agents' preferences (or even that preferences be linear), making the resulting mechanisms attractive when the designer has limited information about preferences.

We next drop the restriction to simultaneous-move games and characterize OSP in arbitrary extensive-form games with perfect recall, subject only to the requirement that agents retain their outside-option rights throughout the mechanism.⁵ Our second main result shows that much of the structure of static OSP mechanisms survives in the general case (Theorem 2). All and only OSP extensive-form mechanisms are equivalent to *extended recursive polarized-ray* (eRPR) mechanisms with *ex-post vetoes*. Loosely, this means that every time an agent moves, each action is associated with a set of possible rays in which their assignment can lie, and any two rays associated with different actions must be polarized. Moreover, if an agent moves for the last time, each action is associated with a single ray, as in a static

⁴Formally, what we describe here are *polarized-ray* mechanisms; the “extended” prefix refers to mechanisms in which an agent may also have multiple actions they can take that ensure them their status-quo assignment. This possibility is not essential for understanding the result, so we defer the details.

⁵It is well-known that obvious strategy-proofness is sensitive to the extensive form of the game; the standard revelation principle of Myerson (1981) does not apply.

polarized-ray mechanism (hence the “recursive” qualifier). After play is completed, agents are asked to simultaneously either “sign off” on the proposed allocation, or opt to retain their status quo (this is the “ex-post veto” step).

Obvious strategy-proofness is well suited for agents who have difficulty engaging in contingent reasoning regarding the actions of others. However, agents are still assumed to form complete plans of action regarding their own future play. Pycia and Troyan (2023b) introduce a solution concept, *strong obvious strategy-proofness* (SOSP), to model simplicity for agents who may have difficulty with such planning. Under SOSP, we show that the canonical class of mechanisms is a refinement of eRPR that more closely resembles the static case: every agent moves at most once, and their action sets have precisely the same structure as in an ePR mechanism (Theorem 3).⁶

The characterizations of OSP and SOSP have useful implications for market design. We study the designer’s problem in settings where their objective does not depend directly on the agents’ preferences. For example, each agent produces some known output when assigned to each task, and the designer seeks to maximize total output. We assume that the designer restricts attention to OSP mechanisms, for the reasons outlined above. However, before analyzing the choice among OSP mechanisms we consider how costly this restriction might be. It turns out that under mild conditions it is possible for the designer to do better (relative to the status quo) by choosing among Bayesian incentive-compatible mechanisms if and only if it is also possible to do so using a static OSP mechanism (Proposition 1). This does not imply that imposing OSP is without loss for the magnitude of achievable gains, merely that the restrictions implied by strategic simplicity do not eliminate the basic value of eliciting agents’ multidimensional private information.⁷

We then return to the designer’s choice among OSP mechanisms. We focus on the static case, both for tractability and because such mechanisms are common in practice, as they allow the designer to collect agents’ choices simultaneously before selecting an allocation. By Theorem 1 we can restrict attention to polarized-ray mechanisms (the “extended” version

⁶The only difference with the static case is that the menu of actions offered to each agent can depend on the actions of those who move earlier.

⁷In fact, if there are only two types of tasks and a continuum of agents, Baron et al. (2026) show that static OSP mechanisms are without loss of optimality among all BIC mechanisms, under standard regularity conditions on the distribution of preferences.

is not needed given the form of the designer’s objective). Even within this narrow class, optimization remains challenging.⁸ We make progress in two ways. First, we characterize the convex structure of polarized-ray menus. We partition the space of such menus into a finite collection of subsets and make a change of variables so that each subset can be represented by a relatively small set of linear constraints, which greatly simplifies the program. Second, we introduce a “large-market” relaxation in which market clearing is required in expectation rather than for every realization of agents’ preferences. This relaxation simplifies the design problem by permitting the optimization to be decomposed across agents. We show that a solution to the large-market program is asymptotically optimal in the original finite-market program as the number of agents grows (Proposition 4). Together, these results provide a procedure for constructing promising OSP mechanisms in applications, even when an exact solution is computationally difficult to obtain.

We apply this procedure to two high-stakes public-sector allocation problems: the assignment of Child Protective Services (CPS) investigations to caseworkers in Michigan and the assignment of misdemeanor cases to prosecutors in Suffolk County, Massachusetts. Both institutions rely on assignment procedures that are approximately rotational, helping equalize caseloads and promote procedural fairness but leaving potential gains from specialization unexploited. We use administrative data and quasi-random assignment to estimate how individual caseworkers and prosecutors perform on different types of cases. We then use empirical Bayes methods to estimate the expected gains from choosing a static OSP mechanism according to the procedure outlined above. Agents’ preferences over assignments are not observed in the administrative data, so we evaluate the mechanisms under calibrated preference distributions; in the CPS application these are informed by survey evidence on investigators’ willingness to handle higher-risk cases, while in the prosecution application we choose the distribution to reflect greater burden from higher-risk cases.

The resulting mechanisms generate economically meaningful gains despite the restrictions imposed by static OSP. In the CPS application, the optimized finite-agent mechanism reduces

⁸If each agent has finitely many possible types and the designer’s objective is linear in the allocation, the problem can be written as a linear program. However, the dimensionality of this program is exponential in the number of agents and tasks unless strong restrictions are imposed. For example, if there are J agents, I tasks, and each agent has two possible values for each task, then an allocation rule has dimension 2^{JI} and the linear-programming approach is computationally intractable (Daskalakis et al., 2014).

both false negatives and false positives relative to the status-quo assignment system—false negatives fall by nearly 4 percent and foster care placements by 3 percent—without increasing investigators’ workloads. Relative to a full-information benchmark in which the designer directly observes investigators’ preferences, the simple OSP mechanism captures roughly 20 percent of the reduction in the planner’s weighted social cost.⁹ In the misdemeanor application, the finite-agent OSP mechanism captures approximately 40 percent of the reduction in social cost achieved under full information; it also reduces false negatives by approximately 22 percent and prosecutions by approximately 1.5 percent. Thus, although OSP sharply constrains the form of the mechanism, suitably chosen OSP mechanisms can obtain a nontrivial share of the potential gains from better matching heterogeneous agents and tasks.

Related literature

Our main contribution is the characterization of obvious strategy-proofness in an important class of multidimensional screening problems.¹⁰ This analysis helps to rationalize the use of simple mechanisms in practice, and may also be useful from a normative market-design perspective. Such problems present a number of technical challenges under weaker solution concepts, but under OSP it suffices to consider mechanisms that grant 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 multidimensional screening, i.e., maximizing an objective subject to incentive and feasibility constraints.¹¹

Obvious strategy-proofness, introduced by Li (2017), has been widely studied due to its normative appeal. 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 task-allocation problem studied here. More gener-

⁹This benchmark is a loose upper bound on the possible gains from any feasible mechanism, as it ignores incentive constraints.

¹⁰In particular, the problem we study here is one of multidimensional screening with type-dependent participation constraints and no transfers.

¹¹This 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.

ally, despite recognition of its promise, the potential of OSP mechanisms in multidimensional screening problems – both as a means of rationalizing the simple mechanisms observed in practice and as a guide for designing new systems – remains underexplored (Palacios-Huerta et al., 2024).¹² The most closely related work is that of Pycia and Troyan (2023a), who show in a general setting that any OSP outcome can be implemented with a “millipede game.” The main differences between our setting and that of Pycia and Troyan (2023a) are (i) the restriction to linear preferences, and (ii) the status-quo constraint. Without either feature our setting would be a special case of their general environment, and adding just the status-quo restriction would further narrow the class of OSP mechanisms. However, the restriction to linear preferences violates the richness assumption in Pycia and Troyan (2023a), and thus expands the class of OSP mechanisms.¹³ As a result, the settings are not nested.

In studying strategically simple mechanisms, we also relate to the growing literature on simplicity in mechanism design. Camara (2025) characterizes preferences under which an agent’s choice problem is computationally tractable given a rich class of menus from which they may be asked to choose. Yang and Dworczak (2026) instead impose a notion of computational tractability as a constraint on the menu the designer can offer. Valenzuela-Stookey (2023) shows that when agents have preferences that are sensitive to choice complexity, optimal menus may take a simple form. In contrast, here we impose notions of strategic simplicity. Under our approach, simplicity of the menu is an implication of OSP, rather than an exogenous constraint.

We also connect to other work on related task-allocation problems (Baron et al., 2026; Valenzuela-Stookey, 2025). The current 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. It turns out that with a continuum of agents and a regularity condition on pref-

¹²Palacios-Huerta et al. (2024) write “We argue that the behavioral implications for future developments moving toward OSP [combinatorial allocation] formats should not be underestimated.”

¹³Relatedly, the restriction to environments “without transfers” in Pycia and Troyan (2023b) precludes the monotonicity assumption that we make here. One notable place where richness is used in Pycia and Troyan (2023a) is in establishing that when a player moves in an OSP mechanism, there can be at most one action that can lead to a possible payoff that cannot be guaranteed by some continuation strategy. This is key to establishing the equivalence between OSP and millipede games (see Claim 1 in the proof of Lemma 1 in Pycia and Troyan (2023a), or similarly Lemma A.5 in Pycia and Troyan (2023b)). This is no longer the case under linear preferences.

erence distributions the solution is a polarized-ray mechanism, so static OSP mechanisms are 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.

Finally, this paper is related to a literature in personnel and labor economics examining ways to improve the performance of the public sector, where common tools such as performance pay and promotion incentives are typically unavailable. Similar to the current paper, this literature has studied mechanisms for allocating public-sector workers across tasks, teams, and employers. Bates et al. (2025), Biasi et al. (2026), Laverde et al. (2025), and Eastmond et al. (2026), for example, examine the allocation of teachers to schools and classrooms. Bergeron et al. (2026) study the assignment of tax collectors to teams and neighborhoods, and Ba et al. (2022) examine alternative mechanisms for assigning police officers to Chicago districts. As in many of these studies, we allow for heterogeneous agent preferences. However, this literature typically estimates preferences structurally from choice data, such as job applications or transfers, and then computes counterfactual assignments. Such data are not always available, and when the status-quo assignment is random, as in the settings we study, workers’ choices over tasks are never observed. Our mechanism-design approach offers an alternative in such contexts: rather than estimating preferences, the mechanism elicits them at the time of assignment, guarantees that no agent is made worse off relative to the status quo, and can be re-run as preferences change over time.

The remainder of the paper is organized as follows. We present the model in Section 1. We next characterize simultaneous-move (Section 2) and general (Section 3) OSP and SOSP mechanisms. We then turn to the designer’s problem. In Section 4, we identify the conditions under which choice-based mechanisms are useful, without restricting to OSP. We then develop a procedure for optimizing over OSP 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). Section 7 concludes.

1 Model

There is a finite set of perfectly divisible tasks (or objects) indexed by $i \in \mathcal{I} = \{1, \dots, I\}$ (with $I \geq 2$) and a finite set of agents, indexed by $j \in \mathcal{J} = \{1, \dots, J\}$. Let $n^i > 0$ be the total supply of task i . An *allocation* is a matrix $\mu \in \mathbb{R}_{\geq 0}^{\mathcal{I} \times \mathcal{J}}$ such that row i sums to n^i . One interpretation is that $i \in \mathcal{I}$ represents a type of task of which there is a mass n^i , and each instance of which must be assigned to exactly 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 agents' preferences 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.¹⁴ 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 , and denote its support by C_j . We say that the preference domain is *rich* if $C_j = \mathbb{R}_{>0}^I$ for all j . For the applications we maintain the assumption that types are independent across agents, but this assumption is not relevant for characterizing OSP.

The agents' outside options are defined by an allocation $\sigma \in \mathcal{Z}$, which we refer to as the status quo.¹⁵ Each agent has the right to demand their assignment from the status quo, σ_j . We clarify the timing of such demands below.¹⁶

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 objective is to minimize the linear function

$$\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \pi_j(i) \mu_j(i), \tag{1}$$

for some known weights $\pi_j(i)$. We can think of $\pi_j(i)$ as representing the (observable) per-

¹⁴The substantive assumption is $c_j(i) > 0$, i.e. all tasks have the same qualitative effect on agents' payoffs.

¹⁵The status quo 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).

¹⁶There is a sense in which the class of OSP UV mechanisms is independent of the status quo. See the discussion following Theorem 1, and Section 5.1.

formance of agent j on task i (where lower $\pi_j(i)$ means better performance). The designer’s problem would be straightforward if they could implement any allocation in $\mathcal{Z}$. However, the designer here is constrained by the fact that (i) the designer does not know the agents’ preferences and, (ii) agents can demand their status-quo assignment.

Remark 1. An alternative interpretation of the model is that tasks are indivisible, and μ_j is the expected assignment of agent j .¹⁷ This model is equivalent from the agents’ and designer’s perspectives, given that all are assumed to have linear preferences.¹⁸ 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 mechanisms. For one, an extensive empirical literature has shown that when players have an obviously dominant strategy they are far more likely to play it, compared to settings where their strategy is merely dominant, especially in complex environments such as the multidimensional problem at hand (Li, 2017; McGee and Levin, 2019). This suggests that OSP mechanisms are not only normatively appealing, but also more likely to be observed in practice. Proposition 1 also provides some indirect evidence that OSP is not overly restrictive.

A more pragmatic reason for focusing on OSP mechanisms is tractability. The large theoretical literature on multidimensional 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 OSP mechanisms. Thus the restriction to OSP allows us to make progress on an otherwise intractable problem.

Linear preferences

Linearity of agents’ preferences over assignments is a restrictive assumption. Under our baseline interpretation of the model as one with divisible tasks, linearity rules out comple-

¹⁷We 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).

¹⁸There is one subtlety regarding the interpretation of μ_j as a lottery in the context of OSP, which we discuss in Section 3.

mentarities. If instead we adopt the perspective of indivisible tasks and interpret agents’ assignments as expected values, linearity also imposes risk neutrality. Linearity is thus most palatable in settings where agents handle many tasks and there are not significant complementarities, for example because tasks are processed at different times.

On the other hand, restricting the domain of preferences expands the set of OSP mechanisms. Thus in terms of the analysis, the relevant question is whether the class of OSP mechanisms that we characterize is too large. In principle, our results can serve as a starting point for analyzing models with more general preference domains. In practice, a designer can simply ask whether the mechanisms that we identify seem reasonable for the application at hand. Moreover, we show that these mechanisms are robust to the form of agents’ preferences: while they may not be OSP if preferences violate linearity, regardless of how agents respond the designer is guaranteed to do no worse than they would by implementing the status quo (see Corollary 2 and the subsequent discussion).

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.¹⁹

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. (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

¹⁹Roth (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).

that the mechanism generate a fixed level of revenue. Relaxing this constraint changes the space of feasible mechanisms. Thus the modifications needed to incorporate transfers do not appear to be large (particularly if tasks are burdensome, payments must be non-negative, and revenue must equal some predetermined level). 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 $\Gamma : \prod_{j \in \mathcal{J}} A_j \rightarrow \mathcal{Z}$, and type-contingent strategies for the agents $s_j : C_j \rightarrow A_j$.²⁰ 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 $\Gamma_j(a_{\mathcal{J}}) \in \mathbb{R}_{\geq 0}^I$ 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. We impose that mechanisms must 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

$$\max_{\mu_j \in \mathcal{O}_j(s_j(c_j))} \mu_j \cdot c_j \leq \min_{\mu_j \in \mathcal{O}_j(a)} \mu_j \cdot c_j \quad \forall c_j \in C_j, \text{ and } s_j(c_j) \neq a \in A_j$$

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

Recalling that agents prefer lower workload, this definition says that the worst possible assignment for j if they follow the prescribed strategy (taken over the actions of other agents) is no worse than the best possible assignment following any deviation.

²⁰It is straightforward to allow for mixed strategies. This does not substantively change the results, but introduces additional notation. Similarly, we can accommodate randomness in Γ , and do so in the general model of Section 3. Alternatively, we can think of $\mathcal{Z}$ as a set of expected allocations (see Remark 1) although this corresponds to a slightly different notion of obvious dominance, as discussed in Section 3. For clarity we focus here on deterministic mechanisms and interpret $\mathcal{Z}$ as deterministic allocations.

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

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

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 $x \in \mathcal{O}_j(a)$ for some $a \in a_j^g(c_j)$. Define the *greedy-cautious* choices for j as

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

That is, $a_j^*(c_j)$ is the subset of $a_j^g(c_j)$ with the best worst-case guarantee when the assignment is selected adversarially from $\mathcal{O}_j(a)$. We say that an agent's strategy is greedy-cautious if they always make such choices.²¹ Then we have the following simple observation.

Lemma 1. If a static mechanism is OSP then agents' strategies are greedy-cautious. Moreover, if there are multiple greedy actions for c_j then any greedy-cautious action induces a degenerate set of possible payoffs, i.e.,

$$c_j \cdot x' = \min_{a \in A_j} \{ \inf \{ c_j \cdot x : x \in \mathcal{O}_j(a) \} \} \quad \forall x' \in \mathcal{O}_j(a') \text{ and } a' \in a_j^*(c_j).$$

Proof. In Section A.1. □

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

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

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

In other words, in a static veto game 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

²¹The set of greedy-cautious choices could in principle be empty, but Lemma 1 implies that this cannot be the case in an OSP mechanism.

the ability of the designer to do so imposes some restrictions on the action sets. We refer to a player's actions as *veto sets*. A static *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 static veto game, the action set for agent j consists of *polarized rays* if

1. Each $a \in A_j$ is a 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 $\kappa(a) \in \mathbb{R}^I$.²² Moreover, $\{\sigma_j\} \in A_j$ and all other rays are non-monotone, i.e. if $\kappa(a) \neq 0$ then $\kappa(a) \not\geq 0$ and $\kappa(a) \not\leq 0$.
2. For any distinct $a', a'' \in A_j$ there exists $\alpha \in [0, 1]$ such that $\alpha\kappa(a') + (1 - \alpha)\kappa(a'') \geq 0$.

We refer to condition (2) as polarization. Intuitively, polarization means that the rays defining an agent's veto sets move the agent's allocation in different directions. Indeed, a necessary (but not sufficient) condition for polarization is that no two rays "take away" the same task, i.e. the sets of negative coordinates for $\kappa(a')$ and $\kappa(a'')$ must be disjoint. Figure 1 depicts an example. In words, polarized rays are diverse menus of narrowly defined rights.

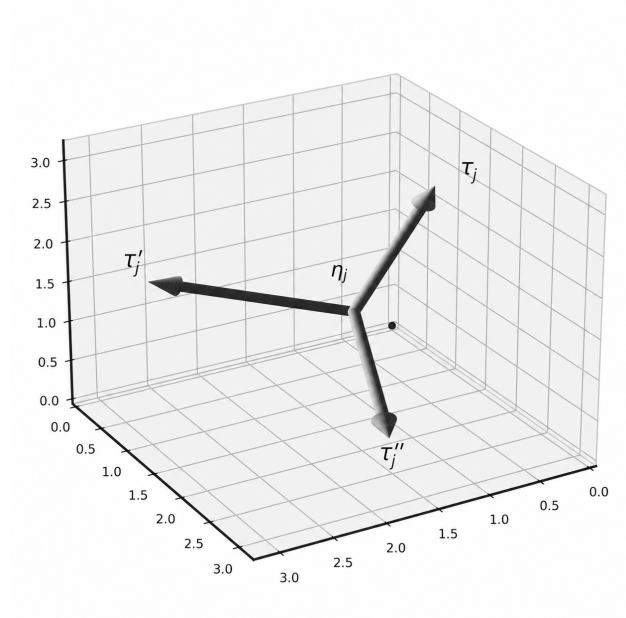


Figure 1: Polarized rays

²²The fact that these rays are not contained in $\mathbb{R}_{\geq 0}^I$ is not important, as the selection rule is restricted to non-negative allocations.

Definition. A *polarized-ray mechanism* consists of a static 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 $\{x \in \mathbb{R}_{\geq 0}^I : x \leq \sigma_j\}$; and
2. j 's strategy is greedy-cautious, 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*.²³ The other option 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 slight 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: $A_j = \hat{A}_j \cup W_j$, where actions $\hat{A}_j$ satisfy the conditions of a polarized-ray mechanism. The selection rule satisfies $\Gamma_j(a_{\mathcal{J}}) \in a_j$ if $a_j \in \hat{A}_j$ and $\Gamma_j(a_{\mathcal{J}}) = \sigma_j$ if $a_j \in W_j$. Strategies are greedy-cautious.

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 additional ways to claim their outside option. These different options are identical from j 's perspective; they are 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, having more than one withdrawal message has no value. However, they are needed to state the full equivalence result.

Theorem 1. Every extended polarized-ray mechanism is OSP. If the preference domain is rich then every static OSP mechanism is equivalent to an extended polarized-ray mechanism.

²³Technically, in order to respect the status quo we might also need to give remedial agents the option to demand σ_j . Adding this option does not affect the set of OSP strategies, so we omit it for simplicity.

Proof. Theorem 1 can be deduced from Theorem 2; however we include an independent proof in Section A.2, as it illustrates many of the key ideas without the notational encumbrance required for general extensive-form games. $\square$

Theorem 1 has a simple economic intuition. Because every other agent can demand their status quo, market clearing implies that an agent's own status-quo assignment remains possible following any action they take. The status quo therefore provides a common anchor against which the consequences of every action can be evaluated. OSP sharply limits the ambiguity that an action can create around this anchor. If the same action could lead to qualitatively different changes in the agent's assignment, then with a sufficiently rich set of linear preferences there would be a type that prefers some of these possibilities to the status quo but dislikes others. Such an action cannot obviously dominate opting out. This forces the set of assignments reachable following an action to be one-dimensional. OSP also restricts how these rays can relate to one another. Polarization guarantees that, for any preference type, at most one ray can offer an improvement over the status quo. An agent therefore needs only to determine whether one narrowly defined adjustment is beneficial; they never need to compare several attractive multidimensional options or forecast how far the designer will move along the selected direction. 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 non-degenerate members, so if the preference domain is rich a static OSP mechanism needs only I non-withdrawal actions.

Proof. Let $\kappa(a) \in \mathbb{R}^I$ be such that $a = \{y\kappa(a) + \sigma_j : y \in \mathbb{R}_{\geq 0}\}$. By definition, if $a' \neq \{\sigma_j\}$ then $\kappa(a') \not\geq 0$, so it must have at least one strictly negative coordinate. Moreover, if $\kappa(a')_i < 0$ then $\kappa(a'')_i \geq 0$ for any $a'' \neq a'$, because otherwise $y\kappa(a') + (1 - y)\kappa(a'') \not\geq 0$ for any $y \in [0, 1]$. Thus there can be at most I such vectors. $\square$

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 greedy strategies, which depend in principle on Γ through $\mathcal{O}_j$. However, in such a mechanism, identifying greedy strategies is particularly easy.

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

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

That is, there is at most one direction that the agent can move that reduces their workload relative to the status quo. This in turn implies that the selection rule does not matter for determining OSP. Say that an extended polarized-ray mechanism is *non-redundant* if there are no rays $a \in \hat{A}_j$ such that $\mathcal{O}_j(a) = \{\sigma_j\}$. In other words, there are no withdrawal actions disguised as rays; if there were, we could simply relabel them as withdrawal actions.

Corollary 2. Let $(A_{\mathcal{J}}, \Gamma, s_{\mathcal{J}})$ be a non-redundant extended polarized-ray mechanism. Then $(A_{\mathcal{J}}, \Gamma', s_{\mathcal{J}})$ is an extended polarized-ray mechanism for any feasible Γ' such that $\Gamma'_j(a_{\mathcal{J}}) \in a_j$ 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 their veto rights, rather than the rule the designer uses to select an allocation consistent with their choices. In light of Corollary 2, it is without loss of optimality for the designer to use an ex-post-optimal allocation rule:

$$\Gamma(a_{\mathcal{J}}) \in \arg \min_{\mu \in \mathcal{Z}} \sum_{j \in \mathcal{J}} \pi_j \cdot \mu_j \quad s.t. \quad \mu_j \in a_j \quad \forall j \in \mathcal{J}. \quad (2)$$

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.

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.²⁴ 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.

A simple example of a polarized-ray mechanism is constrained bilateral trade: the designer specifies a status-quo allocation and then gives agents j and j' the option to trade a vector of tasks κ . If both agree to trade then their assignments become $\sigma_j + \kappa$ and $\sigma_{j'} - \kappa$ respectively, otherwise they retain their status quo. Bilateral trading is common in many settings, such as shift trading among nurses (Holton et al., 2024). In general, this occurs in a decentralized way. On one hand, by constraining the available trades the designer can make the system easier for agents to navigate. However, this comes at the cost of flexibility. To regain some flexibility, the designer may wish to incorporate another feature of real-world trading: the fact that it is generally a dynamic process.

3 General OSP mechanisms

As is well-known, the standard revelation principle does not apply to OSP implementation; the extensive form matters as well. We therefore drop the restriction to simultaneous-move games and consider all extensive-form games with perfect recall. General OSP mechanisms preserve much of the structure of the simultaneous-move case. In particular, whenever an agent is called on to act, the possible assignments in the continuation game following the different actions they could take must be polarized. In other words, we can associate

²⁴If 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.

each action with a *set* of possible assignments, where any two assignments associated with different actions must lie in polarized rays. Unlike in the static setting, not all possible outcomes following a given action need belong to the same ray. However, along the paths of play on which an agent does not move again, their possible assignments must belong to a single ray.

Formally, we define an extensive-form game in the usual way as a tuple

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

with the following components:

- $\mathcal{H}$ is a collection of partially ordered *histories*, i.e. finite sequences of actions. We write $h \preceq h'$ when h precedes h' , i.e. h is a subhistory of h' . The initial history $\emptyset$ is in $\mathcal{H}$, and for any non-empty history $h = (a_1, \dots, a_k, a_{k+1})$, the subhistory $(a_1, \dots, a_k)$ 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 $A(h) := \{a : (h, a) \in \mathcal{H}\}$.
- $P : \mathcal{H} \setminus \bar{\mathcal{H}} \rightarrow \mathcal{J} \cup \{0\}$ 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.²⁵
- $\mathcal{M}_j$ 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 $h' \in M$.
- $\Gamma : \bar{\mathcal{H}} \rightarrow \mathcal{Z}$ is the outcome function. Let Γ_j be the corresponding assignment for j .

²⁵The full-support assumption is relevant under OSP: actions in $A(h)$ that are never taken by nature could still affect a strategy's obvious dominance; dropping them from $A(h)$ weakly expands the implementable set.

To define perfect recall, consider a history $h = (a_1, \dots, a_K)$ and let $h^t = (a_1, \dots, a_t)$ denote its subhistory of length t , with $h^0 = \emptyset$. Let $0 \leq t_1 < \dots < t_L < K$ be the indices at which agent j previously moved along h , so $P(h^{t_\ell}) = j$ for $1 \leq \ell \leq L$. Then define agent j 's *personal history* at h by

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

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

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 $M \in \mathcal{M}_j$.²⁶ Let

$$\Sigma_j := \prod_{M \in \mathcal{M}_j} A(M), \quad \Sigma_{-j} := \prod_{k \in \mathcal{J} \setminus \{j\}} \Sigma_k, \quad \Sigma := \prod_{k \in \mathcal{J}} \Sigma_k$$

denote the corresponding strategy spaces. A type-contingent strategy for agent j is a mapping $s_j : C_j \rightarrow \Sigma_j$ where $s_j(c_j)$ is the strategy prescribed to type c_j , and $s_j(c_j)(M)$ is the action taken at information set M . We write a type-contingent strategy profile as $s_{\mathcal{J}} = (s_j)_{j \in \mathcal{J}}$. A realization of nature's moves is a complete 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 . We say that information set $M \in \mathcal{M}_j$ is *reachable* by strategy s_j if there exists $(s_{-j}, d_0) \in \Sigma_{-j} \times D_0$ such that $M \preceq \zeta(\emptyset; s_j, s_{-j}, d_0)$.

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:

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

²⁶We focus here on pure strategies; the extension to mixed strategies is straightforward.

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

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

We can now define obvious dominance.

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 information set at which j acts, the worst possible workload from following s_j is no larger than the best possible workload from playing some other action at M and following any continuation strategy.

Remark 2. Obvious dominance treats random moves by nature in the same way as moves by other agents. This matters if we want to interpret μ as the expectation of a random allocation, as it implies that the two sources of randomness are treated differently. This approach is suitable if agents struggle with contingent reasoning, but have no trouble evaluating lotteries. Otherwise, the definition of obvious dominance we adopt here should be understood as a relaxation of the one that treats all randomization as a move by nature.

As in the static setting, obviously dominant strategies are greedy-cautious. For any j , $M \in \mathcal{M}_j$, and $a \in A(M)$, define $\mathcal{O}_j^*(M, a) := \{x \in \mathcal{O}_j(M, s'_j) : s'_j \in \Sigma_j, s'_j(M) = a\}$ and

$$\begin{aligned} \bar{\mathcal{O}}_j^*(M, a) &:= \{\Gamma_j(\zeta((h, a); s'_{\mathcal{J}}, d_0)) : h \in M, s'_{\mathcal{J}} \in \Sigma, d_0 \in D_0, \\ &\quad P(h') \neq j \ \forall \ (h, a) \preceq h' \prec \zeta((h, a); s'_{\mathcal{J}}, d_0)\} \end{aligned}$$

In other words, $\mathcal{O}_j^*(M, a)$ is the set of possible assignments j can get by following strategies that take action a at information set M , and $\bar{\mathcal{O}}_j^*(M, a)$ is the set of possible assignments

that j can obtain on terminal histories at which j does not move again.

We define the set of *greedy* choices for c_j at M by

$$a_j^g(c_j, M) = \arg \min_{a \in A(M)} \left\{ \inf \{c_j \cdot x : x \in \mathcal{O}_j^*(M, a)\} \right\}$$

Definition. A type-dependent strategy s_j is *greedy-cautious* if

$$s_j(c_j)(M) \in \arg \min_{a \in a_j^g(c_j, M)} \left\{ \inf_{\{s'_j \in \Sigma_j : s'_j(M) = a\}} \left\{ \sup \{c_j \cdot x : x \in \mathcal{O}_j(M, s'_j)\} \right\} \right\}$$

for every $c_j \in C_j$ and all $M \in \mathcal{M}_j$ that are reachable under $s_j(c_j)$.²⁷

A greedy-cautious strategy chooses among the greedy choices at M by selecting one that, combined with a suitable continuation strategy, has the best worst-case outcome. Just as in the static setting, OSP strategies are greedy-cautious.

Lemma 3. In any OSP mechanism, strategies s_j are greedy-cautious for all $j \in \mathcal{J}$. Moreover, if $M \in \mathcal{M}_j$ is reachable by $s_j(c_j)$ and there are multiple greedy actions for c_j at M then $s_j(c_j)$ induces a degenerate set of possible payoffs, i.e.,

$$c_j \cdot x' = \min_{a \in A(M)} \left\{ \inf \{c_j \cdot x : x \in \mathcal{O}_j^*(M, a)\} \right\} \quad \forall x' \in \mathcal{O}_j(M, s_j(c_j))$$

Proof. In Section A.3. □

As in the static setting, we restrict attention throughout to extensive forms in which agents always have the right to demand their outside option, and we assume that this option is not lost along any path of play. For each $j \in \mathcal{J}$, let

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

By perfect recall, if $R_j(h) = \emptyset$ 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.

Definition. The game form $\mathcal{G}$ *respects the agents' outside options* if for every $j \in \mathcal{J}$

- i. $R_j(\bar{h}) \neq \emptyset$ for every terminal history $\bar{h} \in \bar{\mathcal{H}}$.

²⁷If M is not reachable by $s_j(c_j)$ then the action at M is irrelevant.

- ii. For every $M \in \mathcal{M}_j^0$ there exists $s_j \in \Sigma_j$ such that $\mathcal{O}_j(M, s_j) = \{\sigma_j\}$.
- iii. $\sigma_j \in \mathcal{O}_j(M, s_j)$ for every $s_j \in \Sigma_j$ and $M \in \mathcal{M}_j$.

The first condition says that every agent gets a chance to move along any path of play, which on its own is without loss of generality. The second condition says that whenever an agent moves for the first time, they have the option to secure their status-quo assignment. The third condition says that no agent ever fully relinquishes the possibility of getting their status quo. This notion of respecting the outside options is satisfied, for example, if before the allocation is finalized each agent must sign off, and can otherwise opt to get their status quo. In fact, such mechanisms play a special role.

Formally, given an extensive-form game $\mathcal{G}$ with terminal histories $\bar{\mathcal{H}}$ and outcome function Γ , we define the *augmented game with ex-post vetoes* as follows. Starting at each terminal history $\bar{h} \in \bar{\mathcal{H}}$, we initiate a new subgame. This subgame is a simultaneous-move veto game, in which agent j chooses between two actions, the status quo $\{\sigma_j\}$ and the ray $\{\sigma_j + (\Gamma_j(\bar{h}) - \sigma_j)y : y \in \mathbb{R}_{\geq 0}\}$. The selection rule in this simultaneous-move veto game, $\hat{\Gamma}(\cdot, \bar{h})$, satisfies the usual requirements: $\hat{\Gamma}_j(a_{\mathcal{J}}, \bar{h}) \in a_j$, and the market clears. Moreover, if all agents choose the ray, i.e. sign off on the proposed allocation, then $\hat{\Gamma}_j(a_{\mathcal{J}}, \bar{h}) = \Gamma_j(\bar{h})$.²⁸ In the augmented game, we call $\bar{h} \in \bar{\mathcal{H}}$ a *provisionally-terminal history*. It is easy to see that any game augmented with ex-post vetoes will respect the agents' outside options; and in fact Theorem 2 shows that it is without loss of generality to consider such games.

The critical class of extensive-form games is a generalization of static polarized-ray games. We first extend the notion of a veto game.

Definition. A *veto game* is one in which for every $j \in \mathcal{J}$, $M \in \mathcal{M}_j$, and $a \in A(M)$, we have $a \subset \mathbb{R}^I$ and $\mathcal{O}_j^*(M, a) \subset a$.

Definition. A set of actions, $A(M)$, in a veto game consists of *polarized sets* if

1. Each $a \in A(M)$ is the union of a collection of non-monotone rays with endpoint σ_j .
2. For any distinct $a', a'' \in A(M)$ and any $x' \in a', x'' \in a''$, there exists $\alpha \in [0, 1]$ such that $\alpha(x' - \sigma_j) + (1 - \alpha)(x'' - \sigma_j) \geq 0$ (in which case we say that x' and x'' are polarized).

²⁸A special case would be a unanimity rule: if any agent chooses to veto the proposed allocation then all agents get their status quo. For our purposes it would be equivalent to use the unanimity rule.

Definition. A *recursive polarized-ray mechanism* consists of a veto game $\mathcal{G}$ and strategy profile $s_{\mathcal{J}}$ such that for every $j \in \mathcal{J}$

1. For every $M \in \mathcal{M}_j$ either $A(M)$ consists of polarized sets, or $A(M) = \{\mathbb{R}_{\geq 0}^I\}$;
2. If $A(M) \neq \{\mathbb{R}_{\geq 0}^I\}$ then $\bar{\mathcal{O}}_j^*(M, a)$ is contained in a non-monotone ray originating at σ_j for every $a \in A(M)$.
3. j 's strategy is greedy-cautious.

Condition (1) says that whenever an agent has a non-trivial set of actions to choose from, they consist of polarized sets. Polarization implies that there is at most one action $a \in A(M)$ such that by taking action a it is possible to obtain a strict improvement over the status quo. The difference between this and the static case is that the possible assignments following each action need not belong to a one-dimensional set. The reason is that by moving again in the future, an agent can prevent unfavorable terminal histories from being reached. It is thus only along paths at which the agent does not get to move again that the set of possible assignments must be one dimensional.

Definition. An *extended recursive polarized-ray (eRPR) mechanism* appends a set of opt-out messages to each action set: $A(M) = \hat{A}(M) \cup W(M)$ for each $M \in \mathcal{M}_j$, where actions $\hat{A}(M)$ satisfy the conditions of a recursive polarized-ray mechanism, and $\mathcal{O}_j^*(M, a) = \{\sigma_j\}$ for all $a \in W(M)$. Strategies are greedy-cautious.

An *extended recursive polarized-ray mechanism with ex-post vetoes* is an eRPR mechanism augmented with ex-post vetoes, such that in the ex-post veto round agent j chooses the ray $\{\sigma_j + (\Gamma_j(\bar{h}) - \sigma_j)y : y \in \mathbb{R}_{\geq 0}\}$ if and only if $c_j \cdot \Gamma_j(\bar{h}) \leq c_j \cdot \sigma_j$ (observe that this is greedy-cautious).

Theorem 2. Any extended recursive polarized-ray mechanism with ex-post vetoes is OSP. If the preference domain is rich then any OSP mechanism is equivalent to such a mechanism.²⁹

Proof. In Section A.4. □

²⁹In fact, the proof shows that any OSP mechanism is equivalent to an extended recursive polarized-ray mechanism; ex-post vetoes are not needed. However, adding an ex-post veto round may be necessary to ensure that an arbitrary eRPR mechanism is OSP and respects the outside option.

An extended recursive polarized-ray mechanism preserves much of the structure of its static counterpart. The ex-post veto round ensures that the mechanism respects the status quo; agents do not lose the option to demand their status-quo assignment until the allocation has been finalized.

Following an obviously dominant strategy is simple in the sense that the agent does not need to engage in contingent reasoning regarding the actions of others. However, it is still necessary for the agent to forecast the worst possible outcome from following their prescribed strategy. Doing so requires a different type of sophistication: the ability to plan the agent's own future choices, and trace out the implications of these actions. To account for this dimension of complexity, Pycia and Troyan (2023b) introduce a refinement of obvious dominance, which we can apply in the current context.

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

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

A mechanism is *strongly obviously strategy-proof (SOSP)* if every type plays a strongly obviously dominant strategy.

Equivalently, strategy s_j is strongly obviously dominant for type c_j if for every M that is reachable by s_j and every $a \in A(M) \setminus \{s_j(M)\}$, we have

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

The key difference between this notion and obvious dominance is that the agent is unable to forecast their own future behavior, and takes a worst-case view with regard to their own future actions following the prescribed action at the current information set. Because of this, we also strengthen the requirement for respecting the outside option: say that a mechanism *strongly respects the outside option* if it respects the outside option and, moreover, for every $M \in \mathcal{M}_j^0$ there exists $a \in A(M)$ such that $\mathcal{O}_j^*(M, a) = \{\sigma_j\}$ (the original notion required only that there exist a strategy s_j such that $\mathcal{O}_j(M, s_j) = \{\sigma_j\}$). When imposing strong obvious dominance we also require the mechanism to strongly respect the outside option.

The stronger dominance notion implies a simpler set of canonical mechanisms.

Definition. A *sequential extended polarized-ray game* is an extensive-form game of the following form: for every $j \in \mathcal{J}$

- $A(M) = \hat{A}_j(M) \cup W_j(M)$ is an extended polarized-ray action set for every $M \in \mathcal{M}_j$, where $\hat{A}_j(M)$ is allowed to be empty.
- If M is not the first information set at which j moves along some path of play (i.e., $M \notin \mathcal{M}_j^0$) then every action but one is a withdrawal message (i.e., $|\hat{A}_j(M)| = 1$).
- The final assignment $\Gamma(\bar{h})$ satisfies $\Gamma_j(\bar{h}) \in a_j$ if $(h, a_j) \preceq \bar{h}$ for some $h \in M^0 \in \mathcal{M}_j^0$ and $a_j \in \hat{A}_j(M^0)$; and $\Gamma_j(\bar{h}) = \sigma_j$ if $(h, a_j) \preceq \bar{h}$ for some $h \in M \in \mathcal{M}_j$ and $a_j \in W_j(M)$.

In words, when an agent moves for the first time they choose from an extended polarized-ray menu, and any further moves consist only of withdrawal actions or a trivial continuation action. A *sequential extended polarized-ray mechanism* consists of such a game, together with greedy-cautious strategies.

Theorem 3. Every sequential extended polarized-ray mechanism is SOSP. If the preference domain is rich, any SOSP mechanism is equivalent to such a mechanism.

Proof. In Section A.5. □

As in the static case, providing multiple withdrawal messages is only useful in order to correlate one agent's assignment with another's type. The same is true of withdrawal messages sent after the agent has selected a non-degenerate ray. If the principal has no incentive to do so, as in the applications we consider, then it is without loss of optimality to include a single withdrawal message in each action set. Moreover, Theorem 3 implies that in this case it is without loss of optimality to have each agent move at most once.

The mechanisms identified above as OSP and SOSP bear some resemblance to those currently used in certain settings. For example, many airlines use a layered preferential bidding system (PBS) to organize the schedules of flight attendants and pilots. Agents simultaneously submit several “layers”, each defined by acceptable schedule characteristics such as trip

length, layover destination, and days off. The system constructs monthly schedules in seniority order. It first attempts to complete an agent’s schedule under the constraints imposed by their highest-priority layer, proceeding to subsequent layers – which relax scheduling constraints – as necessary (Association of Professional Flight Attendants, 2026).

Formally, layered PBS is a static mechanism because all agents submit their layers up front. It resembles a static unilateral veto mechanism, in that a layer specifies acceptable schedules. However, the construction of schedules in order of agent seniority adds a dynamic flavor. Our results suggest that the mechanism could be simplified by making it explicitly dynamic, i.e. having agents report layers only after the schedules of their more senior colleagues have been fixed. This would then more closely resemble a sequential unilateral veto mechanism, with each agent moving once. Of course, the alignment between the PBS mechanism and our results is not perfect. For one, there is no clear status quo. Moreover, the ray structure appears to be absent from this setting, perhaps because there are strong complementarities between schedule characteristics.

4 When is choice useful?

Before turning to the problem of choosing among OSP mechanisms, we take a step back and answer a more basic question: does the designer benefit from sacrificing some control of the allocation to the agents? We focus on the case where the designer’s objective takes the form in (1). We are interested in the answer to this question both when the designer is restricted to 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 . Assume that types are independent across agents. 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 is³⁰

$$\begin{aligned}
& \min_m \mathbb{E} \left[\sum_{i=1}^I \sum_{j=1}^J \pi_j(i) m_j(i|\mathbf{c}) \right] & (3) \\
& s.t. \quad \sum_{i \in \mathcal{I}} c_j(i) \mathbb{E}_{-j}[m_j(i|c_j, c_{-j})] \leq \beta_j \sum_{i \in \mathcal{I}} c_j(i) \sigma_j(i) \quad \forall j \in \mathcal{J}, c_j \in C_j & (\text{SQ}) \\
& \quad \sum_{i \in \mathcal{I}} c_j(i) \mathbb{E}_{-j}[m_j(i|c_j, c_{-j})] \leq \sum_{i \in \mathcal{I}} c_j(i) \mathbb{E}_{-j}[m_j(i|c'_j, c_{-j})] \quad \forall j \in \mathcal{J}, c_j, c'_j \in C_j & (\text{IC}) \\
& \quad \sum_{j \in \mathcal{J}} m_j(i|\mathbf{c}) = n^i \quad \forall i \in \mathcal{I}, \mathbf{c} \in \mathbf{C} & (\text{MC})
\end{aligned}$$

The designer here faces a multidimensional screening problem with a type-dependent participation constraint. While the solution to this problem is an open question, 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 strict improvement for the designer over the status quo. Say that *choice is optimal* if all solutions to the designer's program in (3) are non-constant mechanisms.³¹

Say that *the status quo has full support* if $\sigma_j(i) > 0$ for all $i \in \mathcal{I}, j \in \mathcal{J}$; and 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 hyperrectangle. Let $\bar{c}_j^i = \max\{c_j(i) : c_j \in C_j\}$ and $\underline{c}_j^i = \min\{c_j(i) : c_j \in C_j\}$, so $[\underline{c}_j^i, \bar{c}_j^i]$ is the

³⁰It is straightforward to extend Proposition 1 to define sufficient conditions for choice to have value or be optimal for 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.

³¹From 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.

domain of j 's preference in dimension i . Say that *performance is identical* for all agents if $\pi_{j'} = \pi_{j''}$ for all $j', j'' \in \mathcal{J}$.

Proposition 1. 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 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 $\underline{c}_j^i < \bar{c}_j^i$ for all i, j ,³² then the following is also equivalent to *i.* - *iii.*

- iv. Choice is optimal if the designer is restricted to static OSP mechanisms (and a fortiori if they can use any BIC mechanism).³³

Proof. In Section A.6. □

Condition *i.* in Proposition 1 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.

³²As preferences are invariant with respect to positive rescaling, it is always without loss of generality to normalize $\bar{c}_j^i = \underline{c}_j^i$ for some i . While under this normalization the support of c_j is not a hypercube, it is straightforward to adapt the statement of the result to this representation.

³³“Choice is optimal” means that every solution to the program in question is a non-constant mechanism, i.e. responds to preferences.

At a high level, the conclusion that choice is generally valuable should not be surprising. However, it is worth emphasizing that Proposition 1 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 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.

Proposition 1 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 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 OSP mechanisms. We focus on the static case, both for tractability and because such mechanisms are common in practice. We assume that the preference domain is rich, so by Theorem 1 and Corollary 2 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 (2). 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, in a way that allows us to simplify the polarization constraint. 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 menus. Our objective is to simplify the polarization constraint when searching for mechanisms that perform well according to the designer’s objective. We show that this set can be divided into classes, and that within each class a change of variables makes the polarization constraint linear. This does not make the designer’s objective linear or concave in the new variables. It does, however, allow us to optimize over a convex parameter set described by a small number of linear constraints.

As this section concerns the menu offered to a given agent, we suppress the agent subscript. Recall that a non-degenerate ray in a polarized-ray menu has endpoint σ and can be written as $\{\sigma + y\kappa : y \in \mathbb{R}_{\geq 0}\}$, where $\kappa \in \mathbb{R}^I$ has at least one strictly negative and one strictly positive coordinate. Its direction is unchanged if we multiply κ by a strictly positive scalar. Let R denote a menu of such directions, omitting the degenerate ray $\{\sigma\}$ (and for this section ignore the degenerate menu consisting only of $\{\sigma\}$, for which $R = \emptyset$). The menu is polarized if and only if for every distinct $\kappa, \kappa' \in R$, there exists $\alpha \in (0, 1)$ such that

$$\alpha\kappa + (1 - \alpha)\kappa' \geq 0. \tag{4}$$

This condition concerns only the directions, and is independent of σ . We denote the set of polarized-ray menus by $\mathcal{P}$, and the subset consisting of menus with exactly n rays by $\mathcal{P}^n$,

for $1 \leq n \leq I$. For a sign vector $\zeta \in \{-1, 1\}^I$, let

$$N(\zeta) := \{i \in \mathcal{I} : \zeta_i = -1\}.$$

We use ζ to identify the *negative support* of a ray: a direction with this label is strictly negative at every coordinate in $N(\zeta)$, and nonnegative at every other coordinate. In particular, a positive entry of ζ permits a zero entry of the direction. This convention assigns each ray a unique label, including when some coordinates are zero.

A class $P \subset \mathcal{P}^n$ is specified by a set of sign vectors $E(P)$, with $|E(P)| = n$, such that the sets $N(\zeta)$ are nonempty, proper subsets of $\mathcal{I}$ and are pairwise disjoint. In other words, for a polarized-ray menu $R \in \mathcal{P}^n$, we have R in P if and only if for every $\zeta \in E(P)$ there is exactly one $\kappa \in R$ such that $\{i \in \mathcal{I} : \kappa_i < 0\} = N(\zeta)$. With a slight abuse of notation we denote this ray by $R(\zeta)$. Disjointness is necessary for polarization: if two rays were strictly negative in the same coordinate, no convex combination of them could be nonnegative. Thus every polarized-ray menu belongs to exactly one of these classes. Moreover, if $i \in N(\zeta)$, polarization requires $R(\zeta')_i > 0$ for every $\zeta' \neq \zeta$. Indeed, both weights in (4) are strictly positive, so the negative entry of one ray must be offset by a positive entry of the other.

To remove the indeterminacy in the length of each direction, we henceforth choose its unique representative satisfying $\sum_{i \in N(\zeta)} -R(\zeta)_i = 1$ for every $\zeta \in E(P)$. This is only a normalization of the vectors defining the rays; it does not restrict the veto sets or the assignments that the designer may select from them.

We now introduce the change of variables. Let $N(P) := \cup_{\zeta \in E(P)} N(\zeta)$ be the set of all negative coordinates represented by class P . Fix a menu $R \in P$. For every $\zeta \in E(P)$, $i \in N(\zeta)$ and $\zeta \neq \zeta' \in E(P)$, define the ratio

$$t_i(\zeta') := \frac{R(\zeta')_i}{R(\zeta')_i - R(\zeta)_i} \in (0, 1) \quad (5)$$

The ratio $t_i(\zeta')$ is the largest weight that can be placed on $R(\zeta)$ in a convex combination with $R(\zeta')$ without making coordinate $i \in N(\zeta)$ negative. It lies in $(0, 1)$ because $R(\zeta)_i < 0$ by construction, and $R(\zeta')_i > 0$ by polarization. The next lemma shows how these coordinate-specific bounds characterize polarization.

Lemma 4. Consider two non-monotone directions $R(\zeta)$ and $R(\zeta')$ with disjoint negative

supports, each strictly positive on the negative support of the other. They are polarized if and only if $t_i(\zeta') + t_k(\zeta) \geq 1$ for every $i \in N(\zeta)$, $k \in N(\zeta')$.

Proof. In Section A.7. □

To complete the change of variables, for each $\zeta \in E(P)$ and $i \in N(\zeta)$ define

$$d_i := -R(\zeta)_i > 0, \tag{6}$$

which is well-defined because the sets $N(\zeta), N(\zeta')$ for $\zeta \neq \zeta'$ are disjoint. Finally, for coordinates that are negative in no ray, retain the original nonnegative entries:

$$q_i(\zeta) := R(\zeta)_i \geq 0, \quad i \in \mathcal{I} \setminus N(P), \quad \zeta \in E(P). \tag{7}$$

For a class P , let $\mathcal{D}(P)$ be the set of parameter vectors (d, t, q) satisfying

$$d_i > 0 \quad (i \in N(P)), \tag{8}$$

$$\sum_{i \in N(\zeta)} d_i = 1 \quad (\zeta \in E(P)), \tag{9}$$

$$q_i(\zeta) \geq 0 \quad (\zeta \in E(P), i \notin N(P)). \tag{10}$$

$$0 < t_i(\zeta') < 1 \quad (\zeta' \neq \zeta, i \in N(\zeta)), \tag{11}$$

$$t_i(\zeta') + t_k(\zeta) \geq 1 \quad (\zeta \neq \zeta', i \in N(\zeta), k \in N(\zeta')). \tag{12}$$

If $n = 1$, we additionally require

$$\sum_{i \notin N(P)} q_i(\zeta) > 0, \tag{13}$$

so that the single ray has a strictly positive coordinate (when $n \geq 2$, this is implied by polarization).

The original polarization constraint requires finding a convex combination for each pair of rays that is nonnegative in every coordinate. In the new variables, this amounts to checking that the lower bounds on the weight do not exceed the upper bounds. The resulting inequalities are linear.

Proposition 2. For every class $P \subset \mathcal{P}^n$, the normalized polarized-ray menus in P are in

one-to-one correspondence with $\mathcal{D}(P)$. The correspondence is given by (5)–(7), with inverse

$$R(\zeta)_i = \begin{cases} -d_i, & i \in N(\zeta), \\ d_i \frac{t_i(\zeta)}{1 - t_i(\zeta)}, & i \in N(P) \setminus N(\zeta), \\ q_i(\zeta), & i \in \mathcal{I} \setminus N(P). \end{cases} \quad (14)$$

The set $\mathcal{D}(P)$ is convex and is described by the linear constraints (8)–(12), plus (13) if $n = 1$.

Proof. In Section A.8. □

For illustration, suppose there are two tasks and two rays, with negative supports $\{1\}$ and $\{2\}$. The normalized directions take the form $\kappa^1 = (-1, b)$ and $\kappa^2 = (c, -1)$, where $b, c > 0$. Polarization requires $bc \geq 1$. Define $x = c/(1 + c)$, $y = b/(1 + b)$. The same constraint becomes $x + y \geq 1$, with $0 < x, y < 1$. Conversely, every point satisfying these linear restrictions yields a polarized menu through $b = y/(1 - y)$ and $c = x/(1 - x)$. The change of variables therefore replaces the multiplicative constraint $bc \geq 1$ by a linear inequality.

The parametrization is also economical in the number of variables and constraints.

Lemma 5. For $P \subset \mathcal{P}^n$ the parameter domain $\mathcal{D}(P)$ has $n(I - 1)$ degrees of freedom. The number of distinct polarization inequalities from (12) is

$$\frac{1}{2} \sum_{\{(\zeta, \zeta') \in E(P)^2 : \zeta \neq \zeta'\}} |N(\zeta)| |N(\zeta')| = \frac{1}{2} \left(|N(P)|^2 - \sum_{\zeta \in E(P)} |N(\zeta)|^2 \right) \leq \frac{I(I - 1)}{2}.$$

Proof. In Section A.9. □

For example, with ten tasks and ten rays, each negative in a different coordinate, there are 90 free parameters and 45 polarization inequalities. Thus polarization can be imposed directly in a non-linear optimization routine.³⁴

³⁴Alternatively, when $N(\zeta), N(\zeta')$ are large, we can introduce an auxiliary variable $\alpha_{\zeta\zeta'} \in [0, 1]$ and impose

$$t_i(\zeta') \geq \alpha_{\zeta\zeta'} \quad (i \in N(\zeta)), \quad t_k(\zeta) \geq 1 - \alpha_{\zeta\zeta'} \quad (k \in N(\zeta')).$$

By the proof of Lemma 4, existence of this variable is equivalent to the original polarization inequalities for the pair. This replaces $|N(\zeta)| |N(\zeta')|$ inequalities by $|N(\zeta)| + |N(\zeta')|$, at the cost of one variable, and can be useful when it reduces the size of the constraint system.

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

Lemma 6. $C(I) = 2^I - 2 + \sum_{h=2}^I c(I, h)$, and $c(I, h)$ can be written recursively as

$$c(I, h) = \frac{1}{h} \sum_{j=1}^{I-(h-1)} \binom{I}{j} c(I-j, h-1) \quad (15)$$

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

Proof. Proof in Section A.10. □

For example $C(2) = 3, C(4) = 50, C(6) = 875, C(8) = 21,145, C(10) = 678,568$, and $C(12) = 27,644,435$. The gain from Proposition 2 is in the search within each class; it does not eliminate the need to compare classes. However, when the number of tasks is moderate we can thus exploit the convex structure of polarized-ray menus of a given class to reduce the computational burden of identifying (approximately) optimal static OSP mechanisms.

As we discuss below, the designer's objective is generally non-linear in the menu even when it is linear in assignments. If $V(R)$ denotes the value of a menu in a particular optimization problem, and $R_P(d, t, q)$ is the menu reconstructed by (14), the maximization problem over class P can be written as

$$\sup_{(d,t,q) \in \mathcal{D}(P)} V(R_P(d, t, q)).$$

The feasible parameter set is convex, but the composed objective need not be concave. The parametrization consequently simplifies the constraints without guaranteeing that a local solution is globally optimal.

The strict bounds in the parametrization also matter for implementation. One may begin with $d_i \geq \varepsilon$ and $\varepsilon \leq t_i(\zeta) \leq 1 - \varepsilon$, for a sufficiently small $\varepsilon > 0$, and then relax these bounds to assess their effect on the objective. For menus with a single ray, one can also impose $\sum_{i \notin N(P)} q_i(\zeta) \geq \varepsilon$. These bounds restrict the search and should not be treated as part of

the economic model. Alternatively, replacing each ratio by its log odds,

$$z_i(\zeta) := \log \frac{t_i(\zeta)}{1 - t_i(\zeta)},$$

preserves linearity of the polarization constraints: $z_i(\zeta') + z_k(\zeta) \geq 0$. The corresponding positive coordinate in (14) is $d_i \exp(z_i(\zeta))$. This form can be convenient when the relevant trade ratios vary substantially in magnitude. In either parametrization, the reconstructed directions can be rescaled before evaluating the objective without changing the menu of rays.

Remark 3. The closure of $\mathcal{D}(P)$ is a polyhedron. We could alternatively describe it using its vertices and recession directions (the latter are needed when $N(P) \neq \mathcal{I}$, since the q coordinates are unbounded). However, this representation can be much larger than the description by linear constraints. For example, suppose that $n = I$ and each negative support is a singleton. The d coordinates are then fixed, there are no q coordinates, and each pair of rays contributes two ratio variables whose feasible set in the closure is the triangle

$$\{(x, y) \in [0, 1]^2 : x + y \geq 1\}.$$

These pairs involve separate variables, so the closure of $\mathcal{D}(P)$ is a product of $I(I - 1)/2$ triangles. It therefore has $3^{I(I-1)/2}$ vertices, although only $I(I - 1)/2$ polarization inequalities are required, in addition to the coordinate bounds. With five tasks this gives 59,049 vertices and ten polarization inequalities; with ten tasks it gives approximately 2.95×10^{21} vertices and 45 polarization inequalities. Moreover, the objective need not be convex in the parameters, so there is no reason to expect in general that it will be maximized at a vertex. We therefore work directly with the representation by linear constraints.

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. (3) with

$$\mathbb{E} \left[\sum_{j \in \mathcal{J}} m_j(i|\mathbf{c}) \right] = n^i \quad \forall i \in \mathcal{I}. \quad (16)$$

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 (16) as the large-market market-clearing (LM-MC) constraint.

Fix the status quo $\sigma \in \mathcal{Z}$. Let A_j be a collection of veto sets for each j , and let $s_j : C_j \mapsto A_j$ be a greedy strategy for j . Let $\{\Gamma_j : A_j \rightarrow \mathbb{R}_{\geq 0}^I\}_{j \in \mathcal{J}}$ be a collection of *individual-level selection rules* for the designer, where we require that $\Gamma_j(a) \in a$, 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

$$\sum_{j \in \mathcal{J}} \mathbb{E}_j[\Gamma_j(s_j(c_j))] = (n^1, \dots, n^I).$$

Conversely, if we replace the individual-level selection 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 relaxation of the original program of choosing the optimal static OSP mechanism:

$$\min_{(A_j, \Gamma_j)_{j \in \mathcal{J}}} \sum_{j=1}^J \pi_j \cdot \mathbb{E}[\Gamma_j(s_j(c_j))] \quad (17)$$

s.t. either j 's veto sets are polarized rays, or j receives

two veto sets: $\{\sigma_j\}$ and $\{x \in \mathbb{R}_{\geq 0}^I : x \leq \sigma_j\}$; and (OSP)

$$\sum_{j \in \mathcal{J}} \mathbb{E}_j[\Gamma_j(s_j(c_j))] = (n^1, \dots, n^I). \quad \text{(LM-MC)}$$

where s_j is a greedy-cautious strategy for j (which is unique up to zero-measure perturbations if F_j is smooth, i.e. assigns zero mass to any hyperplane). 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 econo-

mize on notation we do not explicitly model randomization in program (17), as it can be represented by redefining $\mathcal{J}$.³⁵ We refer to the program in (17) (allowing for randomization) as the *large-market OSP program*.

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

$$\mathcal{F}_j := \text{cvh} \left\{ \mathbb{E}_j[\Gamma_j(s_j(c_j))] \in \mathbb{R}_{\geq 0}^I : \Gamma_j(a) \in a \ \forall a \in A_j ; \ s_j \text{ is greedy; and either } A_j \right. \\ \left. \text{consists of polarized rays, or } j \text{ receives two veto sets: } \{\sigma_j\} \text{ and } \{x \in \mathbb{R}_{\geq 0}^I : x \leq \sigma_j\} \right\}.$$

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 (17) aside from (LM-MC). Then the large-market OSP program has the same value as

$$\begin{aligned} \min_{\{x_j\}_{j \in \mathcal{J}}} \quad & \sum_{j \in \mathcal{J}} \pi_j \cdot x_j \\ \text{s.t.} \quad & x_j \in \mathcal{F}_j \quad \forall j \in \mathcal{J} \\ & \sum_{j \in \mathcal{J}} x_j(i) = n^i \quad \forall i \in \mathcal{I} \end{aligned} \tag{18}$$

(LM-MC)

We refer to eq. (18) as the *outer program*. It is instructive to study the dual of this program, which can be written as

$$\sup_{\lambda \in \mathbb{R}^I} \sum_{i \in \mathcal{I}} \lambda^i n^i - \sum_{j \in \mathcal{J}} S_j(\lambda - \pi_j) \tag{19}$$

where $S_j(w) := \max\{w \cdot y : y \in \mathcal{F}_j\}$ is the support function of the convex set $\mathcal{F}_j$. Under strong duality the values of the outer program in eq. (18) and its dual in eq. (19) coincide.

Proposition 3. If $\mathcal{F}_j$ is compact for every j then strong duality for the outer program holds. Moreover, if λ_* is a solution to the program in eq. (19) then there exist selections

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

³⁵Technically, this notational trick is only valid if the randomization involves rational weights, but this should not cause confusion.

such that $\{x_j\}_{j \in \mathcal{J}}$ satisfy market clearing, and these solve the program in eq. (18).

Proof. Proof in Section A.11. □

The condition that $\mathcal{F}_j$ is compact 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 4. 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), which implies that 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. (19) 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

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

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 use the change of variables from Proposition 2 to write $S_j(w)$ as a maximization subject to a relatively small set of linear constraints. 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 simplifies the challenging polarization constraint. Still, there are two problems with this technique in practice. First, when I is large there are many classes (Lemma 6). Second, the objective is a complicated non-linear function of the menu. Nonetheless, the change of variables 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, denoted by m^* . For each finite N -replica economy, we implement the mechanism m^* by offering each agent their menu from m^* , and having the designer choose an ex-post optimal selection rule. 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 menu in the support of the randomized mechanism to a proportional fraction of the N group- j agents.³⁶ Let $V_{LM}^*(N)$ be the random variable corresponding to the social cost under the resulting EOPR 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 $\frac{1}{N}(\mathbb{E}[V_{LM}^*(N)] - \mathbb{E}[V^*(N)]) \rightarrow 0$ as $N \rightarrow \infty$.

Proposition 4. The optimal mechanism in the large-market program is asymptotically optimal in the original program as the market grows.

³⁶That is, if under m^* agent j has actions A_j with probability ε , then $\lfloor \varepsilon N \rfloor$ of the group- j agents receive this action set (any residual agents can be allocated to any mechanism in the support of m^* , as this will not affect the asymptotic results).

Proof. Proof in Section A.12. □

5.5 Optimization procedure

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

1. *Estimate the support functions of the OSP-feasible sets for each agent.* This is done by maximizing linear objectives 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 2 to simplify the polarization constraints, 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 2 to optimize over menus of polarized rays) can be used to identify a mechanism.

Details of the numerical implementation used in the empirical applications are provided in Appendix B.6.

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 caseloads 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 systems 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 status-quo 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

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

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

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

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

where $g(k) \in \mathcal{I}$ denotes the type of case k and $(w_{\text{FN}}, w_{\text{FP}}, w_{\text{TN}}, w_{\text{TP}})$ 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. A central empirical challenge in both applications is that $\pi_j(i)$ is not non-parametrically 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 down-

stream 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 and scrutiny and to impose greater emotional burdens, 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. Because each agent-by-type moment is estimated from a limited number of cases, we take an empirical-Bayes perspective and report the expected welfare change of each mechanism given the data. Since the welfare change from any allocation is linear in the performance vector, this expectation equals the welfare change evaluated at the posterior means of the performance estimates, which we obtain through empirical Bayes shrinkage. We use the same posterior means to construct each candidate mechanism and select among the computed candidates according to their posterior expected welfare gains, averaged over simulated preference profiles. We also 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 thus evaluate the mechanisms under calibrated preference distributions. In the CPS application, the distribution is informed by a survey of CPS investigators which elicited their relative willingness to handle high- versus low-risk cases in Baron et al. (2026). We do not have comparable survey evidence in the prosecution application, so we use a stylized preference distribution with increasing expected costs across risk quartiles – which reflects the possibility that higher-risk cases could generate greater private costs to prosecutors. We provide more details regarding the calibration of the preference distributions 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 $c_j = (c_{j1}, \dots, c_{j4})$. 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 apply the procedure of Section 5.5 to identify a static OSP mechanism, using the calibrated preference distribution and estimated performance posterior described above. We then simulate the mechanism by repeatedly drawing preference profiles and report the average gains relative to the status quo. We then 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 expected gains from the alternative 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. All changes are reported in levels. Changes in the weighted social cost are measured in units of the planner’s objective. Percent changes relative to the status quo are reported in brackets for false negatives and placements, whose baseline levels are directly identified.³⁷ See Appendix B.7 for details on the computation of welfare changes. Parentheses report posterior standard deviations of the average gains.

³⁷Because downstream outcomes are observed only for untreated cases, baseline levels of false positives—and hence of the weighted social cost—are not directly identified. Recovering them requires extrapolating outcomes toward a hypothetical agent with a zero treatment rate, as in Arnold et al. (2022) and Baron et al. (2026). While such an extrapolation is plausible in the CPS setting, where placement rates are low, it is not credible in the prosecution setting, where roughly 80% of cases are prosecuted. For comparability across the two applications, we therefore report percent changes only for directly identified outcomes.

The selected mechanisms generate meaningful improvements relative to the status quo. The small-market mechanism in Panel A uses three task groups: it pools quartiles $\{1, 4\}$ and keeps quartiles $\{2\}$ and $\{3\}$ separate. It reduces the weighted social cost by approximately 1,380, while the corresponding large-market approximation reduces it by approximately 2,733. Both mechanisms reduce false negatives and false positives. The small-market mechanism reduces false negatives by 985 cases (3.7% of the status-quo level) and false positives by 1,155 cases, while the large-market approximation reduces false negatives by roughly 1,985 cases (7.5%) and false positives by 2,285 cases. Both mechanisms also slightly reduce overall foster care placements, by 3.3% and 5.8%, respectively. The mechanisms further recover a meaningful share of the gains achievable under full information. The full-information benchmark reduces the weighted social cost by roughly 7,355 and false negatives by nearly 20%, implying that the small-market mechanism captures roughly 20%, and the large-market approximation nearly 40%, 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 increasing investigator workloads.

Panel B reports the corresponding results for the misdemeanor prosecution application. The selected mechanism keeps all four risk quartiles separate. The qualitative patterns are similar. It reduces the weighted social cost by approximately 1,900 in the small market and 2,435 in the large-market approximation. Relative to the status-quo rotational assignment system, both mechanisms reduce false negatives (by 21.5% and 27.7%, respectively), false positives, and overall prosecutions, with the large-market approximation again generating larger gains than the finite-market implementation. 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, which reduces the weighted social cost by roughly 4,526, the small-market mechanism captures 42.0% of the attainable reduction and the large-market approximation captures 53.8%.

Altogether, the two applications suggest that OSP mechanisms can deliver meaningful gains while preserving the participation guarantees provided by the status quo.

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	-1,380.1 (33.8)	-2,732.8 (69.0)	-7,355.5 (158.6)
False negatives	-984.5 (24.7) [-3.7%]	-1,984.8 (50.7) [-7.5%]	-5,276.8 (117.1) [-19.9%]
False positives	-1,154.6 (28.5)	-2,284.9 (58.2)	-6,187.2 (133.4)
Placements	-170.0 (14.4) [-3.3%]	-300.1 (28.5) [-5.8%]	-910.4 (64.1) [-17.5%]
Panel B. Misdemeanor Prosecution			
Weighted social cost	-1,900.4 (109.2)	-2,434.8 (144.0)	-4,525.7 (256.9)
False negatives	-306.5 (19.4) [-21.5%]	-395.8 (25.5) [-27.7%]	-760.8 (45.9) [-53.2%]
False positives	-686.0 (52.6)	-861.9 (69.5)	-1,556.2 (126.0)
Prosecutions	-379.4 (48.9) [-1.5%]	-466.1 (64.6) [-1.9%]	-795.4 (117.3) [-3.2%]

Notes. This table reports estimated gains 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. Empirical Bayes moments are used to construct, select, and evaluate the mechanisms. The selected partition uses the task groups $\{1, 4\}$, $\{2\}$, and $\{3\}$ in Panel A, and uses four separate quartiles in Panel B. The full-information benchmark uses four separate quartiles in both applications. False-positive changes follow $\Delta FP = \Delta P + \Delta FN$. Because the social-cost, treatment, and false-negative moments are shrunk separately, the change in weighted social cost does not exactly equal $\Delta FP + w_{FN}\Delta FN$ computed from the reported outcome changes. Percent changes relative to the status quo are reported in brackets only for false negatives and placements or prosecutions; baseline false-positive and weighted social-cost levels are not required for the reported level changes. Parentheses report posterior standard deviations of the reported average changes. In Columns 1 and 3, allocations are averaged over 1,000 preference profiles before calculating posterior uncertainty. Dispersion across preference profiles is reported separately in Figures C.2 and C.3.

7 Conclusion

We characterize obviously strategy-proof mechanisms in an important class of multidimensional screening problems. We then turn to the more normative market-design question of which mechanism a designer should use. We argue that the class of OSP mechanisms is fairly flexible; for one, under mild conditions this class enables the designer to improve upon the status quo whenever it is possible to do so. Moreover, these mechanisms possess desirable robustness properties. We then exploit the tractable structure of OSP mechanisms to develop tools for guiding the designer’s choice within this set. Finally, we apply these tools to identify promising mechanisms in a pair of high-stakes real-world applications.

A number of interesting questions remain unanswered. For instance, how does the characterization of OSP mechanisms change when we allow for transfers? The characterization is likely to change more in settings where tasks are “goods” (as discussed in Section 1.1). Another question is how we can identify polarized-ray mechanisms that perform well according to the designer’s objective. The dual formulation of the large-market program simplifies this problem, as discussed in Section 5.2. However, this approach requires characterizing the support function S_j , which entails solving a difficult optimization problem in which the choice domain consists of menus of polarized rays. Our characterization of the convex structure of polarized rays eases the computational burden somewhat, but further work in this area would be valuable.

Appendix

A Omitted proofs

A.1 Proof of Lemma 1

Proof. First, suppose j ’s strategy does not always make greedy choices. 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 $x' \in \mathcal{O}_j(s_j(c_j))$. But then $s_j(c_j)$ does not obviously dominate a . Now, suppose that there are multiple greedy choices for c_j .

Let $s_j(c_j) \neq a'' \in a_j^g(c_j)$. In order for $s_j(c_j)$ to obviously dominate a'' , it must be that

$$c_j \cdot x' \leq \min_{a \in A_j} \{ \inf \{ c_j \cdot x : x \in \mathcal{O}_j(a) \} \} \quad \forall x' \in \mathcal{O}_j(s_j(c_j)),$$

and by definition this inequality cannot be strict. Because the set of payoffs associated with $s_j(c_j)$ is a singleton, it is also greedy-cautious. Then any other greedy-cautious action has the same degenerate set of possible payoffs, and so would also be obviously dominant. $\square$

A.2 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 $a_j^0 \in A_j$ 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 5. Assume a rich preference domain. If type-contingent strategy s_j is obviously dominant for j then one of the following holds

1. There is at most one $a \in A_j$ that is remedial and has $\mathcal{O}_j(a) \neq \{\sigma_j\}$. Every other action is redundant.
2. There is a subset of actions $\hat{A}_j \subset A_j$ such that $\mathcal{O}_j(a)$ is contained in a ray originating at σ_j for each $a \in \hat{A}_j$, and the collection of these rays is polarized. 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 $a \in A_j$. $\square$

Claim 2. For any $c_j \in \mathbb{R}_{>0}^I$ 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 . $\square$

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

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 $c_j \in \mathbb{R}_{\geq 0}^I$ 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 . $\square$

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 $y'(x' - \sigma_j) + y''(x'' - \sigma_j) \in \mathbb{R}_{\geq 0}^I$.

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 $c'_j \in \mathbb{R}_{>0}^I$ 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. $\square$

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 $c_j \in \mathbb{R}_{>0}^I$ and $x' \in \mathcal{O}_j(a')$ such that $c_j \cdot x' < c_j \cdot \sigma_j$. Then for any $x \in \mathcal{O}_j(a)$, $x \neq \sigma_j$ we can choose c_j so that $c_j \cdot x < c_j \cdot \sigma_j$ also holds. But this violates Claim 2. $\square$

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 5 $\square$

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 5, Claim 5) $c_j \cdot \kappa(a) < 0$ implies $c_j \cdot \kappa(a') \geq 0$ for all $a' \neq a$. Thus choosing a is obviously dominant for type c_j . If there is no a such that $c_j \cdot \kappa(a) < 0$ then any greedy-cautious choice guarantees a payoff of $c_j \cdot \sigma_j$.

A.3 Proof of Lemma 3

Proof. Suppose that $M \in \mathcal{M}_j$ is reachable by $s_j(c_j)$. If $s_j(c_j)$ is not greedy at M then there exists $a \neq s_j(c_j)(M)$ and $x \in \mathcal{O}_j^*(M, a)$ such that $c_j \cdot x < c_j \cdot x'$ for all $x' \in \mathcal{O}_j^*(M, s_j(c_j)(M))$, and a fortiori for all $x' \in \mathcal{O}_j(M, s_j(c_j))$. Because $x \in \mathcal{O}_j^*(M, a)$, there exists $s'_j \in \alpha_j^{-1}(M, s_j(c_j))$ such that $x \in \mathcal{O}_j(M, s'_j)$, violating obvious dominance of $s_j(c_j)$.

Now suppose there are multiple greedy choices for c_j at M . Let $s_j(c_j)(M) \neq a'' \in a_j^g(c_j, M)$. For $s_j(c_j)$ to obviously dominate a deviation taking a'' at M , it must be that

$$c_j \cdot x' \leq \min_{a \in A(M)} \{\inf\{c_j \cdot x : x \in \mathcal{O}_j^*(M, a)\}\} \quad \forall x' \in \mathcal{O}_j(M, s_j(c_j))$$

and by definition this inequality cannot be strict. Because the set of possible payoffs from following $s_j(c_j)$ is a singleton, it is also greedy-cautious. $\square$

A.4 Proof of Theorem 2

Proof. Fix an OSP mechanism $(\mathcal{G}, s_{\mathcal{J}})$. We begin with some initial claims for any $j \in \mathcal{J}$ and $M \in \mathcal{M}_j$.

Claim 1. For any $c_j \in \mathbb{R}_{>0}^I$ it must be that $c_j \cdot x \geq c_j \cdot \sigma_j$ for all $s'_j \in \alpha_j^{-1}(M, s_j(c_j))$ and $x \in \mathcal{O}_j(M, s'_j)$.

Proof of Claim 1. Suppose there exists $s'_j \in \alpha_j^{-1}(M, s_j(c_j))$ and $x \in \mathcal{O}_j(M, s'_j)$ such that $c_j \cdot x < c_j \cdot \sigma_j$. Because the mechanism respects the outside options, $\sigma_j \in \mathcal{O}_j(M, s_j(c_j))$. So x is possible following s'_j and σ_j following $s_j(c_j)$, violating obvious dominance of $s_j(c_j)$. $\square$

Claim 2. For every $c_j \in \mathbb{R}_{>0}^I$ such that M is reachable by $s_j(c_j)$, if there exists $a \in A(M)$ and $x \in \mathcal{O}_j^*(M, a)$ such that $c_j \cdot x < c_j \cdot \sigma_j$, then $s_j(c_j)(M) = a$.

Proof of Claim 2. That such an a exists means there exists s'_j such that $s'_j(M) = a$ and $x \in \mathcal{O}_j(M, s'_j)$ such that $c_j \cdot x < c_j \cdot \sigma_j$. We can assume moreover that s'_j reaches M , by taking $s'_j = s_j$ at any earlier information sets. By Claim 1, if $a = s'_j(M) \neq s_j(c_j)(M)$ then $c_j \cdot x \geq c_j \cdot \sigma_j$ for all $x \in \mathcal{O}_j(M, s'_j)$, which is a contradiction. $\square$

Claim 3. For any $c_j \in \mathbb{R}_{>0}^I$, if there exists $a \in A(M)$ and $x \in \mathcal{O}_j^*(M, a)$ such that $c_j \cdot x < c_j \cdot \sigma_j$ then M is reachable by $s_j(c_j)$ and $s_j(c_j)(M) = a$.

Proof of Claim 3. We prove the claim by induction. Take any history $h = (a_1, \dots, a_K) \in M$ (by perfect recall, it doesn't matter which). Let $0 \leq t_1 < \dots < t_L = K$ be the indices at which agent j moves along h , so $P(h^{t_\ell}) = j$ for $1 \leq \ell \leq L$. Then $M_j(h^{t_\ell})$ is the ℓ^{th} information set at which j acts along this path. For $\ell < L$ let $a^{t_\ell} \in A(h^{t_\ell})$ be the action such that $(h^{t_\ell}, a^{t_\ell}) \preceq h$. Now observe that $M_j(h^{t_1})$ is reachable by $s_j(c_j)$ for all $c_j \in \mathbb{R}_{>0}^I$. Moreover, $x \in \mathcal{O}_j^*(M_j(h^{t_1}), a^{t_1})$, so by Claim 2 we have $s_j(c_j)(M_j(h^{t_1})) = a^{t_1}$. This establishes the base case. Now suppose that for all $\ell < T \leq L$ we have that $M_j(h^{t_\ell})$ is reachable by $s_j(c_j)$ and $s_j(c_j)(M_j(h^{t_\ell})) = a^{t_\ell}$. That $s_j(c_j)(M_j(h^{t_\ell})) = a^{t_\ell}$ implies h^{t_T} is reachable by $s_j(c_j)$, and by Claim 2 we have $s_j(c_j)(M_j(h^{t_T})) = a^{t_T}$. This establishes the induction step. $\square$

By the pruning principle (Li 2017, Proposition 2) we can assume without loss of generality

that the game $\mathcal{G}$ is pruned with respect to the strategy profile $s_{\mathcal{J}}$, meaning that all histories are reached under $s_{\mathcal{J}}$ for some type profile. Pruning in this way preserves each agent's obviously dominant strategies. Note that after pruning the game may no longer respect the agents' outside options; we deal with this issue below.

Claim 4. Let $\mathcal{G}$ be pruned with respect to $s_{\mathcal{J}}$. If there exists $a \in A(M)$ and $x \in \mathcal{O}_j^*(M, a)$ such that $\sigma_j \neq x \leq \sigma_j$ then M is reachable by $s_j(c_j)$ for all $c_j \in \mathbb{R}_{>0}^I$, and $A(M) = \{a\}$.

Proof of Claim 4. If $x \leq \sigma_j$ and $x \neq \sigma_j$ then $c_j \cdot x < c_j \cdot \sigma_j$ for all $c_j \in \mathbb{R}_{>0}^I$. Then by Claim 3, M is reachable by $s_j(c_j)$ and $s_j(c_j)(M) = a$ for all c_j . Pruning of the game with respect to $s_{\mathcal{J}}$ thus yields $A(M) = \{a\}$. $\square$

Claim 5. Let $\mathcal{G}$ be pruned with respect to $s_{\mathcal{J}}$. There is no $M \in \mathcal{M}_j$, $a \in A(M)$, and $x \in \mathcal{O}_j^*(M, a)$ such that $\sigma_j \neq x \geq \sigma_j$.

Proof of Claim 5. Suppose such M, a , and x exist. Because the game is pruned, there must exist $c_j \in \mathbb{R}_{>0}^I$ and $M^0 \in \mathcal{M}_j^0$ such that $x \in \mathcal{O}_j(M^0, s_j(c_j))$. Because the mechanism respects the outside option, in the original un-pruned game there is also a strategy $s'_j \neq s_j(c_j)$ such that $\mathcal{O}_j(M^0, s'_j) = \{\sigma_j\}$. But then there exists $M^0 \preceq M' \preceq M$ such that $M' \in \alpha_j(s_j(c_j), s'_j)$ and $x \in \mathcal{O}_j(M', s_j(c_j))$. Thus $s_j(c_j)$ is not obviously dominant for c_j . $\square$

We now prove an intermediary result.

Lemma 7. Let $\mathcal{G}$ be pruned with respect to $s_{\mathcal{J}}$. For any $M \in \mathcal{M}_j$, if $x' \in \mathcal{O}_j^*(M, a')$ and $x'' \in \mathcal{O}_j^*(M, a'')$ for $a' \neq a''$ then x' and x'' are polarized.

Proof. Because $|A(M)| \geq 2$ by assumption, Claims 4 and 5 imply that if $x \leq \sigma_j$ or $x \geq \sigma_j$ then $x = \sigma_j$ for either $x \in \{x', x''\}$. If either x' or x'' equals σ_j then x', x'' are trivially polarized, so assume that $x \not\geq \sigma_j$ and $x \not\leq \sigma_j$ for $x \in \{x', x''\}$.

Suppose there exists $c_j \in \mathbb{R}_{\geq 0}^I$ such that $c_j \cdot x' \leq c_j \cdot \sigma_j$ and $c_j \cdot x'' < c_j \cdot \sigma_j$. If this is the case then, because $x' - \sigma_j \not\geq 0$, we can find $c'_j \in \mathbb{R}_{>0}^I$ such that $c'_j \cdot x' < c'_j \cdot \sigma_j$ and $c'_j \cdot x'' < c'_j \cdot \sigma_j$. By Claim 3, M is reachable by $s_j(c'_j)$, $s_j(c'_j)(M) = a'$, and $s_j(c'_j)(M) = a''$, which violates $a' \neq a''$. Thus such a c_j cannot exist. 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 polarization of x', x'' . $\square$

Lemma 8. For any $M \in \mathcal{M}_j$ and $a \in A(M)$, there cannot exist $x', x'' \in \bar{\mathcal{O}}_j^*(M, a)$ and $c_j \in \mathbb{R}_{>0}^I$ such that $c_j \cdot x' < c_j \cdot \sigma_j < c_j \cdot x''$.

Proof. By Claim 3, M is reachable by $s_j(c_j)$ and $s_j(c_j)(M) = a$. Because M is reachable by $s_j(c_j)$, there exists $s_{-j} \in \Sigma_{-j}$, $d_0 \in D_0$, and $h^0 \in M^0 \in \mathcal{M}_j^0$ such that $\zeta(h^0; s_j(c_j), s_{-j}, d_0) \succeq M$. Because the game respects the outside options, there exists s'_j such that $\mathcal{O}_j(M^0, s'_j) = \{\sigma_j\}$. Then there exists $M^0 \preceq M' \preceq M$ such that $M' \in \alpha_j(s_j(c_j), s'_j)$ and $x'' \in \mathcal{O}_j(M', s_j(c_j))$, so $s_j(c_j)$ is not obviously dominant for c_j . $\square$

Lemma 9. Let $\mathcal{G}$ be pruned with respect to $s_{\mathcal{J}}$. Then exactly one of the following holds

1. $A(M) = \{a\}$ and there exists $x \in \mathcal{O}_j^*(M, a)$ with $\sigma_j \neq x \leq \sigma_j$.
2. There does not exist $a \in A(M)$ and $x \in \mathcal{O}_j^*(M, a)$ with $\sigma_j \neq x \leq \sigma_j$; and for every $a \in A(M)$, the set $\bar{\mathcal{O}}_j^*(M, a)$ is contained in a non-monotone ray originating at σ_j , i.e. $\bar{\mathcal{O}}_j^*(M, a) \subset \{x \in \mathbb{R}^I : (x - \sigma_j) = y\kappa, y \in \mathbb{R}_{\geq 0}\}$ for some $\kappa \in \mathbb{R}^I$ with $0 \not\preceq \kappa \not\preceq 0$.

Proof. If there exists $a \in A(M)$ and $x \in \mathcal{O}_j^*(M, a)$ with $\sigma_j \neq x \leq \sigma_j$ then by Claim 4 $A(M) = \{a\}$. Otherwise, by Claims 4 and 5, if $x \in \bar{\mathcal{O}}_j^*(M, a)$ and $x \leq \sigma_j$ or $x \geq \sigma_j$ then $x = \sigma_j$, in which case we are done. Assume then that $x \not\preceq \sigma_j$ and $x \not\preceq \sigma_j$. Given Lemma 8, the proof is the same as that of Claim 4 in the proof of Theorem 1. $\square$

In summary, we have that in the game that is pruned with respect to $s_{\mathcal{J}}$, exactly one of the following holds for each $M \in \mathcal{M}_j$

1. There exists $a \in A(M)$ and $x \in \mathcal{O}_j^*(M, a)$ such that $\sigma_j \neq x \leq \sigma_j$, in which case $A(M) = \{a\}$.
2. For every $a \in A(M)$ and $x \in \mathcal{O}_j^*(M, a)$ either $x = \sigma_j$ or $x \not\preceq \sigma_j \not\preceq x$, and $\bar{\mathcal{O}}_j^*(M, a)$ is contained in a non-monotone ray originating at σ_j . Moreover, for any $a', a'' \in A(M)$, the sets $\mathcal{O}_j^*(M, a')$ and $\mathcal{O}_j^*(M, a'')$ are polarized.

Polarization implies that $\mathcal{O}_j^*(M, a') \neq \mathcal{O}_j^*(M, a'')$ for any distinct $a', a'' \in A(M)$ such that $\mathcal{O}_j^*(M, a') \neq \{\sigma_j\}$. Thus we can uniquely identify any action $a \in A(M)$ such that

$\mathcal{O}_j^*(M, a) \neq \{\sigma_j\}$ with the set $\mathcal{O}_j^*(M, a)$. Any other action is labeled as a withdrawal action. Moreover, if $x \in \mathcal{O}_j^*(M, a)$, polarization implies that the mechanism would also be OSP if the entire ray $\{\sigma_j + (x - \sigma_j)y : y \in \mathbb{R}_{\geq 0}\}$ were also in $\mathcal{O}_j^*(M, a)$. Thus the mechanism is equivalent to an eRPR mechanism. Finally, observe that adding an ex-post veto round does not change the outcome. To see this, note that if terminal history $\bar{h}$ is reachable by strategy $s_j(c_j)$ in the original game, then OSP and respect for the outside option imply $c_j \cdot \Gamma_j(\bar{h}) \leq c_j \cdot \sigma_j$. So in the augmented game it is greedy and obviously dominant to select the ray $\{\sigma_j + (\Gamma_j(\bar{h}) - \sigma_j)y : y \in \mathbb{R}_{\geq 0}\}$ in the veto round following provisionally terminal history $\bar{h}$. Moreover, as observed in cases (1) and (2) above, for any $M \in \mathcal{M}_j$ there is at most one action a such that there exists $x \in \mathcal{O}_j^*(M, a)$ with $c_j \cdot x < c_j \cdot \sigma_j$, and if such an a exists then $s_j(c_j)(M) = a$. Thus $s_j(c_j)$ remains greedy-cautious and obviously dominant at every M in the original non-augmented game.

Conversely, let $(\mathcal{G}, s_{\mathcal{J}})$ be an eRPR mechanism with ex-post vetoes. Observe that in the ex-post veto round, the specified strategy ensures that $c_j \cdot \Gamma_j(\zeta(\bar{h}, s_j(c_j), s_{-j}, d_0)) \leq c_j \cdot \sigma_j$ for any $s_{-j} \in \Sigma_{-j}$, $d_0 \in D_0$, and any provisionally terminal history $\bar{h}$. Take any $c_j \in \mathbb{R}_{>0}^I$ and $M \in \mathcal{M}_j$ that is reachable by $s_j(c_j)$. Suppose first that there exists $a \in A(M)$ and $x \in \mathcal{O}_j^*(M, a)$ such that $c_j \cdot x < c_j \cdot \sigma_j$. Polarization implies that there is at most one such action in $A(M)$ (this is precisely the application of Farkas's Lemma used to prove Lemma 7), so $s_j(c_j)(M) = a$. Moreover, because of the ex-post veto round, $c_j \cdot x \leq c_j \cdot \sigma_j$ for all $x \in \mathcal{O}_j(M, s_j(c_j))$. Thus $s_j(c_j)$ is obviously dominant at M . Suppose instead that $c_j \cdot x \geq c_j \cdot \sigma_j$ for all $x \in \mathcal{O}_j^*(M, a)$ and $a \in A(M)$. The ex-post veto round implies $c_j \cdot x \leq c_j \cdot \sigma_j$ for all $x \in \mathcal{O}_j(M, s_j(c_j))$, so $c_j \cdot x = c_j \cdot \sigma_j$ for all $x \in \mathcal{O}_j(M, s_j(c_j))$. Thus $s_j(c_j)$ is also obviously dominant at M . $\square$

A.5 Proof of Theorem 3

Proof. That every sequential extended polarized-ray mechanism is SOSP follows from the same argument used to show that every extended polarized-ray mechanism is OSP. Conversely, take an SOSP mechanism $(\mathcal{G}, s_{\mathcal{J}})$. The mechanism is also, a fortiori, OSP. Fix an agent j , and let $M^0 \in \mathcal{M}_j^0$ be an information set at which j moves for the first time along some path of play. Then SOSP implies that for any $c_j \in C_j$ and $a \in A(M^0) \setminus \{s_j(c_j)(M^0)\}$

we have

$$\max_{\mu_j \in \mathcal{O}_j^*(M^0, s_j(c_j)(M^0))} \mu_j \cdot c_j \leq \min_{\mu_j \in \mathcal{O}_j^*(M^0, a)} \mu_j \cdot c_j. \quad (20)$$

Because the mechanism must strongly respect the outside option, we have $\sigma_j \in \mathcal{O}_j^*(M^0, a)$ for all $a \in A(M^0)$ and $\mathcal{O}_j^*(M^0, a) = \{\sigma_j\}$ for some $a \in A(M^0)$. Then the condition in (20) is equivalent to the requirement that action $s_j(c_j)(M^0)$ be obviously dominant in a simultaneous move game in which j has action set $A(M^0)$ and possible assignments $\mathcal{O}_j^*(M^0, \cdot)$. By Proposition 5, exactly one of the following holds

1. There is at most one $a \in A(M^0)$ such that $\mathcal{O}_j^*(M^0, a) \subset \{x \in \mathbb{R}_{\geq 0}^I : x \leq \sigma_j\}$ and $\mathcal{O}_j^*(M^0, a) \neq \{\sigma_j\}$. Every other action, a' , is redundant, i.e. $x \geq \sigma_j$ for all $x \in \mathcal{O}_j^*(M^0, a')$.
2. There is a subset of actions $\hat{A} \subset A(M^0)$ such that $(\mathcal{O}_j^*(M^0, a))_{a \in \hat{A}}$ consists of a collection of polarized rays. Every action not in $\hat{A}$ is redundant.

We want to show that there is an equivalent mechanism in which j either does not move at M^0 , or does not move again following the choice at M^0 , unless it is to send a withdrawal message.

Let $\mathcal{G}$ be pruned with respect to $s_{\mathcal{J}}$. Suppose first that case (1) holds. If there is no action $a \in A(M^0)$ and $x \in \mathcal{O}_j^*(M^0, a)$ with $\sigma_j \neq x \leq \sigma_j$ then every action is redundant. No redundant action a' such that $\mathcal{O}_j^*(M^0, a') \neq \{\sigma_j\}$ is ever chosen by any type, which means that every action in $A(M^0)$ (in the pruned game) is a withdrawal action. Consider a subsequent information set $M^0 \preceq M \in \mathcal{M}_j$. If every action in $A(M^0)$ is a withdrawal action then so is every action in $A(M)$. Then we can replace a sequence of withdrawal actions with a single withdrawal action in M^0 , so there is no need to have j choose again. If on the other hand there is some action $a \in A(M^0)$ such that there exists $x \in \mathcal{O}_j^*(M^0, a)$ with $\sigma_j \neq x \leq \sigma_j$ then $A(M^0) = \{a\}$ by Lemma 9. In this case we can construct an equivalent mechanism in which j does not move at M^0 .

Now suppose that case (2) holds. Again, any redundant action $a \notin \hat{A}$ such that $\mathcal{O}_j^*(M^0, a) \neq \{\sigma_j\}$ is dropped from the pruned mechanism, and every other action is a withdrawal action. As in the previous case, there is no need to have an agent move again once they take a withdrawal action.

For each $a \in \hat{A}$ let $\kappa(a)$ be such that $\mathcal{O}_j^*(M^0, a) \subset \{\sigma_j + \kappa(a)y : y \in \mathbb{R}_{\geq 0}\}$. The set $\{c_j \in \mathbb{R}_{>0}^I : c_j \cdot \kappa(a) < 0\}$ is non-empty for every $a \in \hat{A}$ because $\kappa(a)$ is non-monotone by definition, and $s_j(c_j)(M^0) = a$ for all $c_j \in \{c_j \in \mathbb{R}_{>0}^I : c_j \cdot \kappa(a) < 0\}$ because the rays are polarized. Consider a subsequent information set $(h, a) \preceq M \in \mathcal{M}_j$ for $h \in M^0$ and $a \in \hat{A}$. Then because $\mathcal{O}_j^*(M, a') \subset \{\sigma_j + \kappa(a)y : y \in \mathbb{R}_{\geq 0}\}$ for any $a' \in A(M)$, in order for there to exist a strongly obviously dominant strategy for any type c_j with $c_j \cdot \kappa(a) < 0$, there can be at most one non-withdrawal action. $\square$

A.6 Proof of Proposition 1

To understand Proposition 1, 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, $[\underline{c}_j^i, \bar{c}_j^i]$. Define $\hat{c}^i := \max_{j \in \mathcal{J}} \underline{c}_j^i$, so $\hat{c}^i \in [\underline{c}_j^i, \bar{c}_j^i]$ for all i, j if and only if $\cap_{j \in \mathcal{J}} C_j \neq \emptyset$.

Lemma 10. Assume that C_j is a hypercube for all j , and $\cap_{j \in \mathcal{J}} C_j \neq \emptyset$. If an allocation $m \in \mathcal{Z}$ (i.e. a non-choice-based mechanism) satisfies (SQ) and (MC) then

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

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

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

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

$$\begin{aligned} 0 &\geq \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} (m_j(i) - \beta_j \sigma_j(i)) c_j(i) \\ &= \sum_{i \in \mathcal{I}} \hat{c}^i \sum_{j \in \mathcal{J}} (m_j(i) - \beta_j \sigma_j(i)) + \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} (m_j(i) - \beta_j \sigma_j(i)) (c_j(i) - \hat{c}^i) \end{aligned} \quad (21)$$

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. (21) is equal to

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

Then for eq. (21) to hold, it must be that

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

for all $\mathbf{c} \in \mathbf{C}$. Under the assumption that $\cap_{j \in \mathcal{J}} C_j \neq \emptyset$ we have $\bar{c}_j^i \geq \hat{c}^i$ for all $i \in \mathcal{I}, j \in \mathcal{J}$.

Then setting $c_j(i) = \tilde{c}^i$ if $(m_j(i) - \beta_j \sigma_j(i)) < 0$ and $c_j(i) = \bar{c}_j^i$ otherwise, we have

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

as desired. $\square$

Corollary 3. 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 10 for $\beta_j = 1$, (SQ) and (MC) imply that for any non-choice-based mechanism $m \in \mathcal{Z}$,

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

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 $\bar{c}_j^i > \hat{c}^i$ for all i, j , so the above inequality is violated. Thus the only deterministic mechanism satisfying (SQ) and (MC) is the status quo. $\square$

Now consider the equivalence between *i.*, *ii.*, and *iii.* in Proposition 1. Underlying this equivalence is a simple class of *bilateral trading mechanisms*, defined by $\gamma \in \mathbb{R}^I$ and two agents j', j'' as follows: agent j' is offered the option to keep their status quo or trade to $\sigma_{j'} + \gamma$, while agent j'' is offered the option to trade from $\sigma_{j''}$ to $\sigma_{j''} - \gamma$ (so both are given two veto sets to choose from). Trade occurs if and only if both j' and j'' wish to do so. All other agents receive their status quo. The strategies are for agent j' (agent j'') to trade if and only if $\gamma \cdot c_{j'} \leq 0$ ($\gamma \cdot c_{j''} \geq 0$). We begin with the following simple observation.

Lemma 11. The bilateral trading mechanism for agents j', j'' with trade γ is feasible and improves upon the status quo 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. $\max \{\gamma \cdot c_{j''} : c_{j''} \in C_{j''}\} > 0$.

(and these conditions are necessary if the type distributions are smooth).

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), and are necessary if the type distributions are smooth (i.e. assign zero mass to any hyperplane). Condition *ii*. implies that when trade does occur, it reduces the social cost. $\square$

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 Proposition 1.

Proposition 6. 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 12. If $\pi_{j'}^i > \pi_{j''}^i$ for some $i \in \mathcal{I}$; $\sigma_{j'}(i), \sigma_{j''}(i) > 0$ 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 11.

Proof. First, because $\sigma_{j'}(i), \sigma_{j''}(i) > 0$ for all i , if we can find γ satisfying Lemma 11 *ii. – iv.* then we can scale it to satisfy Lemma 11 *i.*

Because $\pi_{j'}^i - \pi_{j''}^i > 0$ and $c' > 0$, 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 11 are satisfied. $\square$

If $\pi_{j'} \neq \pi_{j''}$, there exists $i \in \mathcal{I}$ such that $\pi_{j'}^i - \pi_{j''}^i \neq 0$, so without loss of generality assume $\pi_{j'}^i - \pi_{j''}^i > 0$ (otherwise simply reverse the roles of j' and j''). To prove Proposition 6, 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 12.

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 . $\square$

Proposition 6 shows that bilateral trading mechanisms work to improve upon the status quo under mild conditions. The most restrictive condition of Proposition 6 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.7 Proof of Lemma 4

Proof. For $i \in N(\zeta)$, the i th coordinate of $\alpha R(\zeta) + (1 - \alpha)R(\zeta')$ is nonnegative if and only if $\alpha \leq t_i(\zeta')$. For $k \in N(\zeta')$, the corresponding condition is $\alpha \geq 1 - t_k(\zeta)$. All remaining coordinates are nonnegative in both rays. Thus the admissible weights form the interval

$$\left[\max_{k \in N(\zeta')} (1 - t_k(\zeta)), \min_{i \in N(\zeta)} t_i(\zeta') \right].$$

This interval is nonempty exactly when $t_i(\zeta') + t_k(\zeta) \geq 1$. Since every ratio lies strictly between zero and one, any weight in the interval belongs to $(0, 1)$. $\square$

A.8 Proof of Proposition 2

Proof. It is immediately verified that every normalized menu in P determines unique values of (d, t, q) satisfying the stated conditions. Lemma 4 implies that these values satisfy the polarization inequalities. Conversely, (14) assigns to every point of $\mathcal{D}(P)$ normalized directions with exactly the specified negative supports. Each direction is non-monotone. For $n = 1$, this follows from (13). For $n \geq 2$, each $R(\zeta)$ has a strictly negative coordinate in $N(\zeta)$ and strictly positive coordinates on the negative support of every other ray, since $d_i > 0$ and $0 < t_i(\zeta) < 1$. Thus Lemma 4 implies that every pair is polarized. The two constructions are inverses. Convexity follows because every restriction defining $\mathcal{D}(P)$ is linear. $\square$

A.9 Proof of Lemma 5

Proof. There are $|N(P)|$ variables d_i , $(n - 1)|N(P)|$ variables $t_i(\zeta)$, and $n(I - |N(P)|)$ variables $q_i(\zeta)$. The n negative-magnitude normalizations from (9) are independent, leaving $n(I - 1)$ degrees of freedom. No additional affine restrictions are implied: choose d satisfying (8) and (9), $t_i(\zeta) = 3/4$ for all i, ζ , and all q entries strictly positive. At this point, every inequality defining $\mathcal{D}(P)$ holds strictly, so all sufficiently small perturbations that preserve the normalization equalities remain feasible. Thus $\mathcal{D}(P)$ contains a relatively open neighborhood in the affine subspace defined by these equalities, and its affine hull is that entire subspace. For the inequality count, each pair of labels contributes $|N(\zeta)||N(\zeta')|$ constraints. The displayed identity follows by expanding $|N(P)|^2$, and the bound from $\sum_{\zeta \in E(P)} |N(\zeta)|^2 \geq |N(P)|$

and $|N(P)| \leq I$. □

A.10 Proof of Lemma 6

Proof. With I tasks, we can identify each class P such that $|E(P)| = h$ with a set of sign vectors $\{\zeta^1, \dots, \zeta^h\}$ such that for any coordinate $i \in \{1, \dots, I\}$ there is at most one $k \in \{1, \dots, h\}$ 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 $\{\zeta^2, \dots, \zeta^h\}$). 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 $\{\zeta^2, \dots, \zeta^h\}$ 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 $\{\zeta^2, \dots, \zeta^h\}$ 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 $\{\zeta^2, \dots, \zeta^h\}$ 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. (15).

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 $C(I) = 2^I - 2 + \sum_{h=2}^I c(I, h)$. □

A.11 Proof of Proposition 3

Proof. We first verify that the program in eq. (19) is indeed the dual of eq. (18). Letting λ^i be the multipliers on the LM-MC constraints, we have

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

where the 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 $\{\sigma_j\}$ to agent j is feasible, we have $\sigma_j \in \mathcal{F}_j$. If σ_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

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

for $\varepsilon \geq 0$, and let S_j^ε be the support function of this set. For any $\varepsilon > 0$ there is an allocation μ in a neighborhood of σ such that μ_j is in the interior of $\mathcal{F}_j^\varepsilon$ for all j , so Slater's condition is satisfied 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 $\varepsilon \rightarrow 0$.

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 $P_\varepsilon \rightarrow P_0$ as $\varepsilon \rightarrow 0$, which implies $D_0 = P_0$. $\square$

A.12 Proof of Proposition 4

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

$$\tilde{V}_{LM}^*(N) = \min_{(\mu_j^a)} \sum_{j \in \mathcal{J}} \sum_{a \in A_j} |d_j^a(N)| \pi_j \cdot \mu_j^a \quad s.t. \quad \mu_j^a \in a \cap \mathbb{R}_{\geq 0}^I \quad \forall j \in \mathcal{J}, a \in A_j,$$

$$\frac{1}{N} \sum_{j \in \mathcal{J}} \sum_{a \in A_j} \mu_j^a |d_j^a(N)| = (n^1, \dots, n^I)$$

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 $\frac{1}{N} \tilde{V}_{LM}^*(N) \rightarrow V_{LM}^*(\infty)$ almost surely as $N \rightarrow \infty$. 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

$$\frac{1}{N} \tilde{V}_{LM}^*(N) = \min_{(\mu_j^a)} \sum_{j \in \mathcal{J}} \sum_{a \in A_j} \rho_j^a(N) \pi_j \cdot \mu_j^a \quad (22)$$

$$s.t. \quad \mu_j^a \in a \quad \forall j \in \mathcal{J}, a \in A_j, \quad (23)$$

$$\sum_{j \in \mathcal{J}} \sum_{a \in A_j} \mu_j^a \rho_j^a(N) = (n^1, \dots, n^I) \quad (24)$$

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 $N \rightarrow \infty$ for all j, a . Thus we just need to show that $\tilde{V}_{LM}^*(\rho)$ is upper-semicontinuous at $\rho_j^a = \mathbb{E}_j[\mathbb{1}\{s_j(c_j) = a\}]$.

Any μ_j^a satisfying constraint (23) can be written as $\mu_j^a = \sigma_j + \alpha_j^a \kappa_j(a)$ for some scaling factors $\alpha_j^a \in [0, 1]$; where $\kappa_j(a)$ is the vector defining the ray a if A_j is non-remedial, chosen so that the non-negativity constraint binds at $\sigma_j + \kappa_j(a)$. 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 $\{\sigma_j\}$. 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 σ , and thus $(\delta\hat{\alpha}_j^a)$ is also a valid mechanism in the large-market program. Then for any $\varepsilon > 0$ 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 $\iota > 0$ such that $|\mathbb{E}_j[\mathbb{1}\{s_j(c_j) = a\}] - \rho_j^a| < \iota$ 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 (24), 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

$$\sum_{j \in \mathcal{J}} \sigma_j + \sum_{j \in \mathcal{J}} \sum_{a \in A_j} \alpha_j^a \kappa_j(a) \rho_j^a = (n^1, \dots, n^I)$$

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 the prosecutor-case assignment. 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.³⁸ Given planner-chosen classification weights $(w_{\text{FN}}, w_{\text{FP}}, w_{\text{TN}}, w_{\text{TP}})$, we define the empirical counterpart of the type-specific performance primitive – $\pi_j(i)$ from the theoretical model – as the expected classification loss from assigning a type- i case to agent j :

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

Thus $\pi_j = (\pi_j(1), \dots, \pi_j(I))$ 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 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 $\text{FN}_j(i) = \Pr(D_{kj} = 0, Y_k^* =$

³⁸For 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.

$1 \mid g(k) = i$) 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 an additive 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,

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

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

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

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 $\widetilde{\pi_j(i)} = \pi_j(i) - \pi_{j^0}(i)$ as differences in social cost between agent j and an arbitrary benchmark agent j^0 . Thus, $\widetilde{\pi_j} = (\widetilde{\pi_j(1)}, \dots, \widetilde{\pi_j(I)})$ 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 $\widetilde{\pi_j(i)}$. 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. We then split an investigator’s caseload 50/50 into train and test samples for algorithmic training, and drop cases assigned to an investigator who did not receive at least 25 cases in both the train and test samples 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 then split a prosecutor’s caseload 50/50 into train and test samples for algorithmic training, and drop cases assigned to a prosecutor who did not receive at least 25 cases in both the train and test samples 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 characteristics observed by the decision-maker at the time of assignment.³⁹

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. To guard against overfitting in the risk scores themselves, each model is fit on a random half of the untreated cases; the fitted model then generates predicted risk, and the implied quartiles, for all cases in the analysis sample, and the predictive performance reported below is evaluated on the other half. This random split is used only to fit and evaluate the prediction model.

In the CPS setting, the feature set 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 the held-out half of untreated cases 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.

³⁹Because Y_k^* is observed only for untreated cases, this is not the same as $\Pr(Y_k^* = 1 \mid X_k)$ unless treatment decisions do not depend on risk beyond what is captured by X_k . This does not affect identification of relative performance, which only requires that case types be defined by pre-assignment characteristics. The difference is also small in the CPS application, where only 3% of cases are treated. In the prosecution application, where most cases are treated, the quartiles are best understood as ordering cases by predicted risk among non-prosecuted defendants; the preference distribution we calibrate assumes that agents' costs increase in this ordering.

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) measurement error in agent-specific moments, which we address through empirical Bayes shrinkage as described below, and (ii) the fact that assignment is quasi-random only within institutional strata.

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:

$$D_k = \sum_j \beta_{j1}^D Z_{kj} + \sum_j \sum_{i=2}^4 \beta_{ji}^D \mathbf{1}\{g(k) = i\} Z_{kj} + \mathbf{X}'_k \alpha^D + u_k, \quad (25)$$

$$\text{FN}_k = \sum_j \beta_{j1}^{\text{FN}} Z_{kj} + \sum_j \sum_{i=2}^4 \beta_{ji}^{\text{FN}} \mathbf{1}\{g(k) = i\} Z_{kj} + \mathbf{X}'_k \alpha^{\text{FN}} + \epsilon_k. \quad (26)$$

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 (Arnold et al., 2022). This adjustment relies on an additional additive-separability assumption: that the stratum affects treatment and false-negative rates through a common level shift across case types, so that within-quartile latent-risk distributions do not differ systematically across strata beyond this shift. The assumption is standard in the examiner-design literature (Arnold et al., 2022).

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,

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

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

$$\begin{aligned} \widehat{\widehat{\pi_j(i)}} = & (w_{\text{FP}} - w_{\text{TN}}) \left[(\widehat{\beta_{j1}^D} + \widehat{\beta_{ji}^D}) - (\widehat{\beta_{j^0 1}^D} + \widehat{\beta_{j^0 i}^D}) \right] \\ & + (w_{\text{FN}} + w_{\text{FP}} - w_{\text{TP}} - w_{\text{TN}}) \left[(\widehat{\beta_{j1}^{\text{FN}}} + \widehat{\beta_{ji}^{\text{FN}}}) - (\widehat{\beta_{j^0 1}^{\text{FN}}} + \widehat{\beta_{j^0 i}^{\text{FN}}}) \right]. \end{aligned}$$

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 $\widehat{\widehat{\pi_j(i)}}$. Separately for each case type, we model agents' true relative social costs as normally distributed with a common mean and variance, and each estimate as the true value plus normal sampling error with variance equal to its squared clustered standard error. We estimate the common mean

as the precision-weighted average of the estimates and the prior variance as the precision-weighted variance of the estimates net of the precision-weighted average of their squared standard errors, iterating the precision weights five times. Each agent’s posterior mean is then a weighted average of its own estimate and the common mean, with the weight on the common mean equal to the ratio of its sampling variance to the sum of the sampling and prior variances. The resulting posterior means are the performance vectors used by the mechanism; as discussed in Appendix B.7, the welfare change from any allocation is linear in $\tilde{\pi}$, so evaluating it at the posterior means yields the posterior expected welfare change. We apply the same shrinkage procedure, separately, to the estimated treatment and false-negative moments $\widehat{P_j(i)}$ and $\widehat{FN_j(i)}$, which are used to compute the changes in interventions and false negatives reported in the main tables.

B.5 Preference Distributions

The empirical welfare calculations require the planner to know the distribution of agents’ task-specific assignment costs. However, preference data are 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

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

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 empirical Bayes moments. This keeps the evaluation scale identical across binary, three-group, and quartile mechanisms.

We implement the large-market optimization using column generation: a linear program

chooses lotteries over fixed individual mechanisms, and its dual weights guide searches for additional mechanisms using the parametrization in Proposition 2. The search uses 2,500 fixed preference draws per agent and at most 40 rounds. After the search, we estimate each stored mechanism’s expected assignment using 100,000 new preference draws per agent. We then re-optimize the lottery weights over those same mechanisms. For the finite-market implementation, each agent receives the mechanism with the highest final lottery weight, retaining all menu rays, including those inactive under the large-market selection rule. Following agents’ choices, the designer solves a joint allocation linear program subject to their veto sets and market clearing. We report average gains over 1,000 new preference profile draws.

We select a partition using the small-market criterion. For each of the fourteen candidate partitions, we construct a mechanism using the full-sample empirical Bayes performance moments and evaluate its social-cost change over simulated preference profiles. We select the computed mechanism with the largest average reduction in social cost.

Tables C.3 and C.4 report the small-market criterion for every candidate partition, along with the corresponding large-market and full-information social-cost changes. In the CPS application, the selected mechanism uses the three-group partition $\{1, 4\} \mid \{2\} \mid \{3\}$. In the prosecution application, the selected mechanism uses four separate quartiles.

The main tables report the selected small-market mechanism, the corresponding large-market allocation, and the full-information benchmark, all evaluated using the same full-sample empirical Bayes moments. 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 construct the candidate mechanisms. 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 $q_{jm}^i(b)$ denote the number of quartile- i cases assigned to agent j under mechanism m in simulated preference profile b . Let q_{j0}^i 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.⁴⁰

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

$$\widehat{\Delta C}_m(b) = \sum_j \sum_{i=1}^4 (q_{jm}^i(b) - q_{j0}^i) \widehat{\pi_j(i)}.$$

Since $\widehat{\pi_j(i)}$ is measured relative to an arbitrary benchmark agent, this expression identifies welfare changes rather than welfare levels. When computing welfare changes, we use the same full-sample empirical Bayes posterior means of $\widehat{\pi_j(i)}$ that are used to construct the mechanism. Because $\widehat{\Delta C}_m(b)$ is linear in the performance vector for any fixed allocation, evaluating it at the posterior means gives the posterior expected welfare change conditional on the estimated regression moments and the empirical Bayes prior. Negative values of $\widehat{\Delta C}_m$ therefore correspond to reductions in social cost, or welfare gains.

For the large-market mechanism, we set q_{jm}^i as the expected allocations from solving the large-market problem. For the finite-agent small-market mechanism, allocations $q_{jm}^i(b)$ 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. Let $\widehat{P_j(i)}$ denote the estimated treatment rate for agent j and type i , and let $\widehat{\text{FN}_j(i)}$ denote the corresponding false negative rate. The change in interventions is

$$\widehat{\Delta P}_m(b) = \sum_j \sum_{i=1}^4 (q_{jm}^i(b) - q_{j0}^i) \widehat{P_j(i)},$$

⁴⁰Observed assignments may differ from this equal-split allocation, but under random assignment, this is the expected status-quo assignment.

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

$$\Delta \widehat{FN}_m(b) = \sum_j \sum_{i=1}^4 (q_{jm}^i(b) - q_{j0}^i) \widehat{FN}_j(i).$$

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

$$\Delta \widehat{FP}_m(b) = \Delta \widehat{P}_m(b) + \Delta \widehat{FN}_m(b).$$

Reported percent changes divide each level change by the corresponding status-quo baseline. Baseline false negatives and baseline interventions are directly identified from the estimated $\widehat{FN}_j(i)$ and $\widehat{P}_j(i)$ and the status-quo allocation. Baseline false positives, and hence baseline social costs, depend on the unconditional latent probability of the adverse outcome, $\mathbb{E}[Y_k^*]$, which is not directly identified because outcomes are observed only for untreated cases. Baron et al. (2026), following the “identification at infinity” approach of Arnold et al. (2022), recover this quantity in the CPS setting by extrapolating investigator-specific subsequent maltreatment rates toward a hypothetical investigator with a zero placement rate; the extrapolation is limited there because placement rates are low ($\approx 3\%$). In the prosecution application, roughly 80% of cases are prosecuted, so a comparable extrapolation would extend far beyond the support of the observed treatment-rate distribution and would be highly sensitive to functional form. For comparability across the two applications, we therefore report percent changes only for false negatives and interventions in both settings.

We also report conditional empirical Bayes posterior standard deviations of the welfare changes, holding fixed the fitted prior parameters and mechanism allocations. For each reported outcome, let $\bar{q}_{jm}^i$ denote the average change in agent j ’s allocation of type- i cases and let v_j^i denote the corresponding marginal posterior variance. Under the working assumption of posterior independence across agents and case types, the posterior standard deviation is

$$\left(\sum_j \sum_i (\bar{q}_{jm}^i)^2 v_j^i \right)^{1/2}.$$

For the small-market and full-information mechanisms, allocation changes are averaged over 1,000 preference profiles before computing this quantity; for the large-market approximation,

we use expected allocation changes. These standard deviations measure uncertainty about the average gain conditional on the fitted model. Variation across preference profiles, holding performance at its posterior mean, is reported separately in Figures C.2 and C.3.

References

- 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.
- Association of Professional Flight Attendants. PBS: Basic understanding, 2026. URL <https://www.apfa.org/bidding/monthly-bidding/pbs-basic-understanding/>. Accessed September 22, 2026.
- B. Ba, P. Bayer, N. Rim, R. Rivera, and M. Sidibé. Police officer assignment and neighborhood crime. Working Paper 29243, NBER, 2022.
- 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.
- J. F. Bard and H. W. Purnomo. Preference scheduling for nurses using column generation. *European Journal of Operational Research*, 164(2):510–534, 2005.
- 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. Working Paper 32369, National Bureau of Economic Research, 2026.
- M. Bates, M. Dinerstein, A. C. Johnston, and I. Sorkin. Teacher labor market policy and the theory of the second best. *The Quarterly Journal of Economics*, 140(2):1417–1469, 2025.
- 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. *American Economic Review*, 2026. Conditionally accepted.

- B. Biasi, C. Fu, and J. Stromme. Equilibrium in the market for public school teachers: District wage strategies and teacher comparative advantage. *Journal of Political Economy*, 2026. Forthcoming.
- 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.
- 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.
- M. Camara. Computationally tractable choice. Working paper, 2025.
- 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. Mechanism design via optimal transport. In *Proceedings of the fourteenth ACM conference on electronic commerce*, pages 269–286, 2013.
- 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.
- 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.
- T. S. Eastmond, M. D. Ricks, J. Betts, and N. Mather. Welfare added? Optimal teacher assignment with value-added measures. Working Paper 34768, National Bureau of Economic Research, 2026.
- 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. Approximate revenue maximization with multiple items. *Journal of Economic Theory*, 172:313–347, 2017.
- S. Hart and N. Nisan. Selling multiple correlated goods: Revenue maximization and menu-size complexity. *Journal of Economic Theory*, 183:991–1029, 2019.
- S. Hart and P. J. Reny. Maximal revenue with multiple goods: Nonmonotonicity and other observations. *Theoretical Economics*, 10(3):893–922, 2015.
- S. Holton, B. Rasmussen, K. Long, M. Bellizia, J. C. Mathieson, S. Crowe, D. Mill, H. Pasion, C. Rankin, M. Woodhouse, M. Douglas, N. Glanville, K. Baker, K. Fallon, M. Hoffmann, N. Sliwa, D. Heinjus, L. Fitzpatrick, and P. Gilbert. Developing nurse and midwife centred rostering principles using co-design: A mixed-methods study. *BMC Nursing*, 23(1):938, 2024. doi: 10.1186/s12912-024-02522-7.
- 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.
- M. Laverde, E. Mykerezi, A. Sojourner, and A. Sood. Gains from alternative assignment? Evidence from a two-sided teacher market. Technical report, IZA Discussion Paper No. 17696, 2025.
- 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.
- P. McGee and D. Levin. How obvious is the dominant strategy in an english auction? experimental evidence. *Journal of Economic Behavior & Organization*, 159:355–365, 2019.
- 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.

- 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.
- Q. Valenzuela-Stookey. Subjective complexity under uncertainty. *Mathematical Social Sciences*, 126:76–93, 2023.
- 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.
- F. Yang and P. Dworczak. Searchable menus. *arXiv preprint arXiv:2608.17446*, 2026.

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	Small Market	Large Market	Full Information
Three-group: {1, 4} {2} {3}	-1,380.1	-2,732.8	-7,355.5
Three-group: {1, 2} {3} {4}	-1,192.7	-3,132.3	-7,355.5
Binary: {1, 3} {2, 4}	-1,126.6	-1,863.9	-7,355.5
Binary: {1, 2} {3, 4}	-1,124.6	-2,033.6	-7,355.5
Binary: {1, 4} {2, 3}	-1,085.2	-1,668.2	-7,355.5
Three-group: {1, 3} {2} {4}	-1,081.4	-3,023.0	-7,355.5
Binary: {1, 2, 4} {3}	-996.3	-1,576.4	-7,355.5
Binary: {1, 2, 3} {4}	-877.7	-1,926.8	-7,355.5
Binary: {1, 3, 4} {2}	-846.3	-1,596.9	-7,355.5
Three-group: {1} {2, 4} {3}	-487.7	-3,148.6	-7,355.5
Three-group: {1} {2} {3, 4}	-400.8	-3,161.4	-7,355.5
Binary: {1} {2, 3, 4}	-391.7	-1,820.0	-7,355.5
Three-group: {1} {2, 3} {4}	-378.9	-3,138.2	-7,355.5
Quartile: {1} {2} {3} {4}	-347.0	-4,192.9	-7,355.5

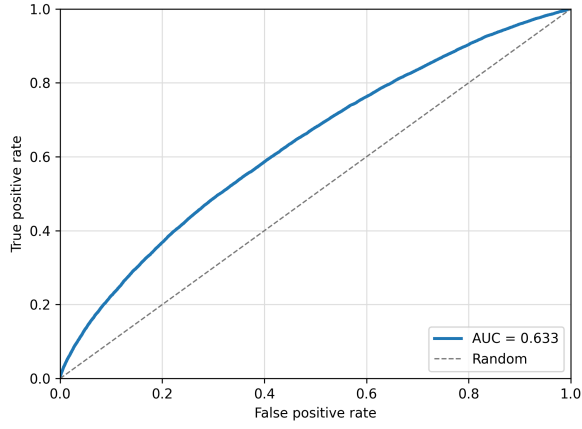
Notes. This table reports full-sample weighted social-cost changes for each task partition. Partitions use predicted-risk quartile indices. Rows are sorted by small-market social-cost changes; more negative values indicate larger estimated reductions. The bold row is selected using the full-sample small-market criterion. The full-information column uses the quartile benchmark and is the same across candidate partitions.

Table C.4: Prosecution Task Partition Comparison

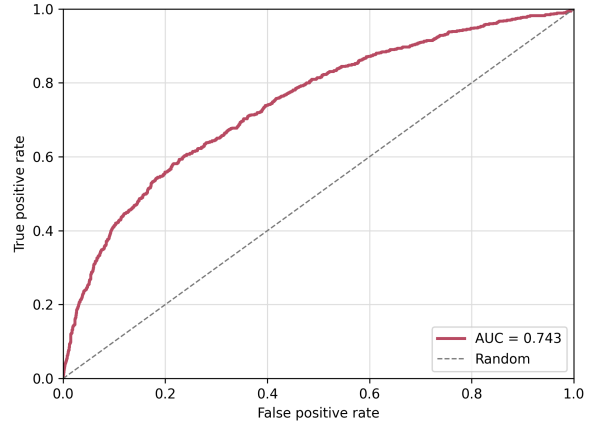
Task Partition	Small Market	Large Market	Full Information
Quartile: {1} {2} {3} {4}	-1,900.4	-2,434.8	-4,525.7
Three-group: {1} {2} {3, 4}	-1,661.7	-2,102.9	-4,525.7
Three-group: {1, 2} {3} {4}	-1,630.3	-1,895.3	-4,525.7
Three-group: {1} {2, 3} {4}	-1,618.3	-1,962.4	-4,525.7
Three-group: {1} {2, 4} {3}	-1,457.2	-1,826.4	-4,525.7
Binary: {1, 2} {3, 4}	-1,355.1	-1,558.4	-4,525.7
Binary: {1} {2, 3, 4}	-1,260.8	-1,613.2	-4,525.7
Three-group: {1, 3} {2} {4}	-1,223.1	-1,402.2	-4,525.7
Binary: {1, 2, 3} {4}	-893.8	-991.7	-4,525.7
Three-group: {1, 4} {2} {3}	-885.7	-1,025.6	-4,525.7
Binary: {1, 3} {2, 4}	-759.0	-849.3	-4,525.7
Binary: {1, 3, 4} {2}	-571.2	-667.3	-4,525.7
Binary: {1, 4} {2, 3}	-488.9	-523.7	-4,525.7
Binary: {1, 2, 4} {3}	-476.3	-536.9	-4,525.7

Notes. This table reports full-sample weighted social-cost changes for each task partition. Partitions use predicted-risk quartile indices. Rows are sorted by small-market social-cost changes; more negative values indicate larger estimated reductions. The bold row is selected using the full-sample small-market criterion. The full-information column uses the quartile benchmark and is the same across candidate partitions.

Figure C.1: Algorithm Performance: ROC Curves



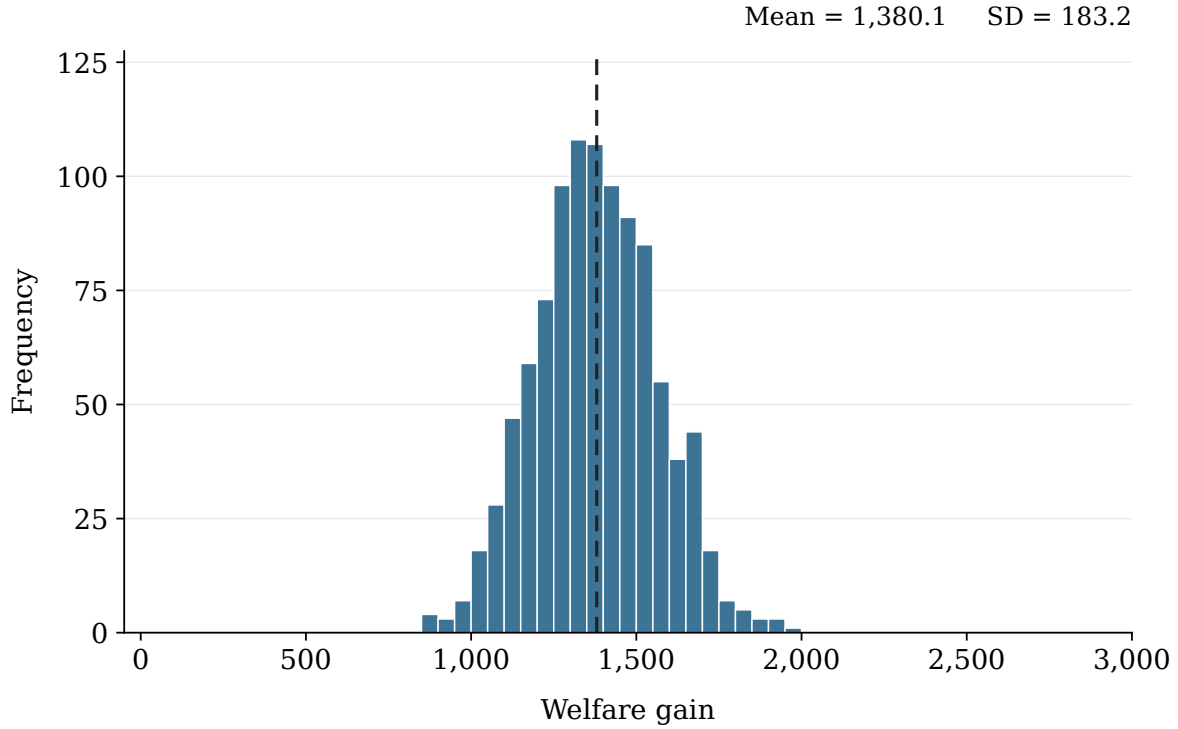
Panel A: CPS Sample



Panel B: Prosecution Sample

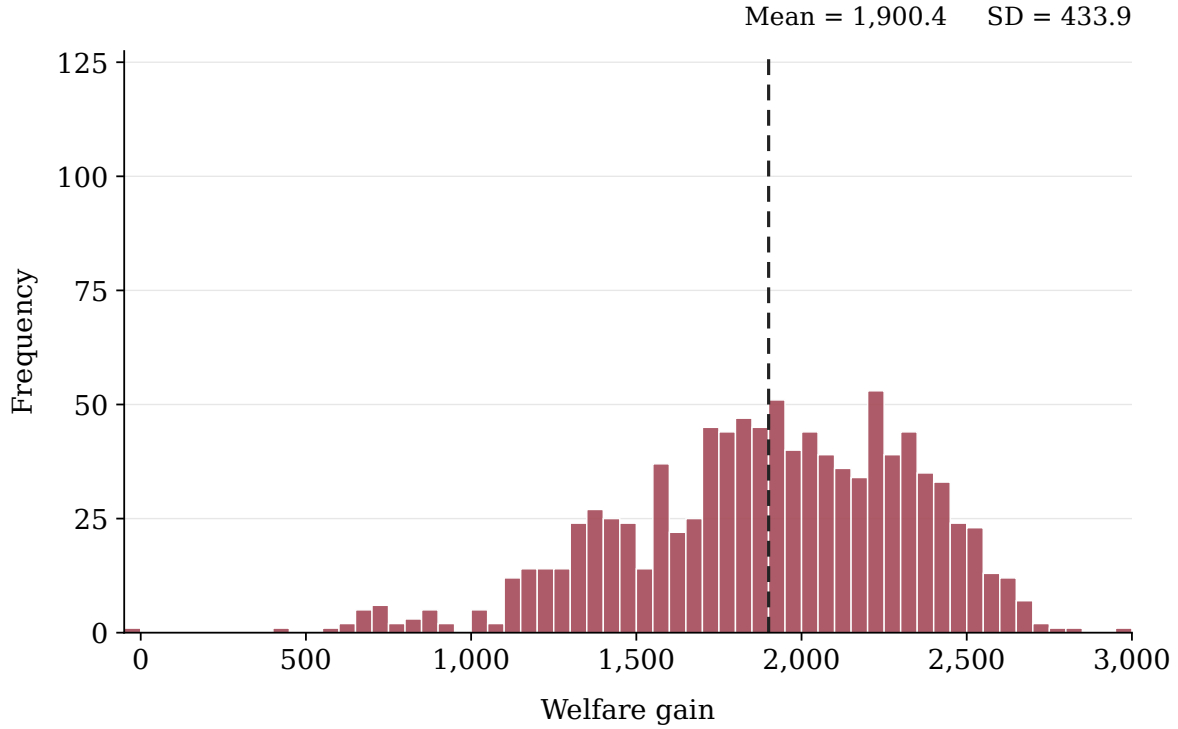
Notes. This figure summarizes the predictive performance of the algorithms $m(X_k)$ in the CPS and prosecution samples, evaluated on the held-out sample. Within each sample, the figure plots the receiver operating characteristic (ROC) curves of $m(X_k)$ as a classifier of Y_k^* and reports the AUC. The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

Figure C.2: CPS: Welfare Gains across Preference Profiles



Notes. This figure reports the distribution of welfare gains from the selected small-market mechanism, which uses the task groups $\{1, 4\}$, $\{2\}$, and $\{3\}$, across 1,000 simulated preference profiles. Gains are reductions in weighted social cost relative to the status quo. Agent performance is held at its empirical Bayes posterior mean. The dashed line marks the mean; the reported SD is the sample standard deviation across preference profiles. It measures preference-driven variation in gains, rather than posterior uncertainty about the average gain reported in Table 1.

Figure C.3: Prosecution: Welfare Gains across Preference Profiles



Notes. This figure reports the distribution of welfare gains from the selected small-market mechanism, which uses four separate quartiles, across 1,000 simulated preference profiles. Gains are reductions in weighted social cost relative to the status quo. Agent performance is held at its empirical Bayes posterior mean. The dashed line marks the mean; the reported SD is the sample standard deviation across preference profiles. It measures preference-driven variation in gains, rather than posterior uncertainty about the average gain reported in Table 1.