Fraud-proof non-market allocation mechanisms *

Eduardo Perez-Richet † Vasiliki Skreta ‡

September 1, 2023

Abstract

We study the optimal design of fraud-proof allocation mechanisms without transfers. An agent's eligibility relies on a score reflecting social value, but gaming generates misallocations, mistrust, unfairness and other negative externalities. We characterize optimal allocation rules that are immune to gaming under two classes of gaming technologies. We examine the impact of demographic changes on allocations within and across identifiable groups, while accounting for resource and quota constraints. Fraud-proof allocation rules enhance fairness and trust in allocation systems at the cost of some allocative efficiency.

Keywords: Mechanism Design, Falsification, Fraud, Manipulation.
JEL classification: C72, D82.

*We thank Ricardo Alonso, Julien Combes, Michael Ostrowsky, James Schummer, Philipp Strack, Andrzej Skrzypacz and Utku Ünver for thought-provoking discussions and suggestions. Eduardo Perez-Richet acknowledges funding by the European Research Council (ERC) consolidator grant 101001694, and thanks the Fernand Braudel visiting program at the European University Institute for its hospitality. Vasiliki Skreta acknowledges funding by the European Research Council (ERC) consolidator grant 682417 "Frontiers In Design." Francesco Conti, Nathan Hancart and Ignacio Núñez provided excellent research assistance at various stages.

†Sciences Po, CEPR – e-mail: eduardo.perez@sciencespo.fr

‡UT Austin, UCL, CEPR – e-mail: vskreta@gmail.com

1 Introduction

Goods, services and rewards[1] are frequently allocated via non-market mechanisms, either due to constraint or because monetary transfers are ineffective at targeting deserving recipients.[2] To target eligible agents, non-market allocation mechanisms must rely on data about their characteristics. For instance, seats in schools are assigned using priorities that combine multiple criteria, green labels are awarded based on measured emissions, and public housing is allocated on the basis of criteria such as household income. Eligibility is often assessed through a *score* measuring characteristics or performance, acting as a proxy for the social value of assigning a unit of the good to an agent.

However, reliance on the score creates strong incentives to game it. Consequently, practices such as falsification, forgery, greenwashing, teaching to the test, or manipulating statistics are commonplace. For example, suggestive evidence indicates that parents fake addresses to gain admission to desirable public schools in Denmark (Bjerre-Nielsen, Christensen, Gandil, and Sievertsen, 2023), French firms underreport workforce size to avoid legal obligations (Askenazy, Breda, Moreau, and Pecheu, 2022), and doctors manipulate their patient's priority in organ transplant waiting lists in the USA and Germany (Bolton, 2018; McMichael, 2022).

Gaming is detrimental. Firstly, it alters achievable assignments unfairly favoring agents with higher gaming ability. Secondly, it generates various negative externalities such as deteriorating the informational content[3] of the score, rendering the mechanism politically unsustainable, or depleting the supply of objects. For instance, authorities in Boston and Chicago abandoned the "Boston" school assignment mechanism due to concerns about its vulnerability to manipulation (Pathak and Sönmez, 2013). In Germany, a scandal involving the manipulation of the liver allocation system by transplant providers led to a 20%-40% erosion in organ donations (Bolton, 2018).

In this paper, we show how to maximize allocative efficiency while ensuring fraud-proofness, meaning that agents cannot benefit by gaming their scores.

[1] Goods include public housing, seats in schools and vaccines; services include training, education and financial assistance programs; rewards include promotions, labels and certificates awarded to businesses meeting certain emissions or social responsibility criteria.

[2] See Condorelli (2013) and Akbarpour, Dworczak, and Kominers (2020) for a theory of when non-market mechanisms are optimal.

[3] This is an instance of Goodhart's law: "when a measure becomes a target, it ceases to be a good measure."

Specifically, we address the problem of allocating a fixed mass of homogeneous objects (or prizes) to a heterogeneous population of agents using non-market mechanisms based on scores. The score is a publicly available but *falsifiable* metric that measures an agent's private characteristics. Throughout the paper, we use the term *falsification* as a broad category that encompasses gaming, manipulation or any other socially wasteful activities agents undertake to artificially alter their observed score compared to their *natural score*. The natural score reflects an agent's true characteristics and emerges when the agent does not engage in such activities. We assume an agent's score is positively correlated with their *worthiness*: the social value of giving them an object.

We characterize the falsification-proof mechanism that maximizes expected social value assuming costly falsification. This mechanism allocates the good randomly with a probability that increases with the score, generating both rejection and allocation errors for an intermediate range of scores around the eligibility threshold. The optimal allocation rule is determined by the agents with the least cost of falsifying, or highest gaming ability.

Our analysis incorporates non-falsifiable public information about agents, which the designer can use in conjunction with the score, as well as private information of agents beyond their natural score. We show, however, that falsification-proof allocation mechanisms cannot condition on such private information, and must exclusively rely on the score and public information. Public information effectively splits the population into groups with the same public characteristics. Additionally, we allow for a resource constraint on the mass of available objects, and for exogenous group-specific quotas. Such quotas may reflect redistributive or fairness concerns (see Abdulkadiroğlu and Sönmez, 2003; Abdulkadiroğlu, 2005; Dur, Pathak, and Sönmez, 2020; Çelebi and Flynn, 2022). The falsification-proofness requirement completes the set of constraints faced by the designer. Our approach, therefore, follows the mechanism design tradition in seeking to maximize expected social value, while imposing desirable or practical criteria as constraints on the designer, akin to the market design literature.

Falsification-proofness in our analysis is analogous to strategy-proofness or the truth-telling constraint in standard mechanism and market design settings. However, in our setup, with falsification costs and no transfers, optimal mechanisms may have to induce falsification, as shown in Perez-Richet and Skreta (2022). Consequently, imposing falsification-proofness entails a loss in allocative efficiency for the designer.

In spite of this loss, we argue for the importance of insisting on falsification-proofness for several reasons. Preventing falsification is desirable because allocation mechanisms may otherwise be politically unsustainable, generate negative externalities, and impose heavy burdens on agents. Even though it may be intended by the designer, the mechanism may come under scrutiny when induced fraud is detected, possibly leading to its collapse. Negative externalities include score distortions and trust erosion. For example, greenwashing may blur our assessment of emissions levels. Trust erosion can deplete supply, but there is also evidence that dishonest behavior spreads in society (see, for example, Rincke and Traxler, 2011; Galbiati and Zanella, 2012; Alm, Bloomquist, and McKee, 2017; Ajzenman, 2021). Moreover, costly falsification may unfairly burden certain agents and raise concerns about fairness, especially when gaming ability varies among individuals. For instance, previous studies (Pathak and Sönmez, 2013; Bjerre-Nielsen et al., 2023) have shown that sophisticated gaming in school choice mechanisms resulted in better assignments for those who engaged in such practices and adversely affected others.

Our optimal design problem can be separated into an *across-groups* problem, that consists in allocating objects across groups while satisfying the allocative constraints, and a series of *within-group* problems, each dealing with fully allocating a specific mass of objects according to scores within a group. To solve the within problem, we use a Lagrangian approach to convert it into an auxiliary problem, eliminating the resource constraint by incorporating its shadow price as an *endogenous outside option value* for allocating an object. We then show that optimal allocation rules must be smooth, monotonic, and flat outside of a *growth interval*. Feasible growth intervals are those for which the expected worthiness of agents with scores within the interval is exactly equal to the value of the outside option. Thus, they form a collection of nested intervals centered around the score that equates conditional expected worthiness with the value of the outside option. We proceed to derive the optimal within-group allocation rule in two steps: first, we solve a *reduced problem* restricted to a feasible growth interval and relaxed of the probability constraint; second, we select the growth interval that makes the probability constraint bind, if applicable, or the maximal one otherwise.

Falsification-proofness puts a bound on the gain in allocation probability between any two scores, equal to the corresponding *least cost* of falsification. We solve the first-step reduced problem in closed form for two broad classes of least-cost functions. If the least-cost function has *upward increasing differences*, (UID), we show that the reduced problem is equivalent to the dual of the classical Monge-Kantorovich optimal transport problem with a known solution. This equivalence allows us to characterize the optimal within-group allocation rule under in Theorem 1. If, instead, the least-cost function has *upward decreasing differences*, (UDD) we use a first-order approach to obtain the solution, which we characterize in Theorem 2. Theorem 3 then outlines how to find the solution to the within problem in each group by adjusting the endogenous outside option in the reduced problem. Finally, in Theorem 4, we characterize the solution of the across problem and provide an algorithm to find it.

We analyze various comparative statics effects. We first consider scenarios in which allocative constraints are not binding, either before or after changes in parameters, so feedback effects between groups are absent. Unsurprisingly, higher gaming ability is detrimental for the designer. The impact on agents is more nuanced. It increases the allocation probability of low-score agents, but decreases that of high-score agents when gaming ability is initially sufficiently low. If gaming ability is initially high, the allocation probability increases for all scores if the average worthiness is sufficiently high, and otherwise decreases for all scores. For the designer, a first-order stochastic dominance shift of the score distribution is beneficial, but not necessarily for agents. If the shift replaces agents with scores below the eligibility threshold by agents scoring above it, then the allocation probability becomes uniformly higher.

Summing up, we derive allocation mechanisms that deter manipulation while maximizing allocative efficiency and analyze the impact of falsification-proofness and various design constraints on the overall efficiency and stability of non-market score-based mechanisms for allocation.

Related literature. We contribute to the literature on optimal allocation mechanisms to privately informed agents which can be categorized along two essential dimensions: the designer's objective, and the tools available for allocation targeting. In the seminal contribution of Myerson (1981), the designer uses monetary transfers to target allocation so as to maximize revenue. However, monetary transfers may lose their effectiveness with a more general designer's objective, as seen in Condorelli (2013), or wishes to maximize a combination of weighted utilitarian and revenue objectives, as demonstrated in Akbarpour, Dworczak, and Kominers (2020). Both studies establish conditions under which the designer optimally refrains from using transfers entirely, with Akbarpour et al. (2020) showing how publicly available data can be a complementary tool for targeting. In this study, we consider a general objective, and exclude transfers altogether. While there may be exogenous reasons to rule out transfers, Condorelli (2013) and Akbarpour et al. (2020) show that this is optimal if the designer's utility from allocation is weakly or negatively correlated with willingness to pay.

In our approach, targeting is enabled by the availability of the (possibly falsified) score. Thus, we contribute to a substantial literature on non-market optimal allocation mechanisms, which studies the use of alternative targeting tools in lieu of transfers. Ben-Porath, Dekel, and Lipman (2019) rely on evidence disclosure. Ben-Porath, Dekel, and Lipman (2014), Lipman (2015), Mylovanov and Zapechelnyuk (2017), Erlanson and Kleiner (2019), Chua, Hu, and Liu (2019), Epitropou and Vohra (2019) and Li (2020) use ex-post (costly) inspection or verification with limited penalties. Hartline and Roughgarden (2008) and Dworczak (2022) consider money (or utility) burning, while Patel and Urgun (2023) combines verification and money burning. In Kattwinkel (2019), the designer has access to private information correlated with the private information of the agent, while in Kattwinkel and Knoepfle (2022) she can additionally verify the agent's type.

In contrast, we consider costly state falsification and impose falsification-proofness. It is similar to money burning in the sense that it is wasteful, but differs in that it is the outcome of the wasteful effort (the falsified score) that is observable rather than the effort itself. Frankel and Kartik (2021) and Ball (2022) study the optimal design of linear scores under a gaming technology that amounts to costly state falsification. Landier and Plantin (2016) characterize optimal tax design under costly income hiding. Kephart and Conitzer (2016), Deneckere and Severinov (2022) and Severinov and Tam (2019) study mechanism design with misreporting costs but focus on settings (mainly with transfers) in which falsification-proofness is without loss. Lacker and Weinberg (1989) investigate the design of risk sharing contracts with costly state falsification. They consider optimal falsification-proof contracts, and are the first to show the constraint may lead to a loss of optimality without characterizing the optimal contract. Our setup builds upon our work in Perez-Richet and Skreta (2022), where we show that inducing falsification is necessary for allocative optimality. Unlike Perez-Richet and Skreta (2022), in this study, we allow for agent heterogeneity in other dimensions than score, introduce public labels, and require falsification-proofness.

In practice, the design of allocation rules based on scores may incentivize both true improvements and score manipulations. For example awarding green certificates to low-emitting firms may prompt them to engage in both greenwashing and abatement. In a closely related framework, Augias and Perez-Richet (2023) study the optimal design of allocation mechanisms when agents can improve their score. In this paper, by contrast, manipulations are socially purely wasteful. Examples of such wasteful activities include wait-list manipulations in the context of organ transplants studied in Schummer (2021) or priority manipulations in school choice settings (Pathak and Sönmez, 2008; He, 2015).

From a technical point of view, we contribute to a growing stream of papers that use optimal transportation techniques in economic theory and econometrics. Early uses of these techniques are surveyed in Carlier (2012) and Galichon (2018). More recently, they have also been applied to mechanism design problems with multidimensional private information (Daskalakis, Deckelbaum, and Tzamos, 2017; Kolesnikov, Sandomirskiy, Tsyvinski, and Zimin, 2022), information design (Arieli, Babichenko, and Sandomirskiy, 2022; Kolotilin, Corrao, and Wolitzky, 2022; Lin and Liu, 2022; Malamud and Schrimpf, 2021), and labor market sorting problems (Boerma, Tsyvinski, and Zimin, 2021).

2 The allocation problem

Framework. The designer seeks to allocate a mass $\bar{\rho} \leq 1$ of indivisible and homogeneous objects to a unit mass of heterogeneous agents without transfers. Each agent is characterized by a private type $\theta = \big(i, s, k\big)$, and a scalar w which captures their *worthiness*, that is the social value (or the designer's value) of allocating an object to them. Without loss of generality, the value of the outside option (not allocating) is normalized to 0 for each object. Agents may know their worthiness (if θ is a sufficient statistics for w) or not. Each agent draws a vector of characteristics (θ, w) i.i.d. from a joint distribution. Hence the different dimensions of an agent's vector of characteristics can, and typically are, correlated; but they are independent from other agents' characteristics.

The first dimension of the type, $i \in I$, encompasses all relevant publicly observable and unfalsifiable (or too costly to falsify) characteristics of an agent. We refer to i as the agent's *group*, and assume I is a finite set. We assume existence of an exogenous one-dimensional metric, the *score*, measuring some private characteristics of the agent. The second dimension, $s \in S_i \subseteq \mathbb{R}$ is the agent's *natural score*, which she obtains when she does not interfere with the measuring technology. The agent may indeed manipulate her score so the designer observes a *falsified score* t instead of her natural score s. The last dimension of type, k is a vector of hidden characteristics drawn from a compact subset of $\mathbb{R}^p$. It includes the agent's value for the good $v > 0$, as well as all privately known characteristics affecting her falsification cost, and may also include characteristics correlated with her worthiness.

The cost of falsifying to t is given by the function $C^\theta(t) \geq 0$. It may capture technical costs, psychological lying costs as well as expected penalties for gaming. In the remainder of the paper, we employ the falsification terminology and refer to the *falsification technology* and the *falsification cost*. We assume the payoff of an agent of type θ choosing to falsify her type to t is $\alpha v - C^\theta(t)$, where α is the probability she gets an object.

Distributional assumptions. Each agent draws her vector of characteristics (θ, w) independently from an identical joint distribution. The mass of group i is $\mu_i > 0$, where $\sum_i \mu_i = 1$. We let F_i denote the cumulative score distribution function conditional on i, which we assume to have full support on an interval $S_i = [\underline{s}_i, \overline{s}_i]$, and no atoms. Conditional on (i, s), the remainder of the type vector is fully supported on $K_{i,s}$, a compact and convex subset of $\mathbb{R}^p$.

We assume that, conditional on (i, s), social worthiness w is bounded and integrable, and denote the corresponding expected worthiness by $w_i(s) = \mathbb{E}(w|i, s)$, and the expected worthiness in group i by $\bar{w}_i = \mathbb{E}(w|i)$. We assume score and worthiness are positively related in the sense that, for every group i, $w_i(s)$ is strictly increasing.

Falsification technologies. We assume not falsifying is costless so, given an agent's type $\theta = (i, s, k)$, $C^\theta(s) = 0$. For every $t \in S_i$, we let $\frac{1}{\gamma_i} c_i(t|s) = \inf_{k \in K_{i,s}} \frac{1}{v} C^{(i,s,k)}(t)$ denote the least cost-to-value ratio of falsifying to t for agents with natural score s in group i. It exists since C_t^θ is bounded below by 0. We assume this bound is tight in the sense that, for each t and every $\varepsilon > 0$, there exists a positive mass of agents from group i with natural score s whose cost of falsifying to t is lower than $\frac{1}{\gamma_i}c_i(t|s) + \varepsilon$. For simplicity, we refer to $\frac{1}{\gamma_i}c_i(t|s)$ as the *least cost*, or even the cost of group i. We use the scalar $\gamma_i > 0$ to study comparative statics with respect to changes in the least-cost function, and refer to it as the *gaming ability* of group i.

We assume the least-cost function is *monotonic* for *upward* falsifications: if $t \geq s$, then $c_i(t|s)$ is (locally) strictly increasing in t and $-s$. We also assume least-cost functions satisfy the following *regularity* assumption.

Definition 1 (Regularity)**.** *A cost function $c(t|s)$ is regular if it is continuously differentiable in t on $[s, \overline{s}]$, and in s on $[\underline{s}, t]$, and there exists $\Lambda > 0$ such that, for every s, t, $c(t|s) \leq \Lambda|t - s|$.*

We denote the partial derivatives of a regular cost function by $c_t(\cdot|\cdot)$ and $c_s(\cdot|\cdot)$.

Depending on the context, the cost function may take different forms, so it is useful to rely on flexible assumptions. We characterize optimal allocation rules for the following two distinct classes of cost functions.

Definition 2 (Upward Differences)**.** *A cost function $c(t|s)$ has upward increasing differences if*

$$\forall s < s' \leq t < t', \quad c(t'|s') - c(t|s') \geq c(t'|s) - c(t|s), \tag{UID}$$

and upward decreasing differences if

$$\forall s < s' \leq t < t', \quad c(t'|s') - c(t|s') \leq c(t'|s) - c(t|s). \tag{UDD}$$

These conditions only bear on upward falsification because we show downward falsification is never beneficial under optimal allocation rules. To gain intuition about the interpretation of the conditions from Definition 2, it is useful to consider the family of *Euclidean cost functions*, $c(t|s) = \xi(|\varphi(t) - \varphi(s)|)$, where $\xi : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing function such that $\xi(0) = 0$, and $\varphi : S \to \mathbb{R}$ is an increasing transformation of the score. A Euclidean cost function satisfies (UID) if ξ is concave (or, more generally, subadditive), and (UDD) if ξ is convex (or, more generally, superadditive). Reasoning as if c were an agent's cost function instead of the least cost, for the sake of intuition, the monotonicity of upward differences captures economies of scale in the amount of falsification $|\varphi(t) - \varphi(s)|$: increasing differences correspond to increasing returns to scale, and decreasing differences to decreasing returns to scale.

Mechanisms. We restrict attention to non-market mechanisms in which the designer can only commit to a *score-based allocation rule* $\boldsymbol{\alpha} = (\alpha_i)_{i \in I}$, where $\alpha_i : S_i \to [0,1]$ is the probability that an object is allocated to an agent from group i conditional on her observed score. The probability constraint plays an important role in the program, so we keep track of it as

$$\forall (i,s) \;\; 0 \leq \alpha_i(s) \leq 1. \tag{PC}$$

The restriction to score-based allocation rules entails two simplifications, but is without loss of generality, as we show in Appendix C. First, an agent's allocation probability is based solely on her score, and not on the score profile of other agents, so the continuum of agents is essentially treated as a single agent. Second, the mechanism is communication-free: Contrary to standard mechanisms, the design does not include a communication protocol in order to extract further information about types. Instead, the designer only relies on observable information: the group and the score.

Falsification-Proofness. For the reasons explained in the introduction, we restrict the designer to *falsification-proof mechanisms*, that is, allocation rules that do not induce any gaming by the agents. Given our *tight bound* assumption on costs, a mechanism is falsification-proof if and only if it satisfies the constraint[4]

$$\forall (i,s,t) \;\; \alpha_i(t) - \alpha_i(s) \leq \frac{1}{\gamma_i} c_i(t|s). \tag{FPC}$$

Allocative constraints. We allow for a *resource constraint* and *quota constraints*. A mechanism is feasible if it satisfies these *allocative constraints*. The resource constraint requires

$$\sum_i \mu_i \int_{\underline{s}_i}^{\bar{s}_i} \alpha_i(s) dF_i(s) \leq \bar{\rho}. \tag{RC}$$

In addition, the designer may have to satisfy a system of exogenous quotas $\phi = (\phi_i)_{i \in I}$, where $\phi_i \in [0,1]$ is a fraction of objects reserved for group i, with

[4]Interestingly, we can also interpret (FPC) as being motivated by inequality awareness as in Akbarpour et al. (2020). The cost $\gamma_i^{-1} c_i(t|s)$ then acts as a bound on allocative inequality between score pairs s and t. We thank Ricardo Alonso for suggesting this interpretation.

$\sum_i \phi_i \leq 1$, and $\phi_i \bar{\rho} \leq \mu_i$. The quota constraints are

$$\forall i, \quad \mu_i \int_{\underline{s}_i}^{\overline{s}_i} \alpha_i(s) dF_i(s) \geq \phi_i \bar{\rho}. \tag{QC}$$

Objective and program. The restriction to falsification-proof mechanisms implies the agent's observed score t can be interpreted at face value by the designer, so we can write the objective function as

$$\sum_i \mu_i \int_{S_i} w_i(s) \alpha_i(s) dF_i(s).$$

The designer's program is to choose an allocation rule $\boldsymbol{\alpha}$ to maximize this objective subject to (PC), (FPC), (RC), and (QC).

We can divide the allocation problem into the *within problem* of optimally allocating a fixed mass of objects within each group, and the *across problem* of optimally choosing the masses of objects accruing to each group while satisfying the allocative constraints.

Within problem. Let ρ_i be the mass of objects allocated to group i. Then the corresponding within problem is

$$W_i(\rho_i) = \max_{\alpha_i} \int_{S_i} \alpha_i(s) w_i(s) dF_i(s) \tag{P}$$

$$\text{s.t. (FPC), (PC)}$$

$$\mu_i \int_{S_i} \alpha_i(s) dF_i(s) = \rho_i, \tag{RC}$$

where the within resource constraint (RC) must hold with equality. The within problem is feasible only if $\rho_i \leq \mu_i$, hence its value function is equal to $-\infty$ otherwise.

Across problem. A *group allocation profile* $\boldsymbol{\rho}$ satisfies the allocative constraints if it belongs to the feasible set $R = \{\boldsymbol{\rho} : \sum_i \rho_i \leq \bar{\rho}, \ \rho_i \geq \phi_i \bar{\rho} \ (\forall i)\}$. The designer's problem is then summarized by the across problem

$$\overline{W}(\boldsymbol{F}, \boldsymbol{\gamma}) = \max_{\boldsymbol{\rho} \in R} \sum_i \mu_i W_i(\rho_i). \tag{$\overline{\text{P}}$}$$

3 Simplifying the within problem

In this section, we simplify the within problem, and we characterize optimal allocation rules in the next section. We drop the i index from the within problem to simplify notations. The simplification proceeds in two steps, which are both applications of the Lagrangian method, first on the resource constraint, and then on the probability constraint.

Auxiliary problem, endogenous outside option and priorities. In this first step of simplification, we obtain an *auxiliary problem*, where the resource constraint is relaxed, and the expected social worthiness function is modified to accommodate an endogenous value of the outside option which is essentially the Lagrange multiplier on the resource constraint, and can be interpreted as its shadow price. Adjusting the endogenous outside option, the mass of allocated objects under the optimal allocation rule spans the full feasible range from no objects to a mass equal to the size of the group μ. The problem then reduces to picking the right value for the outside option.

Letting $\hat{w}/\mu$ be the Lagrange multiplier on the resource constraint, the Lagrangian for (P) is $\int_S \alpha(s)\{w(s) - \hat{w}\}dF(s) + \hat{w}\rho/\mu$. Maximizing the Lagrangian is then equivalent to solving the auxiliary problem

$$\max_{\alpha} \int_S \alpha(s)\{w(s) - \hat{w}\}dF(s) \quad \text{s.t.} \quad (\text{FPC}), (\text{PC}), \tag{$\tilde{\text{P}}$}$$

where $\hat{w}$ is an endogenous outside option value and the resource constraint is relaxed. In the auxiliary program, the social value is $w(s) - \hat{w}$. The outside option value $\hat{w}$ is equal to the Lagrange multiplier (scaled by μ) on the resource constraint, and can be interpreted as the shadow price of marginally tightening the constraint.

Lemma 1 (Lagrangian and auxiliary problem)**.** *The following statements are equivalent:*

(i) α *solves* (P)*.*

(ii) There exists an outside option $\hat{w}(\rho)$ *such that* α *solves the auxiliary problem* ($\tilde{\text{P}}$) *and* $\rho = \mu \int_S \alpha(s)dF(s)$*.*

(iii) There exists $\hat{w}$ such that $(\alpha, \hat{w})$ is a saddle-point for the Lagrangian $\int_S \alpha(s)\{w(s) - \hat{w}\}dF(s) + \hat{w}\rho/\mu$.

Furthermore, the value function of the within problem is concave in ρ, and its derivative $W'(\rho)$ exists almost everywhere, and is then equal to $\hat{w}(\rho)/\mu$.

All points are classical results in optimization theory (see,for example, Luenberger, 1969, chapter 8). Necessity of (i) holds because the initial program is linear in α.

We say the group has *high priority* if expected worthiness exceeds the outside option, $\bar{w} > \hat{w}$, and *low priority* if $\bar{w} < \hat{w}$. In case of equality, $\bar{w} = \hat{w}$, we say the group has *neutral priority*. We let $\hat{s}$ denote the *eligibility threshold* at which the social value is equal to the outside option value $w(\hat{s}) = \hat{w}$. We refer to the quantity $w(s) - \hat{w}$ as the *social surplus* (from allocation). It is positive above, and negative below $\hat{s}$. Next, we proceed to solve the auxiliary problem ($\tilde{\mathrm{P}}$).

Smoothness and monotonicity of optimal allocation rules. If agents are prevented from falsifying ($\gamma \to 0$), the solution to the auxiliary problem allocates with certainty to scores above, and with null probability to scores below the eligibility threshold $\hat{s}$. When falsification is possible, however, the discontinuity of this first-best allocation rule at $\hat{s}$ would lead agents just below to falsify to $\hat{s}$. In fact, any discontinuity in the allocation rule generates falsification. To prevent it, an allocation rule must therefore be continuous. More generally, it must inherit some of the regularity of the cost function. In our case, the *regularity* assumption on the cost function combined with the falsification-proofness constraint (FPC) directly imply Lipschitz continuity of any feasible allocation rule.

Given the score-monotonicity of expected worthiness, it is natural to expect optimal allocation rules must be monotonic. Indeed, by replacing any nonmonotonic feasible allocation rule α by the highest monotonic allocation rule everywhere below α to the left of the eligibility threshold $\hat{s}$, and by the lowest monotonic allocation rule everywhere above α to the right of $\hat{s}$, we obtain a monotonic allocation rule that remains feasible and strictly increases social surplus as it increases the probability of allocation for scores with positive social surplus, and decreases it for scores with negative social surplus.[5]

Monotonicity implies downward falsification-proofness constraints are satisfied, so we only retain upward constraints. Lipschitz and monotonic continuous allo-

[5]These results are formally stated and proved in Lemma 4 of the Appendix.

cation rules are almost everywhere differentiable, with derivative α' bounded on the interval $[0, \Lambda/\gamma]$. Furthermore, we can rewrite them according to either of the following *integral decompositions*

$$\alpha(s) = \underline{\alpha} + \int_{\underline{s}}^{s} \alpha'(z)dz \tag{$\underline{\text{ID}}$}$$

$$= \overline{\alpha} - \int_{s}^{\overline{s}} \alpha'(z)dz, \tag{$\overline{\text{ID}}$}$$

where $\underline{\alpha} = \alpha(\underline{s})$ and $\overline{\alpha} = \alpha(\overline{s})$. In particular, we can rewrite the auxiliary program as an optimization problem over the bounded function $\alpha'(s)$ and either of the scalars $\overline{\alpha}$ or $\underline{\alpha}$, instead of optimizing directly on α. We call any program obtained this way a differential form of the auxiliary program, or for convenience a *differential program*.

Cumulative surplus functions. Before writing the differential form of the auxiliary program, we introduce *cumulative surplus* functions to help us interpret the equations. The *upward cumulative surplus* is the total amount of social surplus contained above some cutoff s, and it corresponds to the marginal gain of uniformly increasing the allocation probability of all scores above s:

$$\mathcal{W}^{+}(s, \hat{w}) = \int_{s}^{\overline{s}} \{w(x) - \hat{w}\} dF(x).$$

When using ($\underline{\text{ID}}$), and constructing the allocation rule from the left, locally increasing the allocation probability at z by $\alpha'(z)dz$ has a marginal social gain of $\mathcal{W}^{+}(z, \hat{w})$. Indeed, this is reflected into the equations by replacing α with ($\underline{\text{ID}}$) in the designer's objective, and integrating by parts, which yields the differential objective function

$$(\bar{w} - \hat{w})\underline{\alpha} + \int_{\underline{s}}^{\overline{s}} \alpha'(z)\mathcal{W}^{+}(z, \hat{w})dz. \tag{$\underline{\text{DOF}}$}$$

The *downward cumulative surplus* is the total amount of negative social surplus contained below some cutoff s, and corresponds to the marginal gain of uniformly decreasing the allocation probability of all scores below$s$:

$$\mathcal{W}^{-}(s, \hat{w}) = - \int_{\underline{s}}^{s} \{w(x) - \hat{w}\} dF(x) = \mathcal{W}^{+}(s, \hat{w}) - \big(\bar{w} - \hat{w}\big).$$

When using ($\overline{\text{ID}}$), locally decreasing the allocation probability at z by $\alpha'(z)dz$ has a marginal social gain of $\mathcal{W}^-(z,\hat{w})$. Rewriting the designer's objective function with ($\overline{\text{ID}}$) and integration by parts then yields

$$(\bar{w}-\hat{w})\overline{\alpha}+\int_{\underline{s}}^{\overline{s}}\alpha'(z)\mathcal{W}^-(z,\hat{w})dz. \tag{$\overline{\text{DOF}}$}$$

In what follows, we use different notions of cumulative surplus depending on priority. Hence, it is useful to define a composite *cumulative surplus function* that encompasses both cases:

$$\mathcal{W}(z,\hat{w})=\mathcal{W}^+(z,\hat{w})\,\mathbb{1}_{\bar{w}<\hat{w}}+\mathcal{W}^-(z,\hat{w})\,\mathbb{1}_{\bar{w}\geq\hat{w}}\,.$$

Cumulative surplus and matching scores. The following lemma lists important properties of the cumulative surplus function which are illustrated on Figure 1.

Lemma 2 (Properties of cumulative surplus)**.** *The cumulative surplus function $\mathcal{W}(\cdot,\hat{w})$ is continuous and single-peaked at $\hat{s}$. For every $\nu\in[0,\mathcal{W}(\hat{s},\hat{w})]$, there exist unique scores $s_*(\nu)\leq\hat{s}\leq s^*(\nu)$ such that $\mathcal{W}\big(s_*(\nu),\hat{w}\big)=\mathcal{W}\big(s^*(\nu),\hat{w}\big)=\nu$. Furthermore:*

(i) $\mathcal{W}(s,\hat{w})\geq\nu$ *if and only if* $s\in\big[s_*(\nu),s^*(\nu)\big]$,

(ii) $s_*(\nu)$ *and* $-s^*(\nu)$ *are continuous and increasing functions,*

(iii) $s^*(0)=\overline{s}$ *under low and neutral priority, and* $s_*(0)=\underline{s}$ *under high and neutral priority,*

(iv) For all $\nu\in[0,\mathcal{W}(\hat{s},\hat{w})]$, $\mathbb{E}\big(w|s_*(\nu)\leq s\leq s^*(\nu)\big)=\hat{w}$.

These properties allow us to define the decreasing matching function $m:[s_*(0),\hat{s}]\to[\hat{s},s^*(0)]$ that to each $s\in[s^*(0),\hat{s}]$ associates the score $m(s)\in[\hat{s},s^*(0)]$ such that $\mathcal{W}(s,\hat{w})=\mathcal{W}\big(m(s),\hat{w}\big)$. We say a pair (s_*,s^*) is a matching pair if $s^*=m(s_*)$. By point (iv) of Lemma 2, matching pairs characterize the set of intervals $[s_*,s^*]$ around $\hat{s}$ that satisfy the following *Zero Average Social Surplus* condition

$$\mathbb{E}\big(w-\hat{w}|s_*\leq s\leq s^*\big)=0. \tag{ZASS}$$

Using the differential program, which we derive in the next paragraph, we show possible growth intervals are characterized by this condition.

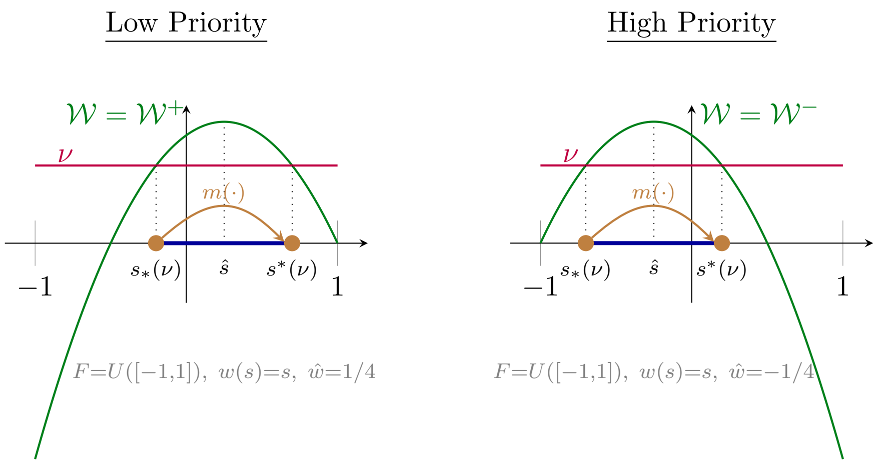

Figure 1: *Cumulative surplus, matching pairs and growth interval.*

Priorities and the differential program. Intuitively, priority determines whether it is more important for the designer to avoid allocation errors for high-score agents (in high priority groups), or low-score agents (in low-priority groups). Indeed, in the absence of information about score, the designer would allocate all objects randomly to agents from a high-priority group, and retain all objects when facing a low-priority group. Hence, the designer should set $\overline{\alpha} = 1$, and construct the test from the right, in high priority groups, but set $\underline{\alpha} = 0$, and construct the test from the left, in low priority groups. This intuition is confirmed by considering, respectively, ($\overline{\text{DOF}}$) for a high priority group, and ($\underline{\text{DOF}}$) for a low priority group. Under neutral priority, the first term is null in both ($\overline{\text{DOF}}$) and ($\underline{\text{DOF}}$), so either approach can be used, and the particular choice of allocation probability at the top or the bottom is irrelevant.

Lemma 3 (Differential program). *In* ($\tilde{\text{P}}$)*, it is optimal to set* $\underline{\alpha} = 0$ *if the group has low priority, and* $\overline{\alpha} = 1$ *if the group has high priority. Under neutral priority, either choice works. In all cases, the differential program simplifies to*

$$\max_{\alpha'} \int_{\underline{s}}^{\overline{s}} \alpha'(z)\mathcal{W}(z, \hat{w})dz$$

$$s.t. \int_{\underline{s}}^{\overline{s}} \alpha'(z)dz \leq 1 \tag{DPC}$$

$$\int_{s}^{t} \alpha'(z)dz \leq \frac{1}{\gamma}c(t|s), \quad \forall s < t \tag{DFPIC}$$

$$0 \leq \alpha'(s), \quad \forall s. \tag{DMC}$$

To complete the argument, we need to ensure the choices of $\underline{\alpha} = 0$, or $\overline{\alpha} = 1$ do not tighten the constraints. The differential falsification-proofness constraint is written as (DFPIC) in all cases and is therefore unaffected by these choices, whereas the *differential probability constraint* can be written either as $\underline{\alpha} + \int_{\underline{s}}^{\overline{s}} \alpha'(z)dz \leq 1$, or as $\overline{\alpha} - \int_{\underline{s}}^{\overline{s}} \alpha'(z)dz \geq 0$. Hence, setting either $\underline{\alpha} = 0$ under low priority, or $\overline{\alpha} = 1$ under high priority relaxes the probability constraint, and collapses it to (DPC) in all cases. We have also added the differential monotonicity constraint (DMC) since it must be satisfied by optimal allocation rule.

Probability constraint, growth interval, and simplified program. The second step of simplification allows us, first, to relax the probability constraint by considering the corresponding Lagrangian, and, second, to argue solutions to ($\tilde{\text{P}}$) must be constant outside of a *growth interval*. We then show possible growth intervals are nested intervals characterized by the *Zero Average Social Surplus* condition (ZASS). To construct a solution, we solve the relaxed problem *reduced* to a growth interval, that is, ignoring scores that lie outside of the growth interval. We call the resulting program a *simplified problem*, and show how to construct a solution to ($\tilde{\text{P}}$) by solving the simplified problem.

The probability constraint (DPC) bounds the total growth of the allocation rule and therefore determines growth intervals. Let $\nu \geq 0$ be the Lagrange multiplier on this constraint. The Lagrangian of the differential program is then

$$\mathcal{L}(\alpha, \nu) = \int_{\underline{s}}^{\overline{s}} \alpha'(z)\big\{\mathcal{W}(z, \hat{w}) - \nu\big\}dz + \nu.$$

By Lemma 2, (i), maximizing this Lagrangian under (DMC) implies setting $\alpha'(s) = 0$ for a.e. s outside of the growth interval $\big[s_*(\nu), s^*(\nu)\big]$ (see Figure 1 for an illustration). In particular, this implies growth intervals must be defined by matching pairs, or equivalently they must satisfy the (ZASS) condition. Since the Lagrange multiplier is non-negative, the growth interval must be contained in $[s_*(0), s^*(0)]$, implying growth intervals also satisfy the following principle: In low priority groups, scores in the interval $[\underline{s}, s_*(0)]$ must receive the good with null probability, whereas in high priority groups, scores in the interval $[s^*(0), \overline{s}]$ must receive the good with probability one. For easy reference, we call this principle Extreme Score Error Avoidance (ESEA). Under low or high priority the (ESEA) principle is asymmetric, and avoiding errors at one extreme may come at the cost of making errors at the other extreme. Under neutral priority, $[s_*(0), s^*(0)] = [\underline{s}, \overline{s}]$, and the designer is indifferent between errors at the top and at the bottom.

Combining these observations and a Lagrange necessity and sufficiency result (see Lemma 5 in Appendix for a precise statement), we show it is possible to construct a solution to the within problem by following the ensuing procedure which focuses on solving a *simplified program* which stems from the auxiliary program by: (i) reducing it to growth intervals, and (ii) relaxing the probability constraint:

Procedure 1 (From the simplified problem to the auxiliary problem)**.** *To construct an optimal allocation rule for* $(\tilde{\mathrm{P}})$*, consider the simplified program:*

$$\max_{\alpha(s)\in\mathbb{R}} \int_{s_*}^{s^*} \{w(s) - \hat{w}\}\alpha(s)dF(s)$$
$$s.t.\ \alpha(t) - \alpha(s) \leq \frac{1}{\gamma}c(t|s) \qquad \forall s_* \leq s < t \leq s^*$$

where (s_*, s^*) *is a matching pair, and proceed as follows:*

- ***Step 1.*** *Find a solution* α *of the simplified program for any matching pair* (s_*, s^*)*. Note that such a solution is always determined up to addition of a constant.*

- ***Step 2.*** *Choose the pair* (s_*, s^*) *so either (i) the probability constraint binds if possible:* $\alpha(s^*) - \alpha(s_*) = 1$*, or (ii) according to the (ESEA) principle:* $(s_*, s^*) = \big(s_*(0), s^*(0)\big)$*.*

- ***Step 3.*** *Set the additive constant so* $\alpha(s_*) = 0$ *in a low priority group, or* $\alpha(s^*) = 1$ *in a high priority group.*

- ***Step 4.*** *Complete the allocation rule by setting* $\alpha(s) = \alpha(s_*)$ *for all* $s < s_*$*, and* $\alpha(s) = \alpha(s^*)$ *for all* $s > s^*$*.*

◇

Hence, to construct a solution of $(\tilde{\mathrm{P}})$, we first solve the simplified program on all possible growth intervals, and then choose the growth interval to be either one that saturates the probability constraint, if possible, or, otherwise, to be the largest possible growth interval $[s_*(0), s^*(0)]$.

4 Optimal within group allocation rules

Next, we follow Procedure 1 to obtain the optimal allocation rule for the auxiliary program. The only remaining constraints are the falsification-proofness constraints. The binding ones determine the shape of the allocation rule over the growth interval. We characterize optimal allocation rules under (UID) and (UDD). Under (UDD), the falsification-proofness constraints bind locally, so the problem is easily solved using a first-order approach. Under (UID), in contrast, the falsification-proofness constraints bind for far apart scores, making the first-order approach inadequate. Instead, we solve the program by drawing a connection with the dual problem of the Monge-Kantorovitch optimal transport problem. We conclude this section by constructing the optimal allocation rule for the within problem from the solution of the auxiliary problem, which essentially amounts to choosing the right value for the outside option $\hat{w}$.

4.1 Upward Increasing Differences

Optimal allocation as optimal transport. We start by drawing a connection between the simplified program and optimal transport theory. For that, let the pair (s_*, s^*) define a possible growth interval. We consider the following relaxation of the reduced problem on $[s_*, s^*]$:

$$\max_{\alpha} \int_{s_*}^{\hat{s}} \alpha(s)\{w(s) - \hat{w}\}dF(s) + \int_{\hat{s}}^{s^*} \alpha(t)\{w(t) - \hat{w}\}dF(t)$$
$$\text{s.t. } \alpha(t) - \alpha(s) \leq \frac{1}{\gamma}c(t|s), \quad \forall s_* \leq s \leq \hat{s} \leq t \leq s^*,$$

in which we only require (FPC) to hold for scores below the eligibility threshold $\hat{s}$ targeting scores above that threshold. We have also separated the objective function between scores below and above that threshold.

In our formulation of the model, we have masses of agents distributed over the space of scores, which we can think of as locations, each endowed with a certain social surplus. Alternatively, we can view the problem in terms of masses of negative or positive social surplus distributed at the different locations. Each location s then harbors a mass $|w(s) - \hat{w}|dF(s)$ of negative social surplus below, and positive social surplus above $\hat{s}$. We frame the program as a problem involving the transportation of negative social surplus to locations that harbor positive social surplus.

To see that, we start by changing the variables of this problem to identify scores (or locations) by their distance to the eligibility threshold, letting $y = \hat{s} - s$ for $s \leq \hat{s}$, and $z = t - \hat{s}$ for $t \geq \hat{s}$. These variables belong, respectively, to the space of negative social surplus locations $Y = [0, \hat{s} - s_*]$, and the space of positive social surplus locations $Z = [0, s^* - \hat{s}]$. By the (ZASS) principle, each of these spaces harbors the same mass of social surplus. We endow each of them with a probability distribution measuring the fraction of this total mass of surplus, as given by the cumulative density functions

$$P(y) = \frac{\mathcal{W}(\hat{s}, \hat{w}) - \mathcal{W}(\hat{s} - y, \hat{w})}{\mathcal{W}(\hat{s}, \hat{w}) - \mathcal{W}(s_*, \hat{w})},$$

and

$$Q(z) = \frac{\mathcal{W}(\hat{s}, \hat{w}) - \mathcal{W}(\hat{s} + z, \hat{w})}{\mathcal{W}(\hat{s}, \hat{w}) - \mathcal{W}(s^*, \hat{w})},$$

where the normalizing factors are equal by (ZASS). Note that $dP(y) \propto |w(\hat{s} - y) - \hat{w}|dF(\hat{s} - y)$, and $dQ(z) \propto |w(\hat{s} + z) - \hat{w}|dF(\hat{s} + z)$.

Finally, we rewrite allocation probabilities as location specific prices $\phi(y) = \alpha(\hat{s} - y)$, and $\psi(z) = \alpha(\hat{s} + z)$ so the program becomes (up to multiplication by the normalizing factor of P and Q)

$$\begin{aligned} \max_{\phi, \psi} \quad & \int_Z \psi(z) dQ(z) - \int_Y \phi(y) dP(y) \\ \text{s.t. } & \psi(z) - \phi(y) \leq \frac{1}{\gamma} c(\hat{s} + z | \hat{s} - y) \quad \forall y, z. \end{aligned}$$

To view this program in terms of transportation, suppose the designer is a planner who wants to support production of a locally produced good (social surplus) at locations in Z, but discourage it at locations in Y. As a result, she wishes to maximize the profit of producers at locations in Z, but minimize the profit of producers in Y. The good costs nothing to produce, but can only be produced in quantity $dQ(z)$ at $z \in Z$, and $dP(y)$ at $y \in Y$. Suppose demand exceeds supply at every location and the economy is entirely regulated so the planner can choose the price at which the good can be sold at each location. However, producers in Y can be tempted to transport their production to locations in Z at a cost if they can profit from it. The designer should then naturally be interested in the least costly routes between Y and Z. Indeed, her program is actually the dual of the optimal transport problem, which seeks to find the least costly way of transporting P to Q:

$$\min_{\zeta \in \mathcal{M}(P,Q)} \frac{1}{\gamma} \int_{Y \times Z} c(\hat{s}+z|\hat{s}-y) d\zeta(y,z),$$

where $\mathcal{M}(P,Q)$ is the set of joint distributions on $Y \times Z$ with marginals P on Y, and Q on Z.

The optimal allocation rule. By (UID), the transportation cost $c(\hat{s}+z|\hat{s}-y)$ is submodular on $Y \times Z$. Under this condition, it is well known from optimal transport theory[6] that the optimal transportation plan is assortative, and precisely given by the matching function m: all surplus at y is transported to location z such that $\hat{s}+z = m\big(\hat{s}-y\big)$. In terms of our original problem, this implies the binding falsification-proofness constraints are between scores $s \in [s_*, \hat{s}]$, and their matching score $t = m(s)$. Optimal transport theory also provides us with closed form formulas for the optimal *price functions* ϕ and ψ, which are uniquely determined up to a constant. We then obtain the optimal allocation rule by following Procedure 1. It is slightly difficult to parse, so we first give the formula, and then explain its different terms:

$$\alpha^*_{uid}(s,\hat{w},r) = \begin{cases} 0 & \text{if } s < s_* \\ \Gamma_{uid} I(\hat{w},r) - \frac{1}{\gamma}\int_{s_*}^{s} c_s\big(m(x)|x\big)dx & \text{if } s \in [s_*, \hat{s}] \\ 1 - \Gamma_{uid}\bar{I}(\hat{w},r) - \frac{1}{\gamma}\int_{s}^{s^*} c_t\big(x|m^{-1}(x)\big)dx & \text{if } s \in [\hat{s}, s^*] \\ 1 & \text{if } s > s^* \end{cases}.$$

On the growth interval, the allocation rule grows at speed $-\frac{1}{\gamma}c_s\big(m(x)|x\big)$ for scores x below the eligibility threshold, and at speed $\frac{1}{\gamma}c_t\big(x|m^{-1}(x)\big)$ for scores x above the eligibility threshold. Duality in the optimal transport problem implies the falsification-proofness is binding between matching scores, and in particular for the pair (s_*, s^*), so

$$\alpha^*_{uid}(s^*,\hat{w},r) - \alpha^*_{uid}(s_*,\hat{w},r) = \frac{1}{\gamma}c\big(s^*|s_*\big).$$

[6] All relevant results from optimal transport theory can be found in Galichon (2018, chapter 4).

Therefore, the growth interval $[s_*, s^*]$ is pinned down by two equations: first, the (ZASS) condition, $s^* = m(s_*)$, and, second, the *boundary condition*

$$s_* = \min\big\{s \in [s_*(0), \hat{s}] : c\big(m(s)|s\big) \leq \gamma\big\}, \tag{B$_{uid}$}$$

which uniquely pins down the growth interval, and the corresponding Lagrange multiplier $\nu = \mathcal{W}(s_*, \hat{w})$. The boundary condition makes the probability constraint bind if possible, and otherwise picks the largest possible growth interval $[s_*(0), s^*(0)]$, and $\nu = 0$.

Letting $\gamma^0_{uid} = c\big(s^*(0)|s_*(0)\big)$, (B$_{uid}$) implies the probability constraint is binding if gaming ability is sufficiently low, that is $\gamma \leq \gamma^0_{uid}$. If instead $\gamma > \gamma^0_{uid}$, the probability constraint does not bind and there exists a positive *probability gap.* $\Gamma_{uid} \in [0, 1]$ measures the size of the *probability gap*, it is equal to 0 if $\gamma \leq \gamma^0_{uid}$, and is positive otherwise,

$$\Gamma_{uid} = 1 - \frac{1}{\gamma} c(s^*|s_*). \tag{G$_{uid}$}$$

Finally $I(\hat{w}, r) = \mathbb{1}_{\hat{w} > \bar{w}} + r\mathbb{1}_{\hat{w} = \bar{w}}$, and $\bar{I}(\hat{w}, r) = 1 - I(\hat{w}, r)$, with $r \in [0, 1]$. $I(\hat{w}, r)$ is an indicator whose role is to allocate the probability gap so as to satisfy the (ESEA) principle, that is: Under low priority, the probability gap is kept from the agents to ensure low-score agents do not receive an object, $I(\hat{w}, r) = 0$. Under high priority, the probability gap is allocated to the agents to ensure high-score agents get the good with certainty, $I(\hat{w}, r) = 1$. This only matters if the probability constraint is not binding so $\Gamma_{uid} > 0$, that is, if gaming ability is sufficiently high $\gamma > \gamma^0_{uid}$.

Under neutral priority, when positive, the probability gap can be allocated in any way, and $I(\bar{w}, r) = r \in [0, 1]$ is then the share of the probability gap that is allocated to the agents: that is, all scores receive the good with probability at least $r\Gamma_{uid}$. Note that γ^0_{uid} is a function of $\hat{w}$, and when $\hat{w} = \bar{w}$ the interval $[s_*(0), s^*(0)]$ is equal to the full support $[\underline{s}, \overline{s}]$. Therefore we define $\bar{\gamma}_{uid} = c(\overline{s}|\underline{s})$ to be the gaming ability threshold in this case. The condition for a slack probability constraint, and therefore a strictly positive probability gap under neutral priority is then $\gamma > \bar{\gamma}_{uid}$. In this case, there are multiple solutions and the set of optimal allocation rules is indexed by the choice of $r \in [0, 1]$. If $\hat{w} \neq \bar{w}$ or $\gamma \leq \bar{\gamma}$, the solution of the auxiliary problem is unique and the allocation rule is independent of r.

Theorem 1 (Solution of auxiliary problem under (UID))**.** *If the cost function* *satisfies* (UID)*, then* $\alpha^*_{uid}(s,\hat{w},r)$ *is an optimal allocation rule for* ($\tilde{\mathrm{P}}$)*. Furthermore, it is the unique optimal allocation rule, and it is independent of* r *under non-neutral priority* ($\hat{w} \neq \bar{w}$)*, or if gaming ability is sufficiently low,* $\gamma \leq \bar{\gamma}_{uid}$*. Otherwise, the set of optimal allocation rules is* $\left\{\alpha^*_{uid}(s,\bar{w},r)\right\}_{r\in[0,1]}$*.*

To complete the proof of the theorem, we show in the appendix that the relaxed falsification-proofness constraints between scores on the same side of the eligibility threshold $\hat{s}$ are indeed satisfied by α^*_{uid}, which is a consequence of (UID).

4.2 Upward Decreasing Differences

Under (UDD), the falsification-proofness constraint can be replaced by the first-order condition for $\alpha(t) - \frac{1}{\gamma}c(t|s)$ to be maximized at s, which is

$$\alpha'(s) \leq \frac{1}{\gamma}c_{t^+}(s|s),$$

where $c_{t^+}(s|s)$ is the right-derivative of c with respect to target t evaluated at $t = s$. Therefore, the simplified problem in its differential form is

$$\max_{0\leq\alpha'(s)\leq\frac{1}{\gamma}c_{t^+}(s|s)} \int_{s_*}^{s^*} \left[\mathcal{W}(s,\hat{w}) - \nu\right]\alpha'(s)ds,$$

where $\nu = \mathcal{W}(s_*,\hat{w}) = \mathcal{W}(s^*,\hat{w})$. Since $\mathcal{W}(s,\hat{w}) - \nu > 0$ on the interior of $[s_*, s^*]$, the only solution is to set $\alpha'(s) = \frac{1}{\gamma}c_{t^+}(s|s)$ for almost every s. Then, using Procedure 1 to build the unique optimal allocation rule, we obtain:

$$\alpha^*_{udd}(s,\hat{w},r) = \begin{cases} 0 & \text{if } s < s_* \\ \Gamma_{udd}I(\hat{w},r) + \frac{1}{\gamma}\int_{s_*}^{s} c_{t^+}(x|x)dx & \text{if } s \in [s_*, s^*] \\ 1 & \text{if } s > s^* \end{cases}.$$

On the growth interval, the allocation rule grows at speed $c_{t^+}(x|x)$ at score x. The growth interval $[s_*, s^*]$ is pinned down by the (ZASS) condition, $s^* = m(s_*)$, and the *boundary condition*

$$s_* = \min\left\{s \in [s_*(0), \hat{s}] : \int_{s}^{m(s)} c_{t^+}(x|x)dx \leq \gamma\right\} \qquad (\mathrm{B}_{udd})$$

plays the same role as in the (UID) case: it picks a value[7] of the Lagrange multiplier that makes the probability constraint bind, or otherwise sets $\nu = 0$.

The probability gap is now given by

$$\Gamma_{udd} = 1 - \frac{1}{\gamma} \int_{s_*}^{s^*} c_{t^+}(x|x)dx. \tag{G$_{udd}$}$$

The indicator $I(\hat{w}, r)$ is defined as in the (UID) case for $r \in [0, 1]$, and plays the same role of allocating the probability gap according to priority and the (ESEA) principle.

In the (UDD) case, the gaming ability thresholds are defined as

$$\gamma^0_{udd} = \int_{s_*(0)}^{s^*(0)} c_{t^+}(x|x)dx,$$

which depends on $\hat{w}$, and $\bar{\gamma}_{udd} = \int_{\underline{s}}^{\overline{s}} c_{t^+}(x|x)dx$ for the case $\hat{w} = \bar{w}$. As in the (UID) case, the probability constraint binds if and only if $\gamma \leq \gamma^0_{udd}$.

Theorem 2 (Solution of auxiliary problem under (UDD))**.** *If the cost function satisfies* (UDD)*, then $\alpha^*_{udd}(s, \hat{w}, r)$ is an optimal allocation rule for* ($\tilde{\mathrm{P}}$)*. Furthermore, it is the unique optimal allocation rule, and it is independent of r under non-neutral priority ($\hat{w} \neq \bar{w}$), or if gaming ability is sufficiently low, $\gamma \leq \bar{\gamma}_{udd}$. Otherwise, the set of optimal allocation rules is $\left\{\alpha^*_{udd}(s, \bar{w}, r)\right\}_{r\in[0,1]}$.*

Note that the optimal allocation rule α^*_{udd} is flat if $c_{t^+}(x|x) = 0$ for almost every x, that is if a marginal falsification is uniformly costless. Then the optimal rule is to allocate to all scores under high priority, and to never allocate under low priority. This is, for example, the case with the quadratic cost function $c(t|s) = (t - s)^2$.

4.3 Solution of the within problem

In this section, we start by studying how the solutions of the auxiliary problem ($\tilde{\mathrm{P}}$) vary with $\hat{w}$ and r, and use these properties to provide a unique solution for the within problem. We denote the solution of ($\tilde{\mathrm{P}}$) by $\alpha^*(s, \hat{w}, r)$, without precision on the assumption on upward differences of the cost function unless necessary.

Intuitively, a higher value of the outside option leads to a lower ex ante probability of allocating objects in the auxiliary problem. We start by showing a

[7] Multiplicity for the Lagrange multiplier cannot be excluded if $c_{t^+}(x|x) = 0$ on a set of positive measure, but even in this case the optimal allocation rule is unique.

stronger result: the optimal allocation rule is decreasing in $\hat{w}$ for every score s.

Proposition 1 (Properties of solutions of the auxiliary problem)**.** *The solution to the auxiliary problem, $\alpha^*(s, \hat{w}, r)$, is decreasing in $\hat{w}$. It is continuous in $\hat{w}$ and independent of r for $\hat{w} \neq \bar{w}$. If $\gamma \leq \bar{\gamma}$, it is also continuous and independent of r at $\hat{w} = \bar{w}$. If instead $\gamma > \bar{\gamma}$, $\alpha^*(s, \bar{w}, r)$ is strictly decreasing and continuous in r. Furthermore, it satisfies $\lim_{\hat{w} \to \bar{w}^-} \alpha^*(s, \hat{w}, r) = \alpha^*(s, \bar{w}, 1)$ and $\lim_{\hat{w} \to \bar{w}^+} \alpha^*(s, \hat{w}, r) = \alpha^*(s, \bar{w}, 0)$.*

Let $A^*(\hat{w}, r) = \int_S \alpha^*(s, \hat{w}, r) dF(s)$ denote the ex ante probability of allocation under the optimal allocation rule $\alpha^*(s, \hat{w}, r)$. The next result is a corollary of Proposition 1.

Corollary 1. *The ex ante allocation probability $A^*(\hat{w}, r)$ is strictly decreasing in $\hat{w}$. It is continuous in $\hat{w}$ and independent of r for $\hat{w} \neq \bar{w}$. If $\gamma \leq \bar{\gamma}$, it is also continuous and independent of r at $\hat{w} = \bar{w}$. If instead $\gamma > \bar{\gamma}$, $A^*(\bar{w}, r)$ is strictly decreasing and continuous in r, and satisfies $\lim_{\hat{w} \to \bar{w}^-} A^*(\hat{w}, r) = A^*(\bar{w}, 1)$ and $\lim_{\hat{w} \to \bar{w}^+} A^*(\hat{w}, r) = A^*(\bar{w}, 0)$.*

By assumption, $w(s)$ is bounded. It is easy to see that, for any $\hat{w}$ below the lower bound on $w(s)$, the unique optimal allocation rule is to allocate an object with certainty regardless of scores, while for any $\hat{w}$ above the upper bound on $w(s)$, the optimal allocation rule is to never allocate an object with positive probability. Hence, by varying $\hat{w}$ between these bounds, we can find an outside option $\hat{w}(\rho)$ such that the optimal allocation rule satisfies the resource constraint, and therefore solves (P). If $\hat{w}(\rho) = \bar{w}$ and $\gamma > \bar{\gamma}$, we also need to adjust r to a unique value $r(\rho)$ so as to allocate exactly ρ objects. The allocation rule $\alpha^*(s, \hat{w}(\rho), r(\rho))$ is then the unique solution to the within problem.

Theorem 3 (Optimal within group allocation)**.** *For any $0 \leq \rho \leq \mu$, there exists a unique outside option value $\hat{w}(\rho)$ and, if $\hat{w}(\rho) = \bar{w}$ and $\gamma > \bar{\gamma}$, a unique value $r(\rho)$, such that $\mu A^*\big(\hat{w}(\rho), r(\rho)\big) = \rho$. Furthermore, $\hat{w}(\rho)$ is continuous, decreasing in ρ if $\hat{w}(\rho) \neq \bar{w}$ or $\gamma \leq \bar{\gamma}$, and constant at $\bar{w}$ otherwise. The function $r(\rho)$ is continuous and strictly decreasing. The allocation rule $\alpha^*(s, \hat{w}(\rho), r(\rho))$ is then the unique solution to the within problem (P). The value function of (P), $W(\rho)$ is strictly concave at ρ if $\hat{w}(\rho) \neq \bar{w}$ or $\gamma \leq \bar{\gamma}$.*

5 Optimal across group allocation

We first characterize the solutions to the across problem, and provide an algorithm to obtain the optimal allocation profile $\boldsymbol{\rho} = (\rho_i)_{i \in I}$. We then analyze the effect of changes in the characteristics of the groups on the designer's payoff interpreted as social welfare. Finally, we study the effects of changes in the characteristics of the groups on the optimal allocation profile and on agents' payoffs.

5.1 Solving the across problem

Recall the across problem is

$$\overline{W}(\boldsymbol{F}, \boldsymbol{\gamma}) = \max_{\boldsymbol{\rho} \in R} \sum_i \mu_i W_i(\rho_i), \tag{$\overline{\mathrm{P}}$}$$

where ρ_i is the mass of objects allocated to group i, $\boldsymbol{\rho} = (\rho_i)_{i \in I}$ is the allocation profile which must belong to the feasible set $R = \{\boldsymbol{\rho} : \sum_i \rho_i \leq \bar{\rho},\ \rho_i \geq \phi_i \bar{\rho}\ (\forall i)\}$, and $W_i(\rho_i)$ is the value function of the within problem.

Theorem 4 (Optimal across group allocation)**.** *The across problem* ($\overline{\mathrm{P}}$) *admits a solution* $\boldsymbol{\rho}$*. Furthermore,* $\boldsymbol{\rho}$ *solves the across problem if and only if a scalar* $\lambda_R \geq 0$ *and, for each* i*, a scalar* $\lambda_i \geq 0$*, an outside option value* $\hat{w}_i(\rho_i)$*, and a gap share* $r_i(\rho_i)$ *exist such that:*

(i) $\lambda_i(\phi_i \bar{\rho} - \rho_i) = 0$ *for all* i*,*

(ii) $\lambda_R(\sum_i \rho_i - \bar{\rho}) = 0$*,*

(iii) $\hat{w}_i(\rho_i) = \lambda_R - \lambda_i$*,*

(iv) $\mu_i A_i^*\big(\hat{w}_i(\rho_i), r_i(\rho_i)\big) = \rho_i$*.*

The solution $\boldsymbol{\rho}$ *is unique if, for each* i*,* $\hat{w}_i(\rho_i) \neq \bar{w}_i$ *or* $\gamma_i \leq \bar{\gamma}_i$*.*

This characterization suggests the following algorithm to find a solution of the across problem. First, we start by computing the solutions to each within problem when setting all outside options to 0. For these solutions, we check which constraints are binding or violated. Next, we adjust outside options to satisfy all the previously violated constraints with equality when recomputing the corresponding solution. That may lead to hitting additional constraints. Indeed, increasing allocation to one group so as to satisfy its quota may violate a previously slack resource constraint, or violate another group's quota constraint if the resource constraint was already binding. If so, we adjust outside options to satisfy with equality all constraints either binding or violated at any previous step. Because the set of constraints that require adjustment at some step is bounded and grows at every additional step, the process must end, and the allocation profile at which it ends is a solution to the across problem.

Algorithm 1 provides a formal version of this algorithm in the simple case where, for every group i, $\gamma_i \leq \bar{\gamma}_i$. In this case, we can control allocated mass solely through the outside options $\hat{\boldsymbol{w}}$, and do not need to use $\boldsymbol{r}$. To understand the algorithm, we first introduce some preliminary definitions. For any allocation profile $\boldsymbol{\rho} \in [0,1]^{|I|}$, $\mathcal{Q}(\boldsymbol{\rho}) = \{i \in I : \rho_i \leq \phi_i \bar{\rho}\}$ is the set of groups whose quota constraint is binding or violated, and $\mathcal{R}(\boldsymbol{\rho}) = \mathbb{1}_{\sum_i \rho_i \geq \bar{\rho}}$ indicates whether the resource constraint is binding or violated. Finally, define $\hat{w}_i^{\phi}$ to be the unique value of $\hat{w}$ such that $\mu_i A_i^*(\hat{w}, r) = \phi_i \bar{\rho}$.

Algorithm 1: Simplified algorithm to solve the across problem

---

For each group i, set ρ_i^0 to be its optimally allocated mass of objects under the initial outside option $\hat{w}_i = 0$, that is $\rho_i^0 \leftarrow \mu_i A_i^*(0,1)$;
Let $R^0 \leftarrow \mathcal{R}(\boldsymbol{\rho}^0)$ indicate whether the resource constraint is then violated or binding, and $Q^0 \leftarrow \mathcal{Q}(\boldsymbol{\rho}^0)$ indicate the set of groups whose quota constraint is violated or binding;
Initiate counter: $k \leftarrow 0$;
repeat
  Iterate counter: $k \leftarrow k+1$;
  For groups with violated or binding quota constraints $\ell \in Q^{k-1}$, set $\hat{w}_\ell^k \leftarrow \hat{w}_\ell^\phi$ so they get exactly their quota ϕ_ℓ;
  if *the resource constraint was not binding* ($R^{k-1} = 0$) **then**
    For groups with a slack quota constraints, $\ell \notin Q^{k-1}$, keep the outside option at 0: $\hat{w}_\ell^k \leftarrow 0$;
  else
    For groups with a slack quota constraint, $\ell \notin Q^{k-1}$, set $\hat{w}_\ell^k \leftarrow \hat{w}$ where $\hat{w}$ is the unique solution of $\sum_{\ell \in Q^{k-1}} \phi_\ell \bar{\rho} + \sum_{\ell \notin Q^{k-1}} \mu_\ell A_\ell^*(\hat{w}, 1) = \bar{\rho}$;
  For all groups compute the corresponding mass of optimally allocated objects $\rho_i^k \leftarrow \mu_i A_i^*(\hat{w}_i^k, 1)$, and let $R^k \leftarrow \mathcal{R}(\boldsymbol{\rho}^k)$ and $Q^k \leftarrow \mathcal{Q}(\boldsymbol{\rho}^k)$ be the corresponding constraint indicators;
until *constraint indicators are stable:* $Q^k = Q^{k-1}$ *and* $R^k = R^{k-1}$;

In Appendix B, we provide the algorithm for the general case and show it finds a solution of the across problem.

Welfare. Next, we show social welfare is decreasing in gaming ability, and increasing with first-order stochastic dominance shifts of the score distribution.

Proposition 2 (Properties of the designer's value function)**.** *The value function of the across problem* $\overline{W}(\boldsymbol{F}, \boldsymbol{\gamma})$ *is nonincreasing in* γ_i*, and nondecreasing in* F_i *with respect to the first-order stochastic dominance ordering.*

The effect of gaming ability is simple to analyze as increasing it for any group tightens the (FPC) constraints, and thus reduces welfare. The effect of score distributions is more difficult to analyze as a first-order stochastic dominance shift in the score distribution of one group may have complicated effects on the optimal allocation rule for that group, as well as cross-group effects on allocation rules. However, an envelope theorem argument implies we can bypass the analysis of the effects of the shift on the allocation rule, and instead focus on the effect on welfare holding the optimal allocation rule fixed. Even then, the effect remains complicated to analyze because the surplus function $w_i(s) - \hat{w}_i$ takes positive and negative values on S_i. However, an argument based on the analysis of cumulative surplus functions and the differential form of the objective function shows improving the score distribution in the first-order stochastic dominance order always increases welfare under a fixed allocation rule.

5.2 Comparative statics and economic lessons

5.2.1 Comparative statics in the unconstrained problem

In this section, we provide comparative statics for the solution to the unconstrained problem, that is $\bar{\rho} = 1$. This setup is a good description of situations in which objects are prizes or labels, for which the limitation is not availability but the agent's *worthiness*, as captured by the social value function $w_i(s)$ and the designer's exogenous outside option worth 0. For any group for which there is no quota ($\phi_i = 0$), an optimal allocation rule is then given by $\alpha^*(s, 0, r)$. It is unique if $\bar{w} \neq 0$ or $\gamma \leq \bar{\gamma}$. We provide the comparative statics on $\alpha^*(s, \hat{w}, r)$ for any fixed outside option value $\hat{w}$, as it works the same way, and this analysis is also the first step in analyzing comparative statics in constrained problems.

Gaming ability. We show higher gaming ability favors low-score agents, and hurts high-score agents, if gaming ability is initially low. When initial gaming ability is sufficiently high, increasing it further affects all scores in the same way, favoring them if the group has high priority, and hurting them under low priority.

These comparative statics can also be interpreted in terms of externalities: because they pin down the falsification-proofness constraint, and therefore the optimal allocation rule, agents with the least falsification costs exert an externality on other agents. We can think of decreasing gaming ability as removing these agents to the pool of candidates, and interpret the effect it has on other agents as a measure of the negative externality of these least cost agents. For example, when gaming ability is initially low, our result says agents with the highest gaming ability have a positive externality on low score agents, but a negative externality on high score agents.

Proposition 3. *Consider increasing gaming ability from γ to $\gamma' > \gamma$. Then:*

(i) ***Low gaming abilities:*** *If $\gamma < \gamma' \leq \gamma^0$, there exists a threshold $\tilde{s}$ such that $\alpha^*(s, \hat{w}, r|\gamma') \geq \alpha^*(s, \hat{w}, r|\gamma)$ for $s \leq \tilde{s}$, and $\alpha^*(s, \hat{w}, r|\gamma') \leq \alpha^*(s, \hat{w}, r|\gamma)$ for $s \geq \tilde{s}$.*

(ii) ***High gaming abilities, low priority:*** *If $\gamma' > \gamma \geq \gamma^0$, and the group has low priority, then $\alpha^*(s, \hat{w}, r|\gamma') - \alpha^*(s, \hat{w}, r|\gamma) \leq 0$ is decreasing in s and equal to 0 at $s_*(0)$.*

(iii) ***High gaming abilities, high priority:*** *If $\gamma' > \gamma \geq \gamma^0$, and the group has high priority, then $\alpha^*(s, \hat{w}, r|\gamma') - \alpha^*(s, \hat{w}, r|\gamma) \geq 0$ is decreasing in s and equal to 0 at $s^*(0)$.*

How returns to scale in fraud shape optimal allocation rules. In this section, we focus on the class $\mathcal{E}$ of Euclidean cost functions $c(t|s) = \xi\big(|\varphi(t) - \varphi(s)|\big)$ such that ξ is either concave or convex. Intuitively, more convexity captures higher economies of scale in the size of falsification. To compare Euclidean cost functions, we will need some normalization, but we first characterize the shape of optimal allocation rules under Euclidean costs in $\mathcal{E}$.

Proposition 4 (Optimal allocation rules under Euclidean cost)**.** *If ξ is convex, then the cost function satisfies* (UDD) *and the optimal allocation rule is linear in $\varphi(s)$ on $[s_*, s^*]$, taking value $\alpha^*(s) = \xi'(0)\big(\varphi(s) - \varphi(s_*)\big)$. If ξ is concave, then the* *cost function satisfies* (UID) *and the optimal allocation rule is convex in* $\varphi(s)$ *on* $[s_*, \hat{s}]$*, and concave in* $\varphi(s)$ *on* $[\hat{s}, s^*]$*.*

Next we seek to compare Euclidean cost functions, while fixing gaming ability to γ. For simplicity, we only consider cost functions in $\mathcal{E}$ such that $\varphi(s) = s$, and normalize the cost functions so that the maximum amount of falsification an agent is willing to undertake to get the good is identical for all cost functions and less than $s^*(0) - s_*(0)$. That is, we consider two cost functions $\xi, \hat{\xi}$ such that $L \equiv \xi^{-1}(\gamma) = \hat{\xi}^{-1}(\gamma) < s^*(0) - s_*(0)$. We say $\hat{\xi}$ is more convex than ξ, and denote $\hat{\xi} \succeq_{vex} \xi$ if either $\hat{\xi}$ is convex and ξ is concave, or both are concave and ξ is more concave than $\hat{\xi}$ in the usual sense, or both are convex and $\hat{\xi}$ is more convex than ξ in the usual sense.[8] We denote the corresponding optimal allocation rules by $\alpha^*, \hat{\alpha}^*$, and the growth intervals of these allocation rules by $I^*, \hat{I}^*$.

Proposition 5 (Effect of lowering economies of scale)**.** *If* $\hat{\varphi}(s) = \varphi(s) = s$ *and* $\hat{\xi}$ *is more convex than* ξ*, then:*

(i) $I^* \subseteq \hat{I}^* \subseteq [s_*(0), s^*(0)]$*. Furthermore,* $I^* = \hat{I}^* \subset [s_*(0), s^*(0)]$ *if both functions are concave.*

(ii) If $I^* \subset [s_*(0), s^*(0)]$*, there exists a threshold* $\tilde{s} \in I^*$ *such that* $\hat{\alpha}^*(s) \leq \alpha^*(s)$ *for* $s > \tilde{s}$*, and* $\hat{\alpha}^*(s) \geq \alpha^*(s)$ *for* $s < \tilde{s}$*.*

(iii) If $I^* = \hat{I}^* = [s_*(0), s^*(0)]$*, then both functions are convex and* $\hat{\alpha}^*(s) \leq \alpha^*(s)$ *for all* s *if the group has low priority, but* $\hat{\alpha}^*(s) \geq \alpha^*(s)$ *for all* s *if the group has high priority.*

In words, lower economies of scale, like higher gaming ability, favor agents with lower scores and hurt agents with higher scores. If diseconomies of scale are too strong, as in (iii), then the effect is the same for all scores, similarly to the case of high gaming abilities in Proposition 3

Distributional effects. Next, we study changes in the score distribution that lead to uniformly increasing the optimal allocation probability. For this purpose, we redefine score so that $s = w(s) - \hat{w}$, which amounts to a transformation of the score distribution F. Hence the eligibility threshold is fixed to 0. We consider

[8] That is, there exists an increasing and concave function $g : [0, 1] \to [0, 1]$ such that $\xi = g \circ \hat{\xi}$ when both are concave, or an increasing and convex function $h : [0, 1] \to [0, 1]$ such that $\hat{\xi} = h \circ \xi$ if both are convex.

two atomless score distributions $\hat{F}$ and $\tilde{F}$ whose common support $[\underline{s}, \overline{s}]$ contains a neighborhood of 0. We let the function $\Delta(s) = \tilde{F}(s) - \hat{F}(s)$ denote the change in the score distribution.

All distributional effects on the allocation rule are transmitted through the matching functions $\hat{m}(s)$ and $\tilde{m}(s)$, so we start by showing that the allocation rules satisfies $\tilde{\alpha}(s) \geq \hat{\alpha}(s)$ for every s, and every cost function that satisfies (UID) or (UDD) if and only if $\tilde{m}(s) \leq \hat{m}(s)$ for every $s \leq 0$. Then we provide a necessary and sufficient condition on the distributions for the matching function to decrease.

Note that the matching function is normally defined on $[s_*(0), 0]$, but the lower end $s_*(0)$ may now depend on which score distribution is used. To ease the exposition, we extend each matching function $\hat{m}(s)$ and $\tilde{m}(s)$ to the left by letting $\hat{m}(s) = \hat{m}\big(\hat{s}_*(0)\big)$ for $s \leq \hat{s}_*(0)$, and $\tilde{m}(s) = \tilde{m}\big(\tilde{s}_*(0)\big)$ for $s \leq \tilde{s}_*(0)$.

Proposition 6. *The following statements are equivalent:*

(a) The allocation rules satisfy $\tilde{\alpha}(s) \geq \hat{\alpha}(s)$ for every s, and every cost function that satisfies (UID) *or* (UDD)*.*

(b) The matching functions satisfy $\tilde{m}(s) \leq \hat{m}(s)$ for every $s \leq 0$.

(c) For every $0 > s \geq \max\{\hat{s}_*(0), \tilde{s}_*(0)\}$,

$$\int_s^{\hat{m}(s)} x d\tilde{F}(x) \geq \int_s^{\tilde{m}(s)} x d\hat{F}(x).$$

(d) For every $0 > s \geq \max\{\hat{s}_*(0), \tilde{s}_*(0)\}$,

$$\int_s^0 \{\Delta(s) - \Delta(x)\} dx + \int_0^{\tilde{m}(s)} \{\Delta\big(\tilde{m}(s)\big) - \Delta(x)\} dx \geq 0.$$

Since it is difficult to relate conditions (c) and (d), we provide a sufficient condition on Δ that is more easily interpretable. We say Δ *divests* an interval $I \subseteq S$ if every score in I (formally, every measurable subset of I) is less likely under $\tilde{F}$ than under $\hat{F}$. That is if, for every, $[s, s'] \subseteq I$,

$$\Delta(s') - \Delta(s) = \big\{\tilde{F}(s') - \tilde{F}(s)\big\} - \big\{\hat{F}(s') - \hat{F}(s)\big\} \leq 0,$$

or, equivalently, if Δ is nonincreasing on I. If instead Δ is nondecreasing on I, we say it *invests* I.

Proposition 7. *Suppose there exists $a \in [\underline{s}, 0)$ and $b \in (0, \overline{s}]$ such that*

1. $\Delta(a) = \Delta(b) = 0$*,* $\Delta(s) \geq 0$ *for all* $s \leq a$*, and all* $s \geq b$*;*

2. Δ *divests* $[a, 0]$ *and invests* $[0, b]$*;*

3. $\int_{\underline{s}}^{0} \Delta(x)dx \leq 0$ *and* $\int_{0}^{\overline{s}} \Delta(x)dx \leq 0$*.*

Then the allocation rules satisfy $\tilde{\alpha}(s) \geq \hat{\alpha}(s)$ *for every* s*.*

In particular, a change of distribution that replaces scores below the eligibility threshold by scores above the eligibility threshold satisfies the conditions of Proposition 7 and therefore uniformly increases the allocation probability.

5.2.2 Comparative statics and feedback effects in the constrained problem.

When some of the allocative constraints are binding, changes in the characteristics of one group can generate both cross-group and within-group feedback effects. Suppose for example that the resource constraint is binding, and the characteristics of group i change with the direct effect of increasing the mass of objects $A_i^*(\hat{w}_i, r_i)$ allocated to group i under a constant outside option $\hat{w}_i \neq \bar{w}_i$ (recall r_i plays no role in this case), corresponding to the value of the solution of the across problem before the change. This could, for example, be the result of increasing gaming ability γ_i, if it is already above $\bar{\gamma}_i$ and $\hat{w}_i < \bar{w}_i$ (see Proposition 3). The direct effect of the change of characteristics of group i then leads to a violation of the resource constraint. In order to compensate for this direct effect, the endogenous outside option of all groups including i must be adjusted upward, leading the mass allocated to all groups but i to decrease (the cross-group effect), and the mass allocated to group i to be readjusted downward (the within group feedback effect). Overall, while the mass allocated to all other groups must decrease due to the feedback effect, the mass allocated to group i must increase when combining the direct and feedback effects. Indeed, the feedback effect is only necessary to keep the resource constraint satisfied. Since the feedback effect lowers the mass allocated to all other groups with an initially slack quota constraint, this frees some resources for group i, so there is no need for the feedback effect to overcompensate the direct effect.

6 Discussion

To conclude this study, we discuss some economic implications and extensions of our framework, as well as some theoretical connections.

Discriminating on observables. Consider two groups $i = 1, 2$, that differ only in their gaming ability $\gamma_1 > \gamma_2$, and assume there are no allocative constraints, or they do not bind in any of the considered situations. The in-group (group 1) has higher gaming ability than the out-group (group 2). Under the naive selection rule which allocates the object to agents with $w_1(s) = w_2(s) \geq \hat{w}$, agents from the in-group are unfairly advantaged as they can more easily falsify to the threshold. In fact, under any allocation rule that induces some falsification but cannot discriminate across groups, the in-group must be doing better than the out-group. Under our optimal falsification-proof rule without discrimination, the outcome is *fair* in the sense that agents with equal scores are treated identically. However, agents from the in-group exert an externality on agents from the out-group. Assuming gaming abilities are sufficiently low ($\gamma_1 \leq \gamma^0$), we can use Proposition 3 to describe how discriminating across groups would change the outcome. Discrimination yields the same optimal falsification-proof allocation rule for the in-group, whereas the out-group would face a different and steeper allocation rule, favoring high scores but harming low-scores from the out-group. Whether the out-group benefits on average depends on the score distribution. However, if we consider worthiness w to be the right standard of social value from allocations, the new allocation rule faced by the out-group is then closer to the first best allocation from a social perspective. In that sense, adding observables that allow to discriminate across different levels of gaming ability is socially valuable.

Designer as a certification intermediary. In our model, we assume the designer can commit to an allocation mechanism conditioning on score and group. We show next she can attain the same outcome with less commitment power. We assume the allocation decision is delegated to a *decision maker* whose preferences may be slightly misaligned with the designer's objective. Specifically, for the decision maker, expected worthiness is given by the increasing function $\tilde{w}_i(s)$ instead of $w_i(s)$. Without loss of generality, we normalize her outside option to 0. We also assume she faces the same allocative constraints as the designer in our initial model. She can observe the group label, but has no access to the score. Instead, she relies on the designer who can commit to selectively communicate information about scores as she wishes. In this version of the model, the designer is a certification intermediary who can design an information structure but has no control of allocation decisions.

With less commitment power, the designer can only be worse off. However, she can try to emulate her full commitment payoff by using a binary-signal information structure that recommends to allocate with probability $\alpha_i^*(s)$, and to reject with probability $1 - \alpha_i^*(s)$. For this to be effective, the decision maker must find it optimal to obey the recommendation. Then we say that α^* is *obedient.* We show this is the case whenever the designer and decision maker preferences are sufficiently well aligned.

Proposition 8. *There exists $\varepsilon > 0$ such that α^* is obedient whenever $||w_i - \tilde{w}_i||_\infty < \varepsilon$ for every $i \in I$.*

Therefore, our optimal allocation rule also solves the information design problem of the designer as a certification intermediary.

Theoretical connections. We consider a hybrid setting where agents have both soft and hard pieces of private information about their characteristics. Methodologically, our approach differs from Myersonian techniques that rely on virtual surplus and work directly with the allocation rule. By contrast, we use cumulative surplus and work with the derivative of the allocation rule to identify the growth interval, which is the region of scores where allocation probability must be distorted to prevent falsification. Notably, our analysis introduces novel connections to the literature using optimal transport theory in economics. Our methodology adds to both the Myersonian techniques in settings with transfers, and the Lagrangian techniques in settings without transfers as in Amador, Werning, and Angeletos (2006) and Amador and Bagwell (2022).

Appendix

A Proofs

Lemma 4 (Smoothness and monotonicity)**.** *If an allocation rule satisfies* (FPC)*, it is Lipschitz continuous. Furthermore, if α is feasible for* ($\tilde{\mathrm{P}}$) *but not monotonic,* *then there exists a nondecreasing allocation rule $\tilde{\alpha}$ that is feasible and strictly better for* ($\tilde{\mathrm{P}}$).

Proof. If α satisfies (FPC), the regularity assumption of Definition 1 directly implies it is Lipschitz continuous. Next, define

$$\tilde{\alpha}(s) = \alpha^-(s)\, \mathbb{1}_{s \le \hat{s}} + \alpha^+(s)\, \mathbb{1}_{s \ge \hat{s}},$$

where $\alpha^- : [\underline{s}, \hat{s}] \to [0,1]$ is the largest nondecreasing function that is everywhere below α on $[\underline{s}, \hat{s}]$, and $\alpha^+ : [\hat{s}, \overline{s}] \to [0,1]$ is the lowest nondecreasing function everywhere above α on $[\hat{s}, \overline{s}]$. So $\tilde{\alpha}$ is nondecreasing.

First, we show $\tilde{\alpha}$ remains feasible. It obviously satisfies (PC). Since $\tilde{\alpha}$ is nondecreasing, we only need to check (FPC) for upward falsification, so let $s < t$, and let $s' = \max\{x \ge s : \tilde{\alpha}(x) = \tilde{\alpha}(s)\}$, and $t' = \min\{x \le t : \tilde{\alpha}(x) = \tilde{\alpha}(t)\}$. We can assume $s \le s' < t' \le t$, for otherwise $\tilde{\alpha}(t) = \tilde{\alpha}(s)$ and the proof is done. Then, we have

$$\tilde{\alpha}(t) - \tilde{\alpha}(s) = \tilde{\alpha}(t') - \tilde{\alpha}(s') = \alpha(t') - \alpha(s') \le \frac{1}{\gamma} c(t'|s') \le \frac{1}{\gamma} c(t|s),$$

where the first equality is by definition of s', t'; the second equality is because $\tilde{\alpha}$ must coincide with α wherever it is not flat, and therefore also at the end of every flat interval. The first inequality is due to falsification-proofness of α, and the last inequality to cost monotonicity.

Scores above (below) $\hat{s}$ are more (less) likely to get an object under $\tilde{\alpha}$ than under α. Hence, $\tilde{\alpha}$ is better than α for ($\tilde{\mathrm{P}}$). Furthermore, if α is not monotonic, there must exist an interval of scores for which α and $\tilde{\alpha}$ do not coincide. Since F has full support, $\tilde{\alpha}$ is therefore strictly better than α. □

Proof of Lemma 2. By strict monotonicity of w, $w(s) - \hat{w}$ and $s - \hat{s}$ have the same sign implying both $\mathcal{W}^+$ and $\mathcal{W}^-$ are increasing on $[\underline{s}, \hat{s}]$ and decreasing on $[\hat{s}, \overline{s}]$, and therefore single-peaked at $\hat{s}$. The existence of $s_*(\nu)$ and $s^*(\nu)$ is then ensured by continuity of both $\mathcal{W}^+$ and $\mathcal{W}^-$, and the fact that both functions take weakly negative values at both ends of the score interval. Then (i) and (ii) are direct consequence of single-peakedness and continuity. For (iii), note that in the low priority case we have $\mathcal{W}(\underline{s}, \hat{w}) = \mathcal{W}^+(\underline{s}, \hat{w}) = 0$, while in the high priority case

$\mathcal{W}(\overline{s}, \hat{w}) = \mathcal{W}^{-}(\overline{s}, \hat{w}) = 0$. Finally, for (iv), $\mathcal{W}(s^*(\nu), \hat{w}) = \mathcal{W}(s_*(\nu), \hat{w})$ implies:

$$\mathbb{E}\big(w | s_*(\nu) \leq s \leq s^*(\nu)\big) = \frac{\mathcal{W}(s^*(\nu), \hat{w}) - \mathcal{W}(s_*(\nu), \hat{w})}{F(s^*(\nu)) - F(s_*(\nu))} + \hat{w} = \hat{w}.$$

$\square$

Proof of Lemma 3. The objective functions ($\overline{\text{DOF}}$) and ($\underline{\text{DOF}}$) are obtained by integration by parts after using ($\overline{\text{ID}}$) and ($\underline{\text{ID}}$). To complete the argument, we also rewrite the constraints in the same way. The differential version of the falsification-proofness constraint (DFPIC) is immediate. We can add the monotonicity differential constraint (DMC) to the program without loss of generality by Lemma 4. Given (DMC) the probability constraint can be written in the two following equivalent manners:

$$0 \leq \underline{\alpha}, \text{ and } \underline{\alpha} + \int_{\underline{s}}^{\overline{s}} \alpha'(x)dx \leq 1,$$

or

$$\overline{\alpha} \leq 1, \text{ and } \overline{\alpha} - \int_{\underline{s}}^{\overline{s}} \alpha'(x)dx \geq 0.$$

Considering the program written as an optimization program on $(\underline{\alpha}, \alpha')$, it appears setting $\underline{\alpha} = 0$ in the low priority case maximizes the objective function ($\underline{\text{DOF}}$), and relaxes the probability constraint on α'. It is therefore optimal. Similarly, considering the program as an optimization program on $(\overline{\alpha}, \alpha')$, setting $\overline{\alpha} = 1$ both maximizes ($\overline{\text{DOF}}$) and relaxes the probability constraint on α'. Having simplified the program in this way results in the differential program of the lemma in each cases. $\square$

Lemma 5. *A (nondecreasing and Lipschitz) allocation rule $\hat{\alpha}$ solves ($\tilde{\text{P}}$) if and only if there exists a Lagrange multiplier $\nu \geq 0$ such that:*

(i) $\hat{\alpha}(s) = \hat{\alpha}\big(s_*(\nu)\big)$ *for every* $s \leq s^*(\nu)$*, and* $\hat{\alpha}(s) = \hat{\alpha}\big(s^*(\nu)\big)$ *for every* $s \geq s^*(\nu)$*,*

(ii) $\hat{\alpha}(t) - \hat{\alpha}(s) \leq \frac{1}{\gamma}c(t|s)$ *for every* $s_*(\nu) \leq s < t \leq s^*(\nu)$*,*

(iii) If $\bar{w} \leq \hat{w}$, $\hat{\alpha}\big(s_*(\nu)\big) = 0$ *and* $\nu\big(1 - \hat{\alpha}\big(s^*(\nu)\big)\big) = 0$*,*

(iv) If $\bar{w} \geq \hat{w}$, $\hat{\alpha}\big(s^*(\nu)\big) = 1$ *and* $\nu\hat{\alpha}\big(s_*(\nu)\big) = 0$*,*

(v) For every nondecreasing Lipschitz allocation rule α that satisfies (i) and (ii),

$$\int_{s_*(\nu)}^{s^*(\nu)} \{w(s) - \hat{w}\}\hat{\alpha}(s)dF(s) \geq \int_{s_*(\nu)}^{s^*(\nu)} \{w(s) - \hat{w}\}\alpha(s)dF(s).$$

Proof. We proceed in two steps. The first step is a standard Lagrangian necessity and sufficiency theorem. The second step ensures the conditions of the lemma are equivalent to the Lagrangian conditions. In this proof, we say an allocation rule α is feasible if it is (Λ/γ)-Lipschitz and nondecreasing, satisfies (FPC) and $\underline{\alpha} = 0$ under low priority, and $\overline{\alpha} = 1$ under high priority. It is immediate to verify that the set of such feasible allocation rules, which we denote by $\mathbb{A}$, is convex.

Step 1: *A feasible allocation rule $\hat{\alpha}$ solves the within problem if and only if there exists $\nu \geq 0$ such that (a) $\nu = 0$ or $\overline{\hat{\alpha}} - \underline{\hat{\alpha}} = 1$, and (b) $\mathcal{L}(\hat{\alpha}, \nu) \geq \mathcal{L}(\alpha, \nu)$ for every feasible allocation rule α.*

$\boxed{\Leftarrow}$ If $\nu = 0$, the conclusion is immediate. Suppose instead $\nu > 0$. Then (a) implies $\int_S \hat{\alpha}'(z)dz = 1$. Hence, for any feasible α that satisfies (DPC),

$$\int_S \hat{\alpha}'(z)\mathcal{W}(z, \hat{w})dz = \mathcal{L}(\hat{\alpha}, \nu) \geq \mathcal{L}(\alpha, \nu) \geq \int_S \alpha'(z)\mathcal{W}(z, \hat{w})dz,$$

where the last inequality is implied by $\nu > 0$ and $\int_S \alpha'(z)dz \leq 1$.

$\boxed{\Rightarrow}$ For every $b \geq 0$, consider the program where we replace the probability constraint (DPC) by the constraint $g(\alpha) \leq b$ where $g(\alpha) = \int_S \alpha'(z)dz$. Let its value be

$$h(b) = \max_{\alpha \in \mathbb{A}} \ \Omega(\alpha) \ \text{ s.t. } \int_S \alpha'(z)dz \leq b,$$

where $\Omega(\alpha) = \int_S \alpha'(z)\mathcal{W}(z, \hat{w})dz$. Since the objective $\Omega(\cdot)$ and the constraint $g(\cdot)$ are both linear in α', and $\mathbb{A}$ is convex, $h(b)$ is a concave function. It is also obviously nondecreasing. Let $\nu \geq 0$ be the left-derivative of h at $b = 1$, which exists by concavity and is nonnegative by monotonicity.

By assumption, we have $h(1) = \Omega(\hat{\alpha})$. If $g(\hat{\alpha}) = 1$, then we also have $\Omega(\hat{\alpha}) = \mathcal{L}(\hat{\alpha}, \nu)$. Otherwise, we must have $g(\hat{\alpha}) < 1$. But then $\hat{\alpha}$ must also solve the program for any $b \in [g(\hat{\alpha}), 1]$, implying h is constant on this interval, and $\nu = 0$. Then again, $\Omega(\hat{\alpha}) = \mathcal{L}(\hat{\alpha}, \nu)$. For all $\alpha \in \mathbb{A}$, we have $\Omega(\alpha) \leq h\big(g(\alpha)\big)$ by definition of h, and $h\big(g(\alpha)\big) \leq h(1) + \nu\big(g(\alpha) - 1\big)$. Hence, $\mathcal{L}(\alpha, \nu) = \Omega(\alpha) - \nu\big(g(\alpha) - 1\big) \leq h(1) = \Omega(\hat{\alpha}) = \mathcal{L}(\hat{\alpha}, \nu)$.

Step 2: *A nondecreasing and Lipschitz allocation rule α satisfies (i)-(v) for some $\nu \geq 0$ if and only if it is feasible and satisfies (a) and (b).*

$\boxed{\Rightarrow}$ It is easy to see (i), (iii) and (iv) imply (a), and $\underline{\hat{\alpha}} = 0$ under low priority and $\overline{\hat{\alpha}} = 1$ under high priority. Next, we show that (i) and (ii) imply $\hat{\alpha}$ satisfies (FPC). Let $s < t$, and define $s' = \max\{s_*(\nu), s\}$ and $t' = \min\{s^*(\nu), t\}$. Then

$$\hat{\alpha}(t) - \hat{\alpha}(s) = \hat{\alpha}(t') - \hat{\alpha}(s') \leq \frac{1}{\gamma} c(t'|s') \leq \frac{1}{\gamma} c(t|s),$$

where the first equality is from (i), the first inequality from (ii), and the last inequality by cost monotonicity. Hence $\hat{\alpha}$ is feasible and satisfies (a).

Suppose α is feasible. Then, let $\tilde{\alpha}(s) = \big[\alpha(s)+a\big]_0^1$, where $a = -\alpha\big(s_*(\nu)\big)\, \mathbb{1}_{\bar{w}<\hat{w}} + \big(1-\alpha\big(s^*(\nu)\big)\big)\, \mathbb{1}_{\bar{w}\geq\hat{w}}$, and $[z]_0^1 = z\, \mathbb{1}_{0\leq z\leq 1} + \mathbb{1}_{z>1}$. Then $\tilde{\alpha}$ satisfies (i) and (ii), and (b) follows from:

$$\begin{aligned}
\mathcal{L}(\hat{\alpha}, \nu) &= \int_S \hat{\alpha}'(z)\mathcal{W}(z, \hat{w})dz && \text{(by (i), and (a))}\\
&= \nu\big(\overline{\hat{\alpha}} - \underline{\hat{\alpha}}\big) + \int_{s_*(\nu)}^{s^*(\nu)} \{w(s) - \hat{w}\}\hat{\alpha}(s)dF(s) \\
& && \text{(by integration by parts and Lemma 2)}\\
&= \nu + \int_{s_*(\nu)}^{s^*(\nu)} \{w(s) - \hat{w}\}\hat{\alpha}(s)dF(s) && \text{(by (a))}\\
&\geq \nu + \int_{s_*(\nu)}^{s^*(\nu)} \{w(s) - \hat{w}\}\tilde{\alpha}(s)dF(s) && \text{(by (v))}\\
&= \nu + \int_{s_*(\nu)}^{s^*(\nu)} \{w(s) - \hat{w}\}\alpha(s)dF(s) + a \underbrace{\int_{s_*(\nu)}^{s^*(\nu)} \{w(s) - \hat{w}\}dF(s)}_{=0} \\
& && \text{(by Lemma 2)}\\
&= \nu + \int_{s_*(\nu)}^{s^*(\nu)} \alpha'(z)\mathcal{W}(z, \hat{w})dz - \nu\big\{\alpha\big(s^*(\nu)\big) - \alpha\big(s_*(\nu)\big)\big\} \\
& && \text{(by integration by parts)}\\
&= \nu + \int_{s_*(\nu)}^{s^*(\nu)} \alpha'(z)\big[\mathcal{W}(z, \hat{w}) - \nu\big]dz \\
&\geq \nu + \int_S \alpha'(z)\big[\mathcal{W}(z, \hat{w}) - \nu\big]dz = \mathcal{L}(\alpha, \nu), \\
& && \text{(as } \mathcal{W}(z, \hat{w}) < \nu \text{ for } z \notin [s_*(\nu), s^*(\nu)])
\end{aligned}$$

where we use the relation $\mathcal{W}\big(s_*(\nu),\hat{w}\big)=\mathcal{W}\big(s^*(\nu),\hat{w}\big)=\nu$ from Lemma 2.

$\boxed{\Leftarrow}$ Feasibility directly implies (ii) as $\hat{\alpha}$ must satisfy (FPC). By Lemma 2, maximizing $\mathcal{L}(\alpha,\nu)$ implies setting $\alpha'(s)$ to 0 for almost every s outside of $[s_*(\nu),s^*(\nu)]$ which implies (i). Feasibility and (i), then imply the first equality in (iii) and (iv). If $\nu=0$, the second equality is automatically satisfied, otherwise, it is satisfied by (a). Consider any nondecreasing and Lipschitz allocation rule α that satisfies (i) and (ii). Then it is feasible, and

$$
\begin{aligned}
\int_{s_*(\nu)}^{s^*(\nu)}\alpha(s)\{w(s)-\hat{w}\}dF(s) &= \int_{s_*(\nu)}^{s^*(\nu)}\alpha'(z)\mathcal{W}(z,\hat{w})dz-\nu\big\{\alpha\big(s^*(\nu)\big)-\alpha\big(s_*(\nu)\big\} \\
&\qquad\qquad\qquad\qquad\text{(by integration by parts)} \\
&= \int_{s_*(\nu)}^{s^*(\nu)}\alpha'(z)\big[\mathcal{W}(z,\hat{w})-\nu\big]dz \\
&= \int_{S}\alpha'(z)\big[\mathcal{W}(z,\hat{w})-\nu\big]dz && \text{(by (i))} \\
&= \mathcal{L}(\alpha,\nu)-\nu \\
&\leq \mathcal{L}(\hat{\alpha},\nu)-\nu && \text{(by (b))} \\
&= \int_{S}\hat{\alpha}'(z)\big[\mathcal{W}(z,\hat{w})-\nu\big]dz \\
&= \int_{s_*(\nu)}^{s^*(\nu)}\hat{\alpha}'(z)\big[\mathcal{W}(z,\hat{w})-\nu\big]dz && \text{(by (i))} \\
&= \int_{s_*(\nu)}^{s^*(\nu)}\alpha'\hat{(}z)\mathcal{W}(z,\hat{w})dz-\nu\big\{\hat{\alpha}\big(s^*(\nu)\big)-\hat{\alpha}\big(s_*(\nu)\big\} \\
&= \int_{s_*(\nu)}^{s^*(\nu)}\alpha(s)\{w(s)-\hat{w}\}dF(s) \\
&\qquad\qquad\qquad\qquad\text{(by integration by parts)}
\end{aligned}
$$

$\square$

Proof of Theorem 1. To keep notations simple, we only indicate the dependence of α^*_{uid} on $\hat{w},r$ when it is useful for the argument. We check that the conditions of Lemma 5 are satisfied. We pick the multiplier $\nu=\mathcal{W}(s_*,\hat{w})\geq 0$. Then α^*_{uid} clearly satisfies (i). To see it satisfies (iii) and (iv), note that $\alpha^*_{uid}(s^*)-\alpha^*_{uid}(s_*)=\frac{1}{\gamma}c(s^*|s_*)$. Hence, by ($\mathrm{B}_{uid}$), either it is equal to 1 and the probability constraint is binding, or it is strictly less than 1, and then $\nu=0$ and $(s_*,s^*)=(s_*(0),s^*(0))$.

α^*_{uid} maximizes the relaxed program of Section 4.1. To show it satisfies (v), we need to show it satisfies the falsification-proofness constraint in (ii) for any pair s, t such that $s, t \in [s_*, \hat{s}]$ or $s, t \in [\hat{s}, s^*]$. Take the first case, for example. Then

$$\begin{aligned}
\alpha^*_{uid}(t) - \alpha^*_{uid}(s) &= -\frac{1}{\gamma}\int_s^t c_s\big(m(x)|x\big)dx \\
&\leq -\frac{1}{\gamma}\int_s^t c_s(m(t)|x)dx && \text{(by (UID))} \\
&= \frac{1}{\gamma}\big\{c\big(m(t)|s\big) - c\big(m(t)|t\big)\big\} \\
&\leq \frac{1}{\gamma}c(t|s). && \text{(by (UID))}
\end{aligned}$$

The argument is similar in the second case.

For uniqueness, first note that $c(s^*(\nu)|s_*(\nu))$ is increasing in ν so there is a single value of the Lagrange multiplier that satisfies (B_{uid}), that is a single value of the Lagrange multiplier such that the necessary and sufficient conditions of Lemma 5 are satisfied. Then for this ν and the corresponding bounds (s_*, s^*), the solution to the optimal transport problem is uniquely determined up to a constant. However, this constant is uniquely determined either by the probability constraint if it binds, that is if $c(s^*|s_*) = \gamma$, or by the requirement that $\alpha^*_{uid}(s_*) = 0$ under low priority, and $\alpha^*_{uid}(s^*) = 1$ under high priority. The only case in which uniqueness fails is if we are in the neutral priority case where $\bar{w} = \hat{w}$, and the probability constraint is slack. In this case, note that $(s_*, s^*) = (s_*(0), s^*(0)) = (\underline{s}, \overline{s})$. Hence, for the probability constraint not to bind, it must be the case that $\bar{\gamma} = c(\overline{s}|\underline{s}) < \gamma$. The designer is then indifferent across all allocation rules $\alpha^*_{uid}(s, \hat{w}, r)$ for any $r \in [0, 1]$. Indeed, for $r' > r$, we have $\alpha^*_{uid}(s, \bar{w}, r') - \alpha^*_{uid}(s, \bar{w}, r) = (r' - r)\Gamma_{uid}$ so the difference in the designer's payoff is

$$(r' - r)\Gamma_{uid}\int_{\underline{s}}^{\overline{s}}\{w(s) - \hat{w}\}dF(s) = (r' - r)\Gamma_{uid}(\bar{w} - \hat{w}) = 0.$$

□

Proof of Theorem 2. To keep notations simple, we only indicate the dependence of α^*_{udd} on $\hat{w}, r$ when it is useful for the argument. Again, we only need to check the conditions of Lemma 5 are satisfied. Picking $\nu = \mathcal{W}(s_*, \hat{w})$, it is clear that (i) holds. For (ii), let $s_* \le s < t \le s^*$, then

$$\begin{aligned}
\alpha^*_{udd}(t) - \alpha^*_{udd}(s) &= \frac{1}{\gamma}\int_s^t c_{t^+}(x|x)dx \\
&\le \frac{1}{\gamma}\int_s^t c_t(x|s)dx && \text{(by (UDD))} \\
&= \frac{1}{\gamma}c(t|s).
\end{aligned}$$

This also shows that the first-order approach is valid. Furthermore, the differential program solved by α^*_{udd} is obtained from the program in (v) by using integration by parts on the objective function. Therefore (v) holds. (iii) and (iv) are immediate to check.

For uniqueness, note that while there may be several values of the Lagrange multiplier ν that work if $c_{t^+}(x|x)$ is equal to 0 both in the neighborhoods of s_* and s^*, the corresponding optimal allocation rules for the program would be identical for all such values, so uniqueness of the solution to the differential program is granted when $\gamma \le \bar{\gamma}_{udd}$ or $\hat{w} \neq \bar{w}$. In the remaining neutral priority case, the designer is indifferent across allocation rules $\alpha^*_{udd}(s, \hat{w}, r)$ for any $r \in [0, 1]$. The argument is the same as in the (UID) case. $\square$

Proof of Proposition 1. Let $\tilde{\alpha}^*(\hat{w})$ denote the correspondence mapping $\hat{w}$ to the set of solutions of the auxiliary problem. By Lemma 4, we can write ($\tilde{\mathrm{P}}$) as an optimization problem over the set of nondecreasing functions from S to $[0, 1]$ satisfying (FPC) and (PC). This space is compact by Helly's theorem, and convex, and the objective function is linear and therefore continuous in α. Hence, Berge's maximum theorem implies $\tilde{\alpha}^*(\hat{w})$ is a continuous correspondence. By Theorem 1 and Theorem 2, the correspondence is singleton-valued for $\hat{w} \neq \bar{w}$, and for $\hat{w} = \bar{w}$ if $\gamma \le \bar{\gamma}$, so the continuity results with respect to $\hat{w}$ follow.

The space of nondecreasing and feasible allocation rules is also a lattice with respect to the partial order $\alpha \succeq \beta \Leftrightarrow \alpha(s) \ge \beta(s),\ \forall s$, with the corresponding strict ordering $\alpha \succ \beta$ if $\alpha \succeq \beta$ and $\alpha(s) > \beta(s)$ for some s. Indeed, it is easy to see that, for two such allocation rules α and β, their meet $\alpha \wedge \beta$ and their join $\alpha \vee \beta$ are also nondecreasing and feasible. Furthermore, the objective function is supermodular in α and has strictly increasing differences in $(-\hat{w}, \alpha)$. Hence, by Milgrom and Shannon's monotone selection theorem (Milgrom and Shannon, 1994), $\alpha^*(\cdot, \hat{w}, r)$ is strictly decreasing in $\hat{w}$ for the $\succeq$ order (recalling the allocation rule is independent of r for $\hat{w} = \bar{w}$, the only role of r is to pin down the selection at $\hat{w} = \bar{w}$).

Furthermore, it is easy to see (by inspection) that $\alpha^*(s, \bar{w}, r)$ is strictly decreasing in r for every s, both in the (UID) and (UDD) cases. Together with the continuity of the correspondence at $\hat{w} = \bar{w}$, this implies the results on the left and right limits of $\alpha(\cdot, \hat{w}, r)$ as $\hat{w} \to \bar{w}$. □

Proof of Corollary 1. This result follows almost directly from Proposition 1. To complete the argument, we only need to notice that, since the solution $\alpha^*(s, \hat{w}, r)$ is continuous in s, the result that $\alpha^*(\cdot, \hat{w}, r) \succ \alpha^*(\cdot, \hat{w}', r)$ for $\hat{w} < \hat{w}'$, implies $\alpha^*(s, \hat{w}, r) > \alpha^*(s, \hat{w}', r)$ for all s on a subinterval of S, so $A^*(\hat{w}, r) > A^*(\hat{w}', r)$. □

Proof of Theorem 3. The function $w(s)$ is bounded by assumption. Let $w^- = w(\underline{s})$ and $w^+ = w(\bar{s})$ be its bounds. Then it is easy to see $\alpha^*(s, w^-, r) = A^*(w^-, r) = 1$ and $\alpha^*(s, w^+, r) = A^*(w^+, r) = 0$. By the continuity and strict monotonicity results of Corollary 1, it follows that there exists a unique value of $\hat{w} \in [w^-, w^+]$, and, if $\hat{w} = \bar{w}$, a unique value of $r \in [0, 1]$, such that $A^*(\hat{w}, r) = \rho/\mu$, for any $\rho \in [0, \mu]$. By Lemma 1, $\alpha^*(\cdot, \hat{w}, r)$ is then the unique solution to the within problem (P).

The continuity and monotonicity results of *Corollary* 1 also imply continuity and monotonicity of $\hat{w}(\rho)$ and $r(\rho)$.

By Lemma 1, $W(\rho)$ is concave on $[0, \mu]$, and since $\hat{w}(\rho)$ is unique, it is differentiable everywhere, and $W'(\rho) = \hat{w}(\rho)/\mu$. In particular, $W(\rho)$ is strictly concave at every ρ such that $\hat{w}(\rho)$ is strictly decreasing, that is whenever $\hat{w}(\rho) \neq \bar{w}$ or $\gamma \leq \bar{\gamma}$. □

Proof of Theorem 4. The objective function of the across problem is continuous and concave in $\boldsymbol{\rho}$ by Theorem 3, and the feasible set is nonempty, compact and convex. Therefore, it admits a solution characterized by the Kuhn-Tucker conditions (i)-(iii), recalling that, by Theorem 3, $W_i'(\rho_i) = \hat{w}_i(\rho_i)/\mu$, and the outside option value $\hat{w}_i(\rho_i)$ is the one that solves the within problem as defined by (iv). The condition for uniqueness holds because the objective function is then strictly concave by Theorem 3. □

Proof of Proposition 2. Increasing γ_i shrinks the set of feasible allocation rules in the original problem, therefore weakly decreases its value function $\overline{W}$. Suppose $\tilde{F}_i$ first-order stochastically dominates F_i, and let $F_i^x = x\tilde{F}_i + (1-x)F_i$. Then F_i^x increases with x in the FOSD order.

Consider the within problem for group i under the score distribution F_i^x. To clarify the dependence on x, we denote its value function by $W_i(\rho_i|x)$ in this proof. By Lemma 1,

$$W_i(\rho_i|x) = \min_{\hat{w}\in[w^-,w^+]} \max_{\alpha} \int_{S_i} \alpha(s)\{w_i(s) - \hat{w}\}dF_i^x(s) + \hat{w}\rho_i/\mu_i,$$

and $\big(\alpha_i^*(\cdot, \hat{w}_i(\rho_i), r_i(\rho_i)|x), \hat{w}_i(\rho_i)\big)$ is the unique solution to this saddle-point problem. In what follows, let $\alpha_i^*(s)$ denote the function $\alpha_i^*(\cdot, \hat{w}_i(\rho_i), r_i(\rho_i)|x)$

Let $\mathcal{L}(\alpha, \hat{w}, x) = \int_{S_i} \alpha(s)\{w_i(s) - \hat{w}\}dF_i^x(s) + \hat{w}\rho_i/\mu_i$ be the objective function. It is continuously differentiable in x since it is linear. Furthermore, the saddle-point problem admits a solution for every $x \in [0, 1]$ by Theorem 3. The interval $[w^-, w^+]$ and the space of nondecreasing continuous functions in which α is taken is also compact by Helly's selection theorem. Therefore, we can apply the envelope theorem for saddle-points of Milgrom and Segal (2002, Theorem 5), and our uniqueness result to obtain that $W_i(\rho_i|x)$ is differentiable in x, and

$$\begin{aligned}\frac{\partial W_i(\rho_i|x)}{\partial x} &= \frac{\partial \mathcal{L}\Big(\alpha_i^*\big(\cdot, \hat{w}_i(\rho_i), r_i(\rho_i)\big), \hat{w}_i(\rho_i), x\Big)}{\partial x} \\ &= \int_{S_i} \alpha_i^*(s)\{w_i(s) - \hat{w}_i(\rho_i)\}d\tilde{F}_i(s) - \int_{S_i} \alpha_i^*(s)\{w_i(s) - \hat{w}_i(\rho_i)\}dF_i(s).\end{aligned}$$

Using the differential form, we have:

$$\frac{\partial W_i(\rho_i|x)}{\partial x} = \int_{S_i} \alpha_i^{*'}(s) \Big(\tilde{\mathcal{W}}_i(s, \hat{w}_i(\rho_i)) - \mathcal{W}_i(s, \hat{w}_i(\rho_i))\Big)\, ds.$$

Next, we show this must be nonnegative. Indeed, note first

$$\begin{aligned}\tilde{\mathcal{W}}_i^+(s, \hat{w}_i(\rho_i)) - \mathcal{W}_i^+(s, \hat{w}_i(\rho_i)) = \big(1 - \tilde{F}_i(s)\big) \int_s^{\overline{s}_i} \{w_i(y) - \hat{w}_i(\rho_i)\}d\frac{\tilde{F}_i(y)}{1 - \tilde{F}_i(s)} \\ - \big(1 - F_i(s)\big) \int_s^{\overline{s}_i} \{w_i(y) - \hat{w}_i(\rho_i)\}d\frac{F_i(y)}{1 - F_i(s)}.\end{aligned}$$

By first-order stochastic dominance, $1 - \tilde{F}_i(s) \geq 1 - F_i(s)$, and the stochastic dominance ordering of the conditional distributions on $[s, \overline{s}_i]$ is preserved since $\frac{\tilde{F}_i(y)}{1-\tilde{F}_i(s)} \leq \frac{F_i(y)}{1-F_i(s)}$. Since $w_i(\cdot)$ is an increasing function, this implies

$$\tilde{\mathcal{W}}_i^+(s, \hat{w}_i(\rho_i)) - \mathcal{W}_i^+(s, \hat{w}_i(\rho_i)) \geq 0.$$

If $\bar{w}_{F_i} \leq \bar{w}_{\tilde{F}_i} \leq \hat{w}_i(\rho_i)$, then the group has low or neutral priority under both distributions. Using the differential version of the objective function (DOF), the difference in welfare is given by

$$\int_{\underline{s}_i}^{\bar{s}_i} \alpha_i^{*'}(s)\big\{\tilde{\mathcal{W}}_i^+(s,\hat{w}_i(\rho_i)) - \mathcal{W}_i^+(s,\hat{w}_i(\rho_i))\big\}ds.$$

Since $\alpha_i^{*'}(s) \geq 0$, and the difference in cumulative surplus is positive, then $\frac{\partial W_i(\rho_i|x)}{\partial x} \geq 0$.

Suppose instead, $\bar{w}_{F_i} \leq \hat{w}_i(\rho_i) \leq \bar{w}_{\tilde{F}_i}$, so the shift in score distributions switches the priority of the group. Then, using (DOF) for F_i, ($\overline{\text{DOF}}$) for $\tilde{F}_i$, and the relationship $\mathcal{W}^-(s,\hat{w}) = \mathcal{W}^+(s,\hat{w}) - (\bar{w} - \hat{w})$, we can write the welfare change as

$$\big(\bar{w}_{\tilde{F}_i} - \hat{w}_i(\rho_i)\big)\left(1 - \int_{\underline{s}_i}^{\bar{s}_i} \alpha_i^{*'}(s)ds\right)\int_{\underline{s}_i}^{\bar{s}_i} \alpha_i^{*'}(s)\big\{\tilde{\mathcal{W}}_i^+(s,\hat{w}_i(\rho_i)) - \mathcal{W}_i^+(s,\hat{w}_i(\rho_i))\big\}ds,$$

where the second term is positive for the same reasons as in the previous case, and the first term is equal to $\big(\bar{w}_{\tilde{F}_i} - \hat{w}_i(\rho_i)\big)\alpha_i^{*'}(\underline{s}_i) \geq 0$. Hence, again, $\frac{\partial W_i(\rho_i|x)}{\partial x} \geq 0$.

Suppose finally $\hat{w}_i(\rho_i) \leq \bar{w}_{F_i} \leq \bar{w}_{\tilde{F}_i}$ so the priority is high under both distributions. Then, using ($\overline{\text{DOF}}$)and the relationship $\mathcal{W}^-(s,\hat{w}) = \mathcal{W}^+(s,\hat{w}) - (\bar{w} - \hat{w})$, we can write the welfare change as

$$\big(\bar{w}_{\tilde{F}_i} - \bar{w}_{F_i}\big)\left(1 - \int_{\underline{s}_i}^{\bar{s}_i} \alpha_i^{*'}(s)ds\right)\int_{\underline{s}_i}^{\bar{s}_i} \alpha_i^{*'}(s)\big\{\tilde{\mathcal{W}}_i^+(s,\hat{w}_i(\rho_i)) - \mathcal{W}_i^+(s,\hat{w}_i(\rho_i))\big\}ds,$$

and both terms are positive, and then $\frac{\partial W_i(\rho_i|x)}{\partial x} \geq 0$.

Now, consider the across problem. Applying the (classical) envelope theorem to this problem, and letting $\boldsymbol{\rho}^*$ denote its unique solution, we obtain

$$\frac{\partial \overline{W}(x)}{\partial x} = \mu_i \frac{\partial W_i(\rho_i^*|x)}{\partial x} \geq 0.$$

$\square$

Proof of Proposition 3. Recall the matching function $m : [s_*(0), \hat{s}] \to [\hat{s}, s^*(0)]$ is decreasing. This implies the growth interval is increasing in γ for the inclusion order. It increases strictly for $\gamma < \gamma^0$, and is equal to $[s_*(0), s^*(0)]$ for $\gamma \geq \gamma^0$.

Consider first $\gamma < \gamma' < \gamma^0$, and the function $\delta(s) = \alpha^*(s|\gamma') - \alpha^*(s|\gamma)$. We denote by $s_*[\gamma]$ and $s^*[\gamma]$ the optimal matching pair under γ, where we use brackets to distinguish them from the functions $s_*(\nu), s^*(\nu)$.

$\delta(s)$ is equal to 0 for $s \leq s_*[\gamma']$ and $s \geq s^*[\gamma']$. It is equal to $\alpha^*(s|\gamma')$, and therefore increasing and positive, on $[s_*[\gamma'], s_*[\gamma]]$, and to $\alpha^*(s|\gamma')-1$, and therefore increasing and negative on $[s^*[\gamma], s^*[\gamma']]$.

If the cost function satisfies (UID), the derivative of δ is

$$\delta'(s) = \left(\frac{1}{\gamma} - \frac{1}{\gamma'}\right) c_s\big(m(s)|s\big) < 0$$

on $[s_*[\gamma], \hat{s}]$, and

$$\delta'(s) = \left(\frac{1}{\gamma'} - \frac{1}{\gamma}\right) c_t\big(m(s)|s\big) < 0$$

on $[\hat{s}, s^*[\gamma]]$.

If, instead, the cost function satisfies (UDD), its derivative is

$$\delta'(s) = \left(\frac{1}{\gamma'} - \frac{1}{\gamma}\right) c_{t^+}\big(s|s\big) < 0$$

on $[s_*[\gamma], s^*[\gamma]]$.

Hence, δ increases from 0, then decreases and becomes negative, and increases back to 0, which proves point (i) of the proposition.

Next, suppose $\gamma' > \gamma > \gamma^0$. If the priority is low, then the growth interval under both γ and γ' is $[s_*(0), \overline{s}]$. The computation of δ' in this interval is the same as above, implying now δ is decreasing on $[s_*(0), \overline{s}]$. Since $\delta\big(s_*(0)\big) = 0$, this proves point (ii).

If the priority is high, then the growth interval under both γ and γ' is $[\underline{s}, s^*(0)]$. The computation of δ' in this interval is the same as above, implying now δ is decreasing on $[\underline{s}, s^*(0)]$. Since $\delta\big(s^*(0)\big) = 0$, this proves point (iii). $\square$

Proof of Proposition 4. For $s < t$

$$c_t(t|s) = \varphi'(t)\xi'\big(\varphi(t) - \varphi(s)\big)$$

is increasing in s if ξ is concave, and decreasing if ξ is convex. Using the formulas from Theorem 1 and Theorem 2, we get

$$\alpha^*(s) = \begin{cases} 0 & \text{if } s < s_* \\ \Gamma_{udd} I(\hat{w}, r) + \frac{1}{\gamma}\xi'(0)\big(\varphi(s) - \varphi(s_*)\big) & \text{if } s_* \leq s \leq s^* \\ 1 & \text{if } s > s^* \end{cases}$$

in the (UDD) case, and

$$\alpha^*(s) = \begin{cases} 0 & \text{if } s < s_* \\ \Gamma_{uid} I(\hat{w}, r) + \frac{1}{\gamma}\int_{s_*}^{s} \varphi'(x)\xi'\big(\varphi(m(x)) - \varphi(x)\big)dx & \text{if } s_* \leq s \leq \hat{s} \\ 1 - \Gamma_{uid}\bar{I}(\hat{w}, r) - \frac{1}{\gamma}\int_{s}^{s^*} \varphi'(x)\xi'\big(\varphi(x) - \varphi(m^{-1}(x))\big)dx & \text{if } \hat{s} \leq s \leq s^* \\ 1 & \text{if } s > s^* \end{cases}$$

in the (UID) case. In the latter case, for $s \in [s_*, \hat{s}]$,

$$\alpha'(s)/\varphi'(s) = \frac{1}{\gamma}\xi'\big(\varphi(m(s)) - \varphi(s)\big),$$

is increasing in $\varphi(s)$ by concavity of ξ and since $m(s)$ is decreasing in s. If instead $s \in [\hat{s}, s^*]$, then

$$\alpha'(s)/\varphi'(s) = \frac{1}{\gamma}\xi'\big(\varphi(s) - \varphi(m^{-1}(s))\big)$$

is decreasing in $\varphi(s)$ by concavity of ξ and since $m^{-1}(s)$ is decreasing in s. □

Proof of Proposition 5. When ξ is concave, the growth interval is determined by the equation $m(s_*) - s_* = L$, and does not vary with the cost function, and it is a subset of $[s_*(0), s^*(0)]$ since we assumed $L < s^*(0) - s_*(0)$. For convex cost functions, the growth interval is given by the equation $m(s_*) - s_* = \min\{s^*(0) - s_*(0), 1/\xi'(0)\}$ by Proposition 4. Furthermore, if both cost functions are convex, the convex ordering and our normalization imply $\hat{\xi}'(0) \leq \xi'(0)$, hence $I^* \subseteq \hat{I}^*$.

Next, let $\delta(s) = \hat{\alpha}^*(s) - \alpha^*(s)$. Suppose first both cost functions are concave and let $I^* = [s_*, s^*]$ be their common growth interval. In particular $\delta(s_*) = \delta(s^*) = 0$. Furthermore, δ is differentiable and

$$\gamma\delta'(s) = \begin{cases} \big\{1 - g' \circ \hat{\xi}\big(m(s) - s\big)\big\}\hat{\xi}'\big(m(s) - s\big) & \text{if } s \in [s_*, \hat{s}], \\ \big\{1 - g' \circ \hat{\xi}\big(s - m^{-1}(s)\big)\big\}\hat{\xi}'\big(s - m^{-1}(s)\big) & \text{if } s \in [\hat{s}, s^*] \end{cases},$$

where g is an increasing and concave bijection of $[0,1]$ such that $\xi = g \circ \hat{\xi}$. As such $g'(0) \geq 1 \geq g'(1)$, and g' is a non-increasing function. Since $\xi' \geq 0$, this implies δ' is single crossing from the positives to the negatives on $[s_*, \hat{s}]$ and from the negatives to the positives on $[\hat{s}, s^*]$. Therefore there exists a single threshold $\tilde{s} \in [s_*, s^*]$ such that $\delta(s) \geq 0$ for $s \leq \tilde{s}$ and $\delta(s) \leq 0$ for $s \geq \tilde{s}$.

If the two cost functions are convex, then for $\hat{\xi}$ to be more convex than ξ, it must be that $\hat{\xi}'(0) \leq \xi'(0)$ which implies (ii).

Let $\tilde{\xi}$ be the unique linear cost function that belongs to our normalized class of functions. Since $\tilde{\xi}$ is both concave and convex, point (ii) is satisfied when comparing $\tilde{\xi}$ to a concave cost function ξ, and also when comparing a convex cost function $\hat{\xi}$ to $\tilde{\xi}$. Since $\hat{\alpha}^* - \alpha^* = \hat{\alpha}^* - \tilde{\alpha}^* + \tilde{\alpha}^* - \alpha^*$, it is also satisfied when comparing $\hat{\xi}$ to ξ.

If $I^* = \hat{I}^* = [s_*(0), s^*(0)]$, then both functions must be convex by (i). If the group has low priority, then $\alpha^*\big(s_*(0)\big) = \hat{\alpha}^*\big(s_*(0)\big) = 0$, and both allocation rules are linear with respective slopes $\xi'(0) \geq \hat{\xi}'(0)$, implying $\hat{\alpha}^*(s) \leq \alpha^*(s)$ for all s. If instead the group has high priority, the slopes compare in the same way, but the allocation rules are tied at $s^*(0)$ instead of $s_*(0)$, implying $\hat{\alpha}^*(s) \geq \alpha^*(s)$ for all s. $\square$

Proof of Proposition 6.

- *(b) $\Rightarrow$ (a)*. Suppose (b) holds.

We start by showing (i) $\tilde{s}_*(0) \leq \hat{s}_*(0)$ and (ii) $\tilde{s}^*(0) \leq \hat{s}^*(0)$.

First suppose $\tilde{s}^*(0) = \overline{s}$. Then (ii) must hold, and (i) also because, otherwise, we would have the following contradiction

$$\overline{s} = \tilde{s}^*(0) = \tilde{m}\big(\tilde{s}_*(0)\big) \leq \hat{m}\big(\tilde{s}_*(0)\big) < \hat{m}\big(\hat{s}_*(0)\big) = \hat{s}^*(0),$$

where the first inequality is by (b), and the second inequality because $\hat{m}$ is decreasing on $[\hat{s}_*(0), 0]$.

Next, suppose $\hat{s}_*(0) = \underline{s}$. Then (i) must hold, and (ii) also because, otherwise, we would have the following contradiction

$$\tilde{s}_*(0) = \tilde{m}^{-1}\big(\tilde{s}^*(0)\big) < \tilde{m}^{-1}\big(\hat{s}^*(0)\big) \leq \hat{m}^{-1}\big(\hat{s}^*(0)\big) = \hat{s}_*(0) = \underline{s},$$

where the first inequality is because $\tilde{m}^{-1}$ is decreasing on $[0, \tilde{s}^*(0)]$, and the second inequality is by (b).

If neither of these cases hold, by Lemma 2, (iii), we must have $\tilde{s}_*(0) = \underline{s}$ and $\hat{s}^*(0) = \overline{s}$, which imply (i) and (ii).

An implication of (i) and (ii) is (iii): if $\hat{F}$ has *high priority*, then so does $\tilde{F}$, and if $\tilde{F}$ has *low priority*, then so does $\hat{F}$.

Next, consider a cost function that satisfies (UDD)*.*

(b) implies $\tilde{m}(\hat{s}_*) \leq \hat{m}(\hat{s}_*) = \hat{s}^*$, therefore

$$\int_{\hat{s}_*}^{\tilde{m}(\hat{s}_*)} c_{t^+}(x|x)dx \leq \int_{\hat{s}_*}^{\hat{s}^*} c_{t^+}(x|x)dx = \gamma.$$

Then, using (B_{udd}) and point (i) we just proved, we must have $\tilde{s}_* \leq \hat{s}_*$.

Then, for every $s \in [\hat{s}_*, \tilde{s}^*]$,

$$\tilde{\alpha}(s) - \hat{\alpha}(s) = \tilde{\Gamma}_{udd}\, \mathbb{1}_{\mathcal{E}} + \int_{\tilde{s}_*}^{\hat{s}_*} c_{t^+}(x|x)dx \geq 0,$$

where $\mathcal{E}$ is the event in which only $\tilde{F}$ has high priority (the event in which only $\hat{F}$ has high priority is impossible by (iii)). This also implies $\tilde{s}^* \leq \hat{s}^*$, so, for any $s \geq \tilde{s}^*$, we also have $1 = \tilde{\alpha}(s) \geq \hat{\alpha}(s)$. Finally, for $s \leq \hat{s}_*$, we have $\tilde{\alpha}(s) \geq \hat{\alpha}(s) = 0$.

Finally, consider a cost function that satisfies (UID)*.*

(b) implies $\tilde{m}(\hat{s}_*) \leq \hat{m}(\hat{s}_*) = \hat{s}^*$, therefore

$$c\big(\tilde{m}(\hat{s}_*)|\hat{s}_*\big) \leq c\big(\hat{s}^*|\hat{s}_*\big) \leq \gamma.$$

Then, using (B_{uid}) and point (i) we just proved, we must have $\tilde{s}_* \leq \hat{s}_*$.

(b) also implies $\hat{m}^{-1}(\tilde{s}^*) \geq \tilde{m}^{-1}(\tilde{s}^*) = \tilde{s}_*$, therefore

$$c\big(\tilde{s}^*|\hat{m}^{-1}(\tilde{s}^*)\big) \leq c(\tilde{s}^*|\tilde{s}_*) \leq \gamma.$$

Then, using (B_{uid}) and point (ii) we just proved, we must have $\tilde{s}^* \leq \hat{s}^*$.

Then, for every $s \in [\hat{s}_*, 0]$,

$$\begin{aligned}\tilde{\alpha}(s) - \hat{\alpha}(s) = \tilde{\Gamma}_{uid}\, \mathbb{1}_{\mathcal{E}} - \frac{1}{\gamma}\int_{\tilde{s}_*}^{\hat{s}_*} c_s\big(\tilde{m}(x)|x\big)dx - \frac{1}{\gamma}\int_{\hat{s}_*}^{s} \big\{c_s\big(\tilde{m}(x)|x\big) - c_s\big(\hat{m}(x)|x\big)\big\}dx \\ \geq 0,\end{aligned}$$

where $\tilde{\Gamma}_{uid} \geq 0$ by definition, the second term is nonnegative by cost monotonicity, and the third term is nonnegative by (UID) and (b).

And for every $s \in [0, \tilde{s}^*]$,

$$\begin{aligned}\tilde{\alpha}(s) - \hat{\alpha}(s) &= \hat{\Gamma}_{uid}\, \mathbb{1}_{\mathcal{E}'} + \frac{1}{\gamma}\int_{\tilde{s}^*}^{\hat{s}^*} c_t\big(x|\hat{m}^{-1}(x)\big)dx \\ &\quad + \frac{1}{\gamma}\int_{s}^{\tilde{s}^*} \big\{c_t\big(x|\hat{m}^{-1}(x)\big) - c_t\big(x|\tilde{m}^{-1}(x)\big)\big\}dx \\ &\geq 0,\end{aligned}$$

where $\hat{\Gamma}_{uid} \geq 0$ by definition, $\mathcal{E}'$ is the event in which only $\hat{F}$ has low priority, the second term is nonnegative by cost monotonicity, and the third term is nonnegative by (UID) and (b).

• $\underline{(a) \Rightarrow (b)}$. Suppose (a) holds, and consider the family of linear cost functions $c(t|s) = \beta\gamma|t - s|$, for $\beta > 0$. By choosing β sufficiently low, we can ensure neither of the allocation rules saturates the probability constraint. In this case, $\hat{\alpha}_\beta(\underline{s}) > 0$ if and only if $\hat{F}$ has high priority, but then (a) implies $\tilde{F}$ must have high priority as well. Similarly $\tilde{\alpha}_\beta(\overline{s}) < 1$ if and only if $\tilde{F}$ has low priority, and then (a) implies $\hat{F}$ has low priority as well.

Then $\tilde{\alpha}_\beta(s) = \beta(s - \tilde{s}_*)$ on $[\tilde{s}_*, \tilde{s}^*]$, and $\hat{\alpha}_\beta(s) = \beta(s - \hat{s}_*)$ on $[\hat{s}_*, \hat{s}^*]$. By varying β from 0 to infinity, we have $\tilde{s}_*$ span $[\tilde{s}_*(0), 0)$, and $\hat{s}_*$ span $[\hat{s}_*(0), 0)$. For β sufficiently large, we have both $\tilde{s}_* > \tilde{s}_*(0)$ and $\hat{s}_* > \hat{s}_*(0)$. Pick such a value of β, then by (a), we have

$$-\beta\tilde{s}_* = \tilde{\alpha}_\beta(0) \geq \hat{\alpha}_\beta(0) = -\beta\hat{s}_*,$$

so $\tilde{s}_* \leq \hat{s}_*$. Furthermore, for such a value of β, we must have

$$\tilde{m}(\tilde{s}_*) = \frac{1}{\beta} + \tilde{s}_* \leq \frac{1}{\beta} + \hat{s}_* = \hat{m}(\hat{s}_*) \leq \hat{m}(\tilde{s}_*).$$

Varying β so $\tilde{s}_*$ spans $[s_*(0), 0]$, this shows (b).

• $\underline{(b) \Leftrightarrow (c) \Leftrightarrow (d)}$. Since, for all $s < 0$, every x between $\tilde{m}(s)$ and $\hat{m}(s)$ is nonnegative, $\tilde{m}(s) \leq \hat{m}(s)$ is equivalent to $\int_{\tilde{m}(s)}^{\hat{m}(s)} x d\hat{F}(x) \geq 0$. By definition of the matching functions,

$$\int_s^{\tilde{m}(s)} x d\tilde{F}(x) = \int_s^{\hat{m}(s)} x d\hat{F}(x) = 0,$$

therefore

$$\int_{\tilde{m}(s)}^{\hat{m}(s)} xd\hat{F}(x) - \int_{s}^{\tilde{m}(s)} xd\tilde{F}(x) = \int_{s}^{\tilde{m}(s)} xd\tilde{F}(x) - \int_{s}^{\tilde{m}(s)} xd\hat{F}(x).$$

This shows the equivalence between (b) and (c). The inequality in (d) results from applying integration by parts to (c). □

Proof of Proposition 7. We show the conditions of the proposition imply that, for every $z < 0 < y$, $\int_z^0 \{\Delta(z) - \Delta(x)\}dx \geq 0$, and $\int_0^y \{\Delta(y) - \Delta(x)\}dx \geq 0$, which implies condition (d) of Proposition 6.

If $z < a$, then

$$\int_z^0 \{\Delta(z) - \Delta(x)\}dx = -z\Delta(z) - \int_z^0 \Delta(x)dx \geq -z\Delta(z) - \int_{\underline{s}}^0 \Delta(x)dx \geq 0,$$

where the first inequality is from (C1), and the second inequality is from (C3) and (C1) as $\Delta(z) \geq 0$. If $z \geq a$, then (C2) implies $\Delta(z) \geq \Delta(x)$ for every $x \in [z, 0]$. The proof is symmetric for the integral from 0 to y. □

Proof of Proposition 8. Consider the problem of the decision maker deciding how to allocate objects. She can only condition her decision on the group label, and the signal provided by the designer's chosen information structure. Under the information structure that recommends allocation with probability $\alpha_i^*(s)$ and rejection otherwise, let $g_i \in [0, 1]$ be the probability that the decision maker allocates an object to members of group i with a positive recommendation, and $b_i \in [0, 1]$ the probability that she allocates an object to members of group i with a negative recommendation. Her problem is

$$\begin{aligned}
\max_{(\boldsymbol{g},\boldsymbol{b})} \ & \sum_i \mu_i \left\{ g_i \int_{S_i} \alpha_i^*(s)\tilde{w}_i(s)dF_i(s) + b_i \int_{S_i} \big(1 - \alpha_i^*(s)\big)\tilde{w}_i(s)dF_i(s) \right\} \\
\text{s.t.} \ & \sum_i \mu_i \{g_i A_i^* + b_i(1 - A_i^*)\} \leq \bar{\rho} \\
& \mu_i \{g_i A_i^* + b_i(1 - A_i^*)\} \geq \phi_i \bar{\rho} \quad \forall i.
\end{aligned}$$

Then, α^* is obedient if choosing $(g_i^o, b_i^o) = (1, 0)$ for every i is a solution to the decision maker's program. Since the program of the decision maker is linear, global optimality is implied by local optimality. So, to check obedience, we only need to verify that $(\boldsymbol{g}^o, \boldsymbol{b}^o)$ is a local optimum.

This is the case if the decision maker is perfectly aligned with the designer, $\tilde{w}_i = w_i$. Indeed, for (g_i, b_i) in the neighborhood of (g_i^o, b_i^o), we have $1 \geq g_i - b_i \geq 0$, therefore the effective allocation rule implemented by the decision maker is $\alpha_i(s) = b_i\big(1 - \alpha_i^*(s)\big) + g_i\alpha_i^*(s)$. It satisfies falsification-proofness since

$$0 \leq \alpha_i(t) - \alpha_i(s) = (g_i - b_i)\big(\alpha_i^*(t) - \alpha_i^*(s)\big) \leq \alpha_i^*(t) - \alpha_i^*(s) \leq \frac{1}{\gamma_i} c_i(t|s),$$

and could therefore have been implemented by the designer in our original problem, so it must be suboptimal.

In fact, $(\boldsymbol{g}^o, \boldsymbol{b}^o)$ is uniquely optimal when preferences are aligned. Again, we only need to check that locally. Indeed, for any (g_i, b_i), the resulting effective allocation rule α_i is in the family of possibly optimal rules $\alpha_i^*(\cdot, \hat{w}, r)$ if and only if $(g_i, b_i) = (g_i^o, b_i^o)$. Indeed, it is true for (g_i^o, b_i^o), and if $(g_i, b_i) \neq (g_i^o, b_i^o)$, then α_i has the same growth interval as α_i^*. However, for $\hat{w} \neq \bar{w}$, each $\alpha_i^*(\cdot, \hat{w}, r)$ has a distinct growth interval. If $\alpha_i^* = \alpha_i^*(\cdot, \bar{w}, r)$ for some r, then the only possibility for α_i to be possibly optimal is if $\alpha_i = \alpha_i^*(\cdot, \bar{w}, r')$ for $r' \neq r$. But then, α_i and α_i^* must differ by an additive constant, which contradicts the definition of α_i.

Suppose then that the decision maker is not perfectly aligned with the designer. We let

$$G_i(\tilde{w}_i) = \int_{S_i} \alpha_i^*(s)\tilde{w}_i(s)dF_i(s),$$

and

$$B_i(\tilde{w}_i) = \int_{S_i} \big(1 - \alpha_i^*(s)\big)\tilde{w}_i(s)dF_i(s),$$

be the linear coefficients corresponding to g_i and b_i in the decision maker's objective function. Then, we have, for every i, $\big|G_i(\tilde{w}_i) - G_i(w_i)\big| < \varepsilon A_i^*$, and $\big|B_i(\tilde{w}_i) - B_i(w_i)\big| < \varepsilon(1 - A_i^*)$. Therefore, we can choose ε sufficiently small to ensure that every strict inequality holding between any pair among the scalars $\{0\} \cup \bigcup_{i \in I}\big\{B_i(w_i), G_i(w_i)\big\}$ also holds for $\{0\} \cup \bigcup_{i \in I}\big\{B_i(\tilde{w}_i), G_i(\tilde{w}_i)\big\}$, regardless of $\tilde{w}_i$.

Suppose, by contradiction, that (g_i^o, b_i^0) is not optimal for the decision maker with preferences given by $\tilde{\boldsymbol{w}}$. Then either one of the following local deviations must be strictly beneficial for the decision maker. For each of them, we show it leads to a contradiction.

(a) Slightly decreasing g_i from $g_i^0 = 1$: For that to be strictly beneficial, it must be that $G_i(\tilde{w}_i) < 0$, therefore $G_i(w_i) \leq 0$. However, this can only be true if the quota constraint is binding for i at $(\boldsymbol{g}^o, \boldsymbol{b}^o)$, or it would contradict the strict optimality of $(\boldsymbol{g}^o, \boldsymbol{b}^o)$ under w_i. But then decreasing g_i is infeasible as it violates the quota for i.

(b) Slightly increasing b_i from $b_i^o = 0$: This is strict beneficial only if $B_i(\tilde{w}_i) > 0$, implying $B_i(w_i) \geq 0$. Then the resource constraint must be binding at $(\boldsymbol{g}^o, \boldsymbol{b}^o)$, or it would contradict the strict optimality of $(\boldsymbol{g}^o, \boldsymbol{b}^o)$ under w_i. Therefore this deviation is not feasible as it would violate the resource constraint.

(c) Decreasing g_i and increasing b_i so as to keep the mass of objects allocated to group i constant: For this to be strictly beneficial, it must be that $G_i(\tilde{w}_i) < B_i(\tilde{w}_i)$, implying $G_i(w_i) \leq B_i(w_i)$. The same deviation would then be feasible and weakly beneficial at w_i contradicting the strict optimality of $(\boldsymbol{g}^o, \boldsymbol{b}^o)$.

(d) Decreasing g_i and increasing b_j for $j \neq i$ while keeping the total mass of objects allocated constant: Then $G_i(\tilde{w}_i) < B_i(\tilde{w}_j)$, implying $G_i(w_i) \leq B_i(w_j)$. This can only hold if the quota constraint of group i is binding at $(\boldsymbol{g}^o, \boldsymbol{b}^o)$, for otherwise it would contradict the strict optimality of $(\boldsymbol{g}^o, \boldsymbol{b}^o)$ at w_i. But then this deviation is infeasible.

$\square$

B Algorithm for the across problem

We present an algorithm that finds a solution to the across problem. We extend the definition of $\hat{w}_i^\phi$ as the unique value of $\hat{w}$ such that $\mu_i A_i^*(\hat{w}, r) = \phi_i \bar{\rho}$ for some r, and let r_i^ϕ be the unique value of r that satisfies this equality if $\hat{w}_i^\phi = \bar{w}_i$ (otherwise let r_i^ϕ be any value on $[0, 1]$).

Algorithm 2: Algorithm to solve the across problem

$\forall i, \quad \rho_i^0 \leftarrow \mu_i A_i^*(0, 1)$;
$R^0 \leftarrow \mathcal{R}(\boldsymbol{\rho}^0)$;
$Q^0 \leftarrow \mathcal{Q}(\boldsymbol{\rho}^0)$;
$k \leftarrow 0$;
repeat
  $k \leftarrow k + 1$;
  $\forall \ell \in Q^{k-1}, \quad \hat{w}_\ell^k \leftarrow \hat{w}_\ell^\phi$ and $r_\ell^k \leftarrow r_\ell^\phi$;
  if $R^{k-1} = 0$ **then**
    $\forall \ell \notin Q^{k-1}, \quad \hat{w}_\ell^k \leftarrow 0$ and $r_\ell^k \leftarrow 1$;
  else
    $\hat{w}, r \leftarrow$ *Solution of:* $\sum_{\ell \in Q^{k-1}} \phi_\ell \bar{\rho} + \sum_{i \notin Q^{k-1}} \mu_i A_i^*(\hat{w}, r\mathbb{1}_{\hat{w}=\bar{w}_i}) = \bar{\rho}$;
    $\forall \ell \notin Q^{k-1}, \quad \hat{w}_\ell^k \leftarrow \hat{w}$ and $r_\ell^k \leftarrow r\mathbb{1}_{\hat{w}=\bar{w}_\ell}$;
  end
  $\forall i, \quad \rho_i^k \leftarrow \mu_i A_i^*(\hat{w}_i^k, r_i^k)$;
  $R^k \leftarrow \mathcal{R}(\boldsymbol{\rho}^k)$;
  $Q^k \leftarrow \mathcal{Q}(\boldsymbol{\rho}^k)$;
until $Q^k = Q^{k-1}$ *and* $R^k = R^{k-1}$;

We did not specify how to find a solution $(\hat{w}, r)$ to

$$\sum_{\ell \in Q^{k-1}} \phi_\ell \bar{\rho} + \sum_{i \notin Q^{k-1}} \mu_i A_i^*(\hat{w}, r\mathbb{1}_{\hat{w}=\bar{w}_i}) = \bar{\rho}$$

within the algorithm. Note, however, that the left-hand side of the equation can be decreased continuously by continuously raising $\hat{w}$ whenever $\hat{w} \neq \bar{w}_i$, for all i, and by continuously raising r from 0 to 1 and keeping $\hat{w}$ constant whenever $\hat{w} = \bar{w}_i$ for some i. Therefore a simple algorithm can solve this equation.

Proposition 9. *Algorithm 2 finds a solution of* ($\overline{\text{P}}$) *in finitely many steps.*

Proof. The sequence (Q^k, R^k) is increasing and bounded above by $(I, 1)$ in the $(\subseteq, \leq)$ order on $2^I \times \{0, 1\}$, so the algorithm stops in finitely many steps. Let k be the step at which it stops. Let $\lambda_R = \hat{w}_i^k$ and $\lambda_i = 0$ for all $i \notin Q^k$. This is consistent since $\hat{w}_i^k$ must be equal across all $i \notin Q^k$. Let $\lambda_\ell = \lambda_R - \hat{w}_\ell^k$ for $\ell \in Q^k$. Then it is easy to verify the vector of multipliers $\boldsymbol{\lambda}$, $\boldsymbol{\rho}^k$, $\hat{w}^k$ and $\boldsymbol{r}^k$ satisfy all the conditions of Theorem 4. Therefore $\boldsymbol{\rho}^k$ is a solution to the across problem. $\square$

C Score-based mechanisms are without loss

In this appendix, we explain why our restriction to score-based mechanisms is without los of generality.

The score-based restriction. First, we define a more general class of mechanisms with communication to use as a benchmark as follows. The designer commits to a communication protocol and an allocation rule, that is, a message space for the agents to communicate, a space of recommendation messages for the designer to communicate back, a recommendation rule contingent on the messages received from the agents, and an allocation rule contingent on the messages received from the agents, the recommendation messages sent privately to the agents, and the agents' observed and possibly falsified score. In this class of mechanisms, communication is available before falsification takes place, and the agent's choice of falsified score is treated as an action. This is reminiscent of Myerson (1982), with the only difference that the agents' actions (their falsified scores) are observed by the designer, and hence *contractible*, meaning the designer can condition the allocation rule on these actions. The revelation principle in Myerson (1982) establishes that every mechanism is replicable by a canonical direct, truthful and obedient mechanism, in which agents report their type and the designer recommends score falsifications that are followed by agents. Leveraging the observability of falsified scores, we can additionally assume that canonical mechanisms are *harsh* in that they punish disobedience maximally by assigning a null probability of obtaining an object.

Next, we need to incorporate the falsification-proofness constraint in this general class of mechanisms with communication. The natural way to do it is to restrict attention to mechanisms such that the agents' equilibrium response is to abstain from falsifying. In the canonical class of mechanisms, falsification-proofness implies that canonical mechanisms recommend that agents generate the score they (truthfully) reported. A falsification-proof canonical mechanism is fully specified by the probability $\alpha(t|i, s, k) \in [0, 1]$ of obtaining the good conditional on belonging to group i (which is observable), reporting private information (s, k)

(and therefore being recommended[9] to "falsify" to s), and actually falsifying to t. The harshness of canonical mechanism additionally implies $\alpha(t|i,s,k) = 0$ if $t \neq s$.

In a falsification-proof canonical mechanism, obedience is directly implied by truthfulness. Indeed, following a truthful report (s,k) by type $\theta = (i,s,k)$, obeying the recommendation yields $\alpha(s|i,s,k)$ whereas, by harshness, disobeying and falsifying to t yields $-C^\theta(t)$. Truthfulness requires that, for every type $\theta = (i,s,k)$, every alternative report (s',k'), and every choice of falsification t the following holds

$$\alpha(s|\theta) \geq \alpha(t|i,s',k') - C^\theta(s').$$

In particular, that type (i,s,k) must not benefit from reporting (s,k'), and type (i,s,k') must not benefit by reporting (s,k), yielding $\alpha(s|i,s,k) = \alpha(s|i,s,k')$, so the allocation probability must be independent of k.

Consequently, canonical mechanisms (and therefore all mechanisms with communication) satisfying the falsification-proofness constraint cannot exploit any information about k for allocation purposes, and it is without generality to use score-based allocation rules. An analogous result appears in Akbarpour et al. (2020) where mechanisms with transfers cannot exploit information about agents types beyond their willingness to pay.[10]

Continuum as a single-agent. The continuum of agents is interpreted as a limit case where the size of the population becomes arbitrarily large. We already discussed why there is no loss of generality in considering allocation rules that only depend on the observed score profile and group identity. In the finite population case, an agent j in group i then receives the good with ex post allocation probability $\alpha_{i,j}(s_j, s_{-j})$. As often in mechanism design, the problem can be reformulated as one of choosing interim allocation probabilities $\alpha_{i,j}(s_j) = \mathbb{E}_{s_{-j}}\, \alpha_{i,j}(s_j, s_{-j})$. Furthermore, given the symmetry of our setup, we can assume symmetry across agents of the same group, so we can write $\alpha_i(s)$ for the interim allocation probability for an agent with score s in group i. Then the *interim problem* of optimizing over symmetric interim allocation rules in any finite population is exactly the program we solve in the continuum.[11] However, to find a solution to the initial program,

[9]Without the falsification-proofness restriction, the allocation rule would have to be conditioned on the recommendation as well.

[10]See, for instance, Lemma 1 in Dworczak, Kominers, and Akbarpour (2021) and Claim 2 in Akbarpour, Dworczak, and Kominers (2020).

[11]The same approach is used in Mylovanov and Zapechelnyuk (2017).

we need to ensure that the interim allocation rules that solve the interim program are feasible in the sense that they can be obtained from an ex post allocation rule. In the finite population case, the exact condition for this to be possible can be derived from Che, Kim, and Mierendorff (2013) which generalizes the condition of Border (1991) to setups with multiple goods and quotas. In the limit case of the continuum, however, the interim rules can be used directly as ex post allocation rules that only depend on each agent's score, so feasibility is automatically satisfied.

References

Abdulkadiroğlu, A. (2005): "College admissions with affirmative action," *International Journal of Game Theory*, 33, 535–549.

Abdulkadiroğlu, A. and T. Sönmez (2003): "School choice: A mechanism design approach," *American economic review*, 93, 729–747.

Ajzenman, N. (2021): "The power of example: Corruption spurs corruption," *American Economic Journal: Applied Economics*, 13, 230–57.

Akbarpour, M., P. Dworczak, and S. D. Kominers (2020): "Redistributive allocation mechanisms," *Available at SSRN 3609182*.

Alm, J., K. M. Bloomquist, and M. McKee (2017): "When you know your neighbour pays taxes: Information, peer effects and tax compliance," *Fiscal Studies*, 38, 587–613.

Amador, M. and K. Bagwell (2022): "Regulating a monopolist with uncertain costs without transfers," *Theoretical Economics*, 17, 1719–1760.

Amador, M., I. Werning, and G.-M. Angeletos (2006): "Commitment vs. Flexibility," *Econometrica*, 74, 365–396.

Arieli, I., Y. Babichenko, and F. Sandomirskiy (2022): "Persuasion as Transportation," in *Proceedings of the 23rd ACM Conference on Economics and Computation, New York, NY, USA: Association for Computing Machinery, EC*, vol. 22, 468.

Askenazy, P., T. Breda, F. Moreau, and V. Pecheu (2022): "Les entreprises sous-déclarent-elles leur effectif à 49 employés pour contourner la loi?" Note IPP 82, Institut des Politiques Publiques.

Augias, V. and E. Perez-Richet (2023): "Non-Market Mechanisms: Optimal Design and Investment Incentives," Working paper.

Ball, I. (2022): "Scoring Strategic Agents," *Working Paper*.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2014): "Optimal allocation with costly verification," *American Economic Review*, 104, 3779–3813.

——— (2019): "Mechanisms with Evidence: Commitment and Robustness," *Econometrica*, 87, 529–566.

Bjerre-Nielsen, A., L. S. Christensen, M. H. Gandil, and H. H. Sievertsen (2023): "Playing the system: address manipulation and access to schools," .

Boerma, J., A. Tsyvinski, and A. P. Zimin (2021): "Sorting with team formation," Tech. rep., National Bureau of Economic Research.

Bolton, L. (2018): "Manipulation of the Waitlist Priority of the Organ Allocation System through the Escalation of Medical Therapies," *OPTN/UNOS EthicsCommittee.*

Border, K. C. (1991): "Implementation of reduced form auctions: A geometric approach," *Econometrica: Journal of the Econometric Society*, 1175–1187.

Carlier, G. (2012): "Optimal transportation and economic applications," *Lecture Notes*, 18.

Çelebi, O. and J. P. Flynn (2022): "Priority Design in Centralized Matching Markets," *The Review of Economic Studies*, 89, 1245–1277.

Che, Y.-K., J. Kim, and K. Mierendorff (2013): "Generalized Reduced-Form Auctions: A Network-Flow Approach," *Econometrica*, 81, 2487–2520.

Chua, G. A., G. Hu, and F. Liu (2019): "Optimal Multi-unit Allocation with Costly Verification," *Available at SSRN 3407031.*

Condorelli, D. (2013): "Market and non-market mechanisms for the optimal allocation of scarce resources," *Games and Economic Behavior*, 82, 582–591.

Daskalakis, C., A. Deckelbaum, and C. Tzamos (2017): "Strong duality for a multiple-good monopolist," *Econometrica*, 85, 735–767.

Deneckere, R. and S. Severinov (2022): "Screening, Signalling and Costly Misrepresentation," *Canadian Journal of Economics*, 55, 1334–1370.

Dur, U., P. A. Pathak, and T. Sönmez (2020): "Explicit vs. statistical targeting in affirmative action: Theory and evidence from Chicago's exam schools," *Journal of Economic Theory*, 187, 104996.

Dworczak, P. (2022): "Equity-efficiency trade-off in quasi-linear environments," Working paper.

Dworczak, P., S. D. Kominers, and M. Akbarpour (2021): "Redistribution through markets," *Econometrica*, 89, 1665–1698.

Epitropou, M. and R. Vohra (2019): "Optimal On-Line Allocation Rules with Verification," in *International Symposium on Algorithmic Game Theory*, Springer, 3–17.

Erlanson, A. and A. Kleiner (2019): "A note on optimal allocation with costly verification," *Journal of Mathematical Economics*, 84, 56–62.

Frankel, A. and N. Kartik (2021): "Improving information from manipulable data," *Journal of the European Economic Association*, jvab017.

Galbiati, R. and G. Zanella (2012): "The tax evasion social multiplier: Evidence from Italy," *Journal of Public Economics*, 96, 485–494.

Galichon, A. (2018): *Optimal transport methods in economics*, Princeton University Press.

Hartline, J. and T. Roughgarden (2008): "Optimal Mechanism Design and Money Burning," *Proceedings of the fourtieth annual ACM symposium on Theory of computing - STOC08*, 1–23.

He, Y. (2015): "Gaming the Boston school choice mechanism in Beijing," .

Kattwinkel, D. (2019): "Allocation with Correlated Information: Too good to be true," .

Kattwinkel, D. and J. Knoepfle (2022): "Costless Information and Costly Verification: A Case for Transparency," Working paper.

Kephart, A. and V. Conitzer (2016): "The revelation principle for mechanism design with reporting costs," in *Proceedings of the 2016 ACM Conference on Economics and Computation*, 85–102.

Kolesnikov, A., F. Sandomirskiy, A. Tsyvinski, and A. P. Zimin (2022): "Beckmann's approach to multi-item multi-bidder auctions," *arXiv preprint arXiv:2203.06837*.

Kolotilin, A., R. Corrao, and A. Wolitzky (2022): "Persuasion with Non-Linear Preferences," Working paper.

Lacker, J. M. and J. A. Weinberg (1989): "Optimal Contracts with Costly State Falsification," *Journal of Political Economy*, 97, 1345–1363.

Landier, A. and G. Plantin (2016): "Taxing the Rich," *The Review of Economic Studies*, 84, 1186–1209.

Li, Y. (2020): "Mechanism design with costly verification and limited punishments," *Journal of Economic Theory*, 186, 105000.

Lin, X. and C. Liu (2022): "Credible persuasion," ArXiv preprint arXiv:2205.03495.

Lipman, B. (2015): "An elementary proof of the optimality of threshold mechanisms," Tech. rep., working paper, July.

Luenberger, D. G. (1969): *Optimization by Vector Space Methods*, New York: John Wiley and Sons.

Malamud, S. and A. Schrimpf (2021): "Persuasion by Dimension Reduction," ArXiv preprint arXiv:2110.08884.

McMichael, B. J. (2022): "Stealing Organs?" *Indiana Law Journal*, 97, 135.

Milgrom, P. and I. Segal (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

Milgrom, P. and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180.

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58.

Myerson, R. B. (1982): "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics*, 10, 67–81.

Mylovanov, T. and A. Zapechelnyuk (2017): "Optimal allocation with ex post verification and limited penalties," *American Economic Review*, 107, 2666–94.

Patel, R. and C. Urgun (2023): "Costly inspection and money burning," Working paper.

Pathak, P. A. and T. Sönmez (2008): "Leveling the playing field: Sincere and sophisticated players in the Boston mechanism," *American Economic Review*, 98, 1636–1652.

——— (2013): "School Admissions Reform in Chicago and England: Comparing Mechanisms by their Vulnerability to Manipulation," *American Economic Review*, 103, 80–106.

Perez-Richet, E. and V. Skreta (2022): "Test design under falsification," *Econometrica*, 90, 1109–1142.

Rincke, J. and C. Traxler (2011): "Enforcement spillovers," *Review of Economics and Statistics*, 93, 1224–1234.

Schummer, J. (2021): "Influencing waiting lists," *Journal of Economic Theory*, 195, 105263.

Severinov, S. and T. Y.-C. Tam (2019): "Screening Under Fixed Cost of Misrepresentation," .