Falsification-proof non-market allocation mechanisms [*]

Eduardo Perez-Richet [†] Vasiliki Skreta [‡]

January 30, 2025

Abstract

Relying on manipulable characteristics to target the allocation of resources is problematic. Naively optimal rules lead to falsification and allocative inefficiency. Optimizing the rule for allocative efficiency while accounting for falsification typically induces rather than eliminates falsification, at a heavy cost to the agents. Furthermore, falsification causes fairness concerns and negative social externalities. We show that optimal falsification-proof rules offer a viable alternative, eliminating negative externalities and, under some conditions on the falsification technology, significantly improving the agents' welfare, at a low cost in allocative efficiency. To do so, we characterize optimal falsification-proof rules under a rich variety of falsification technologies. We also examine the impact of demographic changes on allocations within and across identifiable groups, under resource and quota constraints.

Keywords: Mechanism design, falsification, fraud, manipulation, optimal transport theory, allocation mechanisms, costly misreporting.
JEL classification: C72, D82.

[*]We thank Ricardo Alonso, Julien Combes, Michael Ostrowsky, James Schummer, Philipp Strack, Andrzej Skrzypacz, Utku Ünver and Alexander Wolitzky for thought-provoking discussions and suggestions. We also thank Xiao Lin for an outstanding discussion at the "Theory at Penn State" conference. 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" and by the NSF grant 2242521"Fraud-proof Mechanism Design." Francesco Conti, Nathan Hancart and Ignacio Núñez provided excellent research assistance at various stages. Earlier drafts of this paper were circulated under the tile "Fraud-proof non-market allocation mechanisms".

[†]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 often allocated via non-market mechanisms, either due to institutional constraints 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 example, 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 in many cases assessed through a *score* measuring characteristics or performance, acting as a proxy for the value of assigning an object 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, and manipulating statistics are commonplace. For example, parents fake addresses to gain admission to desirable public schools (Bjerre-Nielsen, Christensen, Gandil, and Sievertsen, 2023), firms underreport their workforce size to avoid legal obligations (Askenazy, Breda, Moreau, and Pecheu, 2022), and doctors manipulate their patients' priority in organ transplant waiting lists[3] (Bolton, 2018; McMichael, 2022). Throughout the paper, we use the term *falsification* as a broad category that encompasses gaming, manipulation, or any other socially wasteful and individually costly activities agents undertake to produce an altered score.

Falsification is not only wasteful, it is socially harmful. First, it distorts achievable assignments, unfairly penalizing agents for whom falsification is more costly.[4] Second, it deteriorates the informational content of the score. This is an instance of Goodhart's law: "when a measure becomes a target, it ceases to be a good measure." For example, greenwashing can blur our assessment of emissions levels. Third, it may render the mechanism politically unsustainable if it comes under scrutiny after falsification is detected. Fourth, it can erode trust and deplete the supply of objects to allocate. In Germany, for example, a scandal involving the manipulation of the liver allocation system by transplant providers led to a 20%-40% erosion in organ donation (Bolton, 2018). Fifth, there is also evidence that dishonest behavior spreads in society (see,

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

[2]See Condorelli (2013) and Akbarpour, Dworczak, and Kominers (2024) for a theory of when non-market mechanisms are optimal.

[3]Schummer (2021) explores the impact of waiting list manipulations in a theoretical model.

[4]For example, Bjerre-Nielsen et al. (2023) show that priority gaming in school choice mechanisms resulted in better assignments for those who engaged in such practices and adversely affected others.

for example, Rincke and Traxler, 2011; Galbiati and Zanella, 2012; Alm, Bloomquist, and McKee, 2017; Ajzenman, 2021).

In this paper, we show how to maximize allocative efficiency while ensuring falsification proofness, meaning that agents cannot gain by falsifying their scores. Falsification proofness eliminates negative externalities, leads to fairer mechanisms, and improves the welfare of agents. Falsification-proof mechanisms are fair in the sense that they guarantee the same assignment probability to all agents with the same score. They can improve the welfare of agents by eliminating the cost they incur from falsifying,[5] providing greater benefits to those with lower gaming abilities, whose costs are higher.

Specifically, we address the problem of allocating a fixed mass of homogeneous objects (or prizes or labels) 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. If agents do not falsify, they produce a *natural score*, which reflects their true characteristics.[6] Agents who *falsify* produce an altered score at a cost. We assume that an agent's natural score is positively correlated with their *worth*: the designer's value of assigning an object to the agent. The designer's *outside option* of retaining an object has value zero. We characterize the falsification-proof mechanism that maximizes the aggregate worth of rewarded agents. It allocates the good stochastically with a probability that increases smoothly with the score, generating both rejection and allocation errors.

What is the welfare impact of falsification proofness? In standard mechanism design, misreporting is costless, and truth-telling is inconsequential due to the revelation principle. However, in Perez-Richet and Skreta (2022) we show that, under costly falsification, optimal mechanisms harness falsification to enhance allocative efficiency.[7] Therefore, requiring falsification proofness results in a loss of allocative efficiency for the designer. However, we find that it has the opposite effect on agents. We show that, in the absence of resource and quota constraints, agents prefer the optimal falsification-proof mechanism to the optimal mechanism, regardless of their natural score (Proposition 4). Furthermore, we show that the loss in allocative effi-

[5]In the computer science literature, Milli, Miller, Dragan, and Hardt (2019) consider optimal binary classifiers when agents are strategic (i.e., falsify). They show that optimal classifiers can generate a significant cost in falsification for the agents, and argue that this cost bears disproportionately on disadvantaged groups.

[6]Frankel and Kartik (2019) introduced the term *natural* action to refer to the unmanipulated action of a given type.

[7]Perez-Richet and Skreta (2022) derive optimal *tests* which correspond to optimal allocation mechanisms in the current framework. See also our discussion on the value of commitment in Section 6 (in particular, Proposition 8) for a related result.

ciency due to falsification proofness can be arbitrarily small and the gain to agents arbitrarily large if the falsification technology has increasing returns to scale (Proposition 5). These results confirm that, even abstracting from externality and fairness concerns, falsification-proof mechanisms can serve as a cost-effective means to enhance agents' welfare. The central practical message of this paper is that falsification-proof mechanisms can offer a valuable alternative for a planner who values both allocative efficiency and agents' welfare.

We formulate a general model with multiple groups of heterogeneous agents, a resource constraint and possible group-specific quotas. While these allocative constraints introduce rich interdependency across groups and interesting economic effects that we describe later, the key force shaping optimal allocation rules is the falsification-proofness requirement. Furthermore, allocative constraints are often absent when objects are immaterial, such as services, labels, certification or awards. We, therefore, start by solving a *baseline problem* that focuses on the falsification-proofness constraint while abstracting from allocative constraints and group multiplicity. The designer's objective in this problem incorporates a fixed arbitrary outside option value which we later use as an adjustment tool to satisfy the resource and quota constraints. Solving the baseline problem is also a key step in solving the full problem.

We solve the baseline problem in closed form for two broad classes of cost functions that capture different falsification technologies. If the cost function has *upward decreasing differences*,[8] the falsification-proofness constraints bind locally, and we use a first-order approach to solve it. If, instead, the cost function has *upward increasing differences*,[9] the falsification-proofness constraints do not bind locally, which precludes the use of the first-order approach. To address this challenge, we transform the baseline problem into a program that is equivalent to the dual of the classical Monge-Kantorovich optimal transport problem, which we leverage to characterize the optimal allocation rule in Theorem 1.[10] While optimal transport is increasingly used in economics, this connection to the dual problem is novel as we explain in the literature review (Section 7).

We study how changes in falsification technology or the distribution of natural scores affect the solution to the baseline problem. We analyze the effects of changing the distribution of scores and returns to scale in the falsification technology in Ap-

[8]UDD technologies include costs that are superadditive in the amount of falsification, capturing decreasing returns to scale in falsification.

[9]UID technologies include costs that are subadditive in the amount of falsification, capturing increasing returns to scale in falsification.

[10]Because the first-order approach is standard, the corresponding theorem for the UDD case appears in Appendix B.

pendix D. In Section 4, we focus on the effect of *gaming ability*, defined as the inverse of a multiplicative scaling factor on the least cost of falsification in the population. This result yields interesting economic insights.

Since agents are heterogeneous, but only those with the lowest falsification cost (or highest gaming ability) shape the optimal allocation rule that applies to all, their presence exerts an externality on others. We can interpret our comparative statics exercise as an evaluation of this externality, since the effect of removing the highest gaming ability agents from the group results in a decrease in the gaming ability level that shapes the optimal rule. We find that the externality exerted by the highest gaming ability agents is nuanced: it is positive on low-score agents and negative on high-score agents if these agents have sufficiently low gaming ability.[11] We can also use our comparative statics exercise to assess the effect of conditioning allocation on observables. With the ability to distinguish between a group of high-gaming ability agents and a group of low-gaming ability agents, the designer may choose to treat them as a single group or as distinct groups. This choice makes no difference for the high-gaming ability group, as it always faces the same optimal falsification-proof rule shaped by high gaming ability. In contrast, the effect on the low-gaming ability group is subtle. Since being insulated from the other group amounts to a decrease in gaming ability for this group, the effect can be assessed through our comparative statics result. Therefore, the effect of discrimination on the low-gaming ability group can be either uniformly beneficial, uniformly harmful, or beneficial for high-score agents while harmful for low-score agents.

We then solve the full problem of the designer with multiple groups and allocative constraints by splitting it into an *across* problem and a series of *within* problems. Each *within* problem deals with the allocation of a fixed mass of objects according to scores within a group. The across problem consists of allocating objects across groups while satisfying the quota and resource constraints. We show in Theorem 2 that the solution to the within problem coincides with the solution of a baseline problem where the designer's outside option has been adjusted to ensure that the desired mass of objects is allocated. We then characterize the solution to the across problem in Theorem 3, and provide an algorithm to find it through mass adjustments.

We show that the designer's welfare decreases with gaming ability and increases with first-order stochastic dominance shifts in score distribution. Comparative statics for the optimal rule display rich within-group and across-group effects. A change

[11] If the gaming ability of these least-cost agents is sufficiently high, the effect becomes uniform but may be either positive or negative depending on expected worth in the population.

in characteristics has *direct effects* on the baseline allocation rule for that group, holding the group's outside option constant. Direct effects can alter the mass of objects assigned to the group, necessitating adjustments in the outside options of *all* groups to maintain constraint feasibility. These adjustments generate within-group and across-group *indirect effects*. In Proposition 6, we show that indirect effects, because they work through outside option adjustments, impact agents of all scores uniformly, either increasing or decreasing their allocation probability.

To illustrate these indirect effects, we study the consequences of lowering gaming ability in one group, while the other group benefits from a quota in a two-group setting. If the quota is low, the indirect effect of the change must be absorbed by both groups. Hence, all agents in the second group are harmed. In the first group, the direct and indirect effects combine, increasing agents' ex ante welfare, but harming low-score agents while benefiting high-score agents. If the quota is high, it binds both before and after the change, shielding the second group from indirect effects, which must therefore be entirely absorbed by the first group. Hence, the first group receives the same mass of goods on aggregate, but low-score agents are harmed while high-score agents gain.

In Section 6, we discuss some robustness properties of optimal falsification-proof rules. Since optimal falsification-proof allocation rules guard against falsification by the least-cost agents, they are also optimal for a max-minimizing designer seeking robustness against unknown falsification technologies. Additionally, we highlight that score-based allocation rules perform just as well as direct recommendation mechanisms, which require more commitment from the designer, but are simpler because score-based rule do not rely on a mediator. Additionally, we highlight that score-based allocation rules are without loss of generality and thus perform just as well as direct recommendation mechanisms. Finally, we show that optimal falsification-proof rules can be implemented by a designer who acts as a certification intermediary and can only design information structures while lacking control over the ultimate allocation decision (i.e., no commitment to the decision). Thus, the optimal rules are robust to varying levels of commitment on the part of the designer.

The remainder of the paper is structured as follows. Section 2 describes the model. Section 3 analyzes the baseline problem. Section 4 studies the effects of changes in gaming ability and includes a welfare analysis. Section 5 solves the full designer's problem, provides a comparative statics analysis, and illustrates across-group feedback effects. We conclude with a discussion in Section 6 and a literature review in Section 7. All proofs and some additional results are in the Appendix.

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 that captures their *worth*, that is, the designer's value of allocating an object to them. Without loss of generality, the value of the outside option (not allocating an object) is normalized to 0. Agents may know their worth if θ is a sufficient statistics for w, or not.

The first dimension of the type $i \in I$ encompasses all relevant publicly observable and unfalsifiable characteristics of an agent. We refer to i as an agent's *group*, and assume I is a finite set. The mass of group i is $\mu_i > 0$, where $\sum_i \mu_i = 1$.

We assume the existence of an exogenous one-dimensional metric, the *score*, measuring some private characteristics of the agent. The second dimension of an agent's type $s \in S_i \subseteq \mathbb{R}$ is their *natural score*, which they obtain when they do not interfere with the measuring technology. An agent can falsify their score at a cost, so the designer observes a *falsified score* t instead of their natural score s.

The last dimension of type $k \in K_{i,s}$ is a vector of privately known characteristics that includes an agent's value for the good $v(k) > 0$, information about their individual falsification cost, and possibly other characteristics correlated with their worth.

Distributional assumptions. 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 be, and typically are, correlated; but they are independent from other agents' characteristics. F_i denotes the cumulative distribution function of natural score 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 $K_{i,s}$.

Designer and agent payoffs. We assume that the worth w is bounded and integrable conditional on (i, s). We denote the corresponding expected worth by $w_i(s) = \mathbb{E}(w|i, s)$, and by $\bar{w}_i = \mathbb{E}(w|i)$ the expected worth in group i. The designer's payoff from assigning an object to a group i agent with score s is $w_i(s)$. We assume that score and worth are positively related in the sense that, for every group i, $w_i(s)$ is strictly increasing. The expected payoff of an agent is $\alpha v - C(t, \theta)$, where α is the probability of getting an object, and $C(t, \theta) \geq 0$ defines the cost for type θ to produce a score t.

Not falsifying is costless so $C(s, \theta) = 0$ for an agent of type $\theta = (i, s, k)$. The cost of producing score t depends not only on the natural score s, but also on k. The falsification cost may reflect technical costs, psychological lying costs as well as expected penalties.

Falsification-proof mechanisms. We restrict the designer to *falsification-proof mechanisms*, that is, mechanisms that incentivize agents to produce their natural scores. Under this assumption, we show in Perez-Richet and Skreta (2023) that it is without loss of generality to restrict attention to *score-based allocation rules* $\boldsymbol{\alpha} = (\alpha_i)_{i \in I}$, where $\alpha_i : S_i \to [0, 1]$ is the probability that an object is allocated to an agent from group i conditional on their produced score. Such rules only condition on the observable dimension i, and the score s, but not on k. An (incomplete) intuition of why it is without loss of generality is that, since all agents desire an object, also conditioning the allocation on k would then lead all agents from the same group i who produce the same score s to pool on claiming k' that maximizes the probability of receiving an object.[12] Among agents from the same group and with the same natural score, falsification is most tempting for those with the highest valuation and lowest falsification costs. Therefore, a mechanism is falsification-proof if and only if it satisfies the following constraint[13]

$$\forall (i, s, t), \quad \alpha_i(t) - \alpha_i(s) \leq c_i(t|s), \tag{FPC}$$

where $c_i(t|s) = \inf_{k \in K_{is}} \frac{1}{v(k)} C(t, i, s, k)$ is the *least cost* for a member of group i with natural score s to falsify to t.[14] We refer to c_i as the *least-cost* or simply cost function for group i.

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

[12] Lemma 1 in Akbarpour et al. (2024) makes a similar point in a setting where all dimensions are costless to misrepresent, and transfers are allowed: the principal can only elicit information on the dimensions that correlate with agents' willingness to pay.

[13] We can also interpret (FPC) as being motivated by inequality awareness as in Akbarpour et al. (2024). The cost $c_i(t|s)$ then acts as a bound on allocative inequality between score pairs s and t. We thank Ricardo Alonso for suggesting this interpretation.

[14] This infimum exists because $C(t, \theta)$ is bounded below by 0. We assume that this bound is tight in the sense that, for every i, s, t and every $\varepsilon > 0$, there exists a strictly positive mass of agents from group i with natural score s whose cost of falsifying to t is lower than $c_i(t|s) + \varepsilon$.

Assumption 1 (Regularity). *The least-cost function $c(t|s)$ 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 $c(t|s)$ by c_t and c_s. 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 salient classes of cost functions.

Definition 1 (Upward Differences). *A cost function $c(t|s)$ has upward increasing differences if, for all $s < s' \leq t < t'$,*

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

and upward decreasing differences if, for all $s < s' \leq t < t'$,

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

These conditions only bear on upward falsification because we show that downward falsification is never beneficial under optimal allocation rules. To gain intuition about their interpretation, it is useful to consider different families of cost functions and to reason as if the least cost function of the group c were a particular agent's cost function instead.

Example 1 (The Euclidean family). *A cost function is* Euclidean *if $c(t|s) = \mathcal{C}((t - s)^+)$, where $\mathcal{C} : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuously differentiable increasing function with $\mathcal{C}(0) = 0$. Euclidean costs are a convenient modeling choice often used in the literature. They satisfy (UID) if $\mathcal{C}$ is concave (or, more generally, subadditive) and (UDD) if $\mathcal{C}$ is convex (or, more generally, superadditive). The monotonicity of upward differences captures economies of scale in falsification $t - s$: (UID) implies increasing returns to scale, while (UDD) implies decreasing returns to scale. Both increasing and decreasing returns to scale in falsification may arise in different contexts. If the designer's characteristic of interest (the score) is, for example, emission levels of cars models, it is reasonable to assume increasing returns to scale in falsification: for small amounts of emissions to conceal, the cost rises sharply, but once the agent has paid the setup costs to develop software that lowers emissions under testing conditions, concealing additional emissions becomes less costly. If the natural score measures an analytical skill, decreasing returns to scale may be more appropriate: an agent can memorize solutions to some test problems, but achieving a significantly higher* *score requires learning progressively more or increasingly difficult problems, raising the marginal cost. Another example is when falsification costs arise from a linear detection probability* πx *and a fine* $\Phi x + \Phi_0$*, both of which increase with the falsification level* x*. In this case, the Euclidean cost* $\mathcal{C}(x) = \pi x(\Phi x + \Phi_0)$ *satisfies* (UDD). ◇

Example 2 (The shifted Euclidean family)**.** *We generalize the Euclidean class by considering* Shifted Euclidean *cost functions of the form* $c(t|s) = \kappa(s)\mathcal{C}\big((t-s)^+\big)$*. This family captures correlations between the natural score and* gaming ability,[15] *which is captured by* $1/\kappa(s)$*. For instance, consider the* shifted linear *cost function where* $\mathcal{C}(x) = x$*. In this case, c satisfies* (UID) *if* $\kappa(s)$ *increases with* s*, and* (UDD) *if* $\kappa(s)$ *decreases with* s*. If the natural score reflects financial need or hardship, it is natural to assume that gaming becomes more challenging at higher natural scores. Conversely, if the natural score reflects skill or aptitude, gaming may become easier at higher natural scores.* ◇

The (UID) class is particularly suited for studying situations in which falsification involves substantial setup costs. Furthermore, we derived the unconstrained optimal allocation rule in Perez-Richet and Skreta (2022). This allows us to analyze the welfare consequences of imposing falsification-proofness under (UID) in Section 4.2.

Allocative constraints. In many applications we consider, allocative constraints are irrelevant because 'objects' are immaterial, such as services, certificates, labels or awards. To cover other applications, we allow for a *resource constraint* and *quota constraints*. The resource constraint requires

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

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

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

A mechanism is feasible if it satisfies these *allocative constraints*. If $\bar{\rho} = 1$ and $\phi = 0$, the problem has no allocative constraints. Feasible mechanisms also need to satisfy

[15] We follow the terminology of Frankel and Kartik (2019).

probability constraints:

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

Designer's program. The restriction to falsification-proof mechanisms implies the agent's observed score is the natural one, so α_i writes as function on the natural score s rather than falsified scores. The designer's program is to choose a score-based allocation rule $\boldsymbol{\alpha}$ that solves:

$$\max_{(\alpha_i)_{i\in I}} \sum_i \mu_i \int_{S_i} w_i(s)\alpha_i(s)dF_i(s) \tag{P}$$
$$\text{s.t. (PC), (FPC), (RC), (QC).}$$

In Section 3, we address the *baseline problem*, a key subproblem ignoring group multiplicity and allocative constraints. In Section 5, we solve (P) by decomposing it into *within* and *across* problems: the former allocates a fixed object mass within a group, while the latter allocates object masses across groups.

3 Baseline problem

We first solve a *baseline program* that abstracts from group multiplicity and allocative constraints. We therefore drop the group index i in this section.

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

The term $w(s) - \hat{w}$ in the objective represents the allocative surplus relative to an exogenous *outside option* $\hat{w}$. Let $\hat{s}$ denote the *eligibility threshold*, defined by $w(\hat{s}) = \hat{w}$. Allocative surplus is positive for *eligible* scores, $s \ge \hat{s}$, and negative for ineligible scores, $s < \hat{s}$. We refer to the solution to the baseline problem as the *baseline allocation rule*.

Solving (BP) is a key step in solving the designer's program (P) and is also of independent interest. First, it isolates the forces introduced by the falsification-proofness constraint. Second, when objects are immaterial, such as labels, it is natural to assume the absence of allocative constraints. In this scenario, the allocation problems become independent between groups and reduce to the baseline problem with an outside option $\hat{w} = 0$ within each group. Finally, a technical contribution of this paper is the connection it establishes between the baseline problem and optimal transport theory.

To solve the designer's program (P), which involves multiple groups and allocative constraints, we use a baseline allocation rule within each group and adjust outside options to meet resource and quota constraints. Thus, eligibility thresholds in each group are endogenous in the general problem.

3.1 Preliminary analysis

Smoothness and monotonicity. The first-best rule for the designer is to allocate objects to agents with eligible scores and withhold them from those with ineligible scores. It is not falsification proof: an agent with score just below the eligibility threshold can gain by falsifying to the threshold. More generally, falsification proofness bounds the growth of allocation probability between scores by the cost of falsifying from one to the other. Falsification-proof rules therefore inherit the smoothness of the cost function. Specifically, we show in Appendix A that they are Lipschitz continuous. We show that optimal falsification proof rules must also be monotonic, and flat outside of a *growth interval* around the eligibility threshold whose size depends on the magnitude of falsification costs.

Priority. A group has *high-score priority* if its expected worth $\bar{w} = \mathbb{E}(w)$ exceeds the outside option $\hat{w}$, *low-score priority* if $\bar{w} < \hat{w}$, and *neutral priority* if $\bar{w} = \hat{w}$. In the absence of information on scores, the designer would allocate objects to all agents in a high-score priority group and not allocate any object in a low-score priority group. In the baseline problem, priority determines whether it is more important for the designer to ensure that top scores obtain an object with certainty (high-score priority), or to ensure that lowest scores do not obtain an object (low-score priority). Since a group's priority is determined by its outside option, priorities are endogenous in the original problem (P).

Cumulative surplus functions. To solve the baseline problem, we introduce *cumulative surplus* functions, which appear in the designer's objective when integrating by parts. The *upward cumulative surplus* at s is the total amount of surplus generated by scores above s. It corresponds to the marginal gain of uniformly increasing the allocation probability of all such scores:

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

The *downward cumulative surplus* at s is the total amount of negative surplus generated by scores below s, and corresponds to the marginal gain of uniformly decreasing the allocation probability of all such scores:

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

In the analysis, we work with upward cumulative surplus under low-score priority and with downward cumulative surplus under high-score priority. To unify notations, we define a composite *cumulative surplus function* that encompasses both cases:

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

Next, we list useful properties of the cumulative surplus functions. Figure 1 illustrates the functions and their properties.

Lemma 1 (Properties of cumulative surplus)**.**

(i) $\mathcal{W}(\cdot,\hat{w})$ *is continuous and single-peaked at* $\hat{s}$*.*

(ii) For every $\nu \in [0, \mathcal{W}(\hat{s},\hat{w})]$*, there exist a unique pair of scores* $s_*(\nu) \le \hat{s} \le s^*(\nu)$ *such that* $\mathcal{W}\big(s_*(\nu),\hat{w}\big) = \mathcal{W}\big(s^*(\nu),\hat{w}\big) = \nu$*.*

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

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

(v) $s^*(0) = \overline{s}$ *under low-score priority, and* $s_*(0) = \underline{s}$ *under high-score priority. Under neutral priority,* $[s_*(0), s^*(0)] = [\underline{s}, \overline{s}]$*.*

Matching ineligible and eligible scores. We use cumulative surplus to build a *matching function* between ineligible and eligible scores, also illustrated in Figure 1. Specifically, we define the decreasing matching function $m : [s_*(0), \hat{s}] \to [\hat{s}, s^*(0)]$ that maps each ineligible score $s \in [s_*(0), \hat{s}]$ to an eligible 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 that a pair (s_*, s^*) is a *matching pair* if $s^* = m(s_*)$.

The matching function plays an important role in the analysis for two reasons. First, it describes the (nonlocal) binding constraints in the (UID) case. Second, together with the Lagrange multiplier, ν, on the probability constraint, it determines the *growth interval* of the allocation rule: Point (ii) of Lemma 1 implies that matching

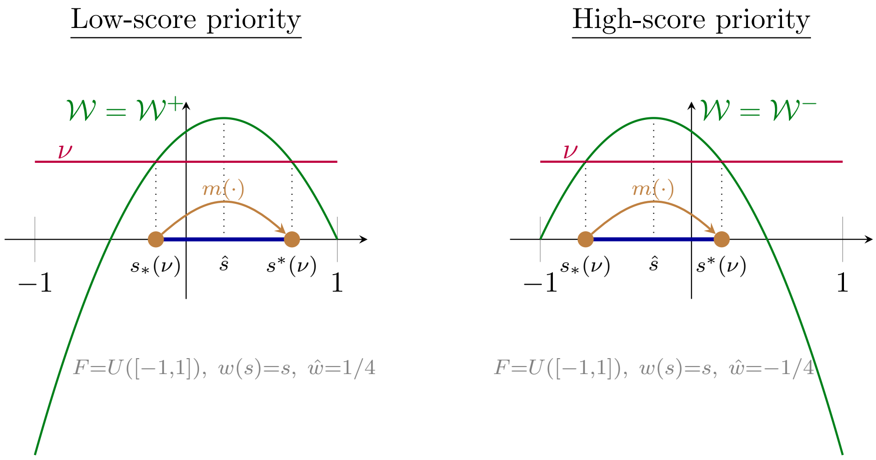

Figure 1: *Cumulative surplus, matching pairs and growth interval.*

pairs characterize a set of intervals $[s_*, s^*]$ around the eligibility threshold $\hat{s}$ that satisfy the following *Zero Average Surplus* condition:

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

We call such intervals (ZAS) intervals. They form a family of nested intervals around the eligibility threshold, ranging from the maximal interval $[s_*(0), s^*(0)]$ to the minimal interval $\{\hat{s}\}$.

We show in Appendix A that the growth interval of a baseline rule is necessarily a ZAS interval. This implies that optimal allocation rules exhibit bunching at the bottom in low-score priority groups, since scores in $[\underline{s}, s_*(0)]$ receive an object with null probability. In contrast, there is bunching at the top in high-score priority groups, since scores in $[s^*(0), \overline{s}]$ receive an object with probability one.

The simplified program. In Appendix A, we show how to construct a solution to the baseline problem by first solving the following *simplified program*

$$\max_{\alpha} \int_{s_*}^{s^*} \big\{w(s) - \hat{w}\big\}\alpha(s)dF(s)$$
$$\text{s.t. } \alpha(t) - \alpha(s) \le c(t|s) \quad \forall s_* \le s < t \le s^*,$$

which limits the integral in the objective function to a ZAS interval $[s_*, s^*]$, only considers falsification-proofness constraints over this interval, and relaxes the probability constraint. We then obtain the baseline rule via the following procedure.

Procedure 1.

- ***Step 1.*** *Solve the simplified program for all* $[s_*, s^*]$. *The solution* α *is determined up to an additive constant.*

- ***Step 2.*** *Select* (s_*, s^*) *to either (i) saturate the probability constraint:* $\alpha(s^*) - \alpha(s_*) = 1$, *or (ii) set* $(s_*, s^*) = \big(s_*(0), s^*(0)\big)$.

- ***Step 3.*** *Set the additive constant so that* $\alpha(s_*) = 0$ *under low-score priority,* $\alpha(s^*) = 1$ *under high-score priority, or to any value such that* $\alpha(s^*) - \alpha(s_*) \leq 1$ *under neutral priority.*

- ***Step 4.*** *Set* $\alpha(s) = \alpha(s_*)$ *for* $s < s_*$, *and* $\alpha(s) = \alpha(s^*)$ *for* $s > s^*$.

◇

Step 1 determines the growth rate and shape on the growth interval. Step 2 identifies the growth interval. Step 3 sets the maximum allocation probability based on group priority. Step 4 extends the rule beyond the growth interval by assigning the minimum probability to scores below s_* and the maximum probability to scores above s^*.

3.2 Baseline rules

The shape of the baseline rule on the growth interval is determined by the binding upward falsification-proofness constraints. Under (UDD), these constraints bind locally, allowing for a first-order approach. In contrast, under (UID), the constraints bind non-locally, rendering the first-order approach inadequate. Instead, we solve the simplified program by establishing a connection with the dual of the Monge-Kantorovich optimal transport problem. We focus on the methodology under (UID) technologies. For (UDD), we provide the formula of the baseline rule with an intuitive explanation, relegating the analysis to Appendix B, as it involves more standard arguments.

Optimal allocation through optimal transport. We start by drawing a connection between the simplified program and optimal transport theory. We consider the following relaxation of the simplified problem:

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

in which we only require falsification proofness to prevent ineligible scores from falsifying to eligible targets, and separate the objective function between ineligible and eligible scores.

In our model, a mass of agents is distributed over the space of scores, which we can think of as locations. Each (infinitesimal) agent in "location" s is endowed with an amount $w(s) - \hat{w}$ of surplus. Alternatively, we can describe the problem in terms of masses of negative or positive surplus available at different locations. Each location s then harbors a mass $|w(s) - \hat{w}|dF(s)$ of negative surplus if s is ineligible, and of positive surplus if s is eligible. We frame the program as a problem involving the transportation of negative surplus from ineligible locations to eligible locations that harbor positive 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 surplus locations $Y = [0, \hat{s} - s_*]$, and the space of positive surplus locations $Z = [0, s^* - \hat{s}]$. By (ZAS), each of these spaces harbors the same total mass of 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 factor is the total mass. 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 the 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)

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

To view this program in terms of transport, suppose that the designer is a planner who wants to support the production of a locally produced good (surplus) at locations in Z, but discourage it at locations in Y. As a result, they wish to maximize the profit of producers at eligible locations in Z, while minimizing 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 that demand exceeds supply at every location and that the economy is entirely regulated so that the planner can choose the price at which the good is 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 is then naturally interested in the least costly routes between Y and Z. Indeed, their program is actually the dual of the optimal transport problem, which seeks to find the least costly way of transporting P to Q:

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

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

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[16] that the optimal transportation plan is deterministic and assortative. It is given precisely by the matching function m: All surplus at y is transported to location z such that $\hat{s}+z = m(\hat{s}-y)$. In terms of our original problem, this implies that the binding falsification-proofness constraints are between scores $s \in [s_*, \hat{s}]$ and their match $t = m(s)$. Optimal transport theory also provides the unique (up to an additive constant) optimal *price functions* ϕ and ψ in closed form.

Baseline rule under (UID). We then obtain the optimal allocation rule by following Procedure 1:

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

Over the growth interval, the allocation rule grows at a speed of $-c_s\big(m(x)|x\big)$ for ineligible scores x, and at a speed of $c_t\big(x|m^{-1}(x)\big)$ for eligible scores x. By duality in the optimal transport problem, the falsification-proofness constraint is binding

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

between matching scores, and in particular for the pair (s_*, s^*). Therefore:

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

Hence, the growth interval $[s_*, s^*]$ is uniquely determined by the *boundary condition*

$$s_* = \min\big\{s \in [s_*(0), \hat{s}] : c\big(m(s)|s\big) \leq 1\big\}. \tag{B}$$

The corresponding Lagrange multiplier on the probability constraint is $\nu = \mathcal{W}(s_*, \hat{w})$. In accordance with step 2 of Procedure 1, the boundary condition makes the probability constraint bind if possible, and otherwise selects the maximal growth interval $[s_*(0), s^*(0)]$, with $\nu = 0$.

Whether or not the probability constraint binds depends on the magnitude of falsification costs, and its degree of slackness is measured by the *probability gap*

$$\Gamma_{uid} = 1 - c(s^*|s_*). \tag{Gap}$$

Then, the probability constraint is slack, and $\Gamma_{uid} > 0$, if and only if the *slackness condition* $c\big(s^*(0)|s_*(0)\big) < 1$ is satisfied.

The probability gap must be managed according to priority. In the low-score priority case, the probability gap is optimally withheld from the agents to ensure $\alpha(s_*(0)) = 0$, thereby preventing low-score agents from receiving an object. In the high-score priority case, it is optimally allocated to agents to ensure $\alpha(s^*(0)) = 1$, ensuring that high-score agents receive an object with certainty. To perform this task, we define the *share* index

$$I(\hat{w}, r) = \mathbb{1}_{\hat{w}<\bar{w}} + r\mathbb{1}_{\hat{w}=\bar{w}},$$

that represents the share of the probability gap allocated to the agents, taking value 0 under low-score priority, and 1 under high-score priority. We refer r as the *neutral gap share*. We define $\bar{I}(\hat{w}, r) = 1 - I(\hat{w}, r)$.

Under neutral priority, the interval $[s_*(0), s^*(0)]$ equals the full support $[\underline{s}, \overline{s}]$, reducing the gap condition to $c(\overline{s}|\underline{s}) < 1$. The designer is then indifferent about the share $r \in [0, 1]$ of the probability gap that is allocated to agents, and the baseline rule is not unique. Multiplicity therefore arises when (i) a group has neutral priority, and

(ii) the cost for the lowest score to falsify to the highest score is less than 1:

$$\hat{w} = \bar{w} \text{ and } c(\bar{s}|\underline{s}) < 1. \tag{Mult}$$

Although the designer is indifferent, the choice of r plays an important role in the solution of the general problem (P), as it allows the designer to adjust the total mass of goods accruing to a group with neutral priority so as to satisfy the allocative constraints.

Theorem 1 (Baseline rule under UID)**.** *If the cost function satisfies* (UID)*, then* α^*_{uid} *solves* (BP)*. It is independent of* r *and the unique baseline rule unless* (Mult) *holds. Otherwise, the set of baseline rules is* $\left\{\alpha^*_{uid}(\cdot, \bar{w}, r)\right\}_{r\in[0,1]}$.

To complete the proof of the theorem, we show in the appendix that (UID) implies that α^*_{uid} satisfies the relaxed falsification-proofness constraints between scores on the same side of the eligibility threshold $\hat{s}$.

Baseline rule under (UDD). The baseline rule under (UDD) has a similar structure but differs by its growth speed, which must be equal to the cost of a marginal upward falsification $\alpha'(s) = c_{t^+}(s|s)$. Consequently, the (UDD) baseline rule is

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

The gap Γ_{udd}, the boundary condition, and the multiplicity condition have different expressions that we provide in Appendix B.

The regimes of baseline rules. Under both classes of costs, the baseline rule α^* has two regimes depending on whether the probability constraint binds. It is in the *binding regime* if falsification costs are high and the probability constraint binds. Allocative distortions are then concentrated on the growth interval, with no distortion at the boundaries of the score set: scores in $[s^*, \bar{s}]$ receive a good with certainty and scores on $[\underline{s}, s_*]$ with null probability. It is in the *slack regime* if falsification costs are low and the probability constraint is slack. In this regime, the baseline rule displays distortions on the growth interval $[s_*(0), s^*(0)]$ as well as at the boundaries of the score set. Under low-score priority, allocation is distorted at the top, $\alpha^*(\bar{s}) < 1$, while it is distorted at the bottom, $\alpha^*(\underline{s}) > 0$, under high-score priority. Under neutral priority, it may be distorted indifferently at either or both boundaries, depending on the choice of the neutral gap share r.

3.3 The shape of baseline rules

We apply our characterization to the Euclidean (Example 1) and Shifted Linear (Example 2) families of cost functions. We show that for both families, baseline rules are S-shaped in the (UID) case: convex over ineligible scores, and concave over eligible ones. This is an interesting property as it boosts the probability of allocation for eligible scores, and deflates it for ineligible ones.[17] In the (UDD) case, by contrast, the baseline rule is linear for Euclidean costs, and concave for shifted linear costs.

First, consider a Euclidean cost function $\mathcal{C}$. The cost function then satisfies (UID) if $\mathcal{C}$ is concave, and (UDD) if it is convex.

Proposition 1 (Baseline rules under Euclidean cost). *If $\mathcal{C}$ is convex, the cost function satisfies (UDD), and the baseline rule is linear in s on $[s_*, s^*]$, with $\alpha^*(s) = \mathcal{C}'(0)\big(s - s_*\big)$. If $\mathcal{C}$ is concave, the cost function satisfies (UID), and the baseline rule is convex in s on $[s_*, \hat{s}]$, and concave on $[\hat{s}, s^*]$.*

In the (UDD) case, the baseline rule is linear with slope $\mathcal{C}'(0)$. This is because the (FPC) constraints bind locally and the marginal cost of falsification is constant for all scores and equal to $\mathcal{C}'(0)$ in the Euclidean case. In the (UID) case, we show that the baseline rule has a convex-concave S shape on its growth interval, switching exactly at the eligibility threshold. If, in addition, the score distribution on the growth interval is symmetric around the eligibility threshold, the baseline rule also has a center of symmetry at the eligibility threshold. This is the case in the example shown in Figure 2.

Next, consider a shifted linear cost function $c(t, s) = \kappa(s)(t - s)^+$, which satisfies (UID) if $\kappa(s)$ is increasing, [18] and (UDD) if $\kappa(s)$ is decreasing.

Proposition 2 (Baseline rules under shifted linear costs). *In the shifted linear cost model, the baseline rule is concave in the (UDD) case. In the (UID) case, it is convex below the eligibility threshold if $\frac{\kappa''(s)}{\kappa'(s)} \leq \frac{2}{\bar{s}-s}$, and always concave above the eligibility threshold.*

[17]For shifted linear costs, convexity below the eligibility threshold requires an additional sufficient condition on κ, whereby it cannot be too convex.

[18]Note that, in order to ensure that $c_s(t|s) < 0$, we need $\kappa'(s)/\kappa(s) \leq 1/(\bar{s} - s)$

Because the S shape can better approximate the first-best allocation rule, which is a step function, Proposition 1 and Proposition 2 suggest that the optimal falsification-proof rule may not be too costly for the designer in the (UID) case, as its shape reduces distortions in allocative efficiency. This intuition is confirmed by our welfare analysis in Section 4.2.

4 Baseline comparative statics and welfare

We analyze how changes in the cost function affect the baseline allocation. We then assess the welfare impact of restricting the designer to falsification-proof mechanisms in the (UID) case, showing that the designer's loss in allocative efficiency is offset by agents' welfare gains across all scores. We explore additional comparative statics, with respect to score distribution and returns to scale, in Appendix D.

In this section, we fix the value of the outside option $\hat{w}$ and the neutral gap share r, and therefore omit the dependence of the baseline rule α^* on these variables in the notations. The comparative statics of baseline allocation rules apply to the solution of the designer's problem when allocative constraints are absent, or nonbinding for the group of interest. When constraints bind, changes in primitives (e.g., changes in falsification costs or score distributions) can cause both *direct* and *indirect* effects. This section describes direct effects. If a change in group i's characteristics directly increases the mass of objects allocated to that group, the resource constraint may be violated, triggering upward adjustments to outside options and resulting in both *within-group* and *across-group* indirect effects. We analyze these indirect effects in Section 5.3.

Default setup. To illustrate our results, we provide examples in a *default setup* defined as follows. The score distribution is uniform on $[-3, 2]$, the worth function is $w(s) = s$ and the outside option is given by $\hat{w} = 0$, so the eligibility threshold is $\hat{s} = 0$, and the example has low-score priority. The first-best allocation is to award an object to the 40% of agents who are eligible, $s \in [0, 2]$. In this default setup, we vary the falsification cost function.

4.1 The effect of falsification cost

Gaming ability. We parameterize the cost function by a scaling *gaming ability* factor γ and perform comparative statics with respect to γ. Specifically, we consider parameterized cost functions

$$c^{\gamma}(t|s) = \frac{1}{\gamma} c(t|s),$$

so higher gaming ability implies lower falsification cost. Therefore, the gap condition holds when gaming ability is sufficiently high. Specifically, let $\hat{\gamma}$ denote the gaming ability threshold such that the gap condition holds for $\gamma > \hat{\gamma}$. This threshold is given by $\hat{\gamma} = \int_{s_*(0)}^{s^*(0)} c_{t+}(x|x)dx$ under (UDD), and by $\hat{\gamma} = c\big(s^*(0)|s_*(0)\big)$ under (UID). Then α^* is in the binding regime if $\gamma \leq \hat{\gamma}$, and in the slack regime otherwise.

Comparative statics. We show that increasing gaming ability starting from an initially low level (binding regime) benefits low-score agents and harms high-score agents. If the initial level is sufficiently high (slack regime), an increase affects all agents in the same way, benefiting them if the group has high-score priority and hurting them under low-score priority. The results hold for both the (UID) and the (UDD) case, and we denote the baseline allocation rule by α^*.

Proposition 3 (Effect of gaming ability)**.** *Consider increasing gaming ability from γ to $\gamma' > \gamma$, and denote the corresponding baseline rules by $\alpha^*_{\gamma'}$ and α^*_{γ}. Then:*

(i) ***Binding regime / Low gaming ability:*** *If $\gamma \leq \hat{\gamma}$, then $\alpha^*_{\gamma}(s) - \alpha^*_{\gamma'}(s)$ is single-crossing from below.*

(ii) ***Slack regime / High gaming ability, low-score priority:*** *If $\gamma \geq \hat{\gamma}$, and the group has low-score priority or neutral priority with $r = 0$, then $\alpha^*_{\gamma}(s) \geq \alpha^*_{\gamma'}(s)$. Furthermore, the difference in probability increases in s and is equal to 0 at $s_*(0)$.*

(iii) ***Slack regime / High gaming ability, high-score priority:*** *If $\gamma \geq \hat{\gamma}$, and the group has high-score priority or neutral priority with $r = 1$, then $\alpha^*_{\gamma'}(s) \geq \alpha^*_{\gamma}(s)$. Furthermore, the difference in probability decreases in s and is equal to 0 at $s^*(0)$.*

This comparative statics is illustrated in Figure 2 for the default setup. The default setup has low-score priority so it illustrates points (i) and (ii) of Proposition 3.

An externality interpretation. These comparative statics can be interpreted in terms of externalities. We use gaming ability to parameterize the least-cost function. Recall that least-cost agents shape the falsification-proofness constraint and, thereby, the baseline allocation rule. Their presence imposes an externality on other agents.

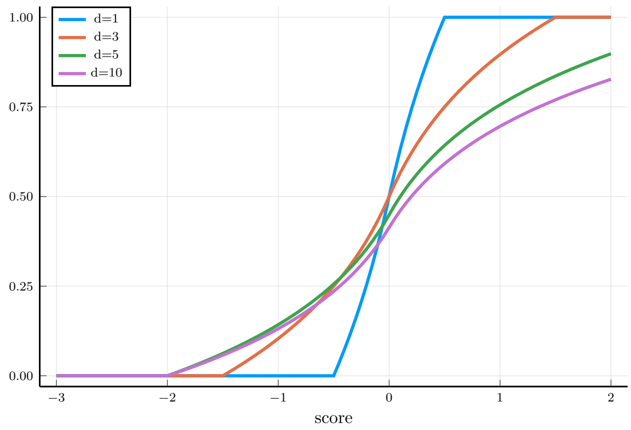

Figure 2: *Effect of gaming ability on* α^*. *Default setup, Euclidean costs:* $\mathcal{C}^d(x) = \frac{\log(1+x)}{\log(1+d)}$, *gaming ability* $\gamma = \log(1+d)$, *binding regime if* $d \leq 4$.

Reducing gaming ability can be interpreted as a removal of least-cost agents, and the resulting impact on remaining agents reflects the externality exerted by these least-cost agents. For example, when gaming ability is low, Proposition 3 implies that the least-cost agents exert a positive externality on low-score agents but a negative one on high-score agents.

Discriminating on observables. As an application of Proposition 3, we examine the effect of discriminating on observable characteristics. Suppose that there are two groups $i = 1, 2$ and that agents within each group are homogeneous except for their score, with falsification costs $c^{\gamma_i}(t|s)$. The groups only differ in their gaming ability, with a higher gaming ability for the *in-group* (group 1) than for the *out-group* (group 2), $\gamma_1 > \gamma_2$. Assuming no allocative constraints or nonbinding ones, the group specific baseline rules under $\hat{w} = 0$ solve the designer's problem (P). We also assume low-score priority in both groups. We consider the effect of using separate rules for each group (discrimination), or a common rule for both.

Naively using the first-best rule, which allocates with certainty to eligible scores, unfairly advantages the in-group, as they falsify more easily. Similarly, any common rule that induces falsification favors the in-group over the out-group. The optimal falsification-proof common rule is fair as equal scores are treated identically. However, the in-group then imposes an externality on the out-group, as it has higher gaming ability. Discriminating between groups necessarily benefits the designer who can use more information. For the in-group, the outcome remains identical since the rule is pinned down by their gaming ability in both cases. For the out-group, all agents benefit regardless of their score if $\gamma_2 > \hat{\gamma}$. If, instead, $\gamma_2 < \hat{\gamma}$, Proposition 3 implies that discrimination leads to a steeper allocation rule, which benefits high-score agents but harms low-score agents.

4.2 Welfare consequences of falsification proofness

Imposing falsification proofness eliminates the negative effects of falsification but comes at the expense of the designer, who could achieve a better allocation by optimizing against agents' falsification behavior. We evaluate the welfare implications of falsification proofness by comparing the welfare of the designer and agents under the optimal unconstrained mechanism versus the optimal falsification-proof rule.

This comparison is feasible in the (UID) case using the characterization of the optimal unconstrained mechanism in Perez-Richet and Skreta (2022). We therefore restrict our framework to fit that of Perez-Richet and Skreta (2022) by assuming a single group of agents that are homogeneous except for scores, and no binding allocative constraints under either allocation rule. Since it is clear that we are in the (UID) case, we drop the index and denote the baseline rule as α^*.

The unconstrained optimal mechanism. We first describe the optimal mechanism. For clarity, we focus on a low-score priority setting for the discussion, but the results are general. The optimal allocation mechanism consists of a nominal allocation rule $\alpha^{**}(s)$ paired with the designer-optimal incentive compatible falsification strategy. The rule α^{**} ensures that agents with scores within its growth interval are indifferent between not falsifying and falsifying to the top of the interval.[19] Designer optimality requires agents to resolve this indifference by falsifying to the top if eligible and not falsifying otherwise. Allocative optimality thus relies on eligible agents incurring a cost from falsification.

The optimal rule has three regimes. It is in the *high-cost regime* if $c(\overline{s}|\hat{s}) > 1$. Then, the rule achieves first-best by allocating objects only to eligible agents, but lower eligible scores must falsify. The rule is in the *intermediate-cost regime* if $c(\overline{s}|\hat{s}) \leq 1$ but $c(\overline{s}|s_*(0)) > 1$. Then, all eligible agents secure an object by falsifying to the top score, while some ineligible agents must also receive objects with positive probability to deter falsification. Finally, the rule is in the *low-cost regime* if $c(\overline{s}|s_*(0)) \leq 1$. The designer then reduces the maximal allocation probability for groups with low-score priority to prevent agents below $s_*(0)$ from falsifying and obtaining objects with

[19]This is achieved by setting $\alpha^{**}(s) = c(s_+|s_-) - c(s_+|s)$ on the growth interval $[s_-, s_+]$. See the proof of Proposition 4 for a detailed description.

positive probability. Figure 3 illustrates both rules in a low-score priority setting.

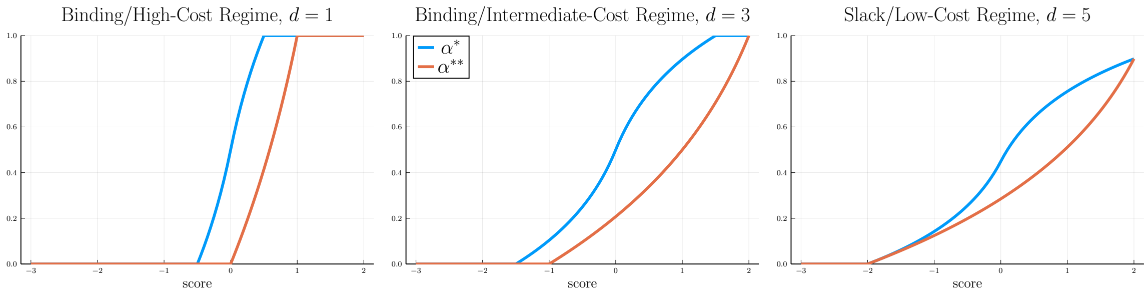

Figure 3: *Default setup, cost* $\mathcal{C}^d(x) = \frac{\log(1+x)}{\log(1+d)}$, *gaming ability* $\gamma = \log(1+d)$. *The optimal rule* α^{**} *is in red, the optimal falsification-proof rule* α^* *in blue.*

Agents' welfare. Designer optimality requires some eligible scores to falsify. The designer is, therefore, strictly better off under α^{**}. In Perez-Richet and Skreta (2022), we show that not falsifying is also a best response for agents who falsify. Therefore, the equilibrium payoff of an agent under α^{**} equals her payoff without falsification, $\alpha^{**}(s)$. Thus, comparing $\alpha^{**}(s)$ with $\alpha^*(s)$ suffices to analyze the agents' welfare. Figure 3 shows an example in which α^{**} lies below α^* everywhere, implying that agents are better off under the falsification-proof rule. This result generalizes: agents benefit from falsification proofness regardless of score.

Proposition 4 (Value of falsification proofness for agents)**.** *In the* (UID) *case, all agents are better off under the optimal falsification proof rule:* $\alpha^*(s) \geq \alpha^{**}(s)$, *with a strict inequality for a positive mass of agents.*

Agents with ineligible scores do not falsify under either rule but receive objects with higher probability under the falsification-proof rule. When α^{**} is coupled with the designer-optimal falsification strategy, eligible agents receive objects with a lower probability under α^* but avoid the falsification costs required by the optimal mechanism. Thus, the loss of allocative efficiency for the designer due to falsification-proofness could be offset by the agents' welfare gains, echoing some of the arguments for falsification-proofness presented in the introduction. We proceed to quantify the relative gains and losses of imposing falsification-proofness on the agents and the designer.

Welfare in the Euclidean case. Consider a concave Euclidean cost function $\mathcal{C}$. Let $A^*(\mathcal{C})$ and $A^{**}(\mathcal{C})$ denote the aggregate payoff of agents under α^* and α^{**}, respectively, while $D^*(\mathcal{C})$ and $D^{**}(\mathcal{C})$ denote the designer's payoff. With Euclidean costs, these payoffs take relatively simple forms whose explicit formulas we give in the proof of Proposition 5. To measure the welfare effect of falsification proofness, we compare the *gain rate of the agents*, $G(\mathcal{C}) = \frac{A^*(\mathcal{C})-A^{**}(\mathcal{C})}{A^*(\mathcal{C})}$, with the loss rate of the designer, $L(\mathcal{C}) = \frac{D^{**}(\mathcal{C})-D^*(\mathcal{C})}{D^*(\mathcal{C})}$.

We show that, controlling for gaming ability, we can find cost functions that make the trade-off arbitrarily favorable to falsification proofness. To understand how we control for gaming ability, let $d_{\mathcal{C}}$ be the maximum amount of falsification an agent would rationally choose, so $\mathcal{C}(d_{\mathcal{C}}) = 1$. This maximum amount $d_{\mathcal{C}}$ is an alternative measure of gaming ability. Hence, we control for gaming ability by holding $d_{\mathcal{C}}$ constant to some value d. The nature of the trade-off depends on whether d is above or below a threshold $\hat{d} = \bar{s} - \hat{s}$, which measures the size of the interval or eligible agents that are required to falsify, and marks the frontier between the intermediate and high-cost regimes of α^{**}.

Proposition 5. *Suppose $d \geq \hat{d}$. Then, for every $\varepsilon > 0$, there exists a Euclidean cost function $\mathcal{C}$ with $d_{\mathcal{C}} = d$, such that $L(\mathcal{C}) \leq \varepsilon$ and $G(\mathcal{C}) \geq 1/\varepsilon$. If $d < \hat{d}$, then, for every $\varepsilon > 0$, there exists a Euclidean cost function $\mathcal{C}$ with $d_{\mathcal{C}} = d$, such that $L(\mathcal{C}) \leq \varepsilon$ and $\left|G(\mathcal{C}) - \frac{F(\hat{s}+d)-F(\hat{s})}{1-F(\hat{s}+d)}\right| \leq \varepsilon$.*

To gain intuition about this result, it is useful to consider the role of returns to scale in falsification. Increasing returns to scale in falsification means making the cost function more concave, which can be done while keeping $d_{\mathcal{C}}$ constant. Suppose we are in the intermediate cost regime with low-score priority for simplicity. Making the cost function more concave makes the unconstrained optimal rule α^{**} more convex on the same growth interval and therefore decreases the agents' payoff, whereas it becomes approximately first-best for the designer. This is because the increasingly convex rule gives vanishingly less allocation probability away to dissuade ineligible agents from falsifying, while eligible agents keep falsifying to the top to obtain an object. At the same time, the optimal falsification rule α^* approaches the designer's first-best because a more concave cost increases the convexity of the rule over ineligible scores and its concavity over eligible scores, allowing the S-shaped rule to approach the first-best step function. This result follows from Proposition D.1, which shows that the S-shape becomes sharper as the cost function becomes more concave.

The condition $d < \hat{d}$ characterizes the high-cost regime for α^{**} in which the designer achieves their first-best allocation. In this case, we cannot make the agents' gain rate arbitrarily large. This is because the agents' payoff does not converge to 0 under the unconstrained optimal mechanism as we make the cost function more concave. Indeed, the unconstrained optimal rule then allocates the good for sure to all agents with a score above $\hat{s} + d < \overline{s}$, and only requires scores between $\hat{s}$ and $\hat{s} + d$ to falsify. Hence, while the payoff of ineligible agents and falsifying eligible agents still converges to 0, agents with a score above $\hat{s}+d$ receive the good with certainty without falsifying, so the aggregate payoff of agents under α^{**} converges to $1 - F(\hat{s} + d)$.

Figure 4: *Payoff trajectories. Default setup, cost* $\mathcal{C}^d(x) = \frac{\log(1+x)}{\log(1+d)}$*, gaming ability* $\gamma = \log(1 + d) : 0 \to \infty$*. The gray zone depicts the set of attainable allocative payoffs (without accounting for falsification costs).*

Quantitative illustration. We illustrate Proposition 5 using our default setup under different Euclidean cost functions. Figure 4 shows the payoff trajectories for the designer and agents under both rules when varying gaming ability $\gamma = \log(1 + d)$ for a family of Euclidean cost functions $\mathcal{C}^d(x) = \frac{\log(1+x)}{\log(1+d)}$. The gray zone denotes the set of attainable allocative payoffs (that is, without accounting for falsification costs). With infinite gaming ability, both allocation rules are to never allocate (leftmost point in both trajectories). As γ approaches 0, both allocation rules converge to the first-best (top-right point in both trajectories). At these extremities, there is no trade-off. For interior gaming ability levels, we see that the optimal rule can be very costly to agents at a moderate gain for the designer. This is especially the case at the frontier between the high and intermediary cost regimes.

We quantify the trade-off by plotting the gain rate of agents and the loss rate of the designer associated with a change from the optimal rule to the falsification-proof optimal rule as in Proposition 5, for two fixed families of cost functions, with varying gaming ability. This exercise in the default setup is shown in Figure 5. Under both cost functions, the agents' gain rate dominates the designer's loss rate for all gaming ability levels. The peak in the gain rate is reached at the frontier between the high

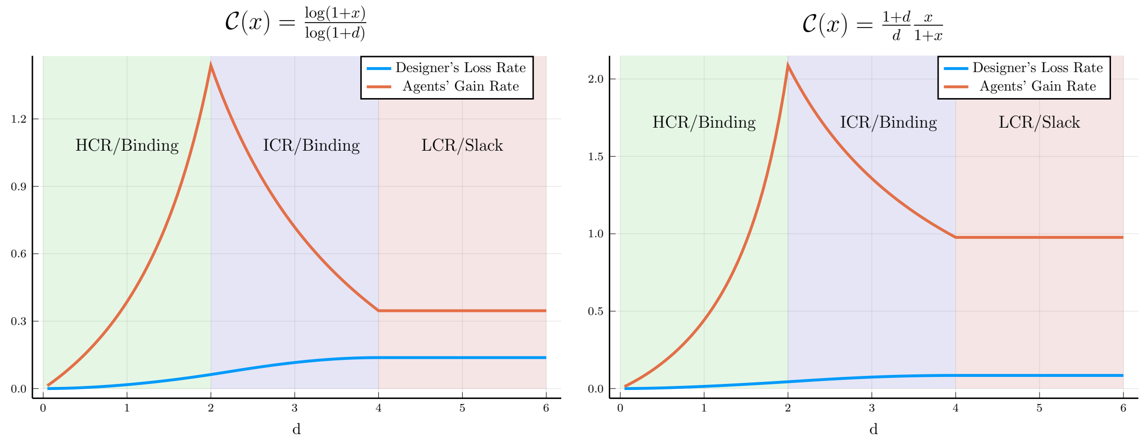

Figure 5: *Welfare impact of falsification proofness. Default setup, $d : 0 \to 6$.*

and intermediate-cost regions. Under the first cost function (left panel), the trade-off is then a 144% gain for a 6.25% loss, while under the second cost function (right panel), it is a 209% gain for a 4.4% loss. The figure also illustrates the role of returns to scale as the gain is higher and the loss lower at every d under the more concave cost function on the right panel.

Proposition 4 and Proposition 5 provide a theoretical foundation for falsification-proofness. Beyond the externality and fairness arguments we invoked as motivation, imposing falsification proofness is an effective way to balance the interests of the designer and the agents, which can, under some configurations, dramatically benefit the agents at a moderate cost in allocative efficiency to the designer. In our discussion section, Section 6, we leverage these results to highlight why falsification proof systems better align with punishment and detection tools compared to institutions relying on falsification.

5 Solution of the designer's problem

To solve the designer's problem, we decompose it into a collection of group specific *within problems*, that consist in optimally allocating a fixed mass of objects within a group, and an overall *across problem* of optimally choosing the masses of objects accruing to the different groups 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

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

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

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

where the within resource constraint (RC$_W$) must hold with equality. Since we require the whole mass ρ_i to be allocated, 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* $\boldsymbol{\rho} = (\rho_i)_{i \in I}$ 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

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

where $\boldsymbol{F} = (F_i)_{i \in I}$ and $\boldsymbol{c} = (c_i)_{i \in I}$ denote profiles.

5.1 Optimal within group allocation

We first derive an optimal allocation rule for the within problem. To simplify notation, we drop the group index i. Let $\hat{w}/\mu$ be the Lagrange multiplier on the resource constraint, where μ is the size of the group. The Lagrangian for (P$_W$) is then

$$\int_S \alpha(s)\{w(s) - \hat{w}\} dF(s) + \hat{w}\frac{\rho}{\mu}.$$

Maximizing the Lagrangian for a fixed value of the multiplier $\hat{w}/\mu$ is therefore equivalent to solving the baseline problem with outside option $\hat{w}$. To solve the within problem, we then need to identify the value of the Lagrange multiplier, i.e., of the outside option, that makes (RC$_W$) hold. To do this, we first study how the baseline allocation $\alpha^*(\cdot, \hat{w}, r)$ varies with the outside option $\hat{w}$, and the neutral gap share r.

The mechanics of indirect effects. Since adjusting outside options and neutral gap shares are the key tools to achieve particular group allocations, understanding how the baseline allocation reacts to such changes reveals the mechanics of how indirect effects operate.

Proposition 6 (Effect of outside option and neutral gap share)**.** *The baseline allocation rule $\alpha^*(s, \hat{w}, r)$ is decreasing in $\hat{w}$. It is continuous at $\hat{w}$, and independent of r unless* (Mult) *holds, in which case it is strictly increasing and continuous in r. Furthermore, it satisfies $\lim_{\hat{w}\to\bar{w}^-} \alpha^*(s, \hat{w}, r) = \alpha^*(s, \bar{w}, 1)$ and $\lim_{\hat{w}\to\bar{w}^+} \alpha^*(s, \hat{w}, r) = \alpha^*(s, \bar{w}, 0)$.*

Intuitively, a higher value of the outside option should reduce the mass of allocated objects in the baseline problem. In fact, a stronger result holds, since the effect is uniform across all scores: for every s, the baseline allocation probability is decreasing in $\hat{w}$. A higher value of the neutral gap share naturally increases the baseline allocation probability for all scores when it is effective, that is, under (Mult).

Let $A^*(\hat{w}, r) = \int_S \alpha^*(s, \hat{w}, r)dF(s)$ denote the per capita mass of objects allocated to the group under the baseline rule $\alpha^*(s, \hat{w}, r)$. Proposition 6 implies:

Corollary 1. *The per capita mass of allocated objects $A^*(\hat{w}, r)$, is strictly decreasing in $\hat{w}$. It is continuous in $\hat{w}$ and independent of r unless* (Mult) *holds, in which case $A^*(\bar{w}, r)$ is strictly increasing and continuous in r, and satisfies $\lim_{\hat{w}\to\bar{w}^-} A^*(\hat{w}, r) = A^*(\bar{w}, 1)$ and $\lim_{\hat{w}\to\bar{w}^+} A^*(\hat{w}, r) = A^*(\bar{w}, 0)$.*

Optimal within allocation. Returning to the within problem, the outside option $\hat{w}$ is equal to the Lagrange multiplier on the resource constraint, scaled by group size μ, and can be interpreted as the shadow price of marginally tightening the constraint. To ensure that (RC_W) holds, we must adjust $\hat{w}$ and r so that $A^*(\hat{w}, r) = \rho/\mu$. Theorem 2 shows that finding such values of $\hat{w}$ and r is always possible:

Theorem 2 (Optimal within group allocation)**.** *For any $0 \le \rho \le \mu$, there exists a unique outside option value $\hat{w}(\rho)$ and, under* (Mult)*, a unique neutral gap share $r(\rho)$, such that $\mu A^*\big(\hat{w}(\rho), r(\rho)\big) = \rho$. Furthermore, $\hat{w}(\rho)$ is continuous, decreasing in ρ unless* (Mult) *holds, in which case it is constant at $\bar{w}$. The function $r(\rho)$ is continuous and strictly increasing. The baseline allocation rule $\alpha^*(s, \hat{w}(\rho), r(\rho))$ is then the unique solution to the within problem* (P_W)*. The value function of* (P_W)*, $W(\rho)$ is strictly concave at ρ unless* (Mult) *holds.*

To see why Theorem 2 holds, note that, 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 baseline allocation rule allocates with certainty to all scores, regardless of scores. Similarly, for any $\hat{w}$

above the upper bound on $w(s)$, the baseline allocation rule never allocates any object. Hence, by varying $\hat{w}$ between these bounds, we can find an outside option $\hat{w}(\rho)$ such that the baseline allocation rule satisfies the resource constraint, and therefore solves (P_W). If $\hat{w}(\rho) = \bar{w}$ and $c(\overline{s}|\underline{s}) \geq 1$, 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.

5.2 Optimal across group allocation

We characterize the solution to the across problem and provide an algorithm to determine the optimal allocation profile $\boldsymbol{\rho} = (\rho_i)_{i \in I}$.

Theorem 3 (Optimal across group allocation)**.** *The across problem* (P_A) *admits a solution* $\boldsymbol{\rho}$*. Furthermore,* $\boldsymbol{\rho}$ *solves the across problem if and only if there exist 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 neutral gap share* $r_i(\rho_i)$ *such that:*

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

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

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

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

The solution $\boldsymbol{\rho}$ *is unique if, for each* i*,* $\hat{w}_i(\rho_i) \neq \bar{w}_i$ *or* $c_i(\overline{s}|\underline{s}) \geq 1$*.*

The characterization in Theorem 3 suggests the following algorithm to find a solution to the across problem. First, we compute the solutions to each within problem, 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. This may lead to hitting additional constraints. Indeed, increasing allocation to one group to satisfy its quota may lead to the violation of a previously slack resource constraint, or another group's quota constraint if the resource constraint was already binding. If so, we adjust outside options to satisfy with equality all constraints that were binding or violated at any previous step. Since, at any step, the set of constraints that have required adjustment at some previous step is bounded, and necessarily increases with each step, the process must eventually end. The allocation profile at which it ends is a solution to the across problem. In Appendix E, we provide the formal algorithm and show that it finds a solution to the across problem.

5.3 Designer welfare and comparative statics

We analyze how changes in the characteristics of the groups affect the designer's payoff. Then, we examine the impact of such changes on the optimal allocation profile and on the agents' payoffs.

Designer Welfare. We show that designer welfare is decreasing in gaming ability, and increasing with first-order stochastic dominance shifts of the score distribution. Let $\boldsymbol{c}^\gamma$ denote a profile of parameterized cost functions $c_i^{\gamma_i} = \frac{1}{\gamma_i} c_i(t|s)$.

Proposition 7 (Properties of the designer's value function)**.** *The value function of the across problem, $\overline{W}(\boldsymbol{F}, \boldsymbol{c}^\gamma)$, is nonincreasing in γ_i, and nondecreasing in F_i with respect to the first-order stochastic dominance order.*

Any increase in gaming ability tightens (FPC), thereby reducing 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 a non-obvious direct effect on the baseline allocation rule for that group (see Proposition D.2), as well as intricate indirect effects. However, an envelope theorem argument implies that we can bypass the analysis of these complex effects and instead focus on the effect of the score distribution on welfare while holding the allocation rule fixed. Even then, the effect remains difficult to analyze because the surplus function $w_i(s) - \hat{w}_i$ takes both positive and negative values on S_i. Nevertheless, an argument based on the analysis of cumulative surplus functions and the differential form of the objective function shows that improving the score distribution in the first-order stochastic dominance order increases the designer's payoff.

Comparative statics: the consequences of indirect effects. Comparative statics in the general problem can have rich effects. There can be numerous scenarios depending on initial conditions regarding gaming abilities and score distributions, the number of groups, quotas, and the mass of objects to be allocated. Instead of listing formal results that exhaust all these cases, we find it more effective to focus on an interesting example. We choose an example that illustrates how a quota can be used to shield a group from the indirect effects of transformations occurring in other groups.

Figure 6 and Figure 7 illustrate our example. We consider two equally sized groups, $\mu_1 = \mu_2 = 0.5$, and gradually lower the gaming ability of group 1 from $\gamma_1 = 2$ to $\gamma_1 = 0.5$ while keeping group 2's gaming ability fixed at $\gamma_2 = 0.8$, all corresponding to a slack regime.. Both groups are otherwise identical, with scores uniformly distributed

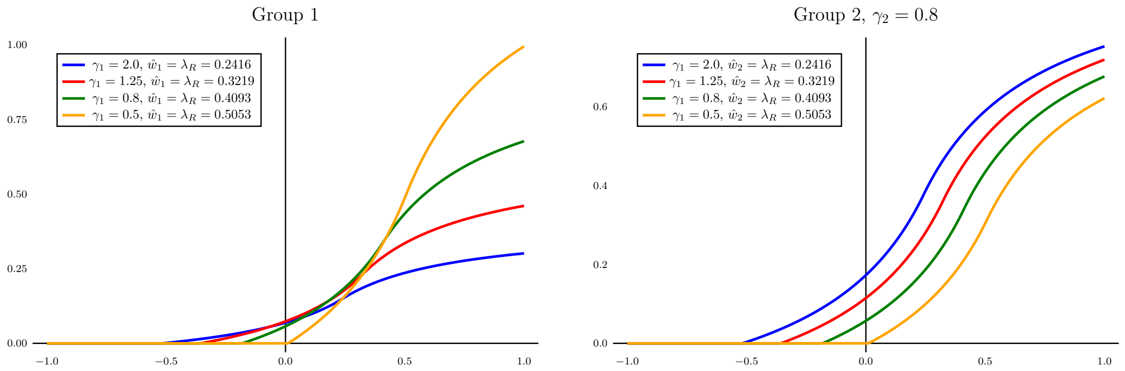

Figure 6: *Group 1's lower gaming ability hurts a low quota group ($\phi_2 = 0.2$). Cost $\mathcal{C}^{\gamma_i}(x) = (1/\gamma_i)x/(1+x)$, score distribution $U(-1,1)$, $\phi_2 = 0.2$, $\gamma_2 = 0.8$, $\rho = 0.2$,* $\mu_1 = \mu_2 = 0.5$

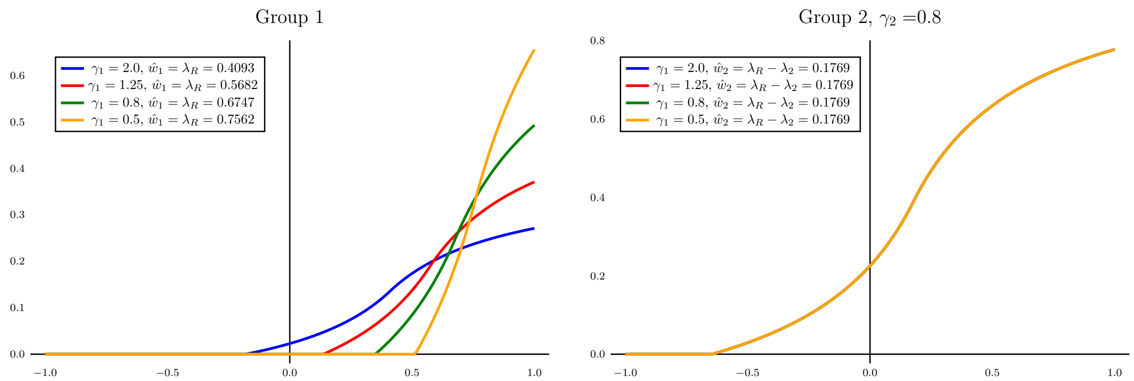

Figure 7: *Group 1's lower gaming ability leaves a high quota group ($\phi_2 = 0.8$) unaffected*

on $[-1,1]$, and a Euclidean cost function $\mathcal{C}^{\gamma}(x) = \frac{1}{\gamma}\frac{x}{(1+x)}$. The mass of objects is $\bar{\rho} = 0.2$ which causes the resource constraint to bind for all parameters considered. There is a quota ϕ_2 of objects reserved for group 2. We consider two scenarios: in the low-quota scenario, the quota $\phi_2 = 20\%$ is never binding, whereas in the high-quota scenario, the quota $\phi_2 = 80\%$ is always binding.

In each case, the decrease in gaming ability of group 1 generates a direct effect that only affects group 1, and an indirect effect that may affect both groups. For each successive decrease in γ_1, since we remain under the slack regime (see Proposition 3), the direct effect on group 1 is to increase the allocation probability for all scores, increasing the mass of objects allocated to this group.

First, consider the low-quota scenario, illustrated in Figure 6. As γ_1 decreases, the direct effect results in a higher mass of objects being allocated to group 1. However, when keeping the outside option $\hat{w}_1$ at its original value, the direct effect of each decrease in γ_1 leads to a violation of the resource constraint. To compensate for this direct effect and restore feasibility, the endogenous outside option, which is common to both groups, adjusts upward, generating the indirect effect. This indirect effect shifts group 2's original allocation rule to the right, reducing the probability of receiving an object for all scores. The final allocation rule for group 1 results from the combination of the upward rotation due to the direct effect and the rightward shift due to the indirect effect. It implies that high-score agents gain while low-score agents lose. The overall effect, however, is to increase the mass of objects allocated to group 1 since both groups bear the allocative costs of the indirect effect.

In the high-quota scenario, illustrated in Figure 7, the quota for group 2 is so high that it always binds. To satisfy this quota and ensure that the mass of objects assigned to group 2 remains at 80%, the outside option of group 2 must remain constant. The direct effect of decreasing γ_1 remains the same as in scenario 1. The direct effect of each decrease in γ_1 still leads to a violation of the resource constraint. To compensate for this direct effect while maintaining the quota for group 2, group 1 must bear the full allocative cost of the adjustment through an increase in its outside option. This implies that the overall mass of objects allocated to group 1 remains constant, while high-score agents still gain and low-score agents still lose.

6 Discussion

Robustness. The designer's problem hinges on the least-cost functions, $c_i(t|s)$. This might suggest limited robustness in our solution. We argue that, on the contrary, our solution satisfies different notions of robustness. For simplicity, we consider the single-group case. Suppose that the designer considers several possible least-cost functions, $c_j(t|s)$, with $j \in J$. A designer facing a falsification-proofness constraint and aiming for robustness in the max-min sense must optimize against the least-cost function $\underline{c}(t|s) = \min_{j\in J} c_j(t|s)$. Indeed, if the allocation rule is not immune to falsification under $\underline{c}$, it implies that it is not immune to falsification under at least one of the possible cost functions. An adversarial nature would then select one of these cost functions, leading to a $-\infty$ payoff for the designer due to constraint violation. Thus, finding a robust falsification-proof solution amounts to solving the problem with a falsification cost given by C which is precisely the problem we solved.[20]

[20]Relatedly, if the designer is uncertain about the true least-cost function, they can always adopt the allocation rule that is optimal under their worst (i.e., lowest) estimate of the cost function, $\underline{c}(t|s) < c(t|s)$. The resulting optimal falsification-proof allocation rule is necessarily immune to falsification under the true least-cost function. Moreover, as the error in the worst estimate van-

Falsification detection and punishment. One argument against the use of falsification-proof mechanisms is that falsification should be resolved outside of the mechanism, through detection and punishment, rather than internalized in the allocation rule. While this is of course desirable, a first remark is that detection and punishment rarely seem to completely eliminate falsification. Second, we note that such anti-falsification policies are equally important under falsification-proof mechanisms as under unconstrained optimal mechanisms. This can be seen on Figure 2 as reducing gaming ability eventually leads payoffs under both rules to converge toward the unconstrained efficient outcome. However, our analysis of Figure 5 also suggest that the use of unconstrained optimal mechanism (the optimum without the falsification proofness constraint) may give pervert incentives to the designer when it comes to investing in anti-falsification policies reducing gaming ability. Indeed, if the designer can reduce gaming ability at a small cost, they will never invest to reduce gaming ability further than the lowest gaming ability level that allows them to implement their first-best outcome. As it turns out, this is exactly the gaming ability level that makes the trade-off with the optimal falsification-proof rule most extreme. Under the falsification-proof rule, in contrast, decreasing gaming ability never stops increasing the designer's payoff.

More commitment: general mechanisms. By restricting the designer to score-based allocation rules that condition the probability of getting an object on observables only (the score produced by the agent and group characteristics), we seemingly restrict the designer's commitment power. The revelation principle in Myerson (1982) implies that a designer with full commitment power chooses among direct, truthful and obedient recommendation mechanisms that include a communication rule and an allocation rule. The communication rule receives (truthful) reports from the agents about their type θ and sends back (obedient) falsification recommendations. The allocation rule conditions on reports, recommended scores and observed scores.

We can think of score-based allocation rules as the analog of tariffs in standard screening model. Tariffs only condition on the quantity purchased, and not on reports. Despite that, the *taxation principle* implies that tariffs can replicate any truthful direct mechanism with deterministic allocation, and the communication phase of the direct mechanism can therefore be bypassed. We show in Proposition 2 and Corol-

ishes, the resulting allocation rule converges to the optimal allocation rule under the true least-cost function. More precisely, if $\underline{c}_n < c$ converges uniformly to c, then the optimal falsification-proof allocation rule under $\underline{c}_n$ converges to the optimal falsification-proof rule under the true least-cost function.

lary 1 of Perez-Richet and Skreta (2023), that as long as score recommendations are deterministic, a similar result applies to our framework, implying that a score-based allocation rule is without loss of generality.

The restriction to score-based allocation rules entails a further simplification, whereby an agent's allocation probability is based solely on their score, rather than the score profile of other agents, effectively treating the continuum of agents as a single agent. This is again without loss, as we explain in Appendix F.

Less commitment: 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 that they 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 worth is given by the increasing function $\tilde{w}_i(s)$ instead of $w_i(s)$. Without loss of generality, we normalize their outside option to 0. We also assume they face the same allocative constraints as the designer in our initial model. They can observe the group label, but has no access to the score.

In this version of the model, the designer acts as a certification intermediary who can design an information structure but lacks control over allocation decisions. With reduced commitment power, the designer is *a priori* worse off. However, Proposition 8 demonstrates that they can replicate full commitment by employing a binary-signal information structure, which recommends allocating with probability $\alpha_i^*(s)$ and rejecting with probability $1 - \alpha_i^*(s)$. Given this straightforward information structure, the decision-maker finds it optimal to follow the recommendation.

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

The proof of Proposition 8 shows that α^* is obedient whenever the designer and decision maker preferences are sufficiently well aligned. Then, our optimal allocation rule also solves the information design problem of the designer as a certification intermediary.

7 Related literature

We contribute to the literature on optimal allocation mechanisms with 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 in order to maximize revenue. However, monetary transfers may lose their effectiveness if the designer has a more general objective, as in Condorelli (2013), or wishes to maximize a combination of weighted utilitarian and revenue objectives, as demonstrated in Akbarpour, Dworczak, and Kominers (2024). Both studies establish conditions under which the designer optimally refrains from using transfers entirely. 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. (2024) show that it is optimal to do so if the designer's utility from allocation is negatively correlated with willingness to pay. By contrast to these works, we consider a setting where agents' private information can be manipulated at a cost. Methodologically, our approach differs from Myersonian techniques that rely on virtual surplus and work directly with the allocation rule. Instead, we use cumulative surplus and work with the derivative of the allocation rule to identify the growth interval.

We add to a sizable literature on non-market optimal allocation mechanisms which study the use of alternative targeting tools in lieu of transfers. In our setting, there are no transfers and targeting is enabled by the availability of the (possibly falsified) score. 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 (2023), 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) combine verification and money burning. In Kattwinkel (2019), the designer has access to a private signal correlated with the agent's private information, while, in Kattwinkel and Knoepfle (2023), they can additionally verify the agent's type. In contrast, we consider costly state falsification and impose falsification proofness. This is similar to money burning in that it is wasteful but differs in that the utility lost through falsification is type-dependent.

We also contribute to the literature on costly screening. Frankel and Kartik (2021) and Ball (2025) study the optimal design of linear scores under a gaming technology that amounts to costly 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. Tan (2023) consider a price discrimination setting in which consumers can distort their data at a cost to avoid high prices. Li and Qiu (2023) study costly screening in a multi-agent setting without transfers and identify conditions under which contests are optimal, as well as situations where random mechanisms dominate contests. Lacker and Weinberg (1989) investigate the design of risk-sharing contracts with costly state falsification, focusing on optimal falsification-proof contracts. They are the first to show that this constraint may lead to a loss of optimality without characterizing the optimal contract. We build on Perez-Richet and Skreta (2022), where we show that optimal mechanisms necessarily induce falsification.

In practice, a score-based allocation rule can incentivize both genuine 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 related framework, Augias and Perez-Richet (2023) study the optimal design of allocation mechanisms when agents can improve their score. In this paper, in contrast, manipulations are purely socially wasteful.

We also add to the growing list of papers using optimal transport theory in economics surveyed in Carlier (2012) and Galichon (2018). More recently, optimal transport theory has 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, 2025; Lin and Liu, 2024; Malamud and Schrimpf, 2021), and labor market sorting problems (Boerma, Tsyvinski, and Zimin, 2021). Most of these papers rely on generalizations of duality characterizations in transport theory.[21] In contrast, our approach relies on the dual of the optimal transport problem which we show to be equivalent to our baseline problem in the (UID) case. This requires constructing the marginal distributions, which we obtain by reinterpreting the cumulative surplus generated by allocation rules as probability measures over the spaces of eligible and ineligible scores. Thus, our approach offers an original solution method that differs from Myersonian techniques in settings with transfers, and from Lagrangian techniques in settings without transfers as in Amador, Werning, and Angeletos (2006) and Amador and Bagwell (2022).

[21] For example, Lin and Liu (2024) rely on properties characterizing optimal coupling for given marginals to establish their characterization of stable credible signals.

References

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 (2024): "Redistributive Allocation Mechanisms," *Journal of Political Economy*, 132, 1831–1875.

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. (2025): "Scoring Strategic Agents," *American Economic Journal: Microeconomics*, 17, 97–129.

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.

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 (2023): "Optimal multi-unit allocation with costly verification," *Social Choice and Welfare*, 61, 455–488.

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.

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

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 (2019): "Muddled information," *Journal of Political Economy*, 127, 1739–1776.

——— (2021): "Improving information from manipulable data," *Journal of the European Economic Association*, 1542–4766.

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 fortieth annual ACM symposium on Theory of computing - STOC08*, 1–23.

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

Kattwinkel, D. and J. Knoepfle (2023): "Costless Information and Costly Verification: A Case for Transparency," *Journal of Political Economy*, 131, 504–548.

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 (2025): "Persuasion and Matching: Optimal Productive Transport," .

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.

Li, Y. and X. Qiu (2023): "Screening Signal-Manipulating Agents via Contests," Working paper.

Lin, X. and C. Liu (2024): "Credible Persuasion," *Journal of Political Economy*, 132, 2228–2273.

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.

Milli, S., J. Miller, A. D. Dragan, and M. Hardt (2019): "The social cost of strategic classification," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 230–239.

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.

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

——— (2023): "Score-based mechanisms," *Working paper*.

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," Working paper.

Tan, T. Y. (2023): "Price Discrimination with Manipulable Observables," *Available at SSRN 4480623*.

A Simplifying the baseline problem

Smoothness and monotonicity of optimal allocation rules. By (FPC), feasible allocation rules inherit the *regularity* of the cost function, implying that they are Lipschitz continuous. Since expected worth is monotonic in score, it is natural that optimal allocation rules are monotonic.

Lemma A.1 (Smoothness and monotonicity)**.** *If an allocation rule satisfies* (FPC)*, it is Lipschitz continuous. Furthermore, if α is feasible for* (BP) *but not monotonic, there exists a nondecreasing allocation rule $\tilde{\alpha}$ that is feasible and strictly better for* (BP)*.*

Proof. If α satisfies (FPC), the regularity assumption of Definition 1 directly implies Lipschitz continuity. Next, define the nondecreasing function

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

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 $[\hat{s}, \overline{s}]$.

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

$$\tilde{\alpha}(t) - \tilde{\alpha}(s) = \tilde{\alpha}(t') - \tilde{\alpha}(s') = \alpha(t') - \alpha(s') \leq c(t'|s') \leq c(t|s),$$

where the first equality is by definition of s' and t', and 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.

Eligible scores are more likely, and ineligible score less likely to get an object under $\tilde{\alpha}$ than under α. Hence, $\tilde{\alpha}$ is better than α for (BP). 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. □

By the fundamental theorem of calculus, we can then rewrite allocation rules 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 $\overline{\alpha} = \alpha(\overline{s})$. The derivative α' exists almost everywhere and is bounded between 0 and Λ. We can therefore rewrite the baseline problem as an optimization problem over the bounded function $\alpha'(\cdot)$ and either of the scalars $\overline{\alpha}$ or $\underline{\alpha}$, instead of optimizing directly on α. We call the resulting problem a *differential program*.

Furthermore, downward falsification-proofness constraints are satisfied by monotonicity, so we only retain upward constraints. Since allocation rules are bounded, monotonicity also implies the existence of a growth interval, to the left of which the allocation probability is null, and to the right of which it is equal to 1. To solve the baseline problem, we first need to characterize possible growth intervals. The differential program allows us to do that.

When using ($\underline{\text{ID}}$), and constructing the allocation rule from the left, locally increasing the allocation probability at z by $\alpha'(z)dz$ yields a marginal gain equal to the upward cumulative surplus $\mathcal{W}^{+}(z, \hat{w})$. Indeed, replacing α with ($\underline{\text{ID}}$) in the designer's objective, and integrating by parts, yields the differential objective function[22]

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

When using ($\overline{\text{ID}}$), locally decreasing the allocation probability at z by $\alpha'(z)dz$ yields a marginal gain equal to the downward cumulative surplus $\mathcal{W}^{-}(z, \hat{w})$. Rewriting the designer's objective function with ($\overline{\text{ID}}$) yields

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

($\overline{\text{DOF}}$) implies that it is optimal to set $\overline{\alpha} = 1$ in high-score priority groups, and ($\underline{\text{DOF}}$) implies that it is optimal to set $\underline{\alpha} = 0$ in low-score priority groups. Under neutral priority, either choice works.

[22]This reformulation is mathematically analogous to obtaining the seller's revenue in terms of the product of the allocation rule and buyer's virtual valuations in standard mechanism design settings with transfers.

We can then write a common differential program for all priorities using the cumulative surplus function $\mathcal{W}$. Note that we add a differential monotonicity constraint (DMC) to the program. Although not necessary, it helps the exposition and is without loss of generality by Lemma A.1.

Lemma A.2 (Differential program)**.** *In* (BP)*, it is optimal to set* $\underline{\alpha} = 0$ *if the group has low-score priority, and* $\overline{\alpha} = 1$ *if the group has high-score priority. Under neutral priority, either choice works. In all cases, the baseline program* (BP) *is equivalent to the differential program*

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

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

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

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

Proof. 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 (DFPC) is immediate. We can add the differential monotonicity constraint (DMC) to the program without loss of generality by Lemma A.1. Given (DMC) the probability constraint can be written equivalently as:

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

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 $(\underline{\alpha}, \alpha')$, setting $\underline{\alpha} = 0$ in the low-score 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 case. $\square$

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

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

By Lemma 1, (iii), maximizing this Lagrangian under (DMC) implies setting $\alpha'(s) = 0$ for almost every s outside of the growth interval $\big[s_*(\nu), s^*(\nu)\big]$. Growth intervals are therefore (ZAS) intervals and are contained within the maximal interval $[s_*(0), s^*(0)]$.

Lemma A.3 (Lagrange necessary and sufficiency result for (BP))**.** *A nondecreasing and Lipschitz allocation rule $\hat{\alpha}$ solves (BP) 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) $\hat{\alpha}(t) - \hat{\alpha}(s) \leq c(t|s)$ *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),

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

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 that an allocation rule α is *feasible* if it is nondecreasing, Λ-Lipschitz, satisfies (FPC), $\underline{\alpha} = 0$ under low-score priority, and $\overline{\alpha} = 1$ under high-score 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, $h(1) = \Omega(\hat{\alpha})$. If $g(\hat{\alpha}) = 1$, 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-score priority and $\overline{\hat{\alpha}} = 1$ under high-score 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

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

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 we use the notation $[z]_0^1 = z\,\mathbb{1}_{0 \leq z \leq 1} + \mathbb{1}_{z > 1}$, for any z, and

$$a = -\alpha\big(s_*(\nu)\big)\,\mathbb{1}_{\bar{w} < \hat{w}} + \big(1 - \alpha\big(s^*(\nu)\big)\big)\,\mathbb{1}_{\bar{w} \geq \hat{w}}\,.$$

Then $\tilde{\alpha}$ satisfies (i) and (ii), and (b) follows from:

$$\begin{aligned}
\mathcal{L}(\hat{\alpha},\nu) &= \int_S \hat{\alpha}'(z)\mathcal{W}(z,\hat{w})dz && \text{(by (i), and (a))}\\
&= \nu\big(\overline{\hat{\alpha}}-\underline{\hat{\alpha}}\big)+\int_{s_*(\nu)}^{s^*(\nu)}\{w(s)-\hat{w}\}\hat{\alpha}(s)dF(s) \\
& && \text{(by integration by parts and Lemma 1)}\\
&= \nu+\int_{s_*(\nu)}^{s^*(\nu)}\{w(s)-\hat{w}\}\hat{\alpha}(s)dF(s) && \text{(by (a))}\\
&\geq \nu+\int_{s_*(\nu)}^{s^*(\nu)}\{w(s)-\hat{w}\}\tilde{\alpha}(s)dF(s) && \text{(by (v))}\\
&= \nu+\int_{s_*(\nu)}^{s^*(\nu)}\{w(s)-\hat{w}\}\alpha(s)dF(s)+a\underbrace{\int_{s_*(\nu)}^{s^*(\nu)}\{w(s)-\hat{w}\}dF(s)}_{=0} && \text{(by (ZAS))}\\
&= \nu+\int_{s_*(\nu)}^{s^*(\nu)}\alpha'(z)\mathcal{W}(z,\hat{w})dz-\nu\big\{\alpha\big(s^*(\nu)\big)-\alpha\big(s_*(\nu)\big)\big\} \\
& && \text{(by integration by parts)}\\
&= \nu+\int_{s_*(\nu)}^{s^*(\nu)}\alpha'(z)\big[\mathcal{W}(z,\hat{w})-\nu\big]dz \\
&\geq \nu+\int_S \alpha'(z)\big[\mathcal{W}(z,\hat{w})-\nu\big]dz=\mathcal{L}(\alpha,\nu), \\
& && \text{(as } \mathcal{W}(z,\hat{w})<\nu \text{ for } z\notin[s_*(\nu),s^*(\nu)])
\end{aligned}$$

where we use the relation $\mathcal{W}\big(s_*(\nu),\hat{w}\big)=\mathcal{W}\big(s^*(\nu),\hat{w}\big)=\nu$ from Lemma 1.

$\boxed{\Leftarrow}$ Feasibility directly implies (ii) as $\hat{\alpha}$ must satisfy (FPC). By Lemma 1, 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

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

□

B Baseline rule: upward decreasing differences

Under (UDD), the falsification-proofness constraint can be replaced by the condition that not falsifying solves the first-order condition of the agent's problem $\max_t \ \alpha(t) - c(t|s)$, that is

$$\alpha'(s) \le c_{t^+}(s|s),$$

where $c_{t^+}(s|s)$ is the right-derivative of c with respect to target t evaluated at $t = s$. Integrating by parts the objective function of the simplified program, we rewrite it in the following differential form:[23]

$$\max_{\alpha'(s)\le c_{t^+}(s|s)} \int_{s_*}^{s^*} \big(\mathcal{W}(s,\hat{w})-\nu\big)\alpha'(s)ds,$$

[23] See Appendix A for details.

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 $\alpha'(s) = c_{t^+}(s|s)$ for almost every s.

Following step 2 of Procedure 1, the growth interval $[s_*, s^*]$ of the baseline rule is pinned down by the *boundary condition*

$$s_* = \min \left\{ s \in [s_*(0), \hat{s}] : \int_s^{m(s)} c_{t^+}(x|x)dx \leq 1 \right\}. \tag{B}$$

As in the (UID) case, we can define the probability gap in the probability constraint as

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

Given the boundary condition (B), the probability constraint is slack, and the gap positive $\Gamma_{udd} > 0$, if and only if the *slackness condition* $\int_{s_*(0)}^{s^*(0)} c_{t^+}(x|x)dx < 1$ is satisfied. When it holds, there is a gap $\Gamma_{udd} > 0$ between the total growth of the baseline rule and its upper bound (equal to 1).

As in the (UID) case, it is optimally withheld in the low-score priority case and optimally allocated to agents in the high-score priority case. The index $I(\hat{w}, r)$ performs this task.

In the (UDD), multiplicity arises under the condition:

$$\hat{w} = \bar{w} \text{ and } \int_{\underline{s}}^{\bar{s}} c_{t^+}(x|x)dx < 1. \tag{Mult}$$

Note that, while we use the same labels as in the (UID) case for the gap equation (Gap), and the multiplicity condition (Mult), they have a different definition under (UDD).

Theorem B.1 (Baseline allocation rule under (UDD))**.** *If the cost function satisfies* (UID)*, then α^*_{udd} solves* (BP)*. It is independent of r and the unique baseline rule unless* (Mult) *holds. Otherwise, the set of baseline rules is $\{\alpha^*_{udd}(\cdot, \bar{w}, r)\}_{r\in[0,1]}$.*

Proof. 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 A.3 are satisfied. Picking $\nu = \mathcal{W}(s_*, \hat{w})$, it is clear that (i) holds. For (ii), let

$s_* \leq s < t \leq s^*$, then

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

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 $\int_{\underline{s}}^{\overline{s}} c_{t^+}(x|x)dx \geq 1$ or $\hat{w} \neq \bar{w}$. In the remaining case, if (Mult) holds, 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. □

Note that the baseline 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-score priority, and to never allocate under low-score priority. This is, for example, the case with the quadratic cost function $c(t|s) = (t - s)^2$.

C Proofs

Proof of Lemma 1. 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 $\hat{s}$, proving (i). 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 (ii), (iii) and (iv) are direct consequences of single-peakedness and continuity. For (v), note that, in the low-score priority, $\mathcal{W}(\underline{s}, \hat{w}) = \mathcal{W}^+(\underline{s}, \hat{w}) = 0$, while, in the high-score priority case, $\mathcal{W}(\overline{s}, \hat{w}) = \mathcal{W}^-(\overline{s}, \hat{w}) = 0$. □

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 A.3 are satisfied. We pick the multiplier $\nu = \mathcal{W}(s_*, \hat{w}) \geq 0$. Then α^*_{uid} clearly satisfies (i). To see that it satisfies (iii) and (iv), note that $\alpha^*_{uid}(s^*) - \alpha^*_{uid}(s_*) = c(s^*|s_*)$ since the falsification-proofness constraint is binding for (s_*, s^*). Hence, by (B), either $\alpha^*_{uid}(s^*) - \alpha^*_{uid}(s_*)$ is equal to 1 and the probability constraint is binding, or it is strictly less than 1, and then $\nu = 0$ and $(s_*, s^*) = (s_*(0), s^*(0))$.

By the optimal transport connection established in Section 3.2, α^*_{uid} solves the relaxed program of that section. To show that it satisfies (v), we need to show that 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, for example, the first case. Then

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

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), that is a single value of the Lagrange multiplier such that the necessary and sufficient conditions of Lemma A.3 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 $c(s^*|s_*) = 1$, or by the requirement that $\alpha^*_{uid}(s_*) = 0$ under low-score priority, and $\alpha^*_{uid}(s^*) = 1$ under high-score priority. The only situation 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 $c(\overline{s}|\underline{s}) < 1$. The designer is then indifferent across all allocation rules $\alpha^*_{uid}(s, \bar{w}, r)$ 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 Proposition 1. Let α denote the baseline rule. Using the formulas from Theorem 1 and Theorem B.1, we get $\alpha'(s) = \mathcal{C}'(0)$ in the (UDD) case. In the (UID) case, for $s \in [s_*, \hat{s}]$, $\alpha'(s) = \mathcal{C}'\big(m(s) - s\big)$ increases in s by concavity of $\mathcal{C}$ and since $m(s)$ is decreasing. If instead $s \in [\hat{s}, s^*]$, then $\alpha'(s) = \mathcal{C}'\big(s - m^{-1}(s)\big)$ decreases in s by concavity of $\mathcal{C}$ and since $m^{-1}(s)$ is decreasing. □

Proof of Proposition 2. Let α denote the baseline rule. For the shifted linear cost family, we have $c_t(t|s) = \kappa(s)$, $c_s(t|s) = -\kappa(s) + (t-s)\kappa'(s)$ and $c_{ts}(t|s) = \kappa'(s)$. If $\kappa'(s) \leq 0$, we are in the (UDD) case. On the growth interval, $\alpha'(s) = c_t(s|s) = \kappa(s)$ and $\alpha''(s) = \kappa'(s) \leq 0$ so the baseline rule is concave along the growth interval.

If $\kappa'(s) \geq 0$, we are in the (UID) case. For ineligible scores on the growth interval,

$$\alpha'(s) = -c_s(m(s)|s) = \kappa(s) - (m(s) - s)\kappa'(s),$$

and

$$\alpha''(s) = (2 - m'(s))\kappa'(s) - (m(s) - s)\kappa''(s).$$

Note that, for the linear shifted cost to satisfy our basic assumptions on costs, we must have $c_s(t|s) = -\kappa(s) + (t-s)\kappa'(s) < 0$ for every t, that is, $\frac{\kappa'(s)}{\kappa(s)} \leq \frac{1}{\bar{s}-s}$. Since $m'(s) < 0$, and $\kappa'(s) \geq 0$, we have $\alpha''(s) \geq 2\kappa'(s) - (m(s) - s)\kappa''(s)$. Then $\alpha''(s) \geq 0$, whenever

$$\frac{\kappa''(s)}{\kappa'(s)} \leq \frac{2}{\bar{s} - s}.$$

For eligible scores, $\alpha'(s) = c_t\big(s|m^{-1}(s)\big) = \kappa(m^{-1}(s))$ and

$$\alpha''(s) = \frac{1}{m'\big(m^{-1}(s)\big)}\kappa'(m^{-1}(s)) \leq 0$$

since m is decreasing. □

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

Consider first $\gamma < \gamma' < \hat{\gamma}$, and define the function $\delta(s) = \alpha^*_{\gamma'}(s) - \alpha^*_{\gamma}(s)$. 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 $\alpha^*_{\gamma'}(s)$, and therefore increasing and positive, on $\big[s_*[\gamma'], s_*[\gamma]\big]$. It is equal to $\alpha^*_{\gamma'}(s) - 1$, and therefore increasing and negative, on $\big[s^*[\gamma], s^*[\gamma']\big]$.

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

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

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

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

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

Next, suppose $\gamma' > \gamma > \hat{\gamma}$. In the low-score priority case, 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 that δ is decreasing on $[s_*(0), \overline{s}]$. Since $\delta\big(s_*(0)\big) = 0$, this proves point (ii).

In the high-score priority case, 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 that δ is decreasing on $[\underline{s}, s^*(0)]$. Since $\delta\big(s^*(0)\big) = 0$, this proves point (iii). □

Proof of Proposition 4. In Perez-Richet and Skreta (2022), we derive optimal allocation rules without a falsification proofness constraint for a cost function that satisfies the following *upper triangular inequality*: for all $s \leq m \leq t$, $c(t|m) + c(m|s) \geq c(t|s)$. We also show in Perez-Richet and Skreta (2022) that this triangular inequality is implied by (UID). Therefore, in the (UID) case, the (unconstrained) optimal allocation rule is

$$\alpha^{**}(s, \hat{w}, r) = \begin{cases} \Gamma^{**} I(\hat{w}, r) & \text{if } s < s_- \\ \Gamma^{**} I(\hat{w}, r) + c(s_+|s_-) - c(s_+|s) & \text{if } s \in [s_-, s_+] \\ \Gamma^{**} I(\hat{w}, r) + c(s_+|s_-) & \text{if } s > s_+ \end{cases},$$

where $I(\hat{w}, r)$ is defined as in Appendix B, and the probability gap is

$$\Gamma^{**} = 1 - c(s_+|s_-).$$

The magnitude of falsification costs, together with score priority, determines the growth interval $[s_-, s_+]$. The relevant cost thresholds are defined as $\hat{c} = c(\overline{s}|\hat{s})$ and $\overline{c} = c\big(\overline{s}|s_*(0)\big)$. By cost monotonicity, $\overline{c} > \hat{c}$.

The unconstrained optimal rule has three regimes:

- The *low-cost regime*, if $\overline{c} < 1$: then, the growth interval is $[s_-, s_+] = [s_*(0), \overline{s}]$, the probability constraint is slack, and the gap $\Gamma^{**} > 0$ is allocated according to priority by $I(\hat{w}, r)$.
- The *intermediate-cost regime*, if $\overline{c} \geq 1 > \hat{c}$: then, $s^+ = \overline{s}$ and s_- solves $c(\overline{s}|s_-) = 1$. The probability constraint is binding.
- The *high-cost regime*, if $\hat{c} \geq 1$: then, $s_- = \hat{s}$ and s^+ solves $c(s_+|\hat{s}) = 1$. The probability constraint is binding.

First, note that, under the low and intermediate-cost regimes, for every eligible score s that lies in the growth interval of both rules, α^* and α^{**},

$$\frac{d\alpha^{**}}{ds}(s) = -c_s\big(\overline{s}|s\big) \leq -c_s\big(m(s)|s\big) = \frac{d\alpha^*}{ds}(s), \tag{1}$$

since, by (UID), $c_s(t|s)$ is nondecreasing in t.

We first treat the *low-score priority* and *neutral priority* cases together:

Low-cost regime. Both rules have the same growth interval $[s_*(0), s^*(0)]$ with $s^*(0) = \overline{s}$. Furthermore, $\alpha^*(s) = \alpha^{**}(s) = I(\hat{w}, r)\Gamma^{**} = I(\hat{w}, r)\Gamma$ for $s \leq s_*(0)$. Then, by (1), $\alpha^*(s) > \alpha^{**}(s)$ for all $s \in (s_*(0), \hat{s})$. For $s \geq \hat{s}$,

$$\begin{aligned}
\alpha^*(s) &= \alpha^*\big(m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big) \\
&\geq \alpha^{**}\big(m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big) \\
&= I(\hat{w}, r)\Gamma^{**} + c\big(\overline{s}|s_*(0)\big) - c\big(\overline{s}|m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big) \\
&\geq I(\hat{w}, r)\Gamma^{**} + c\big(\overline{s}|s_*(0)\big) - c\big(\overline{s}|s\big) = \alpha^{**}(s),
\end{aligned}$$

where the first equality holds because (FPC) binds between $m^{-1}(s)$ and s, the first inequality holds since $m^{-1}(s)$ is ineligible, and the second inequality is due to the *upper triangular inequality.*

Intermediate-cost regime. The growth intervals satisfy $s_* \leq s_- \leq \hat{s} < s^* \leq s^+ = \overline{s}$, and $\alpha^*(s_*) = 0$, while $\alpha^{**}(s_-) = 0$. For all $s \in (s_-, \hat{s})$, (1) holds, implying $\alpha^*(s) \geq$

$\alpha^{**}(s)$. For $s \in (\hat{s}, s^*)$, using the same ideas as above,

$$\begin{aligned}\alpha^*(s) &= \alpha^*\big(m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big)\\ &\geq \alpha^{**}\big(m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big)\\ &\geq 1 - c\big(\overline{s}|m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big)\\ &\geq 1 - c\big(\overline{s}|s\big) = \alpha^{**}(s).\end{aligned}$$

Finally, for $s \geq s^*$, $\alpha^*(s) = 1 \geq \alpha^{**}(s)$.

High-cost regime. The growth intervals satisfy $s_* < \hat{s} = s_- < s^* < s_+ \leq \overline{s}$. Therefore, $\alpha^*(s) > \alpha^{**}(s) = 0$ for $s \in [s_*, \hat{s}]$. Then, for $s \geq \hat{s}$,

$$\begin{aligned}\alpha^*(s) &= \alpha^*\big(m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big)\\ &\geq c\big(s|m^{-1}(s)\big) \geq c\big(s|\hat{s}\big)\\ &\geq c\big(s_+|\hat{s}\big) - c\big(s_+|s\big)\\ &= 1 - c\big(s_+|s\big) = \alpha^{**}(s),\end{aligned}$$

where the second inequality is by cost monotonicity, the third by the triangular inequality, and the remaining equality is by definition of the growth intervals in the hight-cost regime.

Next, we treat the *high-score priority* case.

Low-cost regime. The growth interval is $[s_*(0), s^*(0)] = [\underline{s}, s^*(0)]$ for α^*, and $[\underline{s}, \overline{s}]$ for α^{**}. Furthermore,

$$\alpha^*(\underline{s}) = 1 - c\big(s^*(0)|\underline{s}\big) > 1 - c(\overline{s}|\underline{s}) = \alpha^{**}(\underline{s}).$$

Hence, $\alpha^*(s) > \alpha^{**}(s)$ for all $s \in [\underline{s}, \hat{s}]$ holds by (1). For $s \in \big(\hat{s}, s^*(0)\big)$, using the same ideas as above

$$\begin{aligned}\alpha^*(s) &= \alpha^*\big(m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big)\\ &\geq \alpha^{**}\big(m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big)\\ &\geq 1 - c\big(\overline{s}|m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big)\\ &\geq 1 - c\big(\overline{s}|s\big) = \alpha^{**}(s).\end{aligned}$$

For $s \geq s^*(0)$, $\alpha^*(s) = 1 \geq \alpha^{**}(s)$.

Intermediate-cost regime. There are two possible cases. If $c(s^*(0)|s_*(0)) > 1$, the argument is word for word as in the low-score and neutral priority case. Suppose that $c(s^*(0)|s_*(0)) < 1$. Then the growth interval of α^* is $[s_*(0), s^*(0)]$ with $s_*(0) = \underline{s}$, $s^*(0) \leq \overline{s}$, and $\alpha^*\big(\underline{s}\big) = \Gamma$. The growth interval of α^{**} is such that $s_- > \underline{s}$, and $s^+ = \overline{s}$. Therefore, $\alpha^*(s) > \alpha^{**}(s)$ for all $s \leq s_-$. For $s \in (s_-, \hat{s})$, (1) holds, therefore $\alpha^*(s) > \alpha^{**}(s)$. For $s \in \big(\hat{s}, s^*(0)\big)$, using the same ideas as above

$$\begin{aligned}\alpha^*(s) &= \alpha^*\big(m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big)\\ &\geq \alpha^{**}\big(m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big)\\ &\geq 1 - c\big(\overline{s}|m^{-1}(s)\big) + c\big(s|m^{-1}(s)\big)\\ &\geq 1 - c\big(\overline{s}|s\big) = \alpha^{**}(s).\end{aligned}$$

Finally, for $s \geq s^*(0)$, $\alpha^*(s) = 1 \geq \alpha^{**}(s)$.

High-cost regime. Then, the proof is exactly as under low-score or neutral priority.

□

Proof of Proposition 5. We start by establishing general payoff formulas for a concave Euclidean cost function $\mathcal{C}$ with $d_{\mathcal{C}} = d$. Let $\bar{d} = \overline{s} - \underline{s}$, and $d^*(0) = s^*(0) - s_*(0)$. The length of the growth interval of α^* is equal to $\ell^*_{\mathcal{C}} = \min\{d, d^*(0)\}$. For α^{**}, it is

$$\ell^{**}_{\mathcal{C}} = \ell^*_{\mathcal{C}}\, \mathbb{1}_{\bar{w} \leq \hat{w}} + \min\{d, \bar{d}\}\, \mathbb{1}_{\bar{w} > \hat{w}}\,.$$

In the low and intermediate-cost regimes, all eligible scores falsify, while in the high-cost regime, the scores that falsify are those on the growth interval, which is then a subset of eligible scores. Therefore, we can write the size of the interval of falsifying scores as $f_{\mathcal{C}} = \min\{\ell^{**}_{\mathcal{C}}, \hat{d}\} = \min\{\ell^*_{\mathcal{C}}, \hat{d}\}$, where recall that $\hat{d} = \overline{s} - \hat{s}$ is the size of the interval or eligible agents. Finally, let $p^*_{\mathcal{C}} = \mathcal{C}(\ell^*_{\mathcal{C}})$ and $p^{**}_{\mathcal{C}} = \mathcal{C}(\ell^{**}_{\mathcal{C}})$.

The rule α^{**} is in the low-cost regime if $d \geq \bar{d}\, \mathbb{1}_{\bar{w} > \hat{w}} + d^*(0)\, \mathbb{1}_{\bar{w} \leq \hat{w}}$, in the high-cost regime if $d < \hat{d}$, and in the intermediate-cost regime otherwise.

The rule α^* is in its binding regime if $d \leq d^*(0)$ and in its slack regime otherwise. Note that the slack regime corresponds to the low-cost regime of α^* under low or neutral-score priority, whereas the transition between the slack and binding regimes of α^* is inside the intermediate-cost regime of α^{**} under high-score priority.

We start by defining three adjunct functions. We let $\phi(x) = -m^{-1}(x)$ for $x \in [\hat{s}, s^*(0)]$. Next, ψ denotes the inverse of the strictly increasing function $x \to x + \phi(x)$. Intuitively, to a certain length of a (ZAS) interval, it associates its upper bound.

Finally, for $x \in [\hat{s}, s^*(0)]$,

$$\xi(x) = \frac{F(x) + \phi'(x)F\big(-\phi(x)\big)}{1 + \phi'(x)}.$$

Note that, by the implicit function theorem and the (ZAS) equation, ϕ' exists almost everywhere and satisfies

$$\phi'(x) = -\frac{\big(w(x) - \hat{w}\big)f(x)}{\big(w\big(-\phi(x)\big) - \hat{w}\big)f\big(-\phi(x)\big)}.$$

We first compute the aggregate payoff of agents under α^*. To do this, we start by computing the aggregate payoff of agents with a score in the growth interval.

$$\begin{aligned}
\int_{s_*}^{s^*} \alpha^*(x)dF(x) &= \mathcal{C}(s^* - s_*)F(s^*) - \int_{s_*}^{\hat{s}} F(x)d\alpha^*(x) - \int_{\hat{s}}^{s^*} F(x)d\alpha^*(x) \\
&= \mathcal{C}(s^* - s_*)F(s^*) - \int_{s_*}^{\hat{s}} \mathcal{C}'\big(m(y) - y\big)F(y)dy \\
&\quad - \int_{\hat{s}}^{s^*} \mathcal{C}'\big(x + \phi(x)\big)F(x)dx \\
&= \mathcal{C}(s^* - s_*)F(s^*) - \int_{\hat{s}}^{s^*} \mathcal{C}'\big(x + \phi(x)\big)\Big\{\phi'(x)F\big(-\phi(x)\big) + F(x)\Big\}dx \\
&= \mathcal{C}(s^* - s_*)F(s^*) - \int_{\hat{s}}^{s^*} \mathcal{C}'\big(x + \phi(x)\big)\big(1 + \phi'(x)\big)\xi(x)dx \\
&= \mathcal{C}(s^* - s_*)F(s^*) - \int_{0}^{s^* - s_*} \mathcal{C}'(y)\xi\big(\psi(y)\big)dx
\end{aligned}$$

where the first equality follows from integration by parts, the second from the characterization of baseline rules, the third from the change of variable $y = -\phi(x)$, the fourth from the definition of $\xi(x)$, and the last equality from the change of variable $y = x + \phi(x)$.

Then, plugging in the length of the growth interval $\ell^*_{\mathcal{C}}$, we obtain the following general formula for the agents' payoff

$$A^*(\mathcal{C}) = p^*_{\mathcal{C}} - \int_0^{\ell^*_{\mathcal{C}}} \xi\big(\psi(y)\big)d\mathcal{C}(y) + (1 - p^*_{\mathcal{C}})I(\hat{w}, r).$$

Next, we use similar techniques to obtain the designer's payoff under α^*. As before, we start by computing the designer's payoff over scores on the growth interval $[s_*, s^*]$

$$\begin{aligned}\int_{s_*}^{s^*} \alpha^*(x)\big\{w(x) - \hat{w}\big\}dF(x) &= -\mathcal{C}(s^* - s_*)\mathcal{W}(s^*, \hat{w}) \\ &\qquad + \int_{\hat{s}}^{s^*} \mathcal{W}(x, \hat{w})d\alpha^*(x) + \int_{s_*}^{\hat{s}} \mathcal{W}(x, \hat{w})d\alpha^*(x) \\ &= -\mathcal{C}(s^* - s_*)\mathcal{W}(s^*, \hat{w}) \\ &\qquad + \int_{\hat{s}}^{s^*} \mathcal{W}(x, \hat{w})\big(1 + \phi'(x)\big)\mathcal{C}'\big(x + \phi(x)\big)dx \\ &= -\mathcal{C}(s^* - s_*)\mathcal{W}(s^*, \hat{w}) + \int_0^{s^* - s_*} \mathcal{W}(\psi(y), \hat{w})\mathcal{C}'(y)dy,\end{aligned}$$

where the first equality follows from integration by parts, the second equality is obtained by the change of variable $z = -\phi(x)$ in the second integral and by noticing that $\mathcal{W}(x, \hat{w}) = \mathcal{W}\big(-\phi(x), \hat{w}\big)$ by definition of the surplus and matching functions, and the third equality results from the change of variable $y = x + \phi(x)$.

This yields the following formula for the designer's payoff under the FP rule:

$$D^*(\mathcal{C}) = \big(\bar{w} - \hat{w}\big)^+ + \int_0^{\ell^*_{\mathcal{C}}} \mathcal{W}\big(\psi(y), \hat{w}\big)d\mathcal{C}(y).$$

The next calculation computes the aggregate payoff of agents on the growth interval $[s_-, s_+]$ under α^{**}. Then

$$\begin{aligned}\int_{s_-}^{s_+} \alpha^{**}(x)dF(x) &= \mathcal{C}(s_+ - s_-)F(s_+) - \int_{s_-}^{s_+} F(x)d\alpha^{**}(x) \\ &= \mathcal{C}(s_+ - s_-)F(s_+) - \int_{s_-}^{s_+} F(x)\mathcal{C}'(s_+ - x)dx \\ &= \mathcal{C}(s_+ - s_-)F(s_+) - \int_0^{s_+ - s_-} \mathcal{C}'(y)F(s_+ - y)dy,\end{aligned}$$

where the first equality follows from integration by parts, and the second relies on changing the integration variable from s to $y = s_+ - s$.

This yields the following general formula for the agents' aggregate payoff under α^{**}:

$$A^{**}(\mathcal{C}) = p^{**}_{\mathcal{C}} - \int_0^{\ell^{**}_{\mathcal{C}}} F(f_c - y + \hat{s})d\mathcal{C}(y) + (1 - p^{**}_{\mathcal{C}})I(\hat{w}, r).$$

Finally, we compute the designer's payoff under the unconstrained rule α^{**}. If $d \leq \overline{s} - \hat{s}$, then the designer attains her first-best payoff $\mathcal{W}(\hat{s}, \hat{w})$. Otherwise, her payoff over eligible agents is $p_{\mathcal{C}}\mathcal{W}(\hat{s},\hat{w})$, while ineligible agents yield a negative payoff

$$\begin{aligned}\int_{s_-}^{\hat{s}} \alpha^{**}(x)\big\{w(x)-\hat{w}\big\}dF(x) &= -\mathcal{W}(\hat{s},\hat{w})\alpha^{**}(\hat{s}) + \int_{s_-}^{\hat{s}} \mathcal{W}(x,\hat{w})d\alpha^{**}(x) \\ &= -\mathcal{W}(\hat{s},\hat{w})\left(p^{**}_{\mathcal{C}} - \mathcal{C}(\overline{s}-\hat{s})\right) + \int_{s_-}^{\hat{s}} \mathcal{W}(x,\hat{w})\mathcal{C}'(\overline{s}-x)dx \\ &= -\mathcal{W}(\hat{s},\hat{w})\left(p^{**}_{\mathcal{C}} - \mathcal{C}(f_{\mathcal{C}})\right) + \int_{f_{\mathcal{C}}}^{\ell^{**}_{\mathcal{C}}} \mathcal{W}(\overline{s}-y,\hat{w})\mathcal{C}'(y)dy\end{aligned}$$

This yields the general formula

$$D^{**}(\mathcal{C}) = \big(\bar{w}-\hat{w}\big)^+ + \mathcal{C}(f_{\mathcal{C}})\mathcal{W}(\hat{s},\hat{w}) + \int_{f_{\mathcal{C}}}^{\ell^{**}_{\mathcal{C}}} \mathcal{W}(\overline{s}-y,\hat{w})d\mathcal{C}(y).$$

The cost function $\mathcal{C}(x)$ is an increasing and concave function that defines a cumulative density function over the interval $[0,d]$.

Next, consider a sequence of concave cost functions $\{\mathcal{C}_n\}$ such that $d_{\mathcal{C}_n} = d$ is constant. Since falsifying by more than d is never rational, we consider only the restriction of these cost functions to the interval $[0,d]$. Suppose that $\{\mathcal{C}_n\}$ is an increasing sequence such that $\mathcal{C}_n(x)$ converges to 1 for all $x > 0$. It is easy to construct such a sequence by considering a sequence that increases in the concave order. For example, let $\mathcal{C}_0(x) = x/d$ and $\mathcal{C}_{n+1}(x) = g\big(\mathcal{C}_n(x)\big)$ for any continuously differentiable, strictly increasing and strictly concave bijective function g from $[0,1]$ to itself.

Then the sequence $\{\mathcal{C}_n\}$, viewed as cdfs on $[0,d]$, converges in distribution to the Dirac distribution putting all mass at 0. Furthermore, the sequences $p^*_{\mathcal{C}_n}$ and $p^{**}_{\mathcal{C}_n}$ converge to 1. Finally, the sequence $\mathcal{C}_n(f_{\mathcal{C}_n})$ is either constant at 1, or converges to 1 otherwise. Using these properties, we have the following limits for the payoffs we computed: $\lim_{n\to\infty} A^*(\mathcal{C}_n) = 1 - F(\hat{s})$; $\lim_{n\to\infty} D^*(\mathcal{C}_n) = \big(\bar{w}-\hat{w}\big)^+ + \mathcal{W}(\hat{s},\hat{w})$; $\lim_{n\to\infty} A^{**}(\mathcal{C}_n) = 0$ if $d \geq \overline{s} - \hat{s}$, and $\lim_{n\to\infty} A^{**}(\mathcal{C}_n) = 1 - F(\hat{s}+d)$ otherwise; $\lim_{n\to\infty} D^{**}(\mathcal{C}_n) = \big(\bar{w}-\hat{w}\big)^+ + \mathcal{W}(\hat{s},\hat{w})$.

Putting these together, we obtain that the loss-rate of the designer $L(\mathcal{C}_n)$ always converges to 0, whereas the gain rate of the agents becomes arbitrarily large if $d \geq \overline{s}-\hat{s}$, and converges to $\frac{F(\hat{s}+d)-F(\hat{s})}{1-F(\hat{s}+d)}$ otherwise. □

Proof of Proposition 6. Let $\tilde{\alpha}^*(\hat{w})$ denote the correspondence mapping $\hat{w}$ to the set of solutions of the baseline problem. By Lemma A.1, we can write (BP) as an optimization problem over the set of nondecreasing functions from S to $[0,1]$ satisfying (FPC) and (BP). This space is compact, by Helly's theorem, and convex. Furthermore, the objective function is linear and therefore continuous in α. Hence, Berge's maximum theorem implies that $\tilde{\alpha}^*(\hat{w})$ is a continuous correspondence. By Theorem 1 and Theorem B.1, the correspondence is singleton-valued for $\hat{w} \neq \bar{w}$, and for $\hat{w} = \bar{w}$ if $c(\bar{s}|\underline{s}) < 1$, so the continuity results with respect to $\hat{w}$ follow.

The space of feasible and non-decreasing allocation rules is also a lattice with respect to the partial order $\alpha \succeq \beta \Leftrightarrow \alpha(s) \geq \beta(s),\ \forall s$, 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 that 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 that $\alpha^*(s, \bar{w}, r)$ is strictly increasing in r for every s, both in the (UID) and (UDD) cases, since $I(\bar{w}, r) = r$. 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 6. 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 2. The proof of this theorem relies on the following Lemma:

Lemma C.1 (Endogenizing the outside option)**.** *The following statements are equivalent:*

(i) α *solves* (P_W).

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

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

Furthermore, the value function of the within problem is concave in ρ*, and its derivative* $W'(\rho)$ *exists almost everywhere, and is 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 within problem is linear in α.

Coming to the proof of the theorem, the function $w(s)$ is bounded by assumption. Let $w^- = w(\underline{s})$ and $w^+ = w(\overline{s})$ 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 C.1, $\alpha^*(\cdot, \hat{w}, r)$ is then the unique solution to the within problem (P_W).

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

By Lemma C.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 $c(\overline{s}|\underline{s}) < 1$. □

Proof of Theorem 3. The objective function of the across problem is continuous and concave in $\boldsymbol{\rho}$ by Theorem 2, 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 2, $W_i'(\rho_i) = \hat{w}_i(\rho_i)/\mu$, 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 2. □

Proof of Proposition 7. 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 C.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 2. 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

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

Using the differential form, we have:

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

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

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

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 $[s, \bar{s}_i]$ 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-score 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

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

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 to high scores under both distributions. Then, using (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

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

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

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

□

Proof of Proposition 8. Consider the problem of the decision maker deciding how to allocate objects. They can only condition their 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 they allocate an object to members of group i with a negative recommendation. Their problem is

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

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

$$0 \leq \alpha_i(t) - \alpha_i(s) = (g_i - b_i)\big(\alpha_i^*(t) - \alpha_i^*(s)\big) \leq \alpha_i^*(t) - \alpha_i^*(s) \leq c_i(t|s),$$

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^o) 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.

$\square$

Supplementary appendix

D Additional comparative statics

Returns to scale in falsification We investigate how the shape of the falsification cost function influences optimal allocations. Specifically, we focus on the class $\mathcal{E}$ of Euclidean cost functions, $c(t|s) = \mathcal{C}\big(|t-s|\big)$, where $\mathcal{C}$ is either concave or convex, and examine the effect of increasing the convexity of the cost function. Intuitively, greater convexity captures lower economies of scale in the magnitude of falsification.

To compare cost functions, we need some normalization. We normalize costs 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 $\mathcal{C}, \hat{\mathcal{C}}$ such that $\mathcal{C}^{-1}(\gamma) = \hat{\mathcal{C}}^{-1}(\gamma) < s^*(0) - s_*(0)$. We say that $\hat{\mathcal{C}}$ is more convex than $\mathcal{C}$, and denote $\hat{\mathcal{C}} \succeq_{vex} \mathcal{C}$, if either $\hat{\mathcal{C}}$ is convex and $\mathcal{C}$ is concave, or both are concave and $\mathcal{C}$ is more concave than $\hat{\mathcal{C}}$ in the usual sense, or both are convex and $\hat{\mathcal{C}}$ is more convex than $\mathcal{C}$ in the usual sense.[24] We denote the corresponding baseline allocation rules by $\alpha^*, \hat{\alpha}^*$, and their growth intervals by $I^*, \hat{I}^*$.

Proposition D.1 (Effect of lowering economies of scale)**.** *If $\hat{\mathcal{C}}$ is more convex than $\mathcal{C}$, then:*

(i) $I^* \subseteq \hat{I}^* \subseteq [s_*(0), s^*(0)]$. *Furthermore,* $I^* = \hat{I}^* \subset [s_*(0), s^*(0)]$ *if both cost 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 cost functions are convex and* $\hat{\alpha}^*(s) \leq \alpha^*(s)$ *for all s if the group has low-score priority, but* $\hat{\alpha}^*(s) \geq \alpha^*(s)$ *for all s if the group has high-score priority.*

Proof of Proposition D.1. When $\mathcal{C}$ 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 $m(s_*) - s_* = \min\{s^*(0) - s_*(0), 1/\mathcal{C}'(0)\}$ by Proposition 1. Furthermore, if both cost functions are convex, the convex ordering and our normalization imply $\hat{\mathcal{C}}'(0) \leq \mathcal{C}'(0)$, hence $I^* \subseteq \hat{I}^*$.

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

Next, let $\delta(s) = \hat{\alpha}^*(s) - \alpha^*(s)$. Suppose first that 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

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

where g is an increasing and concave bijection of $[0, 1]$ such that $\mathcal{C} = g \circ \hat{\mathcal{C}}$. As such $g'(0) \geq 1 \geq g'(1)$, and g' is a non-increasing function. Since $\mathcal{C}' \geq 0$, 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 $\hat{\mathcal{C}}$ to be more convex than $\mathcal{C}$, it must be that $\hat{\mathcal{C}}'(0) \leq \mathcal{C}'(0)$ which implies (ii).

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

If $I^* = \hat{I}^* = [s_*(0), s^*(0)]$, then both functions must be convex by (i). If the group has low-score priority, then $\alpha^*\big(s_*(0)\big) = \hat{\alpha}^*\big(s_*(0)\big) = 0$, and both allocation rules are linear with respective slopes $\mathcal{C}'(0) \geq \hat{\mathcal{C}}'(0)$, implying $\hat{\alpha}^*(s) \leq \alpha^*(s)$ for all s. If instead the group has high-score 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. □

In words, lower economies of scale, just like higher gaming ability, benefit low-score agents and hurt high-score agents. However, if the diseconomies of scale become too strong, as in case (iii), the effect is uniform across all scores, resembling the impact of high gaming abilities as described in Proposition 3.

Score distribution We examine what changes in the score distribution result in a uniformly higher optimal allocation probability. To this end, we redefine the score as the surplus, so $s = w(s) - \hat{w}$, which effectively transforms the score distribution F. Consequently, the eligibility threshold is fixed at 0. We consider two atomless score distributions, $\hat{F}$ and $\tilde{F}$, whose common support $[\underline{s}, \overline{s}]$ includes a neighborhood of 0. The function $\Delta(s) = \tilde{F}(s) - \hat{F}(s)$ denotes the change in the score distribution.

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

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

Proposition D.2 (Effect of score distribution)**.** *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 $0 > s \geq \max\{\hat{s}_*(0), \tilde{s}_*(0)\}$,

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

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

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

Proof of Proposition D.2.

- $\underline{(b) \Rightarrow (a)}$. 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 1, (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-score priority, then so does $\tilde{F}$, and if $\tilde{F}$ has low-score 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

$$\int_{\hat{s}_*}^{\tilde{m}(\hat{s}_*)} c_{t^+}(x|x)dx \leq \int_{\hat{s}_*}^{\hat{s}^*} c_{t^+}(x|x)dx = 1.$$

Then, using (B) 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-score priority (the event in which only $\hat{F}$ has high-score 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

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

Then, using (B) 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

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

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

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

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

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

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

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

- *(a) ⇒ (b)*. Suppose (a) holds, and consider the family of linear cost functions $c(t|s) = \beta|t-s|$, 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-score priority, but then (a) implies $\tilde{F}$ must have high-score priority as well. Similarly $\tilde{\alpha}_\beta(\overline{s}) < 1$ if and only if $\tilde{F}$ has low-score priority, and then (a) implies $\hat{F}$ has low-score 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).

- *(b) ⇔ (c) ⇔ (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

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

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

Hence, a simple first-order stochastic dominance shift is not sufficient to increase the allocation probability for all scores. Since it is challenging to interpret conditions (c) and (d), we provide a more easily interpretable sufficient condition on Δ. We say that 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 $\hat{F}$. In other words, for every $[s, s'] \subseteq I$,

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

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

Proposition D.3. *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 in the distribution that shifts mass from ineligible scores to eligible ones satisfies the conditions of Proposition D.3, thereby uniformly increasing the allocation probability.

Proof of Proposition D.3. 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 D.2.

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 condition 1, and the second inequality is from condition 3 and condition 1, as $\Delta(z) \geq 0$. If $z \geq a$, then condition 2 implies $\Delta(z) \geq \Delta(x)$ for every $x \in [z, 0]$. The proof is symmetric for the integral from 0 to y. □

E 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 1: Algorithm to solve the across problem

```
∀i,  ρ_i^0 ← μ_i A_i^*(0, 1);
R^0 ← R(ρ^0);
Q^0 ← Q(ρ^0);
k ← 0;
repeat
    k ← k + 1;
    ∀ℓ ∈ Q^{k−1},  ŵ_ℓ^k ← ŵ_ℓ^φ and r_ℓ^k ← r_ℓ^φ;
    if R^{k−1} = 0 then
        ∀ℓ ∉ Q^{k−1},  ŵ_ℓ^k ← 0 and r_ℓ^k ← 1;
    else
        ŵ, r ← Solution of: Σ_{ℓ∈Q^{k−1}} φ_ℓ ρ̄ + Σ_{i∉Q^{k−1}} μ_i A_i^*(ŵ, r 𝟙_{ŵ=w̄_i}) = ρ̄;
        ∀ℓ ∉ Q^{k−1},  ŵ_ℓ^k ← ŵ and r_ℓ^k ← r 𝟙_{ŵ=w̄_ℓ};
    end
    ∀i,  ρ_i^k ← μ_i A_i^*(ŵ_i^k, r_i^k);
    R^k ← R(ρ^k);
    Q^k ← Q(ρ^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 decreasing r from 1 to 0 and keeping $\hat{w}$ constant whenever $\hat{w} = \bar{w}_i$ for some i. Therefore a simple algorithm can solve this equation.

Proposition E.1. *Algorithm 1 finds a solution of* (P_A) *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 3. Therefore $\boldsymbol{\rho}^k$ is a solution to the across problem. □

F Continuum as a single-agent

In this appendix, we explain why treating the continuum as a single agent is without loss of generality. 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.[25] However, to find a solution to the initial program, 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.

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