Falsification-proof non-market allocation mechanisms *

Eduardo Perez-Richet † Vasiliki Skreta ‡

April 2, 2026

Abstract

Relying on manipulable characteristics to target the allocation of resources is fraught with difficulties. Naively optimal rules lead to falsification and allocative inefficiency. Optimizing the rule for allocative efficiency while accounting for falsification typically induces falsification, at a heavy cost to the agents. Moreover, falsification generates negative externalities. We characterize optimal falsification-proof rules under a broad range of falsification technologies and provide a honesty condition under which they solve the unconstrained problem of a planner who maximizes a weighted sum of allocative surplus, agents' welfare, and the social cost of falsification externalities.

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 title "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 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 based on criteria such as household income. In many cases, eligibility is assessed through a *score* that measures an agent's characteristics or performance and acts as a proxy for their *worth*.

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 (Bolton, 2018; McMichael, 2022).[3] Throughout the paper, we use the term *falsification* as a broad category that encompasses all individually costly and socially wasteful activities that agents undertake to produce an altered score.

Falsification is not only wasteful; it is also socially harmful. First, it distorts achievable allocations, unfairly penalizing agents for whom it is more costly to falsify.[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 the assessment of emissions levels. Third, it may make a mechanism politically unsustainable when evidence of falsification becomes public. Fourth, it can erode trust and deplete the supply of objects to allocate. For example, a scandal involving the manipulation of the German liver allocation system by transplant providers led to a 20%-40% erosion in organ donation (Bolton, 2018). Fifth, there is evidence that dishonest behavior spreads in society (see, e.g., Galbiati and Zanella, 2012; Ajzenman, 2021).

[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.

[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) shows that priority gaming in school choice mechanisms resulted in better assignments for those who engaged in such practices and adversely affected others.

The main contribution of this paper is twofold. First, we characterize allocatively optimal falsification-proof allocation rules under a broad range of falsification technologies, encompassing *decreasing* (Theorem 1) and *increasing differences* (Theorem 2). Falsification-proof allocation rules ensure *honesty* by giving agents no incentive to falsify, thereby eliminating the private[5] and the social costs of falsification. However, requiring honesty has a price, since it constrains possible designs. Our characterization allows us to conduct a cost-benefit analysis of honesty, balancing its allocative price with its value to agents. Second, we show that under meaningful welfare concerns and increasing differences, falsification-proof rules are not merely constrained-optimal but fully optimal. We consider a designer whose objective is a weighted sum of allocative surplus, agents' welfare, and the social cost of falsification externalities. We provide a sufficient condition for honesty to be optimal (Theorem 3). This honesty condition allows us to isolate the four factors that contribute to making honesty optimal in our model. The first two are the welfare weights on agents and externalities. The two concerns are substitutes, and each of them can be sufficient on its own. The remaining factors are high falsification costs and low correlation between an agent's score and their worth.

Specifically, we study the problem of allocating a mass of homogeneous prizes to a heterogeneous population of agents using non-market mechanisms based on *scores*. The score is a publicly available but *falsifiable* measure of an agent's private characteristics. If agents do not falsify, they produce their *natural score*, which reflects their true characteristics.[6] Agents who *falsify* produce an altered score at a cost. The falsification cost can exhibit either increasing or decreasing differences in the natural and altered score. We assume that an agent's natural score is positively correlated with their *worth*: the designer's value of assigning a prize to that agent. We define *allocative surplus* as the aggregate worth of rewarded agents.

Our characterization results make a technical contribution. The constrained optimization problem can be solved with a standard first-order approach under decreasing differences since falsification-proofness constraints bind locally. Under increasing differences, however, the falsification-proofness constraints bind for distant scores. To address this difficulty, we transform the designer's problem into a program that is equivalent to the dual of the classical Monge-Kantorovich optimal transport problem,

[5] Milli, Miller, Dragan and Hardt (2019) show that optimal binary classifiers under falsification can generate a significant cost in falsification for the agents and argue that it 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.

which we leverage to characterize the optimal allocation rule in Theorem 2.

What is the welfare impact of inducing honesty? In standard mechanism design, misreporting is costless by assumption, and honesty is inconsequential due to the revelation principle. In contrast, we show in Perez-Richet and Skreta (2022) that under costly falsification with increasing differences, optimal mechanisms harness falsification to enhance allocative surplus.[7] Therefore, honesty has a price: requiring falsification-proofness results in a loss of allocative surplus for the designer. In return, we show that honesty also has value for agents. Regardless of their natural score, agents are better off under the optimal falsification-proof mechanism than under the unconstrained optimal mechanism (Proposition 4). Furthermore, we show that this (allocative) price of honesty can be arbitrarily small and the value of honesty (for agents) arbitrarily large when the falsification cost has increasing differences (Proposition 5). These results highlight that, even when abstracting from externalities, falsification-proof mechanisms can enhance agents' welfare at a small allocative price.

We also study how changes in *gaming ability* affect optimal falsification-proof rules. Gaming ability is defined as the inverse of a multiplicative scaling factor on the cost of falsification. We find that its effect on the allocation rule is nuanced: it benefits low-score agents and harms high-score agents when gaming ability is sufficiently low, whereas the effect becomes uniform at higher levels, with its direction depending on score priority. Finally, we show how to extend the characterization of allocatively optimal falsification-proof rules when there is a resource constraint.

The Online Appendix contains several extensions and additional results. Section S1 extends the analysis to falsification cost functions with a fixed cost component under increasing differences. Section S2 develops an extended model with multiple identifiable groups, additional sources of heterogeneity across agents, and exogenous quotas. Section S3 provides additional comparative statics results with respect to the score distribution and returns to scale in the falsification technology. Section S4 provides a formal justification for the continuum formulation.

2 The allocation problem

Framework. The designer allocates a mass $\bar{\rho} \leq 1$ of indivisible and homogeneous prizes to a unit mass of heterogeneous agents without transfers. In many applications,

[7] Perez-Richet and Skreta (2022) derive optimal *tests*, which correspond to optimal allocation mechanisms in the current framework.

a resource constraint is irrelevant because prizes are immaterial, such as services, certificates, labels, or awards. In most of the paper, we therefore set $\bar{\rho} = 1$. We extend the analysis to the case $\bar{\rho} < 1$ in Section 6.

An exogenous one-dimensional metric, the *score*, measures certain private characteristics of the agent. Each agent has a private type $s \in S \subseteq \mathbb{R}$, corresponding to their *natural score*—the score they obtain when they do not interfere with the measuring technology. Agents are also characterized by their worth w, a scalar capturing the designer's value of allocating a prize to them. For simplicity, we assume that all agents have a homogeneous value normalized to 1 for prizes,[8] and that they do not observe their worth.[9]

Distributional assumptions. Agents' characteristics are independently and identically distributed draws from a joint distribution over (s, w). Hence, the different dimensions of an agent's characteristics can be, and typically are correlated but are independent across agents. Let F denote the cumulative distribution function of natural scores, which we assume has full support on an interval $S = [\underline{s}, \overline{s}]$ and no atoms.

Designer and agents' payoffs. We assume that the worth w is bounded and integrable, conditional on s. We denote the corresponding expected worth by $w(s) = \mathbb{E}(w|s)$, and by $\bar{w} = \mathbb{E}(w)$ the expected worth in the population. The designer's payoff from assigning a prize to an agent with score s is $w(s)$. We assume that score and worth are positively correlated in the sense that $w(s)$ is strictly increasing. We also assume that $w(\cdot)$ is continuous. We denote the value of the outside option (not allocating a prize) by $\hat{w}$.

Agents can *falsify* their score at a cost $c(t|s) \geq 0$ to produce a *falsified score* t instead of their natural score s. The expected payoff of such an agent receiving an object with probability α is $\alpha - c(t|s)$. We say that agents producing their natural score are being *honest*. Honesty is costless, so $c(s|s) = 0$. The falsification cost may reflect technical costs, psychological lying costs, as well as expected penalties.

Priority. The setup exhibits *high-score priority* if the expected worth exceeds the outside option, $\bar{w} > \hat{w}$; *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 prizes to all agents

[8] We extend our analysis to heterogeneous private values in Online Appendix S2.

[9] In fact, this is irrelevant since we show in Online Appendix S2 that the designer could not elicit this information through the mechanism.

under high-score priority but would not allocate any prizes under low-score priority. We show that priority determines whether it is more important for the designer to ensure that the top score receives a prize with certainty (high-score priority) or that the lowest score does not receive a prize (low-score priority).

Falsification cost. We assume that the cost function is *monotonic* for *upward* falsification: if $t \geq s$, then $c(t|s)$ is (locally) strictly increasing in t and decreasing in s. We also assume that the cost function satisfies a *regularity* assumption.

Assumption 1 (Regularity)**.** *The cost function $c(t|s)$ is continuously differentiable in t on $[s, \overline{s}]$, and in s on $[\underline{s}, t]$.*

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 concern only 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.

Example 1 (The Euclidean family)**.** *A cost function is* Euclidean *if $c(t|s) = \mathcal{C}(t-s)$ for $t \geq s$, 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 (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 cases may arise in different contexts. For instance, emission testing software that becomes cheaper to deploy once the initial investment is made yields increasing returns whereas memorizing increasingly difficult* *test problems yields decreasing returns. Another example is when falsification costs arise from a linear detection probability πx and a fine $\varphi x + \varphi_0$, both of which increase with the falsification level x. In this case, the Euclidean cost $\mathcal{C}(x) = \pi x(\varphi x + \varphi_0)$ satisfies* (UDD). $\diamond$

Example 2 (The shifted Euclidean family)**.** *We generalize the Euclidean class by considering* Shifted Euclidean *cost functions of the form $c(t|s) = \kappa(s)\mathcal{C}(t-s)$ for $t \geq s$. This family captures correlations between the natural score and* gaming ability, *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.* $\diamond$

The (UID) class is particularly suited for studying situations in which falsification involves substantial setup costs. Though our assumptions rule out a fixed cost, we extend our analysis to fixed costs in Online Appendix S1. Furthermore, we derive the unconstrained optimal allocation rule under (UID) in Perez-Richet and Skreta (2022), allowing us to analyze the welfare consequences of imposing falsification-proofness in Section 4.

Falsification-proof allocative surplus optimization. We are interested in mechanisms that incentivize agents to produce their natural score: *falsification-proof* mechanisms. We show in Perez-Richet and Skreta (2023) that it is without loss of generality to restrict attention to *score-based allocation rules* $\alpha : S \to [0,1]$, where $\alpha(t)$ is the probability that a prize is allocated to an agent who produces score t. Then, a mechanism is falsification-proof if and only if it satisfies the constraint[10]

$$\forall (s,t), \quad \alpha(t) - \alpha(s) \leq c(t|s). \tag{FPC}$$

Selecting the *honest* best response of the agents, we can then express the designer's problem as

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

[10]We can also interpret (FPC) as being motivated by inequality concerns. The cost then acts as a bound on inequality between score pairs. We thank Ricardo Alonso for suggesting this interpretation.

with the *probability constraint* (PC): $\forall s,\ 0 \le \alpha(s) \le 1$.

The term $w(s) - \hat{w}$ in the objective represents the surplus relative to the exogenous *outside option* $\hat{w}$. Let $\hat{s}$ denote the *eligibility threshold*, defined by $w(\hat{s}) = \hat{w}$. Surplus is positive for *eligible* scores, $s \ge \hat{s}$, and negative for *ineligible* scores, $s < \hat{s}$. The objective of the designer in this problem is to maximize *aggregate allocative surplus* (or simply allocative surplus), $\int_S \alpha(s)\{w(s) - \hat{w}\}dF(s)$.

The weighted-welfare design problem. A natural designer problem is to maximize an objective that combines three criteria: allocative surplus, agents' welfare, and the negative externality of falsification. This weighted welfare objective is

$$\underbrace{\iint \alpha(t)\big\{w(s) - \hat{w}\big\}\, d\phi(t|s)dF(s)}_{\text{allocative surplus}} + \zeta \underbrace{\iint [\alpha(t) - c(t|s)]\, d\phi(t|s)dF(s)}_{\text{agents' welfare}}$$
$$- \xi \underbrace{\iint D(t, s)\, d\phi(t|s)dF(s)}_{\text{externality cost}},$$

where ϕ is the falsification strategy of the agents, $\zeta \ge 0$ is the weight on agents' welfare, and $\xi \ge 0$ is the weight on the externality cost, which we measure by a distance D on S. For simplicity, we identify $D(t, s)$ with the falsification cost $c(t|s)$, but this is inessential for our results (see Section 5). Collecting terms yields the objective function

$$\Omega(\alpha, \phi, \xi, \zeta) = \iint \alpha(t)\big\{w(s) - \hat{w} + \zeta\big\}\, d\phi(t|s)dF(s) - (\xi + \zeta) \iint c(t|s)\, d\phi(t|s)dF(s).$$

Hence, the weighted welfare objective is equivalent to an allocative surplus problem with a downward shifted outside option $\hat{w} - \zeta$ and a combined weight $\xi + \zeta$ on falsification costs. The weighted welfare design problem is:

$$\max_{\alpha,\phi}\ \Omega(\alpha, \phi, \xi, \zeta) \tag{WWP}$$
$$\text{s.t. (PC)}$$
$$\phi\big(T(\alpha, s)|s\big) = 1, \quad \forall s. \tag{IC}$$

The incentive constraint (IC) requires the agents' falsification strategy $\phi : S \to \Delta(S)$ to be concentrated on the set of *optimal falsification targets*

$$T(\alpha, s) = \arg\max_t\ \alpha(t) - c(t|s).$$

In Section 5, we show that in the (UID) case, the solution to this problem is given by the solution of (P) with the downward-shifted outside option $\hat{w} - \zeta$ whenever the weight on the externality and agents' welfare is sufficiently large. This may occur even without accounting for the externality ($\xi = 0$).

3 Allocatively optimal falsification-proof rules

The first-best rule is a step function that allocates to all eligible scores $s \geq \hat{s}$ and to none below. It is not falsification-proof, since an agent just below $\hat{s}$ can gain by falsifying upward. More generally, (FPC) bounds the growth of α between any two scores. The analysis proceeds as follows. Monotonicity of optimal rules (Lemma 2) allows us to rewrite allocative surplus via integration by parts in terms of *cumulative surplus*—the total surplus generated by scores above or below any given threshold. The probability constraint caps total growth at 1, and its Lagrange multiplier ν pins down the *growth interval* $[s_*, s^*]$, over which α rises from its minimum to its maximum value, as a *Zero Average Surplus* interval (ZAS). We can solve (P) by first solving a *simplified program* that takes this interval as given, then identifying the correct interval via a *boundary condition* encoding the interplay between falsification costs and the probability constraint.

3.1 Cumulative Surplus

We introduce *cumulative surplus* functions that appear when integrating allocative surplus by parts. They play a central role in the analysis, as they determine the nature of growth intervals for optimal falsification-proof rules.

Cumulative surplus functions. The *upward cumulative surplus* at s is the total amount of allocative surplus generated by scores above s. It corresponds to the marginal gain of uniformly increasing the allocation probability of all such scores:

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

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:

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

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

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

We next list key properties of $\mathcal{W}$ illustrated in Figure 1.

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 exists a unique pair of scores $s_*(\nu) \le \hat{s} \le s^*(\nu)$ such that $\mathcal{W}(s_*(\nu),\hat{w}) = \mathcal{W}(s^*(\nu),\hat{w}) = \nu$; (iii) $\mathcal{W}(s,\hat{w}) \ge \nu$ if and only if $s \in [s_*(\nu), s^*(\nu)]$; (iv) $s_*(\nu)$ and $-s^*(\nu)$ are continuous and increasing functions; (v) $s^*(0) = \bar{s}$ under low-score priority, and $s_*(0) = \underline{s}$ under high-score priority. Under neutral priority, $[s_*(0), s^*(0)] = [\underline{s}, \bar{s}]$.*

Matching ineligible and eligible scores. We use cumulative surplus to define 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 $\mathcal{W}(s,\hat{w}) = \mathcal{W}(m(s),\hat{w})$. We say that a pair (s_*, s^*) is a *matching pair* if $s^* = m(s_*)$.

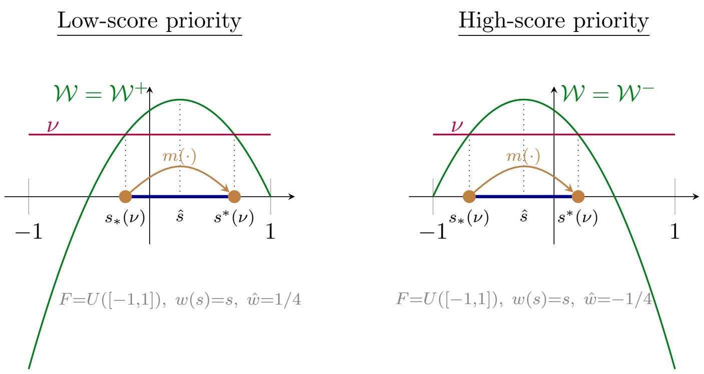

Figure 1: *Cumulative surplus, matching pairs and growth interval.*

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

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

(ZAS) intervals form a nested family of intervals around the eligibility threshold, ranging from $[s_*(0), s^*(0)]$ when $\nu = 0$ to $\{\hat{s}\}$ when $\nu = \mathcal{W}(\hat{s}, \hat{w})$.

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

3.2 Simplifying the program

Monotonicity of optimal rules. We can, without loss of optimality, restrict attention to monotonic allocation rules. Intuitively, replacing a nonmonotonic rule with its lower monotonic envelope below the eligibility threshold and its upper monotonic envelope above the eligibility threshold increases the probability that eligible scores receive a prize while decreasing the probability that ineligible scores do. Furthermore, this transformation preserves falsification-proofness.

Lemma 2 (Monotonicity). *If α is feasible for (P) but not monotonic, there exists a monotonic and feasible rule $\tilde{\alpha}$ that strictly increases allocative surplus.*

We refer to the monotonicity condition on α as (MON). It is also useful to note that (FPC) implies that any feasible allocation rule must be continuous.

Differential program. Relying on the monotonicity of α, we can use integration by parts for Stieltjes integrals to rewrite allocative surplus as a function of either downward or upward cumulative surplus:

$$\int_S \alpha(s)\big\{w(s) - \hat{w}\big\} dF(s) = (\bar{w} - \hat{w})\underline{\alpha} + \int_S \mathcal{W}^+(s, \hat{w})\, d\alpha(s) \tag{$\underline{\text{DOF}}$}$$

$$= (\bar{w} - \hat{w})\overline{\alpha} + \int_S \mathcal{W}^-(s, \hat{w})\, d\alpha(s). \tag{$\overline{\text{DOF}}$}$$

($\overline{\text{DOF}}$) implies that it is optimal to set $\overline{\alpha} = 1$ under high-score priority. Intuitively, an optimal rule must then ensure that the top score receives a prize with certainty. ($\underline{\text{DOF}}$) implies that it is optimal to set $\underline{\alpha} = 0$ under low-score priority. In this case, an optimal rule must ensure that the lowest score does not receive a prize. The *initial condition* that an optimal allocation rule must satisfy is therefore

$$\underline{\alpha} = 0 \text{ if } \bar{w} < \hat{w}; \quad \overline{\alpha} = 1 \text{ if } \bar{w} > \hat{w}; \quad 0 \leq \underline{\alpha} \leq \overline{\alpha} \leq 1 \text{ if } \bar{w} = \hat{w}. \tag{Init}$$

If an allocation rule satisfies (Init) and (MON), we can rewrite the probability constraint (PC) in the differential form (DPC): $\int_S d\alpha(s) \leq 1$. We can then write a common program for all priorities using the composite cumulative surplus function $\mathcal{W}$ from Section 3.1.

Lemma 3 (Differential program)**.** *An allocation rule α solves* (P) *if and only if α satisfies* (Init) *and solves the differential program*

$$\max_{\alpha} \int_S \mathcal{W}(s, \hat{w}) d\alpha(s) \quad \text{s.t. (DPC), (FPC), (MON).} \tag{DP}$$

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

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

By Lemma 1(iii), maximizing this Lagrangian under (MON) implies that α should put no mass outside $\big[s_*(\nu), s^*(\nu)\big]$, where the cumulative surplus is below ν (see Figure 1). Growth intervals are therefore (ZAS) intervals and are contained within the maximal interval $[s_*(0), s^*(0)]$.

The simplified program. We construct a solution to the designer's problem by first solving the *simplified program*

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

which restricts the objective function to a (ZAS) interval $[s_*, s^*]$, only imposing (FPC) over this interval and relaxing the probability constraint. We then obtain the optimal 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 saturate the probability constraint:* $\alpha(s^*) - \alpha(s_*) = 1$ *if possible. Otherwise 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* $\alpha(s^*) \leq 1$ *and* $\alpha(s_*) \geq 0$ *under neutral priority.*
- ***Step 4.*** *Set* $\alpha(s) = \alpha(s_*)$ *for* $s < s_*$*, and* $\alpha(s) = \alpha(s^*)$ *for* $s > s^*$*.*

Lemma A.1 in Appendix A shows formally why Procedure 1 yields the solutions of (P), adapting standard Lagrangian necessity and sufficiency results to our setup.

Step 1 determines the growth rate and shape on the growth interval. Step 2 identifies the growth interval and works like a complementary slackness condition: either the probability constraint is not binding, in which case its Lagrange multiplier is $\nu = 0$ and the growth interval is therefore $\big[s_*(0), s^*(0)\big]$, or it is binding, so the growth interval must satisfy $\alpha(s^*) - \alpha(s_*) = 1$, the Lagrange multiplier satisfies $\nu > 0$, and is determined by the binding condition. Step 3 sets the boundary allocation probability based on priority. Step 4 extends the rule beyond the growth interval by assigning minimum probability to scores below s_* and maximum probability to scores above s^*.

3.3 Optimal rules: characterization

The shape of the optimal rule on the growth interval is determined by the binding 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. We therefore solve the simplified program by establishing a connection with the dual of the Monge-Kantorovich optimal transport problem.

Optimal rule under (UDD). Under (UDD), any allocation rule that satisfies (FPC) must be absolutely continuous and satisfy the necessary first-order condition of the agent's problem: at every s where the derivative $\alpha'(s)$ exists, $\alpha'(s) \leq c_{t^+}(s|s)$, where $c_{t^+}(s|s)$ is the right derivative of c with respect to the target t, evaluated at $t = s$. Integrating by parts the objective function of the simplified program and using the absolute continuity of α, we can rewrite the simplified program in the following differential form:

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

where $\nu = \mathcal{W}(s_*, \hat{w}) = \mathcal{W}(s^*, \hat{w})$. Since $\mathcal{W}(s, \hat{w}) - \nu > 0$ on the interior of $[s_*, s^*]$, the unique 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 optimal rule is determined by the *matching condition* $s^* = m(s_*)$ and the *boundary condition*

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

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 whenever possible and otherwise selects the maximal growth interval $[s_*(0), s^*(0)]$, with $\nu = 0$. Whether the constraint binds depends on the magnitude of falsification costs, and its degree of slackness is measured by the *probability gap* $\Gamma_{udd} = 1 - \int_{s_*}^{s^*} c_{t^+}(x|x)dx$. Then the probability constraint is slack, and $\Gamma_{udd} > 0$, if and only if the *slackness condition* $\int_{s_*(0)}^{s^*(0)} c_{t^+}(x|x)dx < 1$ is satisfied.

The probability gap must be allocated according to priority. In the low-score priority case, it is optimally withheld from agents to ensure $\alpha\big(s_*(0)\big) = 0$, thereby preventing low-score agents from receiving a prize. In the high-score priority case, it is optimally assigned to agents to ensure $\alpha(s^*(0)) = 1$, so that high-score agents receive a prize with certainty. To capture this choice, we define the *share index*

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

which represents the share of the probability gap allocated to agents: it equals 0 under low-score priority and 1 under high-score priority. We refer to r as the *neutral gap share.*

Under neutral priority, the interval $[s_*(0), s^*(0)]$ must be $[\underline{s}, \overline{s}]$, reducing the gap condition to the slackness condition $\int_{\underline{s}}^{\overline{s}} c_{t^+}(x|x)dx < 1$. The designer is then indifferent about the share $r \in [0, 1]$ of the probability gap allocated to agents, and the optimal rule is not unique. Multiplicity therefore arises under neutral priority when the slackness condition holds:

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

Although the designer is indifferent, the choice of r plays an important role in solving the problem with a resource constraint (see Section 6). Indeed, choosing r allows the designer to adjust the total mass of prizes accruing to agents in order to satisfy the constraint.

Consequently, the optimal rule under (UDD) is

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

where (s_*, s^*) satisfies the matching and boundary conditions. Note that the conditions $s < s_*$ and $s > s^*$ are vacuous whenever $s_* = \underline{s}$ and $s^* = \overline{s}$ (respectively) so the allocation rule is always continuous.

Theorem 1 (Optimal rule under UDD)**.** *Suppose the cost function satisfies* (UDD)*. If* (Mult) *does not hold, then α^*_{udd} is independent of r and is the unique solution of* (P)*. If* (Mult) *holds, then the set of optimal rules is $\left\{\alpha^*_{udd}(\cdot, \bar{w}, r)\right\}_{r\in[0,1]}$.*

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

Optimal rule under (UID). Unlike in the (UDD) case, the optimal rule is not pinned down by local slope constraints, but by the matching function m which describes pairs of ineligible-eligible scores for which the (FPC) condition binds:

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

where $\bar{I}(\hat{w}, r) = 1 - I(\hat{w}, r)$. As in the (UDD) case, the conditions $s < s_*$ and $s > s^*$ are vacuous whenever $s_* = \underline{s}$ and $s^* = \overline{s}$ (respectively) so the allocation rule is always continuous. Although they play the same role, the definitions of the boundary condition, the probability gap, and the multiplicity condition differ from the (UDD) case.

Over the growth interval, the allocation rule grows at rate $-c_s\big(m(x)|x\big)$ for ineligible scores x, and at rate $c_t\big(x|m^{-1}(x)\big)$ for eligible scores x. We prove that (FPC) binds between matching scores and, in particular, for the pair (s_*, s^*). Therefore, the growth interval $[s_*, s^*]$ is uniquely determined by the *boundary condition*

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

Whether the probability constraint binds depends on the magnitude of falsification costs, and its degree of slackness is measured by the *probability gap*, which is now defined as

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

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.

As in the (UDD) case, the gap is optimally withheld in the low-score priority case and optimally allocated to agents in the high-score priority case. The index $I(\hat{w}, r)$ captures this choice. Given the boundary condition (B), the probability constraint is slack (and hence $\Gamma_{uid} > 0$) in the neutral-priority case if and only if the slackness condition $c\big(\overline{s}|\underline{s}\big) < 1$ is satisfied. When it holds, there is a gap $\Gamma_{uid} > 0$ between the total growth of the optimal rule and its upper bound (equal to 1). In the (UID) case, multiplicity therefore arises under the condition

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

Theorem 2 (Optimal rule under UID)**.** *Suppose the cost function satisfies* (UID)*. If* (Mult) *does not hold, then* α^*_{uid} *is independent of* r *and is the unique solution of* (P)*. If* (Mult) *holds, then the set of optimal rules is* $\big\{\alpha^*_{uid}(\cdot, \bar{w}, r)\big\}_{r\in[0,1]}$.

A key to the (UID) characterization is identifying the pattern of binding falsification-proofness constraints, which turns out to be pinned down by the matching function m. Our proof identifies the simplified problem under (UID) as the dual of the problem of transporting surplus from ineligible to eligible scores at a minimal (falsification) cost. The (UID) condition implies that the solution to this problem is deterministic and assortative, translating into the matching function. Allocation probabilities simply correspond to the shadow prices of the optimal transport problem. We discuss this proof intuition in depth in Section 3.5, while the remaining details are in Appendix B.

3.4 Optimal rules: properties

We discuss different properties of allocatively optimal rules. First, we discuss the consequence of a binding probability constraint. Second, we discuss the shape of optimal allocation rules, providing intuitions for the welfare properties that we study further in the paper. Finally, we consider comparative statics with respect to gaming ability. All results from this section are illustrated in Figure 2 for the *default setups* with a (UID) cost function. When the distinction between the (UDD) and (UID) cases is not necessary, we denote the optimal allocation rule by α^*.

Default setups. We illustrate our results using two *default setups*. In both of them, the worth function is $w(s) = s$, and the outside option is $\hat{w} = 0$, so the eligibility threshold is $\hat{s} = 0$. The score distribution is uniform, with support $[-3, 2]$ in the low-score priority default setup, and $[-2, 3]$ in the high-score priority default setup.

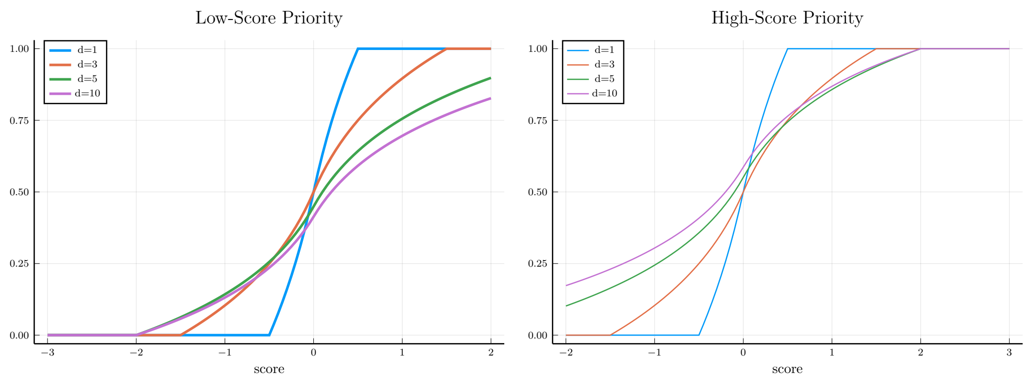

Figure 2: *Default setups, Euclidean costs with* (UID)*:* $\mathcal{C}^d(x) = \frac{\log(1+x)}{\log(1+d)}$ *resulting in an S-shape, gaming ability* $\gamma = \log(1 + d)$ *with* $d = 1, 3, 5, 10$*. Binding regime if* $d \leq 4$*, slack regime with probability gap* $\Gamma > 0$ *otherwise. Under low-score priority,* Γ *is withheld from agents and corresponds to the gap between the* $d = 5, 10$ *curves and 1 at* $\overline{s}$*. Under high-score priority,* Γ *is allocated to agents and corresponds to the gap between the* $d = 5, 10$ *curves and 0 at* $\underline{s}$*.*

The regimes of optimal rules. The optimal 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 bunching and no distortion at the boundaries of the score set: scores in $[s^*, \overline{s}]$ receive a prize with certainty, and scores in $[\underline{s}, s_*]$

receive a prize with zero probability. It is in the *slack regime* if falsification costs are low and the probability constraint is slack. In this regime, the optimal rule displays distortions both on the growth interval $[s_*(0), s^*(0)]$ and at the boundaries of the score set. Under low-score priority, there is bunching at the bottom and distortion at the top, $\alpha^*(\overline{s}) < 1$, while under high-score priority there is bunching at the top and distortion at the bottom, $\alpha^*(\underline{s}) > 0$. Under neutral priority, there is no bunching and distortions may occur at either or both boundaries, depending on the choice of the neutral gap share r.

The shape of optimal rules. We apply our characterization to the Euclidean[11] family of cost functions (Example 1). The cost function satisfies (UID) if $\mathcal{C}$ is concave, and (UDD) if it is convex. We show that optimal rules are S-shaped in the (UID) case: convex over ineligible scores and concave over eligible scores. This property boosts the probability of allocation for eligible scores and reduces it for ineligible ones. In the (UDD) case, by contrast, the optimal rule is linear since the cost of a marginal falsification is constant equal to $\mathcal{C}'(0)$.

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

For a given length of the growth interval, the S-shape concentrates growth close to the eligibility threshold, better approximating the first-best allocation rule, which is a step function at the eligibility threshold, than a linear rule. Proposition 1 suggests that the optimal falsification-proof rule may not be very costly for the designer in the (UID) case, as its shape reduces allocative distortions. This intuition is confirmed by our welfare analysis in Section 5.

Gaming ability. We parameterize the cost function by a scaling factor γ that captures *gaming ability*, 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 that higher gaming ability implies lower falsification cost. Let $\hat{\gamma}$ denote the threshold at which the optimal rule switches from the binding regime $\gamma \leq \hat{\gamma}$ to the slack regime $\gamma > \hat{\gamma}$.[12]

[11] In Online Appendix S3.1, we show that the *shifted linear* family of cost functions (Example 2) also generates S-shaped rules.

[12] Under (UDD) $\hat{\gamma} = \int_{s_*(0)}^{s^*(0)} c_{t^+}(x|x)dx$, and under (UID) $\hat{\gamma} = c\big(s^*(0)|s_*(0)\big)$.

Comparative statics. We show that increasing gaming ability 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 direction, benefiting them under high-score priority and harming them under low-score priority. The results hold for both the (UID) and the (UDD) case.

Proposition 2 (Effect of gaming ability). *Consider increasing gaming ability from γ to $\gamma' > \gamma$, and denote the corresponding optimal rules by $\alpha^*_{\gamma'}$ and α^*_{γ}. Then:*

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

(ii) ***Slack regime, low-score priority:*** *If $\gamma \geq \hat{\gamma}$, and under 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-score priority:*** *If $\gamma \geq \hat{\gamma}$, and under 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 2.

Maximum allocative surplus. The maximum allocative surplus is decreasing in gaming ability because any increase in gaming ability tightens (FPC).

Proposition 3 (Properties of the designer's value function). *The value function of (P), $W(F, c^{\gamma})$, is nonincreasing in γ.*

3.5 Proof of Theorem 2: allocation as transport

We show how to find an optimal rule in the (UID) case 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 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 agent in "location" s is endowed with an amount

$w(s) - \hat{w}$ of surplus. Equivalently, we can describe the problem in terms of masses of negative and positive surplus available at different locations. Each location s then contains a mass $|w(s) - \hat{w}|dF(s)$ of negative surplus if s is ineligible, or of positive surplus if s is eligible. We can view falsification as transporting negative surplus from ineligible locations to eligible locations that contain positive surplus.

To formalize this analogy, we change variables to index 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 contains 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

$$P(y) = \frac{\mathcal{W}(\hat{s}, \hat{w}) - \mathcal{W}(\hat{s} - y, \hat{w})}{\mathcal{W}(\hat{s}, \hat{w}) - \mathcal{W}(s_*, \hat{w})}, \qquad \text{and} \qquad Q(z) = \frac{\mathcal{W}(\hat{s}, \hat{w}) - \mathcal{W}(\hat{s} + z, \hat{w})}{\mathcal{W}(\hat{s}, \hat{w}) - \mathcal{W}(s^*, \hat{w})},$$

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)

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

To interpret 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 therefore interested in the least costly routes between Y and Z. Indeed, the program above is 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[13] 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 (FPC) binds between matched scores $s \in [s_*, \hat{s}]$ and $t = m(s)$. The unique (up to an additive constant) optimal *price functions* ϕ and ψ in closed form are also known, which allows us to solve the simplified problem on any (ZAS) interval. The formula of Theorem 2 is then obtained by applying Procedure 1. To complete the proof of the theorem, we show in Appendix B that (UID) implies that α^*_{uid} satisfies the omitted falsification-proofness constraints between scores on the same side of the eligibility threshold $\hat{s}$.

4 Price of honesty and welfare

We assess the welfare impact of honesty in the (UID) case. To do so, we compare the optimal falsification-proof mechanism to the unconstrained optimal mechanism, which is characterized in Perez-Richet and Skreta (2022). Honesty comes at a price in allocative surplus since it constrains the mechanism. In return, we show that honesty also has value, as it benefits all agents. We use theoretical results and illustrative examples to compare the magnitudes of these effects.

In this section, we fix the outside option $\hat{w}$ and the neutral gap share r. Therefore, we omit the dependence of the optimal rule α^* on these variables in the notation. As we are throughout in the (UID) case, we drop the index and denote the optimal falsification-proof rule by α^*.

The unconstrained optimal mechanism. We first describe the unconstrained allocatively optimal mechanism. For clarity, we focus on a low-score priority setting in the discussion, but the results are general. The unconstrained optimal allocation mechanism consists of a *nominal* allocation rule $\alpha^{**}(s)$ paired with an allocatively

[13]See, for example, Galichon (2018, chapter 4).

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.[14] For the mechanism to be allocatively optimal, agents must resolve this indifference by falsifying to the top if eligible and not falsifying otherwise. The optimal mechanism therefore relies on eligible agents incurring falsification costs.

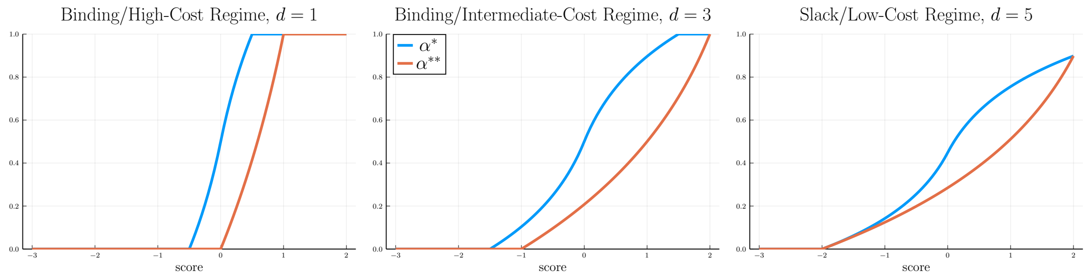

Figure 3: *Low-score priority default setup, cost* $\mathcal{C}^d(x) = \frac{\log(1+x)}{\log(1+d)}$*, gaming ability* $\gamma = \log(1 + d)$.

The unconstrained optimal rule has three regimes. It is in the *high-cost regime* if $c(\bar{s}|\hat{s}) > 1$. Then the rule achieves the first best by allocating prizes only to eligible agents; however, those with lower scores must falsify. The rule is in the *intermediate-cost regime* if $c(\bar{s}|\hat{s}) \leq 1$ but $c(\bar{s}|s_*(0)) > 1$. Then all eligible agents secure a prize by falsifying to the top score, while some ineligible agents must also receive prizes with positive probability to deter falsification. Finally, the rule is in the *low-cost regime* if $c(\bar{s}|s_*(0)) \leq 1$. The designer then reduces the maximal allocation probability under low-score priority to prevent agents below $s_*(0)$ from obtaining prizes with positive probability. Figure 3 illustrates both rules in a low-score priority setting.

The value of honesty for agents. Unconstrained optimality requires some eligible agents to falsify. However, not falsifying is also a best response to α^{**} for agents who falsify. Therefore, the equilibrium payoff of an agent under α^{**} equals her payoff without falsification, namely $\alpha^{**}(s)$. Thus, comparing $\alpha^{**}(s)$ with $\alpha^*(s)$ suffices to analyze 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 honesty regardless of score.

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

Proposition 4 (Value of honesty 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 prizes with a higher probability under the falsification-proof rule. When α^{**} is coupled with the allocatively optimal falsification strategy, eligible agents receive prizes with a lower probability under α^* but avoid the falsification costs required by the optimal mechanism. Thus, the loss of allocative surplus due to falsification-proofness may be offset by gains in agents' welfare. We proceed to quantify the relative gains and losses from honesty.

A cost-benefit analysis of honesty. 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 provide in the proof of Proposition 5. To measure the welfare impact of honesty, we compare the *gain rate of agents*, $G(\mathcal{C}) = \frac{A^*(\mathcal{C}) - A^{**}(\mathcal{C})}{A^{**}(\mathcal{C})}$, with the loss rate in allocative efficiency, $L(\mathcal{C}) = \frac{D^{**}(\mathcal{C}) - D^*(\mathcal{C})}{D^{**}(\mathcal{C})}$. We interpret $G(\mathcal{C})$ as the *value of honesty*, and $L(\mathcal{C})$ as the *price of honesty.*

Controlling for gaming ability, we can make the trade-off arbitrarily favorable to honesty. We control for gaming ability by fixing the maximum amount of falsification an agent would rationally undertake, $d_{\mathcal{C}}$ (defined by $\mathcal{C}(d_{\mathcal{C}}) = 1$) at some value d. The nature of the trade-off depends on how d compares to the size of the interval of eligible scores, $\hat{d} = \overline{s} - \hat{s}$, which is the threshold marking the frontier between the intermediate-cost 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, consider the role of returns to scale in falsification. Higher returns to scale correspond to a more concave cost function, which can be achieved while keeping $d_{\mathcal{C}}$ constant. Suppose we are in the intermediate-cost regime with low-score priority for simplicity. A more concave cost makes α^{**} more convex on the same growth interval, reducing agents' payoffs while pushing the designer closer to first-best. This is because making the rule more convex increasingly lowers the probability of receiving a prize for ineligible agents, while eligible agents continue to falsify to obtain a prize. At the same time, α^* approaches first-best because greater concavity increases the convexity of the rule over ineligible scores and its concavity over eligible scores, allowing the S-shaped rule to get closer to the first-best step function.[15] Hence, Proposition 5 can be read as saying that honesty yields large gains to agents at a small allocative price when returns to scale are strongly increasing.

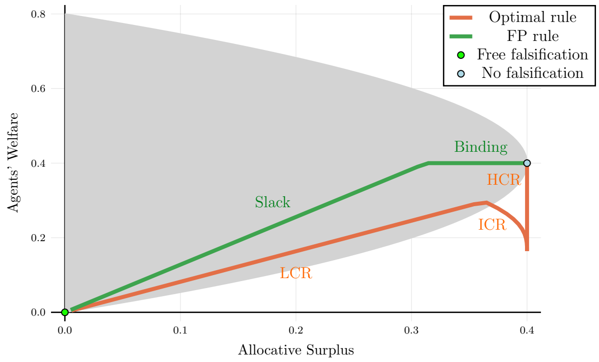

Figure 4: *Payoff trajectories. Low-score priority default setup, cost* $C^d(x) = \frac{\log(1+x)}{\log(1+d)}$*, gaming ability* $\gamma = \log(1+d) : 0 \to \infty$*. The gray zone depicts the set of attainable allocative payoffs (without accounting for falsification costs).*

The condition $d < \hat{d}$ characterizes the high-cost regime for α^{**}, in which the designer achieves the 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 a prize 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 a prize with certainty without falsifying, so the aggregate payoff of agents under α^{**} converges to $1 - F(\hat{s} + d)$.

Quantitative illustration. We illustrate Proposition 5 using the default setups under different Euclidean cost functions. Figure 4 shows the trajectories of allocative surplus and agents' welfare under both rules when varying gaming ability $\gamma = \log(1+d)$ for a family of Euclidean cost functions $C^d(x) = \frac{\log(1+x)}{\log(1+d)}$. With infinite gaming ability, both rules never allocate (leftmost point in both trajectories). As γ approaches 0, both rules converge to the first-best (top-right point in both trajectories). At these extremes, there is no trade-off. For intermediate gaming ability, we see that the

[15]See Online Appendix S3.2 for comparative statics results on the effect of returns to scale

optimal rule can be very costly to agents for a moderate gain in allocative surplus. This is especially true at the frontier between the high-cost and intermediate-cost regimes.

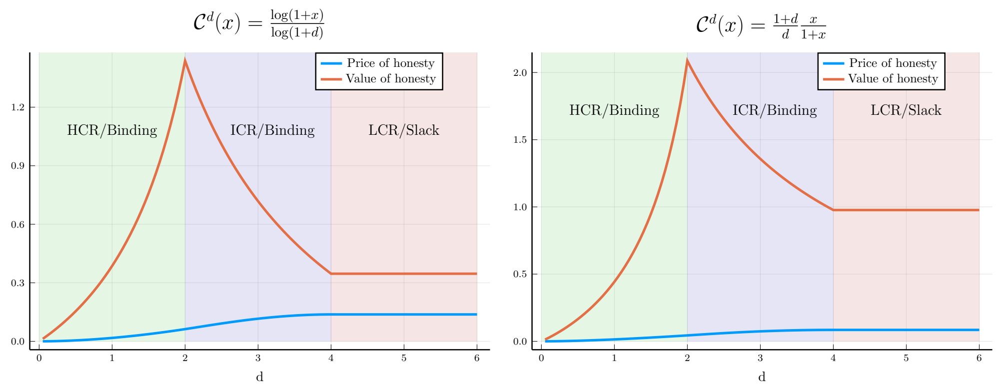

Figure 5: *Welfare impact of honesty. Low-score priority default setup, $d : 0 \to 6$.*

We quantify the trade-off by plotting the value and the price of honesty, as measured in Proposition 5, for two fixed families of cost functions with varying gaming abilities. This exercise is shown in Figure 5. Under both cost functions, the value of honesty dominates its price for all gaming ability levels. The peak in the relative value is reached at the frontier between the high-cost and intermediate-cost regimes. Under the first cost function (left panel), the trade-off is then a 144% value for a 6.25% price of honesty, while under the second cost function (right panel), it is a 209% value for a 4.4% price. The figure also illustrates the role of returns to scale, as the value is higher and the price lower at every d under the more concave cost function in the right panel.

Proposition 4 and Proposition 5 suggest a theoretical foundation for falsification-proofness. Beyond the externality motive, requiring honesty can be an effective way to balance allocative surplus and agents' welfare. It can, under some configurations, generate large gains for agents at a moderate cost in allocative surplus.

5 A welfare foundation for falsification-proofness

We show that honesty can emerge as an unconstrained solution to the weighted welfare problem (WWP) if the planner places sufficient weight on externalities or agents' welfare. Theorem 3 identifies explicit conditions under which honesty is optimal and shows that either concern can, on its own, be sufficient.

Falsification technology. The analysis in this section assumes (UID) throughout. However, we also need to strengthen our assumptions on the falsification technology to handle downward falsification. Specifically, we assume that the cost function satisfies the following *downward triangular inequality*:

$$\forall t < m < s, \quad c(t|m) + c(m|s) \geq c(t|s). \tag{DTRI}$$

Combined with (UID), this assumption ensures that the cost function satisfies the standard triangular inequality (see Lemma C.1).

Optimality of honesty. Recall that, in the weighted welfare problem (WWP), the planner maximizes

$$\Omega(\alpha, \phi, \xi, \zeta) = \iint \alpha(t)\big\{w(s) - \hat{w} + \zeta\big\}\, d\phi(t|s)dF(s) - (\xi + \zeta) \iint c(t|s)\, d\phi(t|s)dF(s)$$

subject to (PC) and (IC).

The first term of this objective is an allocative surplus term with a downward-shifted outside option $\hat{w} - \zeta$. This induces a downward shift of the eligibility threshold to $\hat{\sigma}(-\zeta)$, where $\hat{\sigma}(\cdot)$ is the increasing function $\hat{\sigma}(x) = w^{-1}(\hat{w}+x)$, with the convention that $\hat{\sigma}(x) = \overline{s}$ if $\hat{w}+x \geq w(\overline{s})$, and $\hat{\sigma}(x) = \underline{s}$ if $\hat{w}+x \leq w(\underline{s})$. Note that $\hat{\sigma}(x)$ is simply the eligibility threshold associated with a shift x of the outside option. In particular, $\hat{\sigma}(0) = \hat{s}$.

Then a sufficient condition for honesty to be optimal is

$$c\Big(\hat{\sigma}(\xi) \,|\, \hat{\sigma}(-\zeta)\Big) \geq 1. \tag{H}$$

Theorem 3 (Weighted welfare optimality of honesty)**.** *Under* (UID) *and* (DTRI), (H) *implies that the optimal falsification-proof rule* $\alpha^*(\cdot, \hat{w} - \zeta, r)$ *solves* (WWP)*.*

The *honesty condition* (H) allows us to isolate the factors that can contribute to making honesty welfare-optimal. First, a *lower gaming ability*. Considering a family $c^\gamma(t|s) = \frac{1}{\gamma}c(t|s)$ of cost functions as in Section 3.4, it is clear that decreasing gaming ability γ increases the cost of falsification and makes (H) more likely to hold. Second, a *lower correlation* between score and worth. We can think of this correlation as captured by the slope of w. If $w(s) = \omega s$ is linear, for example, so ω is the coefficient of the score in the linear regression of the worth, then the left-hand side of (H) becomes $c(\hat{s} + \xi/\omega \,|\, \hat{s} - \zeta/\omega)$, which increases as ω decreases. Third, *higher welfare weights* on agents and externalities. The expression of (H) shows that both weights contribute towards honesty. The two rationales are substitutes by (UID), and each of them can, on its own, be sufficient. Concern for externalities alone ($\zeta = 0$) yields honesty whenever $c\big(\hat{\sigma}(\xi)|\hat{s}\big) \geq 1$. Concern for agents' welfare alone ($\xi = 0$) yields honesty whenever $c\big(\hat{s}|\hat{\sigma}(-\zeta)\big) \geq 1$.

Proof overview. The proof has three main steps. First, we show that it is without loss of optimality to restrict attention to allocation rules that satisfy (FPC) and are top-flat in the sense that they allocate with certainty for all scores s such that $c\big(s|\hat{\sigma}(-\zeta)\big) \geq 1$. Note that falsification-proofness does not imply honesty: in particular, optimality for (WWP) might require agents to falsify even though honesty is a best response. Second, we introduce a falsification budget $\Phi \geq 0$, and let $Z(\Phi)$ be the value function of the weighted welfare problem with the additional constraint that the aggregate cost of falsification must be equal to the falsification budget Φ. If $\Phi = 0$, this program is exactly (P) with the shifted outside option $\hat{w} - \zeta$. If Φ is equal to the aggregate cost of falsification when all eligible scores below the cutoff falsify to the cutoff under α^{**}, then this program is the unconstrained allocative surplus optimization problem with the shifted outside option. Then, (WWP) reduces to finding the falsification budget Φ that solves $\max_{\Phi \geq 0} \big[Z(\Phi) - (\xi + \zeta)\Phi\big]$. Third, we show that the top-flatness property implies that, for any $\Phi' > \Phi$, $Z(\Phi') - Z(\Phi) \leq \big(w(s_\zeta^{**}) - \hat{w} + \zeta\big)\big(\Phi' - \Phi\big)$, where s_ζ^{**} solves $c\big(s_\zeta^{**}|\hat{\sigma}(-\zeta)\big) = 1$. Then $\Phi = 0$ (honesty) must be optimal whenever $w(s_\zeta^{**}) - \hat{w} + \zeta \leq \xi + \zeta$, or equivalently $s_\zeta^{**} \leq \hat{\sigma}(\xi)$. By definition of s_ζ^{**}, this is equivalent to (H).

A falsification-proofness principle (without honesty). An interesting byproduct of our proof of Theorem 3 is a new falsification-proofness principle (Proposition C.1): we prove that it is without loss of generality to restrict attention to allocation rules that satisfy (FPC) in the sense that we can make any allocation rule falsification-proof without altering the original optimal falsification strategy or the final allocation. However, falsification-proofness does not imply honesty, and inducing falsification might actually be optimal in (WWP) if (H) does not hold. The falsification-proofness principle is a useful technical analog of the revelation principle, reducing the space of mechanisms that must be considered for cost functions that satisfy the triangular inequality.

Changing the social cost of falsification. For simplicity, we chose to measure the social cost of falsification assuming $D(t, s) = c(t|s)$. The general result of Theorem 3 does not depend on this assumption, but the exact expression of the honesty condition does. First, note that the condition $c\big(\hat{s}|\hat{\sigma}(-\zeta)\big) \geq 1$ is always sufficient for honesty since it yields honesty even in the absence of a concern for externalities. We can do better if we can find a scalar $\beta > 0$ such that, for every s, t, $D(t, s) \geq \beta c(t|s)$. Then the honesty condition becomes $c\big(\hat{\sigma}(\beta\xi) \,|\, \hat{\sigma}(-\zeta)\big) \geq 1$.

6 Resource constraint

If $\bar{\rho} < 1$, the designer faces a resource constraint (RC): $\int_S \alpha(s)dF(s) \leq \bar{\rho}$. Without loss of generality, we normalize the true value of the outside option to 0. The designer's problem with a resource constraint is

$$\max_{\alpha} \int_S w(s)\alpha(s)dF(s) \quad \text{s.t.} \quad (\text{PC}), (\text{FPC}), (\text{RC}). \tag{RCP}$$

Let $\hat{w}$ be the Lagrange multiplier on the resource constraint in (RCP). The Lagrangian for (RCP) is then

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

Maximizing this Lagrangian for a given value of $\hat{w}$ is therefore equivalent to solving (P) with outside option $\hat{w}$. Solving (RCP) then amounts to identifying the value of $\hat{w}$ for which (RC) holds.

Specifically, we show in Appendix D and Proposition D.1 that the total mass of prizes allocated under $\alpha^*(\cdot, \hat{w}, r)$ is smoothly decreasing in the outside option $\hat{w}$ and, when (Mult) holds, increasing in the neutral gap share r. Thus, these two variables can be used to shift the total mass of allocated prizes smoothly. Following a classical Lagrangian approach, this implies that the solution to (RCP) is obtained as follows. If $\alpha^*(\cdot, 0, r)$ satisfies the resource constraint for at least some $r \in [0, 1]$, then $\alpha^*(\cdot, 0, r)$ solves (RCP).

Otherwise, we must find values of $(\hat{w}, r)$ such that the mass allocated under $\alpha^*(\cdot, \hat{w}, r)$ is exactly $\bar{\rho}$. In Theorem D.1, we show that there exists a unique value $\hat{w}(\bar{\rho})$ of the outside option and, under (Mult), a unique value of the neutral gap share $r(\bar{\rho})$ such that the resulting rule allocates mass $\bar{\rho}$.

Hence, the solution to (RCP) is $\alpha^*(\cdot, \hat{w}, r)$ for particular values of $(\hat{w}, r)$ pinned down by the resource constraint. It is therefore useful to understand the effects of $\hat{w}$ and r on the level and shape of the optimal rule. We show in Proposition D.1 that increasing $\hat{w}$ or decreasing r shifts α^* downward.

The presence of a resource constraint affects comparative statics because changes in gaming ability have both direct and indirect effects on the optimal allocation rule. The direct effect is the one analyzed in Proposition 2. If, holding $(\hat{w}, r)$ fixed, this direct effect changes the total mass of allocated prizes, we may need to adjust $(\hat{w}, r)$ to satisfy the resource constraint. The indirect effect therefore shifts the solution to (RCP) upward or downward. Its direction typically depends on the shape of the score distribution.

It is then straightforward to see how indirect effects modify the result of Proposition 2. In the *binding regime*, the result continues to hold, although the crossing point typically changes. In the other two cases, where the direct effect induces an unambiguous change in allocated mass, the indirect effect may offset that change, leading again to single-crossing allocation rules, as in the binding regime. While indirect effects modify comparative statics, they do not alter the forces behind Proposition 5. Hence, it remains true that strongly increasing returns to scale imply that the optimal falsification-proof rule can generate large gains to agents at a small allocative cost to the designer.

7 Related literature

We contribute to the literature on optimal allocation mechanisms with privately informed agents. 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 weighted combination of utilitarian and revenue objectives, as in Akbarpour, Dworczak and Kominers (2024). Both studies establish conditions under which the designer optimally refrains from using transfers altogether. We also consider a general objective and exclude transfers. While there may be exogenous reasons to rule out transfers, Condorelli (2013) and Akbarpour et al. (2024) show that doing so is optimal if the designer's utility from allocation is negatively correlated with willingness to pay. In contrast to these works, we study a setting in which agents' private information can be manipulated at a cost. Methodologically, our approach differs from standard techniques based on virtual surplus that 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 mechanisms that studies the use of alternative targeting tools in lieu of transfers. In our setting, 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), 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 it differs in that the cost of falsification is type-dependent. In a related framework, Augias and Perez-Richet (2024) study the optimal design of allocation mechanisms when agents can improve their scores.

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) considers 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 in which 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 show that this constraint may lead to a loss, without characterizing the optimal contract. We build on Perez-Richet and Skreta (2022), where we show that allocatively optimal mechanisms necessarily induce falsification.

Finally, we add to the growing list of papers using optimal transport in economics, surveyed in Carlier (2012) and Galichon (2018). More recently, it 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), and labor-market sorting problems (Boerma, Tsyvinski and Zimin, 2021). Most of these papers rely on duality-based characterizations in optimal transport.[16] In contrast, our approach relies di-

[16]For example, Lin and Liu (2024) rely on properties characterizing optimal coupling for given

rectly on the dual of the optimal transport problem, which we show to be equivalent to our problem. 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 provides an original solution method that differs both 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).

A Lagrangian for the differential problem

Lemma A.1. *An allocation rule α solves* (P) *if and only if it satisfies* (PC), (MON) *and* (Init)*, and there exists a Lagrange multiplier $\nu \geq 0$ such that:*

(i) ***Growth interval:*** *$\alpha(s) = \alpha(\underline{s})$ if $s < s_*(\nu)$, and $\alpha(s) = \alpha(\overline{s})$ if $s > s^*(\nu)$;*

(ii) ***Continuity:*** *α is continuous at $s_*(\nu)$ and at $s^*(\nu)$;*

(iii) ***Relaxed*** (FPC)***:*** *$\alpha(t) - \alpha(s) \leq c(t|s)$ for every $s_*(\nu) \leq s < t \leq s^*(\nu)$;*

(iv) ***Complementary slackness on*** (PC)***:*** *If $\bar{w} \leq \hat{w}$, $\nu\big(1 - \alpha(\overline{s})\big) = 0$; if $\bar{w} \geq \hat{w}$, $\nu\alpha(\underline{s}) = 0$;*

(v) ***Optimality for the simplified program:*** *For every allocation rule $\hat{\alpha}$ that satisfies (iii),*

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

Proof. We proceed in two steps. The first step is a standard Lagrangian necessity and sufficiency theorem. The second step shows that 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 satisfies (MON), (FPC), and (Init). It is immediate to verify that the set of such feasible allocation rules, which we denote by $\mathbb{A}$, is convex.

Throughout, we let $[x]_y^z = y\,\mathbb{1}_{x<y} + x\,\mathbb{1}_{y\leq x\leq z} + z\,\mathbb{1}_{x>z}$ for any x and $y < z$.

Step 1: *An allocation rule $\alpha \in \mathbb{A}$ solves* (P) *if and only if (A) α satisfies* (PC)*; and there exists $\nu \geq 0$ such that (B) $\nu = 0$ or $\overline{\alpha} - \underline{\alpha} = 1$, and (C) $\mathcal{L}(\alpha, \nu) \geq \mathcal{L}(\hat{\alpha}, \nu)$ for every allocation rule $\hat{\alpha} \in \mathbb{A}$.*

---

marginals to establish their characterization of stable credible signals.

$\boxed{\Leftarrow}$ For any $\hat{\alpha} \in \mathbb{A}$ that satisfies (PC),

$$\int_S \mathcal{W}(z,\hat{w})d\alpha(z) = \mathcal{L}(\alpha,\nu) \geq \mathcal{L}(\hat{\alpha},\nu) = \int_S \mathcal{W}(z,\hat{w})d\hat{\alpha}(z) + \nu\left(1 - \int_S d\hat{\alpha}(z)\right)$$
$$\geq \int_S \mathcal{W}(z,\hat{w})d\hat{\alpha}(z),$$

where (B) implies the first equality, (C) implies the first inequality, and the last inequality holds because $\hat{\alpha}$ satisfies (PC) and $\nu \geq 0$.

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

$$h(b) = \max_{\tilde{\alpha}\in\mathbb{A}} \int_S \mathcal{W}(z,\hat{w})d\tilde{\alpha}(z) \text{ s.t. } g(\tilde{\alpha}) \leq b.$$

Since the objective and the constraint $g(\cdot)$ are both linear in $\tilde{\alpha}$, 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, $h(1) = \int_S \mathcal{W}(z,\hat{w})d\alpha(z)$. If $g(\alpha) = 1$, we also have $\int_S \mathcal{W}(z,\hat{w})d\alpha(z) = \mathcal{L}(\alpha,\nu)$. Otherwise, we must have $g(\alpha) < 1$. But then α must also solve the program for any $b \in [g(\alpha), 1]$, implying that h is constant on this interval, and $\nu = 0$. Then again, $\int_S \mathcal{W}(z,\hat{w})d\alpha(z) = \mathcal{L}(\alpha,\nu)$. For all $\hat{\alpha} \in \mathbb{A}$, $\int_S \mathcal{W}(z,\hat{w})d\hat{\alpha}(z) \leq h\big(g(\hat{\alpha})\big)$, by definition of h, and $h\big(g(\hat{\alpha})\big) \leq h(1) + \nu\big(g(\hat{\alpha}) - 1\big)$, by concavity of h. Hence,

$$\mathcal{L}(\hat{\alpha},\nu) = \int_S \mathcal{W}(z,\hat{w})d\hat{\alpha}(z) - \nu\big(g(\hat{\alpha}) - 1\big) \leq h(1) = \int_S \mathcal{W}(z,\hat{w})d\alpha(z) = \mathcal{L}(\alpha,\nu).$$

Step 2: *An allocation rule α and a scalar $\nu \geq 0$ satisfy* (MON), (PC), (Init), *and (i)-(v) if and only if α and ν satisfy $\alpha \in \mathbb{A}$, (A), (B) and (C).*

$\boxed{\Rightarrow}$ (A) is immediate. (B) must hold for the same scalar ν by (Init) and (iv). Next, we show that (i)-(iii) imply that α satisfies (FPC), so that $\alpha \in \mathbb{A}$ since it already satisfies (MON) and (Init). Let $s < t$, $s' = \big[s\big]_{s_*(\nu)}^{s^*(\nu)}$ and $t' = \big[t\big]_{s_*(\nu)}^{s^*(\nu)}$. Then $\alpha(t) - \alpha(s) = \alpha(t') - \alpha(s') \leq c(t'|s') \leq c(t|s)$, where the first equality is from (i) and (ii), the first inequality is from (iii), and the last inequality is by cost monotonicity. Hence, $\alpha \in \mathbb{A}$ satisfies (A) and (B).

Let $\hat{\alpha} \in \mathbb{A}$. Then, let $\tilde{\alpha}(s) = \big[\hat{\alpha}(s) + a\big]_0^1$, where $a = -\hat{\alpha}\big(s_*(\nu)\big)\,\mathbb{1}_{\bar{w}<\hat{w}} + \big(1 -$

$\hat{\alpha}\big(s^*(\nu)\big)\big)\,\mathbb{1}_{\bar{w}\geq\hat{w}}$. Then $\tilde{\alpha}$ satisfies (iii), and (C) follows from:

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

where we use Lemma 1 (ii), and IBP means integration by parts.

$\boxed{\Leftarrow}$ $\alpha \in \mathbb{A}$ directly implies (MON), (Init), (ii) and (iii). (A) directly implies (PC). (B) and (Init) imply (iv).

Next, we show that (C) implies (i). Suppose α does not satisfy (i). Then α cannot satisfy (C). To show this, suppose we are in a low-score priority setting, so feasibility requires $\alpha(\underline{s}) = 0$ (the high-score priority case is analogous). Suppose first $\alpha\big(s_*(\nu)^-\big) \equiv \lim_{s\to s_*(\nu), s<s_*(\nu)} \alpha(s) > \alpha(\underline{s}) = 0$. Then consider $\hat{\alpha}(s) = \max\big\{\alpha(s) - \alpha\big(s_*(\nu)^-\big), 0\big\}$. Then we have

$$\mathcal{L}(\hat{\alpha},\nu) - \nu = \int_{s_*(\nu)}^{\overline{s}} \big[\mathcal{W}(z,\hat{w})-\nu\big]d\alpha(z) > \int_{\underline{s}}^{\overline{s}} \big[\mathcal{W}(z,\hat{w})-\nu\big]d\alpha(z) = \mathcal{L}(\alpha,\nu)-\nu,$$

since $\mathcal{W}(z,\hat{w}) < \nu$ on $[\underline{s}, s_*(\nu))$. Similarly, if $\alpha\big(s^*(\nu)^+\big) < \alpha(\overline{s})$, a rule $\hat{\alpha}(s) = \alpha(s)\,\mathbb{1}_{s\leq s^*(\nu)} + \alpha\big(s^*(\nu)^+\big)\,\mathbb{1}_{s>s^*(\nu)}$ would yield $\mathcal{L}(\hat{\alpha},\nu) > \mathcal{L}(\alpha,\nu)$, contradicting (C).

Next, we show (v). By the same construction as in Lemma 2, any allocation rule that is nonmonotonic is dominated by a monotonic allocation rule for the simplified program. Consider any allocation rule $\hat{\alpha}$ that is monotonic on $[s_*(\nu), s^*(\nu)]$ and satisfies (iii). Since the simplified program only considers scores in $[s_*(\nu), s^*(\nu)]$, we can assume that $\hat{\alpha}$ is constant on $[\underline{s}, s_*(\nu))$, and on $(s^*(\nu), \overline{s})$. Then,

$$\begin{aligned}
\int_{s_*(\nu)}^{s^*(\nu)} \hat{\alpha}(s)\{w(s)-\hat{w}\}dF(s) &= \int_{s_*(\nu)}^{s^*(\nu)} \big[\mathcal{W}(z,\hat{w})-\nu\big]d\hat{\alpha}(z) && \text{(by IBP)}\\
&= \int_S \big[\mathcal{W}(z,\hat{w})-\nu\big]d\hat{\alpha}(z) = \mathcal{L}(\hat{\alpha},\nu)-\nu\\
&\leq \mathcal{L}(\alpha,\nu)-\nu && \text{(by (C))}\\
&= \int_{s_*(\nu)}^{s^*(\nu)} \big[\mathcal{W}(z,\hat{w})-\nu\big]d\alpha(z) && \text{(by (i))}\\
&= \int_{s_*(\nu)}^{s^*(\nu)} \alpha(s)\{w(s)-\hat{w}\}dF(s). && \text{(by IBP)}
\end{aligned}$$

□

B 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, $\mathcal{W}(\overline{s},\hat{w}) = \mathcal{W}^+(\overline{s},\hat{w}) = 0$, while, in the high-score priority case, $\mathcal{W}(\underline{s},\hat{w}) = \mathcal{W}^-(\underline{s},\hat{w}) = 0$. □

Proof of Lemma 2. Suppose α is feasible but not monotonic. Consider 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 lower nondecreasing envelope of α on $[\underline{s}, \hat{s}]$, and $\alpha^+ : [\hat{s}, \overline{s}] \to [0,1]$ is the higher nondecreasing envelope of α on $[\hat{s}, \overline{s}]$. Then $\tilde{\alpha}$ 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 obtains by definition of s' and t', and the second equality 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 scores less likely to get a prize under $\tilde{\alpha}$ than under α. Hence, $\tilde{\alpha}$ dominates α for (P). Furthermore, if α is not monotonic, there must exist an interval of scores for which α and $\tilde{\alpha}$ do not coincide. Since F has full support, $\tilde{\alpha}$ is therefore strictly better than α. □

Proof of Lemma 3. By Lemma 2, imposing (MON) in (P) is without loss. We can then integrate the objective by parts as in ($\overline{\text{DOF}}$) and ($\underline{\text{DOF}}$) so optimizing either objective under (PC), (FPC) and (MON) is equivalent to solving (P). Then using ($\overline{\text{DOF}}$) under high-score priority and ($\underline{\text{DOF}}$) under low-score priority shows that choosing α to satisfy (Init) strictly increases the objective function. It also relaxes the probability constraint (PC) which can be equivalently written as

$$0 \leq \underline{\alpha}, \text{ and } \underline{\alpha} + \int_S d\alpha(z) \leq 1, \qquad \text{or} \qquad \overline{\alpha} \leq 1, \text{ and } \overline{\alpha} - \int_S d\alpha(z) \geq 0.$$

Given that α must satisfy (Init) and eliminating the constant $\bar{w} - \hat{w}$ from the objective under high-score priority, we obtain the equivalent program (DP). □

Proof of Theorem 1. To keep notations simple, we only indicate the dependence of α^*_{udd} on $\hat{w}, r$ when it is useful for the argument. We only need to check that the conditions of Lemma A.1 are satisfied. Picking $\nu = \mathcal{W}(s_*, \hat{w})$, it is clear that (i) holds and that α^*_{udd} satisfies (MON), (Init), (PC) and (ii). For (iii), let $s_* \leq s < t \leq s^*$, then

$$\alpha^*_{udd}(t) - \alpha^*_{udd}(s) = \int_s^t c_{t^+}(x|x)dx \leq \int_s^t c_t(x|s)dx = c(t|s),$$

where the inequality follows from (UDD). This also shows that the first-order approach is valid. Point (iv) is immediate to check. Furthermore, the differential program solved by α^*_{udd} is obtained from the program in (v) by using integration by parts on the objective function and absolute continuity of α. Therefore, to show that (v) holds, we only need to prove the claim that (FPC) implies that a feasible allocation rule α must be absolutely continuous. For any $s < t$, (FPC) implies $\alpha(t) - \alpha(s) \leq c(t|s) = \int_s^t c_t(x|s)dx$ by the regularity assumption on c. (UDD) implies that this is bounded above by $\int_s^t c_t(x|\underline{s})dx \leq (t - s)\max_{x \in S} c_t(x|\underline{s})$. Hence, α must be Lipschitz.

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 (Mult) does not hold. If, instead, (Mult) holds, the designer is indifferent between the allocation rules $\alpha^*_{udd}(s, \hat{w}, r)$ for any $r \in [0, 1]$. The argument is the same as in the (UID) case. $\square$

Proof of Theorem 2. 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.1 are satisfied. We pick the multiplier $\nu = \mathcal{W}(s_*, \hat{w}) \geq 0$. Then α^*_{uid} clearly satisfies (i). It is easy to check that it satisfies (MON), (Init) and (PC), as well as (ii). To see that it satisfies (iv), note that $\alpha^*_{uid}(s^*) - \alpha^*_{uid}(s_*) = c(s^*|s_*)$ since (FPC) 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 3.3, α^*_{uid} solves the relaxed program of that section. To show that it satisfies (v), we need to show that it satisfies (iii) 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 ν that satisfies (B), and hence satisfies the necessary and sufficient conditions of Lemma A.1. Then for this ν and the corresponding bounds (s_*, s^*), the solution to the optimal transport problem is uniquely determined up to a constant. 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. Uniqueness fails only if we are in the neutral priority case where $\bar{w} = \hat{w}$ and the probability constraint is slack. In this case, $(s_*, s^*) = (s_*(0), s^*(0)) = (\underline{s}, \overline{s})$. 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 $(r' - r)\Gamma_{uid} \int_{\underline{s}}^{\overline{s}} \{w(s) - \hat{w}\} dF(s) = (r' - r)\Gamma_{uid}(\bar{w} - \hat{w}) = 0$. $\square$

Proof of Proposition 1. Let α denote the optimal rule. Using the formulas from Theorem 2 and Theorem 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. $\square$

Proof of Proposition 2. 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

$$\delta'(s) = \begin{cases} \left(\frac{1}{\gamma} - \frac{1}{\gamma'}\right) c_s\big(m(s)|s\big) < 0 & \text{if } s \in \big[s_*[\gamma], \hat{s}\big], \\ \left(\frac{1}{\gamma'} - \frac{1}{\gamma}\right) c_t\big(m(s)|s\big) < 0 & \text{if } s \in \big[\hat{s}, s^*[\gamma]\big]. \end{cases}$$

If, instead, the cost function satisfies (UDD), its derivative on $\big[s_*[\gamma], s^*[\gamma]\big]$ is

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

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). $\square$

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). 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 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: (1) 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 $I(\hat{w},r)$. (2) 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. (3) 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,

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

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

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

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

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 high-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 $[\underline{s}, \bar{s}]$ for α^{**}. Furthermore,

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

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

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

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 $s_*(0) = \underline{s}$, $s^*(0) \leq \bar{s}$, and $\alpha^*\big(\underline{s}\big) = \Gamma$. The growth interval of α^{**} is such that $s_- > \underline{s}$, and $s^+ = \bar{s}$. 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

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

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.

$\square$

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 $\bar{d} = \bar{s} - \underline{s}$, 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 we recall that $\hat{d} = \bar{s} - \hat{s}$ is the size of the interval of 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 within 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 $x \mapsto x + \phi(x)$. Intuitively, to a certain length of a (ZAS) interval, it associates its upper bound. Finally, for $x \in [\hat{s}, s^*(0)]$,

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

By the implicit function theorem and the (ZAS) equation, ϕ' exists almost everywhere and satisfies $\phi'(x) = -\frac{\big(w(x) - \hat{w}\big)f(x)}{\big(w(-\phi(x)) - \hat{w}\big)f(-\phi(x))}$.

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.

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

where the first equality follows from integration by parts, the second from the characterization of optimal 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' aggregate 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 $[s_*, s^*]$

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

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. 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 x to $y = s_+ - x$.

This yields the following 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 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. 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. $\square$

C Proofs of Section 5

As a preliminary, we show that the cost function satisfies the standard triangular inequality

$$\forall s, m, t, \qquad c(t|m) + c(m|s) \geq c(t|s) \tag{TRI}$$

Lemma C.1 (Triangular inequality)**.** *If c satisfies* (DTRI) *and* (UID)*, then it satisfies* (TRI)*.*

Proof. (UID) proves the claim for $s < m < t$, and (DTRI) for $t < m < s$, while cost monotonicity proves it for all other configurations. $\square$

First, we show that we can, without loss of generality, restrict attention to falsification-proof allocation rules.

Proposition C.1 (Falsification-proofness principle)**.** *Assume* (TRI)*. Let α be an allocation rule, and ϕ a best-response of the agents (i.e. ϕ satisfies* (IC)*). Then there exists an allocation rule $\tilde{\alpha}$ that satisfies: (i) falsification-proofness* (FPC)*; (ii) ϕ is a best response to $\tilde{\alpha}$; (iii) for every s, the equilibrium allocation probability is unchanged, $\int \alpha(t) d\phi(t|s) = \int \tilde{\alpha}(t) d\phi(t|s)$.*

Proof. Let $u(s) = \max_t \alpha(t) - c(t|s)$ be the value function of the agents under α. Then we define the new allocation rule $\tilde{\alpha}(s) = u(s)$. It satisfies $\tilde{\alpha}(s) = \alpha(s)$ if $s \in T(\alpha, s)$, and $\tilde{\alpha}(s) > \alpha(s)$ otherwise.

Claim 1: *If $s \notin T(\alpha, s)$, then for every $s' \in S$, $s \notin T(\alpha, s')$.*

Let $t \in T(\alpha, s)$. Since $s \notin T(\alpha, s)$, $\alpha(t) - c(t|s) > \alpha(s)$. Suppose, by contradiction, that $s \in T(\alpha, s')$ for some $s' \neq s$. Then, $\alpha(s) - c(s|s') \geq \alpha(t) - c(t|s')$. Combining these two inequalities, we obtain $c(t|s) + c(s|s') < c(t|s')$, which contradicts (TRI) .

In particular, claim 1 implies that for any score t that is a target under α, so $t \in \cup_{s \in S} T(\alpha, s)$, we have $\tilde{\alpha}(t) = \alpha(t)$.

Claim 2: $T(\alpha, s) \cup \{s\} \subseteq T(\tilde{\alpha}, s)$.

Let $t \in T(\alpha, s)$, so $\alpha(t) = \tilde{\alpha}(t)$ by Claim 1. Consider any $t' \neq s$. If t' is a target under α, then $\tilde{\alpha}(t') = \alpha(t')$ by Claim 1, and $\alpha(t) - c(t|s) \geq \alpha(t') - c(t'|s)$ (since $t \in T(\alpha, s)$) implies $\tilde{\alpha}(t) - c(t|s) \geq \tilde{\alpha}(t') - c(t'|s)$. Then suppose that t' is not a target. Then, there exists t'' such that $\tilde{\alpha}(t') = \alpha(t'') - c(t''|t')$. Then, $t \in T(\alpha, s)$ and (TRI) imply

$$\tilde{\alpha}(t) - c(t|s) = \alpha(t) - c(t|s) \geq \alpha(t'') - c(t''|s) = \tilde{\alpha}(t') + c(t''|t') - c(t''|s) \geq \tilde{\alpha}(t') - c(t'|s).$$

Next, we show that $s \in T(\tilde{\alpha}, s)$. Suppose $s \notin T(\alpha, s)$ since we can otherwise use the same argument as above. Consider any $t' \neq s$. If t' is a target under α, then $\tilde{\alpha}(t') = \alpha(t')$ and $\tilde{\alpha}(s) = u(s) \geq \alpha(t') - c(t'|s) = \tilde{\alpha}(t') - c(t'|s)$. Suppose instead that t' is not a target under α. Then, there exists t'' such that $\tilde{\alpha}(t') = \alpha(t'') - c(t''|t')$. Then, (TRI) implies

$$\tilde{\alpha}(s) = u(s) \geq \alpha(t'') - c(t''|s) = \tilde{\alpha}(t') + c(t''|t') - c(t''|s) \geq \tilde{\alpha}(t') - c(t'|s).$$

Conclusion: Claim 2 implies that ϕ is a best-response to $\tilde{\alpha}$ since, for every s, $T(\alpha, s) \subseteq T(\tilde{\alpha}, s)$. It also implies that $\tilde{\alpha}$ satisfies the falsification-proofness constraint since, for every s, $s \in T(\tilde{\alpha}, s)$. Finally, (iii) holds since $\alpha(t) = \tilde{\alpha}(t)$ for every $t \in T(\alpha, s)$, and $\operatorname{supp} \phi(\cdot|s) \subseteq T(\alpha, s)$ by (IC). $\square$

We define $s^{**}_\zeta = \max\{t \in [\hat{s}_\zeta, \overline{s}] : c(t|\hat{\sigma}(-\zeta)) \leq 1\}$. It is the cutoff above which the unconstrained optimal rule α^{**} under the shifted outside option allocates with maximum probability. Then, s^{**}_ζ is decreasing in ζ and $s^{**}_\zeta < \overline{s}$ if and only if $c(\overline{s}|\hat{s}_\zeta) > 1$. We show that we can restrict attention to allocation rules that are flat above s^{**}_ζ.

Lemma C.2 (Top-Flatness)**.** *Assume* (TRI)*. Let α be an allocation rule, and ϕ a best-response to α. Then there exists an allocation rule $\tilde{\alpha}$ that satisfies* (FPC) *and is flat above s^{**}_ζ, and a best-response $\tilde{\phi}$ to $\tilde{\alpha}$ such that: (i) $\iint \tilde{\alpha}(t)w(s)d\tilde{\phi}(t|s)dF(s) \geq \iint \alpha(t)w(s)d\phi(t|s)dF(s)$ and $\iint c(t|s)d\tilde{\phi}(t|s)dF(s) \leq \iint c(t|s)d\phi(t|s)dF(s)$; (ii) $\tilde{\phi}(s) = \delta(s)$ for all $s \geq s^{**}_\zeta$ and all $s \leq \hat{s}_\zeta$, and $\tilde{\phi}([\hat{s}_\zeta, s^{**}_\zeta]|s) = 1$ for all $s \in (\hat{s}_\zeta, s^{**}_\zeta)$.*

Proof. We start by applying the falsification-proofness principle of Proposition C.1 to α, so we might as well assume that α satisfies (FPC). We can also assume that ineligible scores do not falsify under ϕ as it is a best-response to α by (FPC); it lowers $\iint c(t|s)d\phi(t|s)dF(s)$ and increases $\iint \alpha(t)w(s)d\phi(t|s)dF(s)$.

If $s_\zeta^{**} = \overline{s}$ there is nothing to prove. Suppose, therefore, that $s_\zeta^{**} < \overline{s}$. Then, we let

$$\hat{\alpha}(s) = \begin{cases} 1 & \text{if } s \geq s_\zeta^{**} \\ 1 - c(s_\zeta^{**}|s) & \text{if } s < s_\zeta^{**} \text{ and } c(s_\zeta^{**}|s) \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

And we let $\tilde{\alpha}(s) = \max\{\alpha(s), \hat{\alpha}(s)\}$. By definition of s_ζ^{**}, $c(s_\zeta^{**}|\hat{s}_\zeta) = 1$, hence $\tilde{\alpha}(s) = \alpha(s)$ for all $s \leq \hat{s}_\zeta$. Scores above s_ζ^{**} have no incentive to falsify and receive a higher allocation probability under $\tilde{\alpha}$ than under α.

We partition $[\underline{s}, s_\zeta^{**})$ into two sets: $\hat{S}^+ = \{s \in [\underline{s}, s_\zeta^{**}) : \alpha(s) \geq \hat{\alpha}(s)\}$, and $\hat{S}^- = \{s \in [\underline{s}, s_\zeta^{**}) : \alpha(s) < \hat{\alpha}(s)\}$. Hence, $\hat{S}^-$ is the set where $\hat{\alpha}$ replaces α in $\tilde{\alpha}$.

Consider the falsification strategy:

$$\tilde{\phi}(s) = \begin{cases} \phi(s) & \text{if } s \in \hat{S}^+ \\ \pi(s)\delta_s + \big(1 - \pi(s)\big)\delta_{s_\zeta^{**}} & \text{if } s \in \hat{S}^- \\ \delta_s & \text{if } s \geq s_\zeta^{**} \end{cases}.$$

For $s \geq s_\zeta^{**}$, not falsifying is optimal under $\tilde{\alpha}$. This yields less falsification from these scores and a higher allocation probability.

If $s \in \hat{S}^+$, the only new potential targets for s under $\tilde{\alpha}$ are s_ζ^{**} and scores in $\hat{S}^-$. We show that there is no gain from falsifying to any such score. By definition, we have $\tilde{\alpha}(s) = \alpha(s) \geq 1 - c(s_\zeta^{**}|s)$, so not falsifying dominates falsifying to s_ζ^{**}. But then, (TRI) implies that, for any $\tilde{s} \in \hat{S}^-$,

$$\tilde{\alpha}(s) = \alpha(s) \geq 1 - c(s_\zeta^{**}|s) \geq 1 - c(s_\zeta^{**}|\tilde{s}) - c(\tilde{s}|s) = \tilde{\alpha}(\tilde{s}) - c(\tilde{s}|s) \geq \alpha(\tilde{s}) - c(\tilde{s}|s).$$

Hence, not falsifying also dominates falsifying as $\tilde{s}$ both under α and under $\tilde{\alpha}$. Therefore, $\tilde{\phi}(s) = \phi(s)$ is an optimal strategy for s under $\tilde{\alpha}$.

If $s \in \hat{S}^-$, then s must be eligible, as noted above. We choose $\pi(s) = 1$ if $\tilde{\alpha}(s) \geq \int \alpha(t)d\phi(t|s)$. This yields a higher allocation probability for s and a lower cost of falsification. Otherwise, we choose $\pi(s)$ so that

$$\pi(s)\tilde{\alpha}(s) + 1 - \pi(s) = \int_{T(\alpha,s)} \alpha(t)d\phi(t|s),$$

leaving the final expected allocation probability unchanged. By (FPC), $\alpha(t) = \alpha(s) + c(t|s)$ for all $t \in T(\alpha, s)$. Hence,

$$\pi(s) = \frac{1 - \int_{T(\alpha,s)} \alpha(t) d\phi(t|s)}{c(s_\zeta^{**}|s)} = \frac{1 - \alpha(s) - \int_{T(\alpha,s)} c(t|s) d\phi(t|s)}{c(s_\zeta^{**}|s)}.$$

The cost of falsification for s under $\tilde{\phi}$ is

$$\begin{aligned}\int c(t|s) d\tilde{\phi}(t|s) &= \big(1 - \pi(s)\big) c(s_\zeta^{**}|s) \\ &= c(s_\zeta^{**}|s) - 1 + \alpha(s) + \int_{T(\alpha,s)} c(t|s) d\phi(t|s) \\ &< \int_{T(\alpha,s)} c(t|s) d\phi(t|s),\end{aligned}$$

since $s \in \hat{S}^-$.

Overall, we have obtained a falsification strategy $\tilde{\phi}$ that is optimal under $\tilde{\alpha}$; induces a higher allocation probability for eligible scores and the same allocation probability for ineligible scores; and has a lower cost of falsification than ϕ. If $\tilde{\alpha}$ does not satisfy (FPC), we can apply to it the transformation in Proposition C.1 which clearly preserves flatness above s_ζ^{**}. □

Proof of Theorem 3. By Lemma C.2, it is without loss of optimality for (WWP) to restrict attention to the set $\mathcal{A}$ of allocation rules that satisfy (FPC), and the top-flatness property. Introducing the aggregate cost of falsification Φ as an auxiliary optimization variable, we can rewrite (WWP) as

$$\begin{aligned}\max_{\Phi \geq 0} \max_{a \in \mathcal{A}} \max_{\phi} \iint \alpha(t)\big\{w(s) - \hat{w} + \zeta\big\} d\phi(t|s) dF(s) - (\xi + \zeta)\Phi \\ \text{s.t. (IC)}, \iint c(t|s) d\phi(t|s) dF(s) = \Phi\end{aligned}$$

This allows us to isolate the following optimization problem

$$\begin{aligned}Z(\Phi) = \max_{\alpha \in \mathcal{A}} \max_{\phi} \Pi(\alpha, \phi) \\ \text{s.t. (IC)}, \iint c(t|s) d\phi(t|s) dF(s) = \Phi\end{aligned} \tag{FixFals}$$

where we defined $\Pi(\alpha, \phi) = \iint \alpha(t)\big\{w(s) - \hat{w} + \zeta\big\} d\phi(t|s) dF(s)$ to be the allocative surplus with shifted outside option. The problem consists in choosing an allocation rule and an optimal falsification strategy of the agent so as to maximize this (modified) allocative surplus at a fixed aggregate falsification cost Φ. Note that α satisfies (FPC) so any amount of falsification can be shut down without affecting agent optimality.

The original problem then becomes $\max_{\Phi \geq 0} Z(\Phi) - (\xi + \zeta)\Phi$. To prove the theorem, we show that Z has bounded upward differences with bound $w\big(s_\zeta^{**}\big) - \hat{w} + \zeta$, so that the solution to this problem must be $\Phi = 0$ whenever $\xi + \zeta \geq w\big(s_\zeta^{**}\big) - \hat{w} + \zeta$ (uniquely if the inequality is strict), implying that α^* then solves the weighted welfare problem. This yields the condition $\xi \geq w\big(s_\zeta^{**}\big) - \hat{w}$, which is equivalent to $s_\zeta^{**} \leq \hat{\sigma}(\xi)$. By definition of s_ζ^{**}, this is equivalent to (H).

Bounded upward differences: Let $\Phi' > \Phi \geq 0$ and let (α', ϕ') solve (FixFals) under Φ', and (α, ϕ) under Φ. Then

$$\iint c(t|s) d\phi'(t|s) dF(s) = \Phi' > \Phi = \iint c(t|s) d\phi(t|s) dF(s).$$

As noted above, because α' satisfies (FPC), we can shut down any amount of falsification in ϕ' without affecting agent optimality. Therefore, there exists an optimal falsification strategy $\tilde{\phi}$ of the agents under α' such that $\iint c(t|s) d\tilde{\phi}(t|s) dF(s) = \Phi$. Then we have

$$\begin{aligned}
Z(\Phi') - Z(\Phi) &= \Pi(\alpha', \phi') - \Pi(\alpha', \tilde{\phi}) + \Pi(\alpha', \tilde{\phi}) - \Pi(\alpha, \phi) \\
&\leq \Pi(\alpha', \phi') - \Pi(\alpha', \tilde{\phi}) \\
&= \int_S \int_{T(\alpha', s)} \alpha'(t) \big\{ w(s) - \hat{w} + \zeta \big\} \big[d\phi'(t|s) - d\tilde{\phi}(t|s) \big] dF(s) \\
&= \int_S \int_{T(\alpha', s)} \big(\alpha'(s) + c(t|s)\big) \big\{ w(s) - \hat{w} + \zeta \big\} \big[d\phi'(t|s) - d\tilde{\phi}(t|s) \big] dF(s) \\
&= \int_S \big\{ w(s) - \hat{w} + \zeta \big\} \int_{T(\alpha', s)} c(t|s) \big[d\phi'(t|s) - d\tilde{\phi}(t|s) \big] dF(s) \\
&\leq \Big(w(s_\zeta^{**}) - \hat{w} + \zeta \Big) \iint c(t|s) \big[d\phi'(t|s) - d\tilde{\phi}(t|s) \big] dF(s) \\
&= \Big(w(s_\zeta^{**}) - \hat{w} + \zeta \Big) \big(\Phi' - \Phi \big).
\end{aligned}$$

Since (α, ϕ) is optimal for (FixFals) at Φ, $\Pi(\alpha', \tilde{\phi}) \leq \Pi(\alpha, \phi)$, implying the first inequality. The second equality is because both $\phi'(s)$ and $\tilde{\phi}(s)$ are supported on $T(\alpha', s)$. The third equality is due to (FPC) which implies that $\alpha'(t) - c(t|s) = \alpha'(s)$ for all $t \in T(\alpha, s)$. The fourth equality is because $\phi'(s)$ and $\tilde{\phi}(s)$ are probability distributions fully supported on $T(\alpha', s)$, so $\int_{T(\alpha', s} d\phi'(t|s) = \int_{T(\alpha', s} d\tilde{\phi}(t|s) = 1$. The second inequality is because $\phi'(s) = \tilde{\phi}(s) = \delta_s$ for $s \geq s_\zeta^{**}$ by the top-flatness property.

□

D Resource constraint

Constrained problem. We first consider the constrained problem of optimally allocating a fixed mass of prizes ρ

$$\overline{W}(\rho) = \max_{\alpha} \int_S \alpha(s)w(s)dF(s) \tag{$\overline{\text{RCP}}$}$$
$$\text{s.t. (FPC), (PC),}$$
$$\int_S \alpha(s)dF(s) = \rho. \tag{$\overline{\text{RC}}$}$$

The designer's problem (RCP) is then simply $\max_{\rho \leq \bar{\rho}} \overline{W}(\rho)$. Let $\hat{w}$ be the Lagrange multiplier on the resource constraint in ($\overline{\text{RCP}}$). The Lagrangian for ($\overline{\text{RCP}}$) is then $\int_S \alpha(s)\{w(s) - \hat{w}\}dF(s) + \hat{w}\rho$.

Maximizing the Lagrangian for a fixed value of the multiplier $\hat{w}$ is therefore equivalent to solving (P) with outside option $\hat{w}$. To solve the constrained problem, we then need to identify the value of $\hat{w}$ that makes ($\overline{\text{RC}}$) hold. To do this, we first study how the allocation $\alpha^*(\cdot, \hat{w}, r)$ varies with $\hat{w}$ and r. Since adjusting $\hat{w}$ and r are the key tools to achieve a particular allocation ρ, understanding how the solution to (P) reacts to such changes reveals the mechanics of how indirect effects operate.

Proposition D.1. *The 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 $\lim_{\hat{w}\to\bar{w}^+} \alpha^*(s, \hat{w}, r) = \alpha^*(s, \bar{w}, 0)$.*

Proof. Let $\tilde{\alpha}^*(\hat{w})$ denote the correspondence mapping $\hat{w}$ to the set of solutions of (P). By Lemma 2, we can write (P) as an optimization problem over the set of nondecreasing functions from S to $[0, 1]$ satisfying (FPC) and (PC). This space is compact, by Helly's theorem, and convex. 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 2 and Theorem 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 $\alpha \succeq \beta \Leftrightarrow \forall s \;\; \alpha(s) \geq \beta(s)$, 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 except when $\hat{w} = \bar{w}$, where r pins down the selection.

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 $\hat{w} \to \bar{w}$. □

Intuitively, increasing the value of the outside option reduces the mass of allocated prizes in (P). In fact, a stronger result holds, since the effect is uniform across scores: For every s, the allocation probability is decreasing in $\hat{w}$. A higher value of the neutral gap share naturally increases the 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 mass of allocated prizes under the optimal rule $\alpha^*(s, \hat{w}, r)$. Proposition D.1 implies:

Corollary D.1. *The mass of allocated prizes $A^*(\hat{w}, r)$ 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 $\lim_{\hat{w} \to \bar{w}^+} A^*(\hat{w}, r) = A^*(\bar{w}, 0)$.*

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

Optimal constrained allocation. Returning to ($\overline{\text{RCP}}$), $\hat{w}$ is the Lagrange multiplier on ($\overline{\text{RC}}$) and can be interpreted as the shadow price of marginally tightening the constraint. To ensure that ($\overline{\text{RC}}$) holds, we must adjust $\hat{w}$ and r so that $A^*(\hat{w}, r) = \rho$. Theorem D.1 shows that this is possible:

Theorem D.1. *For any $\rho \in [0, 1]$, there exists a unique outside option value $\hat{w}(\rho)$ and, under* (Mult)*, a unique neutral gap share $r(\rho)$, such that $A^*\big(\hat{w}(\rho), r(\rho)\big) = \rho$. 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 allocation rule $\alpha^*(s, \hat{w}(\rho), r(\rho))$ is then the unique solution to the constrained problem* ($\overline{\text{RCP}}$)*. The value function of* ($\overline{\text{RCP}}$)*, $\bar{W}(\rho)$ is strictly concave at ρ unless* (Mult) *holds.*

Proof. The proof relies on the following Lemma which collects classical results in optimization theory (see, for example, Luenberger, 1969, chapter 8). Necessity of (i) holds because ($\overline{\text{RCP}}$) is linear in α.

Lemma D.1. *The following statements are equivalent:* (i) *α solves* ($\overline{\text{RCP}}$)*;* (ii) *There exists an outside option $\hat{w}(\rho)$ such that α solves* (P) *and $\int_S \alpha(s)dF(s) = \rho$;* (iii) *There exists $\hat{w}$ such that $(\alpha, \hat{w})$ is a saddle-point for the Lagrangian of the within problem $\int_S \alpha(s)\{w(s) - \hat{w}\}dF(s) + \hat{w}\rho$.*

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

The expected worth 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 D.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 $A^*(\hat{w}, r) = \rho$. By Lemma D.1, $\alpha^*(\cdot, \hat{w}, r)$ is then the unique solution to the within problem ($\overline{\text{RCP}}$). The continuity and monotonicity results of Corollary D.1 also imply continuity and monotonicity of $\hat{w}(\rho)$ and $r(\rho)$. By Lemma D.1, $\overline{W}(\rho)$ is concave on $[0, 1]$, and since $\hat{w}(\rho)$ is unique, it is differentiable everywhere, and $\overline{W}'(\rho) = \hat{w}(\rho)$. In particular, $\overline{W}(\rho)$ is strictly concave at every ρ such that $\hat{w}(\rho)$ is strictly decreasing, that is whenever $\hat{w}(\rho) \neq \bar{w}$ or $c(\overline{s}|\underline{s}) \geq 1$. $\square$

Then it is easy to find the solution to (RCP).

Corollary D.2. *If there exists $r \in [0, 1]$ such that $A^*(0, r) \leq \bar{\rho}$, then $\alpha^*(\cdot, 0, r)$ solves* (RCP)*. Otherwise, the unique solution to* (RCP) *is $\alpha^*\big(\cdot, \hat{w}(\bar{\rho}), r(\bar{\rho})\big)$.*

References

Ajzenman, Nicolás (2021) "The power of example: Corruption spurs corruption," *American Economic Journal: Applied Economics*, 13 (2), 230–57.

Akbarpour, Mohammad, Piotr Dworczak, and Scott Duke Kominers (2024) "Redistributive Allocation Mechanisms," *Journal of Political Economy*, 132 (6), 1831–1875, 10.1086/728111.

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

Amador, Manuel, Ivan Werning, and George-Marios Angeletos (2006) "Commitment vs. Flexibility," *Econometrica*, 74 (2), 365–396.

Arieli, Itai, Yakov Babichenko, and Fedor 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*, 22, 468.

Askenazy, Philippe, Thomas Breda, Flavien Moreau, and Vladimir Pecheu (2022) “Les entreprises sous-déclarent-elles leur effectif à 49 employés pour contourner la loi?” Note IPP 82, Institut des Politiques Publiques.

Augias, Victor and Eduardo Perez-Richet (2024) “Non-Market Screening with Investment,” working paper.

Ball, Ian (2025) “Scoring Strategic Agents,” *American Economic Journal: Microeconomics*, 17 (1), 97–129, 10.1257/mic.20230275.

Ben-Porath, Elchanan, Eddie Dekel, and Barton L Lipman (2014) “Optimal allocation with costly verification,” *American Economic Review*, 104 (12), 3779–3813.

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2019) “Mechanisms with Evidence: Commitment and Robustness,” *Econometrica*, 87 (2), 529–566.

Bjerre-Nielsen, Andreas, Lykke Sterll Christensen, Mikkel Høst Gandil, and Hans Henrik Sievertsen (2023) “Playing the system: address manipulation and access to schools,” working paper.

Boerma, Job, Aleh Tsyvinski, and Alexander P Zimin (2021) “Sorting with team formation,”Technical report, National Bureau of Economic Research.

Bolton, Lee (2018) “Manipulation of the Waitlist Priority of the Organ Allocation System through the Escalation of Medical Therapies,” *OPTN/UNOS EthicsCommittee*.

Carlier, Guillaume (2012) “Optimal transportation and economic applications,” *Lecture Notes*, 18.

Chua, Geoffrey A., Gaoji Hu, and Fang Liu (2023) “Optimal multi-unit allocation with costly verification,” *Social Choice and Welfare*, 61 (3), 455–488.

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

Daskalakis, Constantinos, Alan Deckelbaum, and Christos Tzamos (2017) “Strong duality for a multiple-good monopolist,” *Econometrica*, 85 (3), 735–767.

Deneckere, Raymond and Sergei Severinov (2022) “Screening, Signalling and Costly Misrepresentation,” *Canadian Journal of Economics*, 55 (3), 1334–1370.

Dworczak, Piotr (2022) “Equity-efficiency trade-off in quasi-linear environments,” working paper.

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

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

Frankel, Alex and Navin Kartik (2019) "Muddled information," *Journal of Political Economy*, 127 (4), 1739–1776.

——— (2021) "Improving information from manipulable data," *Journal of the European Economic Association*, 1542–4766, 10.1093/jeea/jvab017.

Galbiati, Roberto and Giulio Zanella (2012) "The tax evasion social multiplier: Evidence from Italy," *Journal of Public Economics*, 96 (5-6), 485–494.

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

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

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

Kattwinkel, Deniz and Jan Knoepfle (2023) "Costless Information and Costly Verification: A Case for Transparency," *Journal of Political Economy*, 131 (2), 504–548, 10.1086/721618.

Kephart, Andrew and Vincent 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, Alexander, Fedor Sandomirskiy, Aleh Tsyvinski, and Alexander P Zimin (2022) "Beckmann's approach to multi-item multi-bidder auctions," *arXiv preprint arXiv:2203.06837*.

Kolotilin, Anton, Roberto Corrao, and Alexander Wolitzky (2025) "Persuasion and Matching: Optimal Productive Transport."

Lacker, Jeffrey M. and John A. Weinberg (1989) "Optimal Contracts with Costly State Falsification," *Journal of Political Economy*, 97 (6), 1345–1363.

Landier, Augustin and Guillaume Plantin (2016) "Taxing the Rich," *The Review of Economic Studies*, 84 (3), 1186–1209, 10.1093/restud/rdw033.

Li, Yingkai and Xiaoyun Qiu (2023) "Screening Signal-Manipulating Agents via Contests," Working paper.

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

Lin, Xiao and Ce Liu (2024) "Credible Persuasion," *Journal of Political Economy*, 132 (7), 2228–2273.

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

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

Milgrom, Paul and Chris Shannon (1994) "Monotone Comparative Statics," *Econometrica*, 62 (1), 157–180.

Milli, Smitha, John Miller, Anca D Dragan, and Moritz Hardt (2019) "The social cost of strategic classification," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 230–239.

Myerson, R.B. (1981) "Optimal Auction Design," *Mathematics of Operations Research*, 6 (1), 58.

Mylovanov, Tymofiy and Andriy Zapechelnyuk (2017) "Optimal allocation with ex post verification and limited penalties," *American Economic Review*, 107 (9), 2666–94.

Patel, Rohit and Can Urgun (2023) "Costly inspection and money burning," working paper.

Perez-Richet, Eduardo and Vasiliki Skreta (2022) "Test design under falsification," *Econometrica*, 90 (3), 1109–1142.

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

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

Severinov, Sergei and Terry Yin-Chi Tam (2019) "Screening Under Fixed Cost of Misrepresentation," working paper.

Tan, Teck Yong (2023) "Price Discrimination with Manipulable Observables," *Available at SSRN 4480623.*