GRAPE Working Paper #83

Comparison of screening devices

Frank Yang, Piotr Dworczak, and Mohammad Akbarpour

FAME | GRAPE, 2023

Foundation of Admirers and Mavens of Economics
Group for Research in Applied Economics

Comparison of screening devices

Frank Yang
Stanford University

Piotr Dworczak
Northwestern University
and FAME | GRAPE

Mohammad Akbarpour
Stanford University

Abstract

A designer relies on a costly screening device to allocate a set of goods, aiming to maximize a social welfare function. We provide conditions for one screening device to dominate another. We show that the performance of a screening device depends on two channels: (i) targeting effectiveness which measures the alignment between the implemented and desired assignments, and (ii) rent provision which determines the utilities of agents receiving the goods net of the screening costs. We link these two channels to distinct properties of the joint distribution of agents' characteristics, leading to a number of simple empirical tests for comparing screening devices.

Keywords:

mechanism design, stochastic orders, costly screening, inequality-aware market design

JEL Classification

D47, D61, D63, D82

Corresponding author

Piotr Dworczak, piotr.dworczak@northwestern.edu

Acknowledgements

We would like to thank various seminar audiences and anonymous referees from the program committee of the EC'23 conference for useful comments. Dworczak gratefully acknowledges the support received under the ERC Starting grant IMD-101040122.

Published by: FAME | GRAPE
ISSN: 2544-2473

Foundation of Admirers and Mavens of Economics
Koszykowa 59/7
00-660 Warszawa
Poland

W | grape.org.pl
E | grape@grape.org.pl
TT | GRAPE_ORG
FB | GRAPE.ORG
PH | +48 799 012 202

1 Introduction

Public agencies are often tasked with allocating scarce resources (such as public housing, financial aid, or medical treatment) to a target population. In many such cases, the goal is to maximize social welfare, which requires identifying agents with the highest social values for the resource. This task is typically challenging due to information asymmetries: While public agencies may be able to access some data about potential beneficiaries (e.g., through means testing), they generally lack the information necessary to achieve perfect targeting. At the same time, the conventional economic solution to the problem of private information—relying on prices to screen for agents with the highest values—may fail in such contexts. Monetary transfers are useless for screening when the good being allocated is money itself (e.g., when allocating financial aid, disability insurance, or unemployment benefits). Moreover, prices may be unavailable as a tool for political or ethical reasons (e.g., monetary transfers were generally not used to allocate vaccines during the Covid-19 pandemic). Finally, prices may fail to identify agents with the highest *social* values when private willingness to pay diverges from the social objective (e.g., in problems where the designer has redistributive preferences).

For these and other reasons, public agencies may rely on "ordeals" to improve targeting. An ordeal is a task that an agent needs to complete in order to be considered for the allocation of the resource; ordeals are costly for the agent but do not provide any direct social benefit to anyone. Natural examples include standing in line, filing out complicated forms, dealing with "red tape," waiting, visiting an office at an inconvenient time, or traveling to a registration site.[1] Such ordeals serve the role of costly screening devices: They decrease the utility of the eventual beneficiaries of social programs but may raise social welfare on the net if they improve targeting substantially. For instance, Alatas et al. (2016) analyze the design of Indonesia's Conditional Cash Transfer program, and show that imposing an ordeal of traveling to a registration site may improve the targeting of financial aid. Deshpande and Li (2019) demonstrate that the level of congestion at Social Security Administration field offices affects the targeting of disability programs. Costly screening may also be induced inadvertently when scarce resources are allocated for free—the classical example is waiting time in access to public health care (see, e.g., Rose 2021, Zeckhauser 2021). During the Covid-19 pandemic in the United States, costly screening played a role in determining the allocation of vaccines among agents eligible for vaccination within a given phase—getting a vaccine early often required spending time

[1] Some of these examples are not "pure" ordeals in that they may serve an additional purpose; for example, filling out forms is an inconvenience for the agent but it also provides some information to the agency; see Kleven and Kopczuk (2011) for an analysis of this case and Section 5.2 for an additional discussion.

on the phone, monitoring online queuing systems, or traveling to a far-away pharmacy (or even a different state).

The pervasiveness of costly screening devices in resource allocation problems motivates our main question: What makes one costly screening device better than another? The primary contribution of the paper is to provide a sufficient condition for a robust ranking of two screening devices, where robustness is with respect to the level of supply of the resource being allocated. Our sufficient condition identifies two properties of a screening device that together determine its welfare consequences. The first one is *targeting effectiveness* which measures whether using a given screening device results in allocating the resources to the "right" agents, that is, those with the highest social values. The second one is *rent provision* which measures the surplus of agents that receive the resource net of the cost of completing the ordeal.

Overview of the model. To explain these findings in more detail, we first sketch our formal model. A mechanism designer has a fixed supply of goods that she allocates to a set of agents. Allocation can be stochastic and contingent on completing an ordeal, potentially with varying degrees of difficulty or arduousness (e.g., if the ordeal is to wait, the designer may decide how much the agent needs to wait to obtain a resource of a given quality). Agents have quasi-linear utilities, and are heterogeneous along three dimensions: their private value for receiving the good, their social welfare weight, and their per-unit-of-difficulty cost of completing the ordeal. These characteristics are agents' private information but we assume that the designer knows their joint distribution. The goal of the designer is to maximize social welfare defined as the sum of utilities of all agents weighted by their welfare weights, net of the costs of completing the ordeal. In particular, the first best is to allocate the resources to agents with the highest social value—the highest product of the private value with the welfare weight. In the second best, the designer maximizes social welfare over the entire set of incentive-compatible mechanisms, with the ordeal being the only instrument allowing the designer to screen agents.

In this framework, a costly screening device is modeled as a joint distribution of private values, welfare weights, and costs. Comparing two screening devices consists in comparing the performance of the optimal allocation mechanisms for two joint distributions that have the same marginal distribution of private values and welfare weights. That is, a given ordeal is identified with the marginal distribution of costs for completing it and the correlation of these costs with private values and welfare weights. We say that screening device A dominates screening device B, if the optimal mechanism under the joint distribution induced by device A achieves a higher level of social welfare than the

optimal mechanism under the joint distribution induced by device B, regardless of the level of supply of the good being allocated.

Our main formal result is a sufficient condition for screening device A to dominate screening device B. The sufficient condition quantifies the effects of the two channels mentioned previously: targeting effectiveness and rent provision. Formally, targeting effectiveness is captured by the conditional expectation of social values conditional on a given *rate of substitution*—the ratio of the private value of the good to the cost of completing the ordeal. Rent provision is determined by the dispersion of the distribution of the rate of substitution. These two statistical objects—the conditional expectation of social values and the marginal distribution of the rate of substitution—together pin down the performance of a screening device.

The importance of targeting effectiveness for the performance of a costly screening device is very intuitive: Incurring the deadweight loss associated with using an ordeal can be justified only if screening is improved relative to the untargeted allocation of the resource.[2] Below, we illustrate the importance of the less intuitive rent-provision channel using an example; we show that the designer may opt for a device that achieves worse targeting provided that it has a more dispersed distribution of the rate of substitution.

An illustrative example. The good being allocated is money, with every recipient awarded an amount that we normalize to \$1. (In this special case of our model, the private value for the good is by definition 1 for every agent, because we measure private values and costs in terms of willingness to pay in dollars.) The population of agents has mass 1, but the designer only has a mass $s \in (0,1)$ of money to allocate. Each agent has one of three possible types, p(oor), m(iddle class), or r(ich), and each type is equally likely. The designer puts a social welfare weight $\lambda = 3$ on agents with type p, $\lambda = 1$ on agents with type m, and $\lambda = 0$ on agents with type r.

There are two screening devices available, A and B, with per-unit-of-difficulty costs c_i, $i \in \{A, B\}$, as in Table 1, where ε is a small positive number.

To simplify calculations, suppose that $s = 2/3$, so that the designer has enough money to give a dollar to each agent in two out of the three groups. The first best (under perfect information) would be to give money to agents in groups p and m, generating a total social surplus of $(1/3)\cdot 3 + (1/3)\cdot 1 = 4/3$. When types are unobserved, without a screening device, the designer can allocate money randomly generating a total social surplus of $(2/3)((1/3)\cdot 3 + (1/3)\cdot 1 + (1/3)\cdot 0) = 8/9$.

Next, consider using a costly screening device $i \in \{A, B\}$: The designer gives money

[2] This point was made in the literature at least as early as Nichols and Zeckhauser (1982).

	c_A	c_B	λ
p	$1-2\varepsilon$	ε	3
m	$1-\varepsilon$	2	1
r	1	1	0

Table 1: Screening device A is more correlated with the social values, while screening device B has a higher dispersion of costs

conditional on completing an ordeal with some difficulty $y > 0$. Since an agent's private benefit from getting a dollar is 1, she will complete the ordeal if her utility $1 - c_i y$ is positive, generating $\lambda(1 - c_i y)$ of social surplus. (Note that the rate of substitution in this example is simply $1/c_i$.)

Ordeal A achieves perfect targeting: The designer is able to target monetary transfers to agents in groups p and m by setting the difficulty of ordeal A to be $y = 1$ (the lowest difficulty that disincentivizes agents with type r to apply). Ordeal B, in contrast, allocates money to groups p and r (by setting the difficulty of the ordeal to be $y = 1/2$), since in this case agents in group m find it more costly to apply than agents in group r.

However, it is easy to see that screening device B is better for the designer than screening device A: Using ordeal A leads to total social surplus $(1/3) \cdot 3 \cdot (2\varepsilon) + (1/3) \cdot \varepsilon = (7/3)\varepsilon$ (in fact, ordeal A is worse than random allocation for small enough ε), while using ordeal B leads to total social surplus $(1/3) \cdot 3 \cdot (1 - \varepsilon(1/2)) = 1 - \varepsilon/2$.

Intuitively, screening device B outperforms device A because it gives more net utility to the agents conditional on allocation. The key observation is that costs are more *dispersed* under device B. This means that low-cost agents can be separated from high-cost agents at a low utility cost to the former; in contrast, device A achieves separation at the cost of suppressing the net utility of the recipients almost to zero. This effect—which has nothing to do with targeting—is what we called the rent-provision channel.

We conclude the example with a few comments. First, the rent-provision channel is invariant to the absolute level of the costs. If we multiply all costs under ordeal B by an arbitrary constant $b > 0$, the only change is that the difficulty is set to $1/(2b)$ instead of $1/2$, resulting in exactly the same welfare consequences. Second, we assumed that the difficulty of the ordeal is set to "clear the market"; however, the ranking is the same when the designer uses an optimal mechanism instead. Third, while assuming $s = 2/3$ was convenient for calculations, as long as ε is low enough, ordeal B strictly outperforms ordeal A for any supply $s \in (0, 1)$.

Economic implications. Our results formalize the intuition of the simple example by defining a notion of dispersion of the rate of substitution. Holding fixed the targeting effectiveness of a screening device, increasing the dispersion of the rate of substitution raises the maximal social surplus that can be generated by a screening device, for any supply level. In particular, using a costly screening device with low dispersion can lead to arbitrarily low welfare regardless of its targeting effectiveness, while costly screening devices with high dispersion may even approach the first-best level of welfare.

The analysis of the two channels affecting the performance of a screening device leads to two types of insights. First, we provide a few empirical tests that may be used to establish a robust ranking between two screening devices observed in practice. If some additional structure is imposed on the joint distribution of values, costs, and welfare weights, these tests take the form of simple regressions. Second, we derive some high-level prescriptions for policymakers. For example, we establish conditions under which giving agents a choice between two ordeals (or requiring completion of both) is dominated by using one of them.

Related literature. The idea that ordeals can be useful in screening in the absence of monetary transfers is a classical one within economic theory. Nichols and Zeckhauser (1982) were among the first to point out that, in settings where the designer does not observe individual characteristics, it may be optimal to use a screening device that is costly, as long as the costs are sufficiently lower for the target population. To the best of our knowledge, the sizeable literature on costly screening (for example, Hartline and Roughgarden 2008, Condorelli 2012, Chakravarty and Kaplan 2013, among many others) focused on the optimal design of an allocation mechanism for a *fixed* costly screening device.[3] Thus, our contribution to the theoretical literature is to provide an intuitive condition for comparing different screening devices. Additionally, we clarify circumstances under which a costly screening device is outperformed by, or outperforms, monetary transfers as a screening device.[4] Our focus on maximizing social welfare (rather than allocative efficiency) follows a growing body of papers on inequality-aware market design (see, in particular, Condorelli 2013 and Dworczak ⓡ al. 2021).[5]

There is also a sizable empirical literature on targeting through ordeals (see, for example, Alatas et al. 2016; Finkelstein and Notowidigdo 2019; Deshpande and Li 2019).

[3]Looking at the problem of allocating financial aid, Dworczak (2023) provides a condition under which a given costly screening device outperforms a lump-sum payment.

[4]Unlike Yang (2022, 2023), we do not study how to optimally combine price and nonprice screening.

[5]Our setting is closest to the one studied by Akbarpour ⓡ al. (2023).

This literature successfully quantified the targeting effectiveness of different ordeals in the context of various welfare programs using (natural) experiments that perturb the application costs. For example, Deshpande and Li (2019) analyzed how quasi-random closures of SSA field offices—leading to increased travel and congestion costs—affected the composition and acceptance rates of applicants for disability insurance; Finkelstein and Notowidigdo (2019) implemented an experimental design in which randomly selected elderly individuals received assistance in applying for SNAP benefits (a reduction in the application costs).

We complement the empirical literature in two ways. First, we identify rent provision as an important channel determining the welfare consequences of using a costly screening device. Empirical work used various ways to measure welfare impacts in their respective contexts; however, to the best of our knowledge, the importance of the dispersion in rates of substitution through the rent-provision channel has not been recognized. Measuring the quantitative effects of this channel is relatively straightforward—our analysis reveals that it is analogous to estimating a demand curve in a standard market with money.

Second, empirical analysis is typically limited to studying "local" perturbations to application costs that do not affect the welfare program itself. This means that they must parameterize the extensive margin by putting a social value on the marginal dollar or unit of resource allocated by the program. Theoretical analysis allows us to focus on the intensive margin by assuming that the planner selects an optimal allocation mechanism for any given screening device: A switch to a new screening device induces a change in the program rules, so that the same number of resources is allocated.[6] This perspective casts new light on what makes a screening device attractive. For example, the absolute level of the cost becomes irrelevant for welfare (doubling the cost results in halving the requirements for successful applicants), and what matters instead is the dispersion determining the actual utility cost of obtaining the resource in the optimal mechanism.

2 Model

A designer allocates a set of objects with (potentially) heterogeneous quality to a unit mass of agents. Each agent has a *welfare weight* $\lambda \geq 0$ and *private value* $v \geq 0$. The distribution of (λ, v) is hold fixed and is commonly known. The designer observes neither v nor λ for any individual agent; the assumption that the welfare weight is unobservable

[6] We show that the comparison of screening devices remains unchanged if—instead of the optimal mechanism—the planner selects a mechanism that simply "clears the market" by adjusting the program requirements so that total demand equals the number of resources available for allocation.

captures a situation in which the designer has redistributive preferences but does not perfectly observe who is most in need.[7]

There is a unit mass of objects. Each object has a *quality* $q \in [0,1]$. The cumulative mass of objects with quality less than or equal to q is given by the *quality distribution* F on $[0,1]$.[8]

If assigned an object of quality q, an agent with type (λ, v) gets a payoff vq and the designer gets a payoff λvq. The value v is measured in monetary units and λ can be thought of as converting monetary units to social-utility units, that is, λ is the value for the designer of giving the agent a dollar.[9] We will sometimes refer to λv as the agent's *social value*.

If the good being allocated is money, then $v \equiv 1$, by definition; q should be interpreted as quantity rather than quality (with an upper bound on an individual allowance normalized to 1), and F can be taken to have mass only at 0 and 1.

We assume that the designer can condition allocations of resources on agents performing some costly task—an "ordeal." Crucially, agents may differ in their costs of completing different types of ordeals, and these costs may be correlated with their other characteristics. We thus identify a costly screening device with the conditional distribution of costs that it induces in the target population. We assume that these costs are pinned down up to an intensity (difficulty) level of the task, denoted y (e.g., if the ordeal is waiting, y measures the amount of time the agent must wait). An agent with cost parameter c incurs an additive cost cy for completing the ordeal with intensity y, with c interpreted as the per-unit-of-difficulty cost measured in monetary units.

Formally, a *screening device* D is a (regular) conditional distribution

$$D : (\lambda, v) \to \Delta(\mathbb{R}_{++})$$

that maps a given pair of welfare weight λ and private value v to a distribution of costs c. That is, $D(\lambda, v)$ is the distribution of costs for the group of agents with welfare weight λ and private value v. We may equivalently identify a screening device with the joint distribution of (λ, v, c) (with the joint distribution of λ and v held fixed across different screening devices).

[7] If the designer has access to observable information about agents, the private information in our model can be interpreted as the residual uncertainty conditional on observable information.

[8] The assumption that F is a CDF is without loss of generality since we can always add a set of goods with quality 0.

[9] We use money to measure values even though monetary transfers are not available to the designer as a tool. As long as λ is interpreted correctly, it does not matter in which units individual utility is measured—using monetary units (effectively, willingness to pay) is convenient for interpretations.

For any screening device D, let

$$r = v/c$$

be the induced *rate of substitution* between the value of allocations and the value of avoiding the costly activity. Let G denote the marginal distribution of r, which we assume has a continuous, positive density on an interval. Note that a screening device is uniquely identified by the joint distribution of (λ, v, r). This representation is convenient because r will fully summarize an agent's preferences over choices available in any mechanism.

Conditional on a given screening device, the designer chooses, without loss of generality,[10] a *(direct) mechanism*

$$(\Gamma, y) : r \to \Delta([0,1]) \times \mathbb{R}_+$$

that maps a reported r to a *probabilistic assignment* Γ over the set of qualities $[0,1]$ and an *ordeal intensity* $y \geq 0$. The assumption that y is any non-negative number means that the designer can freely adjust the difficulty of the costly activity. For a probabilistic assignment Γ, let $x^\Gamma(r) := \int_0^1 q \, \mathrm{d}\Gamma(q \mid r)$ denote the *expected quality* that type r receives.

A mechanism (Γ, y) is *feasible* if it satisfies *(i)* incentive-compatibility (IC) and individual-rationality (IR) constraints:

$$\begin{aligned} rx^\Gamma(r) - y(r) &\geq rx^\Gamma(\hat{r}) - y(\hat{r}), && \text{for all } r, \hat{r}, \\ rx^\Gamma(r) - y(r) &\geq 0, && \text{for all } r, \end{aligned}$$

and *(ii)* the feasibility constraint:

$$F(q) \leq \int_{\underline{r}}^{\overline{r}} \Gamma(q \mid r) \, \mathrm{d}G(r) \qquad \text{for all } q \in [0,1].$$

The feasibility constraint states that the distribution of assigned qualities must be first-order stochastic dominated by the available qualities (i.e., free disposal is allowed).

Let $\mathcal{M}(D, F)$ be the space of feasible mechanisms for a given costly screening device D and supply of quality F. The designer wants to maximize total social welfare over all feasible mechanisms:

$$\mathrm{OPT}(D, F) := \sup_{(\Gamma, y) \in \mathcal{M}(D,F)} \mathbb{E}_D \Big[\lambda \Big(v x^\Gamma(r) - c y(r) \Big) \Big],$$

[10] Looking at direct revelation mechanisms that only elicit information about r is without loss of generality in our setting because r fully pins down an agent's preference—see Akbarpour ⓡ al. (2023) for a similar setup and a proof of this claim.

where $\mathrm{OPT}(D, F)$ denotes the optimal value achievable under screening device D and quality distribution F, and $\mathbb{E}_D$ indicates that the conditional distribution over c is given by D. The designer maximizes a standard utilitarian welfare function, with welfare weights λ applied to the utility of each agent (net of screening costs) expressed in monetary units.

Finally, we say that a screening device D_1 *dominates* another screening device D_2 if

$$\mathrm{OPT}(D_1, F) \geq \mathrm{OPT}(D_2, F)$$

for every quality distribution F on $[0, 1]$.[11]

We require robustness of the comparison between two screening devices to the available supply of the good. This is motivated by practical considerations since institutions for allocating public resources are typically persistent and used across multiple instances of the allocation process. In addition, this robustness notion implies that even when the designer can choose among different quality distributions F subject to production costs $\mathcal{C}(F)$, the designer prefers the dominating screening device for any production cost function $\mathcal{C} : \Delta([0, 1]) \to \mathbb{R}$.

It will also turn out that the ranking is partially robust to relaxing the assumption that the designer uses an optimal mechanism; for example, our results continue to hold if the designer uses a mechanism that "clears the market" given the screening device (i.e., when difficulty is set to equalize demand with the supply of available resources).

3 Main Results

For any screening device D, we define the *welfare index* at parameter $s \in [0, 1]$ to be

$$W(D, s) := \int_s^1 \left(1 - \frac{G^{-1}(s)}{G^{-1}(t)}\right) \cdot \mathbb{E}\left[\lambda v \mid r = G^{-1}(t)\right] \mathrm{d}t.$$

This expression quantifies a combination of two effects, *(i) rent provision* and *(ii) targeting effectiveness*. For intuition, note that the first term,

$$1 - \frac{G^{-1}(s)}{G^{-1}(t)},$$

[11] Our results are unaffected if we only consider distributions F supported on $\{0, 1\}$, which is the right assumption when the good being allocated is homogeneous.

can be interpreted as the rent—utility net of screening costs—of an agent at the t-quantile of the distribution of the rate of substitution from obtaining quality 1 of the good when the agent at the s-quantile receives zero utility. The rent is measured in "raw" units of the allocation probability (the utility of receiving the good at no cost is normalized to 1). The second term,

$$V(t) := \mathbb{E}\Big[\lambda v \mid r = G^{-1}(t)\Big],$$

converts these rents into social-welfare units. Thus, $V(t)$ measures how well the allocation of rents aligns with social preferences. Mathematically, $V(t)$ is the expected contribution to social welfare of an agent in the t-quantile of the distribution of the rate of substitution.

Our first result states that these two channels together determine when a screening device dominates another.

Theorem 1. *For any two screening devices D_1 and D_2, if for all $s \in [0,1]$*

$$W(D_1, s) \geq W(D_2, s), \tag{1}$$

then screening device D_1 dominates screening device D_2.

The proof is in the appendix. Theorem 1 relies on two ideas. First, we can redefine mechanisms to depend on quantiles rather than values of the agents' rates of substitution. This transformation enables us to compare mechanisms *across* screening devices. Second, we exploit the mathematical fact that monotone allocation functions can be represented as probability distributions over a set of one-step functions.

If both screening devices induce assortative matching between rates of substitution and quality in the optimal mechanism for any supply level, then condition (1) is also *necessary* for these screening devices to be robustly ranked.[12] However, in general, the optimal mechanism may feature regions of random allocation; for these cases, condition (1) need only hold for a subset of $s \in [0,1]$ for D_1 to dominate D_2. Theorem 5 in Appendix C defines a weakening of condition (1) that is both necessary and sufficient for two screening devices to be robustly ranked.

Our second main result connects the two channels (rent provision and targeting effectiveness) directly to the structural properties of screening devices. In particular, we show that *(i)* rent provision is characterized by the dispersion of the distribution of rates of substitution on the log scale, and *(ii)* targeting effectiveness is characterized by the

[12] Assortative matching is defined by $x^\Gamma(r) = F^{-1}(G(r))$, where F^{-1} is the generalized inverse of the distribution F of quality. Assortative matching is an optimal mechanism for any F when, for example, the distribution G has a non-increasing inverse hazard rate and contains 0 in its support, and $\mathbb{E}[\lambda v|r]$ is non-decreasing in r, implying that $W(D, s)$ is concave in s.

correlation between the rates of substitution and the social values. To state the result, we use two partial orders. Let $\preceq_{disp}$ denote the *dispersive order* (Müller and Stoyan 2002): For any two random variables X and Y with CDFs F_X and F_Y, we write $X \preceq_{disp} Y$ if

$$F_X^{-1}(t) - F_X^{-1}(s) \leq F_Y^{-1}(t) - F_Y^{-1}(s) \quad \text{for all } 0 < s < t < 1 \ .$$

Intuitively, the quantiles of the distribution that is higher in the dispersive order increase more quickly than the corresponding quantiles of a less dispersed distribution.[13] Let $\preceq_{maj}$ denote (an extension of) the *majorization order*: For two functions $h_1, h_2 \in L^1(0,1)$, we write $h_2 \preceq_{maj} h_1$ if

$$\int_s^1 h_2(t)dt \leq \int_s^1 h_1(t)dt, \quad \text{for all } s \in (0,1], \text{ with equality at } s = 0.$$

When h_1 and h_2 are non-decreasing, this is the usual majorization order,[14] which is equivalent to $h_1(U)$ being a *mean-preserving spread* of $h_2(U)$ for a uniform random variable U.

Recall that $V_i(t) = \mathbb{E}\left[\lambda v \mid r_i = G_i^{-1}(t)\right]$.

Theorem 2. *For any two screening devices D_1 and D_2, if*

(i) $\log(r_2) \preceq_{disp} \log(r_1)$,

(ii) $V_2 \preceq_{maj} V_1$,

then screening device D_1 dominates screening device D_2.

The proof is in the appendix. Crucially, conditions *(i)* and *(ii)* are orthogonal in the sense that condition *(i)* depends only on the marginal distribution of the rates of substitution and condition *(ii)* depends only on the distribution of social values conditional on a given rate of substitution.

Condition *(i)* shows that the suitable measure of dispersion in our model is on the log scale. This is intuitive given the form of the rent-provision channel in the welfare index: What matters is the *ratio* of the quantiles of the distribution, rather than their difference. Multiplying r by a constant $\beta > 1$ increases the dispersion of r but has no effect on the optimal welfare (the optimal mechanism will simply adjust the difficulty

[13] The dispersive order is location-invariant in the sense that $X \preceq_{disp} Y \iff X \preceq_{disp} Y + a$ for any $a \in \mathbb{R}$. Provided that $\mathbb{E}[X] = \mathbb{E}[Y]$, the dispersive order implies that Y is second-order stochastically dominated by X. See Appendix A.1 for further characterizations and properties of the dispersive order.

[14] Some authors, e.g., Müller and Stoyan (2002), extend the majorization order to functions that are not non-decreasing by imposing the condition on their monotone rearrangements; that definition would not work for our purposes.

level). In contrast, shifting r by a constant has no effect on the dispersion of r, but it affects the dispersion of $\log(r)$ and hence the effectiveness of the screening device. In particular, uniformly increasing r by a constant makes a screening device *worse*. The proof of Theorem 2 directly links the dispersion in $\log(r)$ to the rent-provision channel.

Condition *(ii)* states that a screening device becomes more effective if agents in any top quantile of the distribution of r have higher average social values (holding fixed the unconditional distribution of social values). Intuitively, an increase in the majorization order means that the rate r becomes a more informative signal in the sense that the posterior mean distribution of social values λv (conditional on observing r) is more spread out. However, unlike in a statistical learning problem, where signals can always be relabeled, the designer in our problem prefers r to be more *positively* correlated with λv, which is taken into account by our extension of the majorization order. To delineate this intuition, let $\preceq_{cor}$ denote the *correlation order* (Müller and Stoyan 2002): For any two random vectors (X_1, X_2) and (Y_1, Y_2), we write $(X_1, X_2) \preceq_{cor} (Y_1, Y_2)$ if

$$C_X(s,t) \leq C_Y(s,t) \quad \text{for all } s,t \in [0,1]\,,$$

where C_X denotes the *copula* of (X_1, X_2).[15] Intuitively, a pair of random variables increases in the correlation order if their correlation becomes more positive at all quantiles. As we show in Appendix A, $(\lambda v, r_2) \preceq_{cor} (\lambda v, r_1)$ implies $V_2 \preceq_{maj} V_1$. That is, condition *(ii)* of Theorem 2 is weaker than requiring an increase in the distribution of social values and rates of substitution in the correlation order. The proof of Theorem 2 directly links the correlation of λv and r to the targeting-effectiveness channel.

For a simple illustration of Theorem 2, consider the following example:

Example 1. Suppose that (λ, v, c), and hence $(r, \lambda v)$, is jointly log-normal:

$$\begin{pmatrix} \log(r) \\ \log(\lambda v) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right).$$

By Theorem 2, a screening device becomes more effective when *(i)* σ_1 gets higher (because $\log(r)$ increases in the dispersive order), or *(ii)* ρ gets higher (because $(\lambda v, r)$ increases in the correlation order).

[15] For any joint distribution of (X_1, X_2) with continuous marginals, their copula is uniquely defined by $C_X(s,t) = \mathbb{P}\left(X_1 \leq F_{X_1}^{-1}(s), X_2 \leq F_{X_2}^{-1}(t)\right)$ for all $s,t \in [0,1]$. More generally, Sklar's theorem implies that a copula of (X_1, X_2) always exists. See Appendix A.2 for the general definition and relevant properties of the correlation order.

4 Economic Implications

In this section, we explore the economic consequences of our main technical results.

4.1 The limits of costly screening

We first use Theorem 2 to explore the boundaries of how effective costly screening can be in allocation problems. This analysis expands on the example in the Introduction by highlighting the role played by the rent-provision channel.

If the supply s of a homogeneous-quality good is fixed, it is easy to see that there exists a costly screening device that approximates the first-best welfare—defined as the level of welfare achievable if the designer could directly observe the social values λv (and would hence give the good to s agents with the highest social values). Suppose that $D^s_\varepsilon(\lambda, v)$ is a Dirac delta on ε if λv lies in the top s-quantile of the distribution of social values, and $D^s_\varepsilon(\lambda, v)$ is a Dirac delta on 1 otherwise. Intuitively, the ordeal is "cheap" for agents who receive the good in the first-best allocation, and costly for all agents who do not. Then, screening device D^s_ε implements the first-best allocation at a total social cost that is of order ε.

However, D^s_ε would perform poorly if the supply of the good was different from s, and in particular would not rank highly in our partial dominance order that requires robustness to the level of supply. It may therefore come as a surprise that we can construct a single screening device that approximates the first best *regardless* of the distribution of quality (and hence supply level).

Proposition 1. *For any $\varepsilon > 0$, there exists a screening device D_ε such that, for any quality distribution F, D_ε achieves welfare that is within ε of first-best welfare.*

We emphasize that the approximation of first-best welfare is uniform in the distribution of quality: Fixing any $\varepsilon > 0$, D_ε implements the first-best allocation at a total social cost that is less than ε, no matter what the distribution of quality is. Of course, the optimal mechanism will depend on the distribution of quality, but for the screening device D_ε, it suffices to always use the "market-clearing mechanism" which, for any level of quality, sets the difficulty of the ordeal to the level that equates supply and demand, achieving assortative matching between r and q (see Footnote 12 for a formal definition).

The proof (in the appendix) constructs D_ε explicitly by exploiting the two channels that we identified in Theorem 2: *(i)* D_ε achieves perfect targeting by making r a strictly increasing function of λv, and *(ii)* the distribution of r becomes increasingly dispersed (on the log scale) as ε decreases. For intuition, note that, for any screening device D, if

we consider a screening device $\widetilde{D}$ defined by $\tilde{r} = r^{\beta}$ where $\beta > 1$, then $\widetilde{D}$ will dominate D (since the copula of the joint distribution of the rates and social values is unaffected, but $\tilde{r}$ has a higher dispersion on the log scale).

Finally, we establish a partial converse to Proposition 1 by showing that the rent-provision channel alone can make any screening device (even one achieving perfect targeting) arbitrarily bad.

Proposition 2. *For any screening device D and any $\varepsilon > 0$, there exists a screening device D_ε such that (i) D_ε achieves (weakly) better targeting effectiveness than D in that $V \preceq_{maj} V_\varepsilon$; and (ii) for any quality distribution F, D_ε achieves social welfare that is less than ε under the assortative-matching allocation rule.*

The proof uses a similar construction as in the proof of Proposition 1 but takes the dispersion of $\log(r)$ to be sufficiently small rather than sufficiently large. As the dispersion vanishes, the rents of all agents under assortative matching converge to 0, and so does welfare regardless of how well the allocation is targeted.[16]

4.2 Empirical tests

The two key statistical objects, *(i)* the dispersion of $\log(r)$ and *(ii)* the correlation of r with λv, can be estimated given sufficient data, under an appropriate structural assumption about social preferences (e.g., tying welfare weights λ to some observable characteristics). In this section, we provide a series of simple empirical tests as applications of our main results.

First, we show that measuring the dispersion of $\log(r)$ is analogous to estimating the price elasticity of a demand curve in a standard market with money. For any screening device D, let

$$P(D,s) := G^{-1}(1-s)$$

be the "inverse demand curve" for quality 1 of the good in the units of ordeal intensity y. The "ordeal elasticity" of demand, analogous to the price elasticity of demand, is then given by

$$\eta(D,s) := \left[\frac{\mathrm{d}\log P(D,s)}{\mathrm{d}\log s}\right]^{-1}.$$

For an illustration, suppose that the ordeal is waiting, and y measures the amount of time the agent must wait. Then, the ordeal elasticity η measures the percentage reduction in demand for quality-1 goods given a 1% increase in the wait time.

[16] Of course, for ε small enough, the designer will not induce assortative matching under the screening device D_ε; the optimal mechanism will instead allocate goods randomly without any ordeal.

Proposition 3. *For any two screening devices D_1 and D_2, $\log(r_2) \preceq_{disp} \log(r_1)$ if and only if $\eta(D_2, s) \leq \eta(D_1, s)$ for all $s \in (0,1)$.*

The proof is in the appendix. Proposition 3 shows that an empirical researcher could leverage standard demand estimation techniques in our setting by exploiting the variation in the quantity demanded given different levels of ordeal requirements.[17]

Next, we provide a series of simple regression tests.

Proposition 4. *If two screening devices D_1 and D_2 satisfy*

$$\log(r_2) = \beta \log(r_1) + \varepsilon$$

for some $\beta \in [0,1]$ and random variable ε that is independent of $(\lambda v, r_1)$, then screening device D_1 dominates screening device D_2.

Proposition 4 says that a researcher can run a simple one-sided empirical test by regressing $\log(r_2)$ on $\log(r_1)$ to determine the relative performance of two screening devices, provided that the two screening devices sort the agents in the same way up to idiosyncratic preference shocks.

The proof is in the appendix. Intuitively, a smaller coefficient β in Proposition 4 leads to a smaller dispersion of $\log(r_2)$. However, the effect of adding an idiosyncratic preference shock ε is ex-ante ambiguous: It increases the dispersion of $\log(r_2)$ but decreases the correlation of r_2 and λv. In general, when the rent-provision and targeting-effectiveness channels go in opposite directions, whether a screening device dominates another depends on the precise trade-off that is characterized in Theorem 1. Nevertheless, the proof shows that when an increase in the dispersion of rates of substitution is due to idiosyncratic preference shocks that are unrelated to social values, the trade-off is always resolved in favor of the targeting-effectiveness force. Intuitively, under screening device D_2, one can consider a "relaxed problem" in which the designer *observes* the noise term ε (which can only improve the performance of D_2). Conditional on observing ε, for each of these subproblems, Theorem 2 implies that the designer does no better than using screening device D_1.

Proposition 5. *Consider any screening devices D_1 and D_2 that satisfy*

$$\log(r_i) = \beta_i \log(\lambda v) + \varepsilon_i$$

[17] To the best of our understanding, the (natural) experiments of Alatas et al. (2016); Finkelstein and Notowidigdo (2019); Deshpande and Li (2019) would be sufficient to measure the ordeal elasticity at some points s.

for random variables $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ *that are independent of* λv*. If*

(i) $\beta_1 \geq \beta_2$

(ii) $\frac{\sigma_1}{\beta_1} \leq \frac{\sigma_2}{\beta_2}$

then screening device D_1 *dominates screening device* D_2*.*

The proof is in the appendix. Note that in part *(ii)* of Proposition 4, the ratio σ/β is smaller if and only if the *R-squared* $\text{Var}(\beta \log(\lambda v))/\text{Var}(\beta \log(\lambda v)+\varepsilon)$ of the regression equation is higher. Proposition 4 says that a researcher can also run a simple empirical test by regressing $\log(r)$ on $\log(\lambda v)$ and looking at both *(i)* the slope and *(ii)* the R-squared to determine the relative performance of two screening devices, provided that the error term in the regression is independent of the social values. Both the slope and the R-squared must be taken into account in this setting because of the two forces we identified. Intuitively, a larger slope means more variation in the rates of substitution *across* groups of different social values, and a larger R-squared means less variation in the rates of substitution *within* groups of different social values. This is consistent with the intuition provided in Proposition 4 and further refines the intuition in Theorem 2: A screening device is better if it induces more dispersion in the rates of substitution that is due to systematic variation across groups of different social values rather than due to idiosyncratic preference shocks.

4.3 Combining screening devices

In practice, different types of ordeals may be used jointly to guide the allocation of resources. For example, applying for a driver's license in the US requires traveling to a DMV site and then waiting in line. In other contexts, such as applying for Social Security Disability Insurance, the applicants can *choose* between visiting an office, making a phone call, or filling out an online form.[18]

We can accommodate combinations of screening devices in our model as follows. For two screening devices D_1, D_2, with units of difficulty y_1 and y_2, respectively, an *additive scoring rule* scores the agents by creating a new unit of difficulty

$$y_3 = y_1 + \beta y_2$$

[18] As pointed out in Deshpande and Li (2019), filing a claim online may be quite costly, especially for applicants with low education and earnings levels; the forms tend to be complicated, and the first two methods allow the applicant to be assisted by a social security officer.

where $\beta > 0$ is a constant. Given an additive scoring rule, the two tasks are *substitutes* and hence the agents only perform the cost-effective task for them. Thus, an additive scoring rule induces a new screening device D_3 defined by

$$c_3 = \min\Big(c_1, c_2/\beta\Big).$$

Alternatively, a *minimum scoring rule* scores the agents by creating a unit of difficulty

$$y_3 = \min\Big(y_1, \beta y_2\Big),$$

where $\beta > 0$ is a constant. Given a minimum scoring rule, the two tasks are *complements* and hence the agents always perform both tasks. Thus, a minimum scoring rule induces a new screening device D_3 defined by

$$c_3 = c_1 + c_2/\beta\,.$$

The following result provides conditions under which social welfare is *reduced* by giving agents a choice between screening devices or by bundling them together.

Proposition 6. *Consider two screening devices D_1 and D_2 that satisfy $r_2 = \psi(r_1,\, \varepsilon)$ for some random variable ε independent of $(\lambda v, r_1)$ and some continuously differentiable function ψ such that $0 \leq \frac{\mathrm{d}\log\psi(r,\varepsilon)}{\mathrm{d}\log r} \leq 1$ for all r, almost surely in ε. Let D_3 be any screening device induced by an additive scoring rule (choosing between tasks) or a minimum scoring rule (bundling tasks). Then, screening device D_1 dominates screening device D_3.*

The proof is in the appendix. The condition in Proposition 6 is a generalization of that in Proposition 4, and implies that, conditional on any realization of the idiosyncratic noise ε, *(i)* $\log(r_2)$ is less dispersed than $\log(r_1)$, and *(ii)* (λ, r_2) has the same copula as (λ, r_1). By Theorem 2, this implies that screening device D_1 dominates screening device D_2. The proof then shows that every additive or minimum scoring rule *preserves* this condition.

Proposition 6 is limited in that it only states that a dominating screening device does not improve when it is combined with a dominated one (either by giving agents a choice, or by requiring them to complete both tasks). In general, taking the sum or the minimum of costs associated with different (potentially non-ranked) ordeals will tend to reduce the dispersion in the left tail of the distribution, and thus have a negative effect through the rent-provision channel. However, a comprehensive analysis of combinations of screening devices requires studying the joint distribution of $(c_1,\, c_2,\, \lambda,\, v)$, and is beyond the scope of

this paper.

4.4 Allocating money

In many allocation problems involving ordeals, the good being allocated is money. In our model—because we chose to measure all values in monetary units—allocating money corresponds to assuming that $v \equiv 1$. In this special case, some of our results become easier to apply and interpret.

Corollary 1. *Suppose that $v \equiv 1$. Then, if two screening devices D_1 and D_2 satisfy*

$$\log(c_2) = \beta \log(c_1) + \varepsilon$$

for some random variable ε that is independent of (λ, c_1), and some $\beta \in [0,1]$, then screening device D_1 dominates screening device D_2.

Corollary 1 is a special case of Proposition 4. When the planner allocates money, running the regression only requires estimating agents' costs for completing the ordeal, and the conditions for the validity of the test become more plausible. Moreover, the result is easier to interpret. For example, in the special case of $\beta = 1$, Corollary 1 states that adding independent noise to the cost makes a screening device worse. Suppose that the ordeal is waiting, and the planner considers introducing some amenities that will make the experience of waiting less unpleasant. As long as the associated reduction in individual waiting costs is not systematically related to welfare-relevant characteristics, introducing the amenities will tend to reduce welfare.[19]

Corollary 2. *Suppose that $v \equiv 1$. Consider two screening devices D_1 and D_2 that satisfy $c_2 = \psi(c_1, \varepsilon)$ for some random variable ε independent of (λ, c_1) and some continuously differentiable function ψ such that $0 \leq \frac{\mathrm{d}\log\psi(c,\varepsilon)}{\mathrm{d}\log c} \leq 1$ for all c, almost surely in ε. Let D_3 be any screening device induced by an additive scoring rule (choosing between tasks) or a minimum scoring rule (bundling tasks). Then, screening device D_1 dominates screening device D_3.*

Corollary 2 is a special case of Proposition 6. The assumptions for the result hold, for example, when either *(i)* c_2 is independent of (λ, c_1) or *(ii)* c_2 is comonotonic with c_1 and $\log(c_2) \preceq_{disp} \log(c_1)$. For example, forcing agents to complete an additional ordeal whose cost is unrelated to welfare-relevant characteristics leads to a decrease in social welfare.

A simple consequence of Corollary 2 is as follows.

[19] Of course, this is because a reduction in costs in our model implies that the difficulty of the ordeal (waiting time) must be increased to equalize supply and demand.

Corollary 3. *Suppose that $v \equiv 1$. Consider screening devices D, D_1, and D_2 satisfying $c_1 = c - k_1$, and $c_2 = \min(c, k_2)$, where $k_1, k_2 > 0$. Then, screening device D_1 dominates screening device D which in turn dominates screening device D_2.*

Corollary 3 further highlights the fact that decreasing the costs of the ordeal can have ambiguous effects on welfare; what matters is how the reduction in costs affects the rent-provision channel. For intuition, recall that costs are measured in monetary units, so we can interpret constants k_1 and k_2 as representing monetary outcomes. Reducing the costs of the ordeal by a constant k_1 (e.g., by eliminating a fee that agents had to pay to a third party) is welfare-enhancing. Letting agents pay an amount of money k_2 instead of completing the ordeal (e.g., by "delegating" completing the ordeal to a third party, when the market price for such a service is k_2) is welfare-reducing. This is because these two changes have a different effect on the dispersion of costs on the log scale.

5 Concluding Remarks

We provided conditions to robustly compare different screening devices. We highlighted two main forces that drive the performance of a screening device: *(i)* targeting effectiveness and *(ii)* rent provision. We showed that these two forces are respectively characterized by *(i)* the correlation between the rates of substitution and the social values and *(ii)* the dispersion in the marginal distribution of the rates of substitution. We applied this decomposition to provide a series of simple empirical tests that could be useful in practice. We conclude the paper by discussing a few additional points and extensions.

5.1 Costly screening versus money

In the vast literature on screening in market-design contexts, costly screening (sometimes referred to as "money burning") is typically motivated by citing various institutional, ethical, and technological reasons for why monetary transfers might not be available (see, for example, Hartline and Roughgarden 2008, Condorelli 2012, Chakravarty and Kaplan 2013). In this section, we show that screening via ordeals in a setting with welfare weights should not be regarded as a "third-best" solution that may only be used when monetary transfers are infeasible; in fact, costly screening may *dominate* monetary transfers as a screening device.[20]

[20] We do not consider the problem of whether and how monetary transfers can be combined with costly screening—see Yang (2022) for a recent contribution along these lines.

Monetary transfers can be considered as a screening device D_M that has costs c identically equal to 1 (because we measure costs in monetary units). The only qualitative difference to a costly screening device in our framework is that using money has the additional benefit of raising revenue which can then be redistributed back to agents through a lump-sum transfer.[21] Effectively, costs incurred by agents can be recovered, which is what makes money "not costly" as a screening device. To model this, when using money to screen, we can let the designer solve

$$\mathrm{OPT}_M(F) := \sup_{(\Gamma,y)\in\mathcal{M}(D_M,F)} \mathbb{E}\Big[\lambda\Big(vx^\Gamma(v) - y(v)\Big)\Big] + \mathbb{E}\Big[\lambda\Big]\cdot\mathbb{E}\Big[y(v)\Big],$$

where the first term is the same as for any other screening device (with $c \equiv 1$) and the second term captures the benefit of redistribution through lump-sum transfers. We say that screening device D *strictly dominates monetary transfers* if

$$\mathrm{OPT}(D,F) \geq \mathrm{OPT}_M(F)$$

for every quality distribution F on $[0,1]$ and the inequality is strict for at least some F.

Proposition 7. *There exists a distribution of (λ, v) such that using monetary transfers is strictly dominated by some costly screening device D. There also exists a distribution of (λ, v) such that using monetary transfers strictly dominates every costly screening device D.*

The proof is in the appendix. Consider part *(i)* of Proposition 7 first. A simple example in which monetary transfers are strictly dominated is when $v \equiv 1$ and there is some dispersion in λ—money is useless as a screening device when the designer allocates money itself. Beyond this trivial case, the proof provides an example in which there is variation in v (so that money can be used to screen), but there is a strong positive correlation between λ and v. Because agents with high welfare weights end up paying the most (while only receiving part of the money back as a lump-sum transfer), welfare is bounded away from the first-best welfare achievable *without* monetary transfers. On the other hand, using a similar construction as in Proposition 1, we show that there exists a costly screening device with sufficiently good targeting effectiveness and sufficiently dispersed rates of substitution that can get arbitrarily close to the first best.

In light of Proposition 1, the second part of Proposition 7 may actually be seen as more surprising. The key observation is that—with monetary transfers available—allocating the good to agents with the highest λv at no cost is no longer the first best in general.

[21] We could allow for other productive uses of money without affecting our conclusions.

The proof considers a case when λ and v are non-degenerate random variables, but they are negatively correlated so that $\lambda v \equiv 1$. Then, the designer is indifferent between agents when allocating the good, and therefore cannot benefit from using any costly screening device. However, the designer is not indifferent between agents when it comes to allocating money, since there is dispersion in λ. When using monetary transfers to screen, the designer sells to agents with high v and redistributes the revenue as a lump-sum payment, which means that agents with high λ receive a net positive monetary transfer. Thus, the designer improves welfare by redistributing money.

5.2 Extensions

Nonoptimal mechanisms. When comparing screening devices, we assumed that the designer chooses an optimal allocation mechanism for each screening device. However, our results also apply if the designer uses certain market and rationing mechanisms that need not be optimal. Consider the class of α*−market mechanisms*, for any $\alpha \in [0, 1]$, that allocate a mass α of goods with the highest qualities under distribution F using assortative matching between qualities and rates of substitution (the *market mechanism*), and allocate the remaining goods uniformly at random to agents who do not complete any ordeal (the *rationing mechanism*). In particular, the market mechanism arises if difficulty levels of the ordeal are adjusted endogenously to equalize demand and supply for every quality level.[22] Both Theorem 1 and Theorem 2 continue to hold if the designer uses an α−market mechanism. In fact, condition (1) in Theorem 1 is both sufficient and *necessary* for one screening device to dominate another when the designer uses the market mechanism.

Endogenous quality distributions. In the model, the quality distribution was assumed to be exogenously given. However, as explained earlier, our results are unaffected if the quality distribution F is optimally chosen (for any screening device) by the designer subject to a production cost function $\mathcal{C} : \Delta([0,1]) \to \mathbb{R}$. All of our results hold for any production cost functions $\mathcal{C} : \Delta([0,1]) \to \mathbb{R}$.

For one screening device to dominate another, we required that they are ranked in the same way for all quality distributions $F \in \Delta([0,1])$ (or, equivalently, for all supply levels of a homogeneous-quality good). More generally, one may define the notion of dominance to be with respect to a subset of quality distributions $\mathcal{F} \subseteq \Delta([0,1])$, which is particularly natural when producing many goods of high quality is prohibitively costly. Moreover, the designer may have prior knowledge about supply at the time of choosing between ordeals.

[22] This is a particularly natural assumption for ordeals involving waiting time; see Ashlagi et al. (2022).

Because our analysis focused on *sufficient* conditions, our results continue to hold for any $\mathcal{F}$. However, this extension would lead to a more complete ordering of costly screening devices. For example, if $\mathcal{F}$ is the set of distributions on $\{0,1\}$ such that the mass on 1 does not exceed $\bar{s}$ (supply is at most $\bar{s}$), then we only need to compare the welfare index $W(D,s)$ in Theorem 1 for all $s \in [1-\bar{s}, 1]$.

Partially productive ordeals. Ordeals in our analysis were assumed not to serve any productive purpose. In particular, the social cost of an agent with welfare weight λ completing an ordeal with difficulty 1 is λc. One may consider a more general model in which y units of the task generate a social benefit $u \cdot y$, where u can be correlated with (λ, v, c). For example, in guaranteed employment programs (such as the one created by India's National Rural Employment Guarantee Act), y is the number of hours worked, and u is the agent's productivity (that can be correlated, for example, with welfare weights). In this more general model, the designer would solve

$$\sup_{(\Gamma,y)\in\mathcal{M}(D,F)} \mathbb{E}_D\Big[\lambda\Big(vx^\Gamma(r) - cy(r)\Big)\Big] + \mathbb{E}_D\Big[u \cdot y(r)\Big],$$

where D is now a mapping from each (λ, v) to a distribution over (c, u).

While we leave this question to future studies, we note that our results can accommodate some special cases. When $u = \mathbb{E}[\lambda]$ and $c \equiv 1$, this setting reduces to that in Section 5.1 where y is interpreted as a monetary transfer. Moreover, our results continue to hold when $u = \alpha\lambda c$ for some constant $\alpha \in (0,1)$, which can be interpreted as saying that the benefit is proportional to the social cost.[23]

References

Akbarpour, M. ⓡ P. Dworczak ⓡ S. D. Kominers (2023): "Redistributive Allocation Mechanisms," *Journal of Political Economy (forthcoming)*.

Alatas, V., R. Purnamasari, M. Wai-Poi, A. Banerjee, B. A. Olken, and R. Hanna (2016): "Self-targeting: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy*, 124(2), 371–427.

[23] The welfare index in Theorem 1 would be modified to be $W(D,s) := \int_s^1 \left(1 - (1-\alpha)\cdot\frac{G^{-1}(s)}{G^{-1}(t)}\right)\cdot V(t)\,\mathrm{d}t$. All of the remaining results remain unchanged.

Ashlagi, I., J. Leshno, P. Qian, and A. Saberi (2022): "Price Discovery in Waiting Lists," *Available at SSRN: https://ssrn.com/abstract=4192003*.

Blackwell, D. (1953): "Equivalent Comparisons of Experiments," *Annals of Mathematical Satistics*, 24(2), 265–272.

Chakravarty, S. and T. R. Kaplan (2013): "Optimal Allocation without Transfer Payments," *Games and Economic Behavior*, 77(1), 1–20.

Condorelli, D. (2012): "What Money Can't Buy: Efficient Mechanism Design with Costly Signals," *Games and Economic Behavior*, 75(2), 613 – 624.

——— (2013): "Market and Non-Market Mechanisms for the Optimal Allocation of Scarce Resources," *Games and Economic Behavior*, 82, 582–591.

Deshpande, M. and Y. Li (2019): "Who Is Screened Out? Application Costs and the Targeting of Disability Programs," *American Economic Journal: Economic Policy*, 11(4), 213–48.

Dworczak, P. (2023): "Equity-Efficiency Trade-Off in Quasi-linear Environments," *Working Paper*.

Dworczak, P. ⓡ S. D. Kominers ⓡ M. Akbarpour (2021): "Redistribution through Markets," *Econometrica*, 89(4), 1665–1698.

Finkelstein, A. and M. J. Notowidigdo (2019): "Take-Up and Targeting: Experimental Evidence from SNAP," *Quarterly Journal of Economics*, 134(3), 1505–1556.

Hartline, J. D. and T. Roughgarden (2008): "Optimal Mechanism Design and Money Burning," in *Proceedings of the 40th Annual ACM SIGACT Symposium on Theory of Computing*, STOC '2008, 75–84.

Kleven, H. J. and W. Kopczuk (2011): "Transfer Program Complexity and the Take-Up of Social Benefits," *American Economic Journal: Economic Policy*, 3(1), 54–90.

Müller, A. and D. Stoyan (2002): *Comparison Methods for Stochastic Models and Risks*, Wiley.

Nichols, A. L. and R. J. Zeckhauser (1982): "Targeting Transfers through Restrictions on Recipients," *American Economic Review*, 72(2), 372–377.

Rose, J. L. (2021): "Rationing with Time: Time-cost Ordeals' Burdens and Distributive Effects," *Economics & Philosophy*, 37(1), 50–63.

Shaked, M. (1982): "Dispersive Ordering of Distributions," *Journal of Applied Probability*, 19(2), 310–320.

Shaked, M. and J. G. Shanthikumar (2007): *Stochastic Orders*, Springer.

Tchen, A. H. (1980): "Inequalities for Distributions with Given Marginals," *Annals of Probability*, 8(4), 814–827.

Yang, F. (2022): "Costly Multidimensional Screening," *Working Paper*.

——— (2023): "The Simple Economics of Optimal Bundling," *Working Paper*.

Zeckhauser, R. (2021): "Strategic Sorting: The Role of Ordeals in Health Care," *Economics & Philosophy*, 37(1), 64–81.

A Preliminaries on Stochastic Orders

A.1 Variability orders

For a CDF F, define the generalized inverse of F by

$$F^{-1}(t) := \inf\{x \in \mathbb{R} : F(x) \geq t\}.$$

Let X and Y be two random variables with CDFs F_X and F_Y, respectively. X is said to be smaller than Y in the *dispersive order* ($X \preceq_{disp} Y$) if

$$F_X^{-1}(t) - F_X^{-1}(s) \leq F_Y^{-1}(t) - F_Y^{-1}(s) \quad \text{for all } 0 < s < t < 1\,.$$

The dispersive order is location-invariant in the sense that

$$X \preceq_{disp} Y \iff X + a \preceq_{disp} Y \quad \text{for any } a \in \mathbb{R}.$$

Provided that $\mathbb{E}[X] = \mathbb{E}[Y]$, we have (see e.g., Shaked and Shanthikumar 2007)

$$X \preceq_{disp} Y \implies X \preceq_{mps} Y \iff Y \preceq_{ssd} X \iff F_X^{-1} \preceq_{maj} F_Y^{-1} \iff F_Y \preceq_{maj} F_X\,,$$

where $\preceq_{mps}$ is the *mean-preserving spread order* (or the *convex order*)[24] and $\preceq_{ssd}$ is the standard *second-order stochastic dominance*.[25]

A useful fact about the dispersive order is the following characterization:

Theorem 3 (Shaked 1982)**.** *For any two random variables X and Y, we have that $X \preceq_{disp} Y$ if and only if*

$$X \overset{d}{=} \phi(Y)$$

for some non-decreasing function ϕ that satisfies $\phi(y') - \phi(y) \leq y' - y$ for all $y < y'$.

From Theorem 3, it follows easily that for any random variable X, we have

$$X \preceq_{disp} aX \qquad \text{whenever } a \geq 1\,, \tag{A1}$$

and that, for any random variable X, and two differentiable, non-decreasing functions ϕ, ψ, we have

$$\phi'(x) \leq \psi'(x) \text{ for all } x \implies \phi(X) \preceq_{disp} \psi(X). \tag{A2}$$

The above implies that for any random variable X, and any two differentiable, non-decreasing functions $\phi > 0, \psi > 0$, we have

$$\frac{\phi'(x)}{\phi(x)} \leq \frac{\psi'(x)}{\psi(x)} \text{ for all } x \implies \log \phi(X) \preceq_{disp} \log \psi(X). \tag{A3}$$

A.2 Positive dependence orders

Let $X = (X_1, X_2)$ be a random vector. Let

$$H(x_1, x_2) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2)$$

be the joint distribution function of X. Let F_1 and F_2 be the marginal cumulative distribution functions of X. The *copula* $C : [0,1]^2 \to [0,1]$ of X is a joint distribution function such that, for all (x_1, x_2),

$$H(x_1, x_2) = C\Big(F_1(x_1), F_2(x_2)\Big).$$

Sklar's theorem asserts that for any joint distribution H, *(i)* there exists a copula C and *(ii)* the copula is unique on $\mathrm{Ran}(F_1) \times \mathrm{Ran}(F_2)$.[26] If the marginal distributions F_1, F_2 are

[24] That is, $X \preceq_{mps} Y$ if $\mathbb{E}[h(X)] \leq \mathbb{E}[h(Y)]$ for all convex functions h provided the expectations exist.

[25] That is, $Y \preceq_{ssd} X$ if $\int_{-\infty}^{z} F_Y(y)\,\mathrm{d}y \geq \int_{-\infty}^{z} F_X(x)\,\mathrm{d}x$ for all $z \in \mathbb{R}$.

[26] When the copula is not unique on $[0,1]\times[0,1]$, it suffices to arbitrarily choose a copula for our purposes.

continuous, then the copula is uniquely defined by

$$C(x_1, x_2) = H\Big(F_1^{-1}(x_1), F_2^{-1}(x_2)\Big).$$

For any two random vectors $X = (X_1, X_2)$, $Y = (Y_1, Y_2)$, X is said to be smaller than Y in the *correlation order* if

$$C_X(s,t) \leq C_Y(s,t) \quad \text{for all } s,t \in [0,1].$$

This order is also known as the *concordance order* or the *positive dependence order* (Shaked and Shanthikumar 2007).[27]

A useful fact about the correlation order is the following characterization:

Theorem 4 (Tchen 1980)**.** *For any random vectors X, Y that have the same marginal distributions, we have that $(X_1, X_2) \preceq_{cor} (Y_1, Y_2)$ if and only if*

$$\mathbb{E}\Big[\phi(X_1, X_2)\Big] \leq \mathbb{E}\Big[\phi(Y_1, Y_2)\Big] \text{ for all supermodular functions } \phi,$$

provided the expectations exist.

From Theorem 4, it follows that correlation measures such as Pearson's r, Kendall's τ, and Spearman's ρ are all monotone with respect to the correlation order.

From Theorem 4, it also follows that the correlation order is invariant under monotone transformations. Specifically, for any random vectors X, Y and any two strictly increasing functions φ, ψ, we have

$$(X_1, X_2) \preceq_{cor} (Y_1, Y_2) \implies \Big(\varphi(X_1), \psi(X_2)\Big) \preceq_{cor} \Big(\varphi(Y_1), \psi(Y_2)\Big). \tag{A4}$$

By Theorem 4, note that condition *(ii)* of Theorem 2, $V_2 \preceq_{maj} V_1$, which is equivalent to

$$\mathbb{E}[\lambda v \mid r_2 \geq G_2^{-1}(s)] \leq \mathbb{E}[\lambda v \mid r_1 \geq G_1^{-1}(s)] \quad \text{for all } s \in [0,1],$$

is implied by

$$(\lambda v, r_2) \preceq_{cor} (\lambda v, r_1)$$

because $\phi(x_1, x_2) := x_1 \cdot \mathbb{1}_{x_2 \geq s}$ is a supermodular function for all $s \in [0,1]$.

[27] This order is usually defined for two random vectors that have the same marginals, which we extend to potentially different marginals by comparing the copulas rather than the joint distribution functions.

B Proofs

B.1 Proof of Theorem 1

Fix any screening device D. First, observe that instead of the rate of substitution r, the designer can equivalently elicit the corresponding *quantile* t because there is a one-to-one mapping $r = G^{-1}(t)$. Second, because both the designer and the agents are risk neutral, any feasible probabilistic assignments that give the same expected quality would be payoff equivalent. Thus, we can directly work with *expected quality quantile allocation* $x : [0,1] \to [0,1]$. An agent's payoff can be written as $G^{-1}(t) \cdot x(t) - y(t)$, after normalizing by c. By the envelope theorem, letting $U[x](t)$ denote the utility of the agent at the $t-$quantile under allocation rule x, we have that

$$U[x](t) = \int_0^t (G^{-1})'(s)x(s)\,\mathrm{d}s + G^{-1}(0)x(0), \tag{B1}$$

where we used the observation that the designer would never want to set $y(0) > 0$ and thus $U(0) = G^{-1}(0)x(0)$. This, in particular, implies that $U[x]$ is a linear functional of x.

Note that we can write the objective as

$$\mathbb{E}\left[\lambda v\left(x(t) - \frac{1}{G^{-1}(t)}y(t)\right)\right] = \int_0^1 \mathbb{E}\left[\lambda v \mid r = G^{-1}(t)\right] \cdot \frac{1}{G^{-1}(t)} U[x](t)\,\mathrm{d}t. \tag{B2}$$

Moreover, observe also that under the allocation rule $x(t) = \mathbb{1}_{t \geq k}$ for any $k > 0$, we have

$$U[x](t) = \mathbb{1}_{t \geq k}\left(G^{-1}(t) - G^{-1}(k)\right),$$

and for $k = 0$, we have $U[x](t) = G^{-1}(t)$. With (B2), these imply that for $x(t) = \mathbb{1}_{t \geq k}$ we may write the objective as (recalling that $V(t) = \mathbb{E}[\lambda v \mid r = G^{-1}(t)]$)

$$\int_k^1 V(t) \cdot \left(1 - \frac{G^{-1}(k)}{G^{-1}(t)}\right)\mathrm{d}t \tag{B3}$$

for any $k > 0$, and $\mathbb{E}[\lambda v]$ for $k = 0$.

Suppose that condition (1) in Theorem 1 holds for all $s \in [0,1]$. Then by (B3), we have that

$$\int_0^1 V_1(t) \cdot \frac{1}{G_1^{-1}(t)} U_1[\mathbb{1}_{t \geq s}](t)\,\mathrm{d}t \geq \int_0^1 V_2(t) \cdot \frac{1}{G_2^{-1}(t)} U_2[\mathbb{1}_{t \geq s}](t)\,\mathrm{d}t \tag{B4}$$

holds for all $s \in (0,1]$. For $s = 0$, the above also holds since they both equal $\mathbb{E}[\lambda v]$. Now

consider any feasible expected quality quantile allocation x under screening device D_2. Note that x is also feasible under screening device D_1 since we are working in the quantile space. By standard arguments, the IC constraints imply that $x : [0,1] \to [0,1]$ is a monotone function. Recall that, by Choquet's theorem, every monotone function can be represented as a mixture over functions $\mathbb{1}_{t \geq s}$. In particular, we have the following:

Lemma 1. *For every monotone function $x : [0,1] \to [0,1]$, there exists a probability measure $\mu \in \Delta([0,1])$ such that for all $t \in [0,1]$,*

$$x(t) = \int \mathbb{1}_{t \geq s} \, \mathrm{d}\mu(s).$$

Let μ be the measure that represents the monotone allocation rule $x : [0,1] \to [0,1]$. By linearity of the functionals $U_1[x]$ and $U_2[x]$, we immediately have that

$$\begin{aligned}
\int_0^1 V_1(t) \cdot \frac{1}{G_1^{-1}(t)} U_1[x](t) \, \mathrm{d}t &= \int_{[0,1]} \left(\int_0^1 V_1(t) \cdot \frac{1}{G_1^{-1}(t)} U_1[\mathbb{1}_{t \geq s}](t) \, \mathrm{d}t \right) \mathrm{d}\mu(s) \\
&\geq \int_{[0,1]} \left(\int_0^1 V_2(t) \cdot \frac{1}{G_2^{-1}(t)} U_2[\mathbb{1}_{t \geq s}](t) \, \mathrm{d}t \right) \mathrm{d}\mu(s) \\
&= \int_0^1 V_2(t) \cdot \frac{1}{G_2^{-1}(t)} U_2[x](t) \, \mathrm{d}t.
\end{aligned}$$

Therefore, by (B2), we must have that the designer's payoff with the quantile allocation rule x under D_1 is weakly higher than that under D_2. Since this holds for every feasible expected quality quantile allocation rule x, device D_1 must dominate device D_2.

B.2 Proof of Theorem 2

First, observe that $\log(r_2) \preceq_{disp} \log(r_1)$ implies

$$\log G_2^{-1}(t) - \log G_2^{-1}(s) \leq \log G_1^{-1}(t) - \log G_1^{-1}(s) \quad \text{for all } 0 < s < t < 1,$$

and hence

$$1 - \frac{G_2^{-1}(s)}{G_2^{-1}(t)} \leq 1 - \frac{G_1^{-1}(s)}{G_1^{-1}(t)} \quad \text{for all } 0 < s < t < 1.$$

Thus, for all $s \in [0,1]$, we have

$$W(D_2, s) = \int_s^1 \left(1 - \frac{G_2^{-1}(s)}{G_2^{-1}(t)}\right) \cdot V_2(t) \, \mathrm{d}t \leq \int_s^1 \left(1 - \frac{G_1^{-1}(s)}{G_1^{-1}(t)}\right) \cdot V_2(t) \, \mathrm{d}t.$$

Now fix any $s \in [0,1]$ and write

$$\int_s^1 \left(1 - \frac{G_1^{-1}(s)}{G_1^{-1}(t)}\right) \cdot V_2(t)\,\mathrm{d}t = \int_0^1 \mathbb{1}_{t \geq s}\left(1 - \frac{G_1^{-1}(s)}{G_1^{-1}(t)}\right) \cdot V_2(t)\,\mathrm{d}t.$$

Since

$$h(t) := \mathbb{1}_{t \geq s}\left(1 - \frac{G_1^{-1}(s)}{G_1^{-1}(t)}\right)$$

is a monotone function from $[0,1]$ to $[0,1]$, by the representation theorem again (Lemma 1), there exists a probability measure μ over $[0,1]$ such that for all $t \in [0,1]$

$$h(t) = \int_{[0,1]} \mathbb{1}_{t \geq k}\,\mathrm{d}\mu(k).$$

This implies that

$$\begin{aligned}
W(D_2, s) \leq \int_0^1 h(t) V_2(t)\,\mathrm{d}t &= \int_{[0,1]} \left(\int_0^1 \mathbb{1}_{t \geq k}\, \mathbb{E}\left[\lambda v \mid r_2 = G_2^{-1}(t)\right] \mathrm{d}t \right) \mathrm{d}\mu(k) \\
&= \int_{[0,1]} \left(1 - k\right) \mathbb{E}\left[\lambda v \mid r_2 \geq G_2^{-1}(k)\right] \mathrm{d}\mu(k) \\
&\leq \int_{[0,1]} \left(1 - k\right) \mathbb{E}\left[\lambda v \mid r_1 \geq G_1^{-1}(k)\right] \mathrm{d}\mu(k) \\
&= \int_{[0,1]} \left(\int_0^1 \mathbb{1}_{t \geq k}\, \mathbb{E}\left[\lambda v \mid r_1 = G_1^{-1}(t)\right] \mathrm{d}t \right) \mathrm{d}\mu(k) \\
&= \int_0^1 h(t) V_1(t)\,\mathrm{d}t = W(D_1, s),
\end{aligned}$$

where the inequality follows from the majorization condition $V_2 \preceq_{maj} V_1$. Since this holds for all $s \in [0,1]$, screening device D_1 must dominate screening device D_2 by Theorem 1.

B.3 Proof of Proposition 1

Let

$$c = \frac{v}{\exp(\gamma H(\lambda v))},$$

where H is the CDF of λv and $\gamma > 0$ is a parameter. Then, $(\lambda v, r)$ are comonotonic, and $G^{-1}(t) = \exp(\gamma t)$. This implies that

$$W(D_\gamma, s) = \int_s^1 \left(1 - e^{-\gamma(t-s)}\right) \cdot H^{-1}(t)\,\mathrm{d}t.$$

Let $K_\gamma(s) := W(D_\gamma, s)$. Let $K(s) = \int_s^1 H^{-1}(t)\,\mathrm{d}t$. First, by the monotone convergence theorem, for any s, we have

$$\lim_{\gamma\to\infty} K_\gamma(s) = \int_s^1 \lim_{\gamma\to\infty} \left(1 - e^{-\gamma(t-s)}\right) \cdot H^{-1}(t)\,\mathrm{d}t = \int_s^1 H^{-1}(t)\,\mathrm{d}t = K(s).$$

Thus K_γ converges to K pointwise. Now, note that *(i)* $K_\gamma(s)$ is continuous in $s \in [0,1]$, *(ii)* $K_\gamma(s)$ is monotone in γ, and *(iii)* $K(s)$ is continuous in $s \in [0,1]$. By Dini's theorem, the functions K_γ converge uniformly to K. Thus, for any $\varepsilon > 0$, there exists γ such that for all $s \in [0,1]$, we have

$$\left|K_\gamma(s) - K(s)\right| < \varepsilon.$$

Let $x(t) = F^{-1}(t)$ and μ be the probability measure that represents $x(t)$, i.e., $x(t) = \int \mathbb{1}_{t\geq s}\,\mathrm{d}\mu(s)$ (by Lemma 1). The first-best welfare is given by

$$\int_0^1 K(s)\,\mathrm{d}\mu(s).$$

By the proof of Theorem 1, the welfare under screening device D_γ and allocation rule $x(\cdot)$ is given by

$$\int_0^1 \widetilde{K}_\gamma(s)\,\mathrm{d}\mu(s),$$

where $\widetilde{K}_\gamma(s) := K_\gamma(s)$ if $s \in (0,1]$ and $\widetilde{K}_\gamma(s) := \mathbb{E}[\lambda v] = K(s)$ if $s = 0$. Clearly, for all $s \in [0,1]$, we also have $|\widetilde{K}_\gamma(s) - K(s)| < \varepsilon$. Thus, we have

$$\left|\int_0^1 \widetilde{K}_\gamma(s)\,\mathrm{d}\mu(s) - \int_0^1 K(s)\,\mathrm{d}\mu(s)\right| \leq \int_0^1 \left|\widetilde{K}_\gamma(s) - K(s)\right|\mathrm{d}\mu(s) < \varepsilon,$$

proving the claim.

B.4 Proof of Proposition 2

We use a similar construction as in the proof of Proposition 1. We will construct D_ε to achieve perfect targeting. Let

$$c = \frac{v}{\exp(\gamma H(\lambda v))},$$

where H is the CDF of λv and $\gamma > 0$ is a parameter. Then, $(\lambda v, r)$ are comonotonic (i.e., maximal positive correlation), and $G^{-1}(t) = \exp(\gamma t)$. This also implies that

$$W(D_\gamma, s) = \int_s^1 \left(1 - e^{-\gamma(t-s)}\right) \cdot H^{-1}(t)\,\mathrm{d}t\,.$$

Let $K_\gamma(s) := W(D_\gamma, s)$. First, by the monotone convergence theorem, for any s, we have

$$\lim_{\gamma\to 0} K_\gamma(s) = \int_s^1 \lim_{\gamma\to 0}\left(1 - e^{-\gamma(t-s)}\right) \cdot H^{-1}(t)\,\mathrm{d}t = 0\,.$$

Thus K_γ converges to 0 pointwise. Now, note that *(i)* $K_\gamma(s)$ is continuous in $s \in [0,1]$, *(ii)* $K_\gamma(s)$ is monotone in γ. By Dini's theorem, the functions K_γ converge uniformly to 0.

Let $x(t) = F^{-1}(t)$ and μ be the probability measure that represents $x(t)$, that is, $x(t) = \int \mathbb{1}_{t\geq s}\,\mathrm{d}\mu(s)$ (by Lemma 1). Note that $F^{-1}(0) = 0$ and hence $x(0) = 0$. By the proof of Theorem 1, this implies that the welfare under screening device D_γ and allocation rule $x(\cdot)$ is given by

$$\int_0^1 K_\gamma(s)\,\mathrm{d}\mu(s)\,.$$

The claim then follows by the same argument as in the proof of Proposition 1.

B.5 Proof of Proposition 3

Note that

$$\log(r_2) \preceq_{disp} \log(r_1)$$

if and only if

$$\log G_1^{-1}(s) - \log G_2^{-1}(s) \text{ is non-decreasing in } s\,,$$

which holds if and only if

$$\frac{\mathrm{d}}{\mathrm{d}s}\log G_1^{-1}(s) \geq \frac{\mathrm{d}}{\mathrm{d}s}\log G_2^{-1}(s) \text{ for all } s \in (0,1)\,,$$

which holds if and only if

$$\frac{\mathrm{d}}{\mathrm{d}\log s}\log G_1^{-1}(1-s) \leq \frac{\mathrm{d}}{\mathrm{d}\log s}\log G_2^{-1}(1-s) \text{ for all } s \in (0,1)\,,$$

which holds if and only if

$$\left[\frac{\mathrm{d}}{\mathrm{d}\log s}\log G_1^{-1}(1-s)\right]^{-1} \geq \left[\frac{\mathrm{d}}{\mathrm{d}\log s}\log G_2^{-1}(1-s)\right]^{-1} \text{ for all } s \in (0,1).$$

Note that the left-hand side is exactly $\eta(D_1, s)$ and the right-hand side is exactly $\eta(D_2, s)$.

B.6 Proof of Proposition 4

We will prove a stronger result that we will also need in the proof of Proposition 6.

Lemma 2. *Consider two screening devices D_1 and D_2 that satisfy*

$$r_2 = \psi\big(r_1,\, \varepsilon\big)$$

for some random variable ε independent of $(\lambda v, r_1)$ and some continuously differentiable function ψ such that $0 \leq \frac{\mathrm{d}\log\psi(r,\varepsilon)}{\mathrm{d}\log r} \leq 1$ for all r, almost surely in ε. Then, screening device D_1 dominates screening device D_2.

Proof of Lemma 2. Let t_1 be the quantile for r_1. Let $\mathcal{E}$ be the set of ε at which $0 \leq \frac{\mathrm{d}\log\psi(r,\varepsilon)}{\mathrm{d}\log r} \leq 1$. By assumption, $\mathbb{P}(\varepsilon \in \mathcal{E}) = 1$. Consider the mechanism design problem under screening device D_2. We can let $x(t_1, \varepsilon)$ be the allocation rule that elicits reports on (t_1, ε). Note that since x is monotone in $\psi(G_1^{-1}(t_1), \varepsilon)$, we have that $x(\cdot, \varepsilon)$ is monotone for all $\varepsilon \in \mathcal{E}$. Suppose that we let the designer *observe* ε and then optimize over y. This provides an upper bound on the designer's payoff under allocation rule x. By the proof of Theorem 1, the designer's payoff in this case is given by

$$\mathbb{E}\left[\lambda v \cdot \frac{1}{\psi(G_1^{-1}(t_1),\varepsilon)}\left(\int_0^{t_1} \partial_s\left[\psi(G_1^{-1}(s),\varepsilon)\right]\cdot x(s,\varepsilon)\,\mathrm{d}s + \psi(G_1^{-1}(0),\varepsilon)x(0,\varepsilon)\right) \mid \varepsilon\right].$$

By the proof of Theorem 2, this value is bounded from above by

$$\mathbb{E}\left[\lambda v \cdot \frac{1}{G_1^{-1}(t_1)}\left(\int_0^{t_1} (G_1^{-1})'(s)\cdot x(s,\varepsilon)\,\mathrm{d}s + G_1^{-1}(0)x(0,\varepsilon)\right) \mid \varepsilon\right].$$

because for any $\varepsilon \in \mathcal{E}$, by (A3), we have

$$\log\psi(r,\varepsilon) \preceq_{disp} \log(r).$$

Thus, the designer's payoff with x under screening device D_2 is bounded from above by

$$\mathbb{E}\left[\lambda v \cdot \frac{1}{G_1^{-1}(t_1)}\left(\int_0^{t_1}(G_1^{-1})'(s)\cdot x(s,\varepsilon)\,\mathrm{d}s + G_1^{-1}(0)x(0,\varepsilon)\right)\right].$$

Now, note that for any feasible and implementable $x(t_1,\varepsilon)$, we have that $\tilde{x}(t_1) := \mathbb{E}_\varepsilon[x(t_1,\varepsilon)]$ is feasible and monotone. Moreover, the designer's payoff with $\tilde{x}$ under screening device D_1 is given by

$$\mathbb{E}\left[\lambda v \cdot \frac{1}{G_1^{-1}(t_1)}\left(\int_0^{t_1}(G_1^{-1})'(s)\cdot \mathbb{E}_\varepsilon[x(s,\varepsilon)]\,\mathrm{d}s + G_1^{-1}(0)\,\mathbb{E}_\varepsilon[x(0,\varepsilon)]\right)\right],$$

which, by independence, equals the upper bound on the designer's payoff with x under screening device D_2. Therefore, screening device D_1 dominates screening device D_2. □

Proposition 4 is a special case of Lemma 2 with $\psi(r,\varepsilon) = \exp\left(\beta\log(r)+\varepsilon\right)$.

B.7 Proof of Proposition 5

First, consider the case where $\beta_2 \le 0$. Note that by Proposition 4, screening device D_2 is dominated by screening device $\widetilde{D}_2$ defined by

$$\log(\tilde{r}_2) = \beta_2\log(\lambda v).$$

We claim that the optimal mechanism under $\widetilde{D}_2$ is always full randomization without costly screening. Indeed, note that

$$\mathbb{E}[\lambda v \mid \tilde{r}_2] = \exp\left(\frac{1}{\beta_2}\log(\tilde{r}_2)\right).$$

Thus, $\mathbb{E}[\lambda v \mid \tilde{r}_2]$ is a non-increasing function of $\tilde{r}_2$. By the FKG inequality, for any implementable (and hence monotone) x^Γ, we have

$$\int \mathbb{E}[\lambda v \mid \tilde{r}_2]\left(x^\Gamma(\tilde{r}_2) - \frac{1}{\tilde{r}_2}y(\tilde{r}_2)\right)\mathrm{d}\tilde{G}_2(\tilde{r}_2) \le \int \mathbb{E}[\lambda v \mid \tilde{r}_2]x^\Gamma(\tilde{r}_2)\,\mathrm{d}\tilde{G}_2(\tilde{r}_2) \le \mathbb{E}[\lambda v]\,\mathbb{E}[x^\Gamma(\tilde{r}_2)],$$

where the welfare $\mathbb{E}[\lambda v]\,\mathbb{E}[x^\Gamma(\tilde{r}_2)]$ is achievable by random allocations without costly screening. So screening device $\widetilde{D}_2$, and hence D_2, is dominated by screening device D_1.

Now, consider the case where $\beta_2 > 0$. By assumption, we have

$$\sigma_2^2 - \Big(\frac{\beta_2}{\beta_1}\sigma_1\Big)^2 \geq 0\,.$$

We construct a coupling of $(\lambda v, r_1, r_2)$ in the following way:

$$\log(r_2) = \frac{\beta_2}{\beta_1}\log(r_1) + \xi$$

where ξ is drawn from

$$\mathcal{N}\Big(0,\, \sigma_2^2 - \Big(\frac{\beta_2}{\beta_1}\sigma_1\Big)^2\Big)$$

independent of λv and ε_1 (and hence independent of $(\lambda v, r_1)$).

We claim that under this coupling, we have that the joint distribution of $(\lambda v, r_2)$ is unchanged. To see this, note that

$$\frac{\beta_2}{\beta_1}\log(r_1) + \xi = \frac{\beta_2}{\beta_1}\Big(\beta_1\log(\lambda v) + \varepsilon_1\Big) + \xi = \beta_2\log(\lambda v) + \frac{\beta_2}{\beta_1}\varepsilon_1 + \xi$$

where

$$\frac{\beta_2}{\beta_1}\varepsilon_1 + \xi =: \tilde{\varepsilon}_2$$

is independent of λv and follows a normal distribution with mean 0 and variance

$$\Big(\frac{\beta_2}{\beta_1}\sigma_1\Big)^2 + \sigma_2^2 - \Big(\frac{\beta_2}{\beta_1}\sigma_1\Big)^2 = \sigma_2^2\,.$$

Therefore, we have

$$\Big(\log(\lambda v), \log(r_2)\Big) \overset{d}{=} \Big(\log(\lambda v), \frac{\beta_2}{\beta_1}\log(r_1) + \xi\Big).$$

Finally, note that by Proposition 4, any screening device D satisfying

$$\log(r) = \frac{\beta_2}{\beta_1}\log(r_1) + \xi$$

is dominated by screening device D_1 since *(i)* $\frac{\beta_2}{\beta_1} \in [0,1]$ and *(ii)* ξ is independent of $(\lambda v, r_1)$. The result follows.

B.8 Proof of Proposition 6

We will show that the condition in Lemma 2 (stated in the proof of Proposition 4) is preserved under any additive or minimum scoring rules.

Additive scoring rule: Let D_3 be induced by any additive scoring rule. Then, we have

$$c_3 = \min\Big(c_1, c_2/\beta\Big)$$

and hence

$$\frac{1}{r_3} = \frac{c_3}{v} = \min\Big(\frac{c_1}{v}, \frac{1}{\beta}\cdot\frac{c_2}{v}\Big) = \min\Big(\frac{1}{r_1}, \frac{1}{\beta}\cdot\frac{1}{r_2}\Big).$$

Let

$$\widetilde{\psi}(r,\varepsilon) := \frac{1}{\min\Big(\frac{1}{r}, \frac{1}{\beta}\cdot\frac{1}{\psi(r,\varepsilon)}\Big)}.$$

Then

$$\log\widetilde{\psi}(r,\varepsilon) = -\min\Big(-\log(r), -\log(\beta) - \log\psi(r,\varepsilon)\Big) = \max\Big(\log(r), \log(\beta) + \log\psi(r,\varepsilon)\Big).$$

Let $\mathcal{E}$ be the set of ε at which $0 \le \frac{\mathrm{d}\log\psi(r,\varepsilon)}{\mathrm{d}\log r} \le 1$. Fix any $\varepsilon \in \mathcal{E}$. By the envelope theorem, the above implies that we have almost everywhere

$$\frac{\mathrm{d}\log\widetilde{\psi}(r,\varepsilon)}{\mathrm{d}\log r} \in \Big\{1, \frac{\mathrm{d}\log\psi(r,\varepsilon)}{\mathrm{d}\log r}\Big\}.$$

Thus, we have almost everywhere

$$0 \le \frac{\mathrm{d}\log\widetilde{\psi}(r,\varepsilon)}{\mathrm{d}\log r} \le 1\,.$$

This implies that screening device D_1 dominates screening device D_3 by Lemma 2.

Minimum scoring rule: Let D_3 be induced by any minimum scoring rule. Then, we have

$$\frac{1}{r_3} = \frac{c_3}{v} = \frac{c_1}{v} + \frac{1}{\beta}\cdot\frac{c_2}{v} = \frac{1}{r_1} + \frac{1}{\beta}\cdot\frac{1}{\psi(r_1,\varepsilon)}.$$

Let

$$\widetilde{\psi}(r,\varepsilon) := \frac{1}{\frac{1}{r} + \frac{1}{\beta}\cdot\frac{1}{\psi(r,\varepsilon)}}.$$

Then

$$\log\widetilde{\psi}(r,\varepsilon) = -\log\Big[\frac{1}{r} + \frac{1}{\beta}\cdot\frac{1}{\psi(r,\varepsilon)}\Big].$$

Fix any $\varepsilon \in \mathcal{E}$. Note that

$$0 \leq \frac{\mathrm{d}\log\widetilde{\psi}(r,\varepsilon)}{\mathrm{d}\log r} = \frac{\mathrm{d}\log\widetilde{\psi}(r,\varepsilon)}{\mathrm{d}r}\cdot\frac{\mathrm{d}r}{\mathrm{d}\log r} = \frac{-(\frac{1}{r})' - \frac{1}{\beta}\cdot\partial_r\left[\frac{1}{\psi(r,\varepsilon)}\right]}{\frac{1}{r}+\frac{1}{\beta}\cdot\frac{1}{\psi(r,\varepsilon)}}\cdot r \leq \frac{1}{r}\cdot r = 1\,.$$

where the second inequality follows by two observations: First, we have

$$\frac{-(\frac{1}{r})'}{\frac{1}{r}} = -\frac{\mathrm{d}}{\mathrm{d}r}\log(\frac{1}{r}) = \frac{\mathrm{d}}{\mathrm{d}r}\log(r) \geq \frac{\mathrm{d}}{\mathrm{d}r}\log\psi(r,\varepsilon) = -\frac{\mathrm{d}}{\mathrm{d}r}\log(\frac{1}{\psi(r,\varepsilon)}) = \frac{-\frac{1}{\beta}\cdot\partial_r\left[\frac{1}{\psi(r,\varepsilon)}\right]}{\frac{1}{\beta}\cdot\frac{1}{\psi(r,\varepsilon)}}$$

since $0 \leq \frac{\mathrm{d}\log\psi(r,\varepsilon)}{\mathrm{d}\log r} \leq 1$. Second, recall that the mediant inequality implies that

$$\frac{a+b}{c+d} \leq \max\left\{\frac{a}{c},\frac{b}{d}\right\}$$

for any positive numbers a,b,c,d.

Thus, screening device D_1 dominates screening device D_3 by Lemma 2.

B.9 Proof of Proposition 7

We first prove the second claim. Suppose that $v \sim U[1,2]$ and $v = \frac{1}{\lambda}$. Note that since $\lambda v \equiv 1$, any screening device D must induce r that is independent of λv. Therefore, under any screening device D, the optimal mechanism is always full randomization (with $y = 0$). Thus, it suffices to show that there exists some F such that money does strictly better. Consider a setting in which there is a mass $s \in (0,1)$ of quality-1 goods to be allocated. Let the price be $2-s$. The payoff to the designer is then

$$\int_1^2 \frac{1}{v}(v-2+s)\mathbb{1}_{v\geq 2-s}\,\mathrm{d}v + s\cdot(2-s)\cdot\mathbb{E}\left[\frac{1}{v}\right]$$

$$= s - \int_1^2 \frac{2-s}{v}\mathbb{1}_{v\geq 2-s}\,\mathrm{d}v + s\cdot(2-s)\cdot\mathbb{E}\left[\frac{1}{v}\right].$$

Note that

$$\mathbb{E}\left[\frac{1}{v}\right] > \frac{1}{s}\int_{2-s}^2 \frac{1}{v}\,\mathrm{d}v$$

and therefore the payoff of this mechanism to the designer is strictly higher than s, which is the payoff to the designer under full randomization. This proves the second claim.

Now, we prove the first claim. Consider the setting where $v = \lambda = \sqrt{t}$ and $t \sim U[0,1]$. The welfare of using money to allocate a mass $1-s$ of quality-1 goods under the market mechanism $x(t) = \mathbb{1}_{t \geq s}$ is given by

$$M(s) := \int_s^1 \sqrt{t}\Big(\sqrt{t} - \sqrt{s}\Big)\mathrm{d}t + \Big(\int_0^1 \sqrt{t}\,\mathrm{d}t\Big)\sqrt{s}(1-s) = \frac{1}{6}(s^2 - 4\sqrt{s} + 3) + \frac{2}{3}\sqrt{s}(1-s).$$

The welfare of using screening device D_α defined by

$$c = \frac{v}{(\lambda v)^\alpha}$$

for parameter $\alpha > 0$ under the market mechanism $x(t) = \mathbb{1}_{t \geq s}$ is given by:

$$W(D,s) = \int_s^1 \left(1 - \frac{G^{-1}(s)}{G^{-1}(t)}\right) \cdot \mathbb{E}\Big[\lambda v \mid r = G^{-1}(t)\Big]\mathrm{d}t = \int_s^1 t - t \cdot \left(\frac{s}{t}\right)^\alpha \mathrm{d}t = \frac{1}{2}\Big(1 - \frac{\alpha s^2 - 2s^\alpha}{\alpha - 2}\Big).$$

One can verify that for sufficiently large α, the concave envelope of $W(D,s)$ is everywhere above the concave envelope of $M(s)$ and strictly so for some $s \in (0,1)$. By the proof of Theorem 5 in Appendix C, this example then proves the first claim.

C Additional Results

For any function $h : [0,1] \to \mathbb{R}$, let $\mathrm{co}(h)$ be the *concave envelope* of h, i.e.,

$$\mathrm{co}(h)(s) := \inf\Big\{k(s) : k \text{ is concave and } k \geq h\Big\}.$$

The next result provides a necessary and sufficient condition for one screening device to dominate another.

Theorem 5. *A screening device D_1 dominates screening device D_2 if and only if, for all $s \in [0,1]$,*

$$\widehat{W}(D_1, s) \geq \widehat{W}(D_2, s), \tag{C1}$$

where

$$\widehat{W} = \mathrm{co}(W_0)$$

and

$$W_0(D,s) = \begin{cases} W(D,s) & \text{if } s \in (0,1] \\ \mathbb{E}[\lambda v] & \text{if } s = 0. \end{cases}$$

Proof of Theorem 5. We follow the notation in the proof of Theorem 1.

The "if" part: Suppose that condition (C1) in Theorem 5 holds for all $s \in [0,1]$. Note that, by (B3) in the proof of Theorem 1, we have

$$W_0(D,s) = \int_0^1 V(t) \cdot \frac{1}{G^{-1}(t)} U[\mathbb{1}_{t \geq s}](t)\,\mathrm{d}t,$$

which is the designer's payoff under a quantile allocation rule $\mathbb{1}_{t \geq s}$.

Now, consider any feasible expected quality quantile allocation rule $x : [0,1] \to [0,1]$. As in the proof of Theorem 1, the set of feasible allocation rules x is the same across the two screening devices D_1 and D_2. Moreover, as in the proof of Theorem 1, by Lemma 1, there exists a probability measure $\mu \in \Delta([0,1])$ such that for every $t \in [0,1]$

$$x(t) = \int_{[0,1]} \mathbb{1}_{t \geq s}\,\mathrm{d}\mu(s).$$

Thus, we can identify a quantile allocation rule x by a probability measure μ. Thus, by the argument in the proof of Theorem 1, it suffices to show that

$$\sup_{\mu_2 \in \mathcal{L}} \int_{[0,1]} W_0(D_2,s)\,\mathrm{d}\mu_2(s) \leq \sup_{\mu_1 \in \mathcal{L}} \int_{[0,1]} W_0(D_1,s)\,\mathrm{d}\mu_1(s)$$

where $\mathcal{L}$ is the set of probability measures that would satisfy the feasibility constraint. For any $\mu \in \mathcal{L}$, we claim that if $\mu \preceq_{mps} \tilde{\mu}$, then we also have that $\tilde{\mu} \in \mathcal{L}$. That is, any mean-preserving spread of the measure μ also induces a feasible allocation rule. To see this, note that for any $z \in [0,1]$, we have

$$\int_z^1 x(t)\,\mathrm{d}t = \int_z^1 \int_0^1 \mathbb{1}_{t \geq s}\,\mathrm{d}\mu(s)\,\mathrm{d}t = \int_0^1 1 - \max\{z,s\}\,\mathrm{d}\mu(s) \geq \int_0^1 1 - \max\{z,s\}\,\mathrm{d}\tilde{\mu}(s) = \int_z^1 \tilde{x}(t)\,\mathrm{d}t$$

where the inequality is due to that $1 - \max\{s,z\}$ is concave in s and $\mu \preceq_{mps} \tilde{\mu}$. Moreover, note that the above holds with equality when $z = 0$. Thus, we have $\tilde{x} \preceq_{maj} x$. Let X be the random expected quality induced by the allocation rule x, and $\tilde{X}$ be the random expected quality induced by the allocation rule $\tilde{x}$. Then, note that $\tilde{X} \preceq_{mps} X$. By Blackwell (1953), there exists a coupling of $(X, \tilde{X})$ such that $\mathbb{E}[X \mid \tilde{X}] = \tilde{X}$. This implies that $\tilde{x}$ is a feasible expected quality allocation rule, because it can be constructed by a compound lottery where one further randomizes the quality lotteries that implement x according to the conditional distribution $X \mid \tilde{X}$. Therefore, we have $\tilde{\mu} \in \mathcal{L}$. This implies that

$$\sup_{\mu_2 \in \mathcal{L}} \int_{[0,1]} W_0(D_2,s)\,\mathrm{d}\mu_2(s) = \sup_{\mu_2 \in \mathcal{L}} \int_{[0,1]} \widehat{W}(D_2,s)\,\mathrm{d}\mu_2(s)$$

$$\leq \sup_{\mu_1 \in \mathcal{L}} \int_{[0,1]} \widehat{W}(D_1, s)\, \mathrm{d}\mu_1(s) = \sup_{\mu_1 \in \mathcal{L}} \int_{[0,1]} W_0(D_1, s)\, \mathrm{d}\mu_1(s)$$

where *(i)* the two equalities use that for any $\mu \in \mathcal{L}$ and $\mu \preceq_{mps} \tilde{\mu}$, we have $\tilde{\mu} \in \mathcal{L}$ and *(ii)* the inequality uses the condition (C1).

The "only if" part: Note that

$$\widehat{W}(D_2, 0) = W_0(D_2, 0) = \mathbb{E}[\lambda v] = W_0(D_1, 0) = \widehat{W}(D_1, 0).$$

Thus, it suffices to show that (C1) holds for all $s \in (0,1)$. Suppose for contradiction that (C1) is violated at some $s \in (0,1)$. Let $F(q)$ be the function $s + (1-s)\mathbb{1}_{q=1}$, representing a fixed supply $1-s$ of quality 1 objects. We claim that with this choice of F,

$$\mathrm{OPT}(D_1, F) < \mathrm{OPT}(D_2, F).$$

Indeed, under this choice of F, an expected quality quantile allocation rule x is feasible if and only if

$$\int_0^1 x(t)\, \mathrm{d}t \leq 1 - s.$$

Now, fix the screening device to be D_1. Again, by the representation theorem (Lemma 1), we may then optimize over probability measures μ such that

$$\int_{[0,1]} k\, \mathrm{d}\mu(k) \geq s.$$

By (B3) and the argument in the proof of Theorem 1, the objective value with measure μ under screening device D_1 is equal to

$$\mathbb{E}_{k\sim\mu}\Big[W_0(D_1, k)\Big] \leq \mathbb{E}_{k\sim\mu}\Big[\widehat{W}(D_1, k)\Big] \leq \widehat{W}\Big(D_1, \mathbb{E}_{k\sim\mu}[k]\Big) \leq \widehat{W}\Big(D_1, s\Big) < \widehat{W}\Big(D_2, s\Big),$$

where the second inequality is due to Jensen's inequality, and the third inequality is due to that $W_0(D_1, \cdot)$, and hence $\widehat{W}(D_1, \cdot)$, is non-increasing. By the proof of the "if" part, the designer under screening device D_2 can achieve the welfare of $\widehat{W}(D_2, s)$ using random allocations. Therefore, we must have $\mathrm{OPT}(D_1, F) < \mathrm{OPT}(D_2, F)$. A contradiction. □