Screening with Tolls and Damages*

Filip Tokarski
Stanford GSB

Abstract

A welfare-maximizing designer allocates two kinds of goods using two screening instruments: *tolls*, whose costs are separable from agents' values, and *damages*, which are more costly to agents whose values for the goods are higher. Tolls include payments, queues, and administrative burdens; damages include quality reductions, delays, and restrictions on use. When agents differ only in their value for one type of good, the designer can never gain from damaging it. However, using damages can be optimal when valuations for both goods are heterogeneous, as the two instruments can "sort" agents across the available options in different ways. I provide conditions under which the optimal mechanism includes a damaged option, as well as sufficient conditions under which it does not; in the latter case, the optimal mechanism posts "market-clearing" tolls for each good. Intuitively, damages are more likely to be optimal when values for the two goods are positively affiliated, and less likely when high value for one good predicts low value for the other.

1 Introduction

Compare two mechanisms for scheduling appointments, such as clinic visits, passport renewals, or public-benefit interviews. The first mechanism offers walk-ins allocated through a *queue*: the first n people to arrive after the office opens are served. In practice, this means that an applicant may need to arrive hours ahead of time to secure a place. The second mechanism is a *waitlist*: people sign up for an appointment in advance and are given the first available slot. This, in turn, entails waiting weeks for one's scheduled time. Both mechanisms impose a burden on participants: in the queue, it is the time spent waiting before the office opens; in the waitlist, it is the delay before service. While both burdens may discourage some applicants from using the system, they differ crucially in how they interact with people's values for the appointment: the disutility of arriving ahead of time is largely separable from one's need for the service, whereas delay is more costly for those whose need is higher. Indeed, a patient with an untreated injury

*I am grateful to Rafael Berriel, Benjamin Brooks, Rebecca Diamond, Laura Doval, Piotr Dworczak, Joey Feffer, Zi Yang Kang, Jacob Leshno, Federico Llarena, Michael Ostrovsky, Ilya Segal, Andrzej Skrzypacz, and Sam Wycherley for their helpful comments and suggestions.

suffers more from a month-long wait than someone seeking a routine check-up; a delayed caseworker appointment is more costly for a family that relies exclusively on benefits than for one with other sources of support.

Waitlists and queues exemplify the two kinds of screening instruments studied in this paper. The former are more costly to agents whose value for the allocated good is higher; I call these instruments *damages*. By contrast, the costs imposed by the latter class, which I call *tolls*, are separable from agents' values for the allocated good. Both categories should be understood broadly: damages include literal reductions in quality, but also delays in allocation, restrictions on use, or other policies that reduce the value of the good itself. Tolls include monetary payments, queues, bureaucratic processes, and other ordeals that burden participants without directly reducing the value of the good they receive. The distinction between tolls and damages appears in a variety of economic settings; I provide several examples below.

Social broadband tariffs. Some governments require internet providers to offer eligible low-income households cheap plans with limited service quality. For instance, Portugal's social internet tariff fixes the monthly price of the required plan and requires providers to offer a basic service with low speeds and a 15 GB monthly data allowance.[1] In this setting, the monthly fee is a toll, whereas the speed restriction and data cap are damages.

Referrals vs. copayments in healthcare. Many conditions admit both low- and high-intensity treatments: for instance, musculoskeletal injuries may often be treated with physical therapy or surgery. Because capacity is limited, health systems often ration access to more intensive treatment. This can be done by charging copayments, which constitute tolls. Healthcare providers can also require referrals, prior authorization, or documentation that less intensive treatments have been tried first. Indeed, World Health Organization (2023) notes that such measures are intended to "ensure patient access to specialist healthcare when needed, while maintaining resource efficiency." However, they also delay access to the treatment itself, and are therefore especially costly for patients whose injuries are more severe. In this sense, referral and authorization requirements act as damages.

Self-targeting in social programs. Goods provided through public programs are often less desirable than their private-market counterparts. Social housing is frequently located in disadvantaged or peripheral neighborhoods, Medicaid covers fewer providers than private insurance, and food assistance delivered through in-kind or voucher programs such as SNAP restricts choice.[2] These restrictions are costly to recipients, but can also serve as self-targeting devices (Nichols and Zeckhauser, 1982; Besley and Coate, 1991; Currie and Gahvari, 2008). For

[1] https://digital.gov.pt/en/estrategia-digital/plano-de-acao-para-a-transicao-digital/01-capacitacao-e-inclusao-digital-das-pessoas/tarifa-social-de-acesso-a-internet.

[2] Some quality reductions may of course reflect lower provision costs. I interpret them as damages only to the extent that quality is reduced below the efficient level, so that the reduction serves as a screening device.

instance, if subsidized housing is built in less desirable neighborhoods, families with better alternatives may decide not to apply, so the subsidy disproportionately benefits poorer households. However, quality reductions and usage restrictions are also more costly for households that value the program more. A sicker patient is more burdened by a narrow provider network, and a family relying more heavily on food vouchers is more constrained by restrictions on eligible products. Thus, in my framework, these instruments constitute damages. Nevertheless, self-targeting can also be achieved without diminishing the value of the allocated good: for example, social programs often require applicants to wait in line, complete paperwork, travel to a distant office, or recertify eligibility (Nichols et al., 1971; Nichols and Zeckhauser, 1982; Besley and Coate, 1992; Kleven and Kopczuk, 2011; Dupas et al., 2016; Alatas et al., 2016; Deshpande and Li, 2019). Since the costs of completing such ordeals are not tied to one's value for the allocated good, they constitute tolls.

Other environments in which the distinction between tolls and damages is relevant include appointment scheduling, discussed above, congestion pricing, discussed in Section 4, and, more generally, settings where goods can be allocated either through passive waitlists or through queues.[3] I microfound this interpretation in Appendix B, where I show that delays arising endogenously in a steady-state waitlist can act as damages.

Motivated by these settings, I consider a model in which a welfarist designer allocates scarce goods using deterministic mechanisms that may combine tolls and damages. Damages are purely wasteful; tolls, by contrast, may represent either monetary payments or ordeals, so I allow their social value to vary. I first study the case with one scarce good and a common-value outside option. I find that in this setting, using damages is suboptimal in a strong sense: any feasible mechanism that uses damages is weakly Pareto-dominated by one that uses only tolls. Intuitively, damages are inefficient because they burden high-value recipients more than marginal ones. Every feasible allocation can be implemented with tolls alone, while also leaving higher rents to inframarginal agents.

I then turn to the case where the designer allocates two kinds of goods and agents have heterogeneous values for both. The two-good case is the simplest tractable setting in which screening instruments serve not only to exclude low-value agents, but also to direct them toward options in a socially efficient way. This "sorting" motive is important in many settings to which the model applies. Affordable housing programs, for instance, offer units that vary in location and size, with households' preferences over them having a strong horizontal component (Waldinger, 2021). These apartments are often allocated through waitlists whose lengths differ substantially across developments, so households trade off their values for particular units against the delay required to obtain them (Van Ommeren and Van der Vlist, 2016). In healthcare, copays and referral requirements affect whether patients pursue less or more intensive care; appointment-scheduling systems allocate slots across clinics or offices in different locations, with better-located offices frequently being more congested.

[3] For flow goods that can be enjoyed in every period after receipt, long waitlists cause delays that deprive recipients of periods of use. For consumable goods, they lead to temporal discounting of the good's value.

I show that tolls and damages induce different patterns of sorting into the available options, and that, unlike in the single-dimensional case, the optimal pattern sometimes requires the use of damages. Nevertheless, my first theorem provides conditions on the distribution of values under which damaging goods is suboptimal. I show that if each good's value has an increasing inverse anti-hazard rate conditional on the value of the other good, the optimal mechanism is the *market-clearing toll mechanism*: the designer offers undamaged goods at tolls set so that both supplies are exhausted once agents choose their favorite option. Theorem 2 then gives a partial converse: it states conditions under which the optimal mechanism does offer damaged options. Intuitively, whether damages are optimal is closely related to the dependence between values for the two goods in the population. When these values are negatively affiliated, damages are likely to be suboptimal, and the designer should use the market-clearing toll mechanism. Conversely, offering damaged options is often optimal when values exhibit sufficiently strong positive affiliation.

I then extend the analysis by allowing agents to differ in how costly they find the toll. For instance, when screening is done with monetary payments, poorer agents whom the program tries to target may find them more burdensome. Conversely, the same agents may be more willing to wait, travel, or endure other inconveniences in order to get the good (Dupas et al., 2016). I show that this extension leaves the main logic of the results unchanged: after appropriately reweighting the type distribution to account for heterogeneous ordeal costs, the same forces determine when damages can improve welfare.

My results have two main market-design implications. First, when the allocated good is homogeneous, screening should be done with tolls rather than damages: delays, usage restrictions, and quality reductions should be replaced, where possible, by instruments such as fees, application requirements, or queues. Second, when goods are heterogeneous, the appropriate instrument depends on how agents' values for them are related. If high value for one good tends to predict low value for another, toll instruments alone are likely to be optimal: for instance, popular appointment locations should carry higher booking fees or use walk-in queues rather than long waitlists. If values are instead strongly positively related, as may be the case for geographically clustered public housing projects, differentiated waitlists could be optimal, possibly in combination with toll instruments like differential rent subsidies.

From a technical perspective, my model is an instance of a tractable multidimensional screening problem. By restricting attention to deterministic mechanisms, I am able to characterize them as pairs of toll and damage menus for the two goods. This in turn allows me to represent two-dimensional mechanisms as interconnected single-dimensional screening problems. The interaction between them is summarized by a boundary in the type space that separates the sets of types who choose each good. The multidimensional problem can then be broken up into two stages: first, determining the optimal way to implement a given boundary, and second, solving an optimal control problem to select the best boundary among all implementable ones. While my paper applies this method to the problem of a welfarist designer, similar ideas could be useful for studying other settings, such as the problem of a two-good monopolist choosing deterministic mechanisms for selling to unit-demand consumers.

The rest of the paper is organized as follows. The next section discusses the related literature and Section 3 presents the model. Section 4 studies the case where agents differ only in their value for one type of good. Section 5 extends the analysis to two-dimensional heterogeneity, introduces the boundary representation of mechanisms, and gives conditions under which damages are and are not optimal. Section 6 presents the key steps in the proof of Theorem 1. Section 7 extends the analysis to heterogeneous toll costs. Finally, Section 8 discusses the implications for market design.

2 Related literature

My paper contributes to the literature on using costly screening devices and money-burning to maximize welfare. Hartline and Roughgarden (2008) and Condorelli (2012) show that when goods are allocated without monetary transfers, asking agents to undertake socially wasteful actions can sometimes improve welfare: although these actions destroy surplus directly, they can help screen agents by value and thereby improve the allocation of goods. Dworczak (2026) asks when a redistributive designer would like to hand out money in exchange for completing a costly ordeal. However, this literature focuses primarily on the allocation of homogeneous goods. An exception is Noda and Okada (2024), who study a symmetric environment with many good varieties and show that screening through money burning becomes less efficient as the number of varieties grows; nevertheless, they restrict attention to mechanisms that offer agents only their favorite variety. By allowing for richer heterogeneity, I explore an additional role for wasteful screening devices: rather than only affecting the participation margin by determining who is excluded, they can also direct agents toward different goods in a socially efficient manner. When one variety is significantly overdemanded, for instance, the designer can damage it or increase its toll to redirect agents with weaker preferences to less scarce alternatives. My paper also relates to Akbarpour et al. (2023), who ask when one wasteful screening device dominates another for a planner aiming to maximize a social welfare function. Unlike them, I allow the designer to *combine* instruments and show that, under certain distributional conditions, screening with ordeals alone dominates any mechanism using both devices.

A related literature studies wasteful screening by profit-maximizing sellers. Deneckere and McAfee (1996) give conditions under which a monopolist may want to damage goods in order to price discriminate. Yang (2021) studies a more general problem in which the monopolist has access to both wasteful and non-wasteful instruments, and describes cases in which the wasteful instrument should not be used. I instead ask how similar instruments should be used by a welfare-maximizing regulator. In doing so, my work relates to Bulow and Klemperer (2012), who study price controls from the perspective of consumer surplus. In my framework, prices are tolls, and their objective corresponds to the case in which the designer places no value on toll revenue. They observe that price controls can be useful when buyers' values are clustered near the market-clearing price, leaving little surplus net of payments. I extend this logic to heterogeneous goods: when sorting with tolls clusters many recipients' values right above the market-clearing price, damages can improve welfare by inducing a different sorting pattern and preserving more surplus for inframarginal recipients.

The effects of the screening instruments I classify as tolls and damages have also been documented empirically in several settings to which my paper could be applied. In healthcare, Manning et al. (1987) show that higher patient cost-sharing—a monetary toll—substantially reduces medical care use, in part by discouraging low-value care. More closely related to the sorting forces in my model, Brot-Goldberg et al. (2023) show that prior authorization restrictions in Medicare Part D substantially reduce the use of restricted drugs and lead many marginal patients to switch to cheaper options. Other work considers non-monetary ordeals in social programs. For example, Finkelstein and Notowidigdo (2019) find that administrative burdens can improve targeting in SNAP. Finally, a related literature studies self-targeting through offering inferior goods and restrictions on use (Nichols and Zeckhauser, 1982; Besley and Coate, 1991; Currie and Gahvari, 2008). Since these policies reduce the value of the allocated good, I consider them damages.

My paper also relates to a literature on waitlist design. While no paper has studied combining waitlists with payments or ordeals in settings with heterogeneous goods, a substantial literature examines the design of waitlists without transfers. Arnosti and Shi (2020) and Waldinger (2021) study the effects of restricting recipients' choice on targeting. Barzel (1974), Bloch and Cantala (2017), and Leshno (2022) observe that in environments with homogeneous waiting costs, wait times may "act as prices," screening for agents with higher valuations. I refine this intuition by showing that the screening properties of wait times are different when the cost of waiting stems from delayed receipt.

3 Model

A designer distributes two types of goods, A and B. Their supplies are equal to $s_A, s_B > 0$. There is a unit mass of agents whose values for the two goods are given by a and b, respectively. Agents' values (a, b) are distributed according to a nonatomic distribution F on $[0,1]^2$. The designer chooses a menu of qualities and tolls for each of the goods. That is, an agent can choose which good she wants to get and then choose a quality and toll option from the relevant good's menu. She can also choose not to participate, which gives her utility 0. When a type-(a, b) agent participates and receives good y, her utility is:

$$x \cdot a - c \ \text{ if } \ y = A,$$
$$x \cdot b - c \ \text{ if } \ y = B,$$

where $c \in \mathbb{R}_+$ is the toll the agent incurs and $x \in [0,1]$ is her good's quality. Whenever $x < 1$, we say the good has been *damaged*. The designer chooses the menu to maximize welfare, counting a fraction $\gamma \in [0,1]$ of each toll as socially valuable. We can then reduce her problem to picking allocation rules for tolls, $c : [0,1]^2 \to \mathbb{R}_+$, qualities, $x : [0,1]^2 \to [0,1]$, and goods, $y : [0,1]^2 \to \{\varnothing, A, B\}$ to maximize:

$$\int u_{a,b}(a,b) \; + \; \gamma \, c(a,b) \, dF(a,b), \tag{O}$$

subject to (IC) and (IR) constraints, and the supply constraint (S):

$$\text{for all } (a,b),(a',b') \in [0,1]^2, \quad u_{a,b}(a,b) \geq u_{a,b}(a',b'), \tag{IC}$$

$$\text{for all } (a,b) \in [0,1]^2, \quad u_{a,b}(a,b) \geq 0, \tag{IR}$$

$$\int \mathbb{1}_{y(a,b)=A} \, dF(a,b) \leq s_A, \quad \int \mathbb{1}_{y(a,b)=B} \, dF(a,b) \leq s_B. \tag{S}$$

Here $u_{a,b}(a',b')$ denotes the utility type (a,b) gets from reporting (a',b') in the mechanism (c,x,y). I call a mechanism (c,x,y) satisfying (IC),(IR), (S) *feasible*.

3.1 Discussion of the model

Restriction to deterministic mechanisms. My model does not allow the designer to offer lotteries.[4] Still, introducing some forms of randomization turns out to be without loss. Suppose, for example, that agents who pay a toll can enter a lottery that gives them one of the goods with some probability, and that unsuccessful agents can re-enter by paying the same cost again. When represented as a direct-revelation mechanism, a menu of such pure lotteries with re-entry is equivalent to deterministic mechanisms allowed by my model (see Appendix B for a related discussion). Nevertheless, my restriction still precludes mechanisms that give agents some probability of receiving either good.

The restriction to deterministic mechanisms reflects practical constraints and considerations present in many settings. Sellers rarely use randomized allocations, presumably because customers may fear that the lottery can be manipulated. Similar worries arise even in the context of public programs: allocating affordable housing randomly sometimes raises concerns about corruption or draw-faking, prompting calls to replace lotteries with more transparent mechanisms such as first-come-first-served waitlists.[5] Relatedly, even honestly executed lotteries for public housing are sometimes perceived as unfair; for example, Whistler, Canada "allocates units based on a waitlist, a method chosen due to its perceived fairness and ease of administration, though lottery and points schemes have been used in the past" (City of Vancouver, 2016).

[4] From the agent's perspective, getting a damaged good with quality $x < 1$ is equivalent to receiving it with probability x. However, randomization and damages differ from the perspective of the designer, as they enter the supply constraint differently: allocating a good with probability x takes up only x of the supply, while damaging it to quality x still uses up a whole unit. In this sense, damages are strictly inferior to randomization.

[5] See, for instance, https://www.camara.leg.br/noticias/523091-projeto-veda-sorteio-na-selecao-dos-beneficiarios-do-minha-casa-minha-vida/ and https://citymeetings.nyc/meetings/new-york-city-council/2025-04-29-1000-am-committee-on-housing-and-buildings/chapter/consideration-of-moving-from-lottery-system-to-universal-waiting-list-for-affordable-housing.

Finally, the restriction is useful analytically. It isolates the comparison between tolls and damages, which is the focus of the paper (for a model of welfare-maximizing screening with lotteries, see Tokarski (2026)). It also gives the model enough structure to make a multidimensional screening problem tractable: together with the assumption of unit demand, it lets me represent mechanisms using boundaries in the type space, which mitigates some of the difficulties posed by multidimensional screening, as studied, among others, by Rochet and Choné (1998); Manelli and Vincent (2006) and Daskalakis, Deckelbaum, and Tzamos (2017).

Value on tolls. To capture different interpretations of tolls, I use the parameter γ to represent their social value. Consider first the case where tolls are monetary payments. If the designer can costlessly rebate revenue to participants, it is natural to set $\gamma = 1$. Intermediate cases with $\gamma \in (0, 1)$ may describe government programs where rebating revenue to participants is possible but administratively costly, or where distributing cash undermines the screening benefits of in-kind transfers.[6] Setting $\gamma = 0$ may be appropriate, for example, when the designer regulates a seller and cares only about consumer surplus. Beyond monetary tolls, the case of $\gamma = 0$ can also be interpreted as "money burning" à la Hartline and Roughgarden (2008), capturing ordeals like queueing.

The revelation principle. Although the designer's problem is written as a choice over direct mechanisms, the usual revelation-principle argument is not immediate because I restrict attention to deterministic options. Appendix A shows that the direct-revelation formulation used above is nevertheless without loss.

4 One-dimensional heterogeneity

I first consider the case where agents differ only with respect to their value for good A; good B plays the role of a common-value outside option. The following result shows that using damages in this setting is suboptimal in a strong sense.

Proposition 1. *Suppose $s_A < 1$, $s_B \geq 1$, and all agents have the same value $b > 0$ for good B. Then the optimal mechanism never uses damages. Moreover, any feasible mechanism is weakly Pareto-dominated by a mechanism that uses only tolls, i.e. one with $x(a, b) \equiv 1$.*

The proof is in the appendix. To understand the intuition behind this result, note that under any mechanism, good A will go to agents whose values for it lie in some upper interval $[\underline{a}, 1]$. The designer's problem therefore boils down to selecting a cutoff $\underline{a}$ and choosing how to enforce it.

[6] When the designer provides a free or subsidized inferior good, only relatively poor agents will want to participate, since wealthier agents can afford higher-quality alternatives. Thus, the subsidy is automatically targeted to those who need it most (Besley and Coate, 1991). Once the designer provides cash, this form of targeting disappears, since money is desired by everyone regardless of wealth.

This requires deterring some agents from choosing good A, and can be done by damaging it or by pairing it with a toll. Note, however, that damages are more burdensome to *inframarginal types* than the types below $\underline{a}$ they are actually meant to deter. Tolls, on the other hand, are equally burdensome to everyone. Thus, enforcing the cutoff $\underline{a}$ with tolls yields higher welfare, as it leaves more rents to inframarginal takers of A (Figure 1).

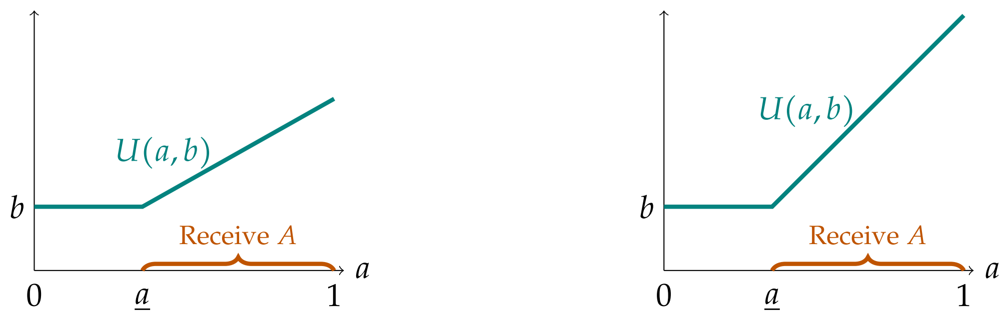


Figure 1: Indirect utilities $U(a,b) = b + \int_{\underline{a}}^{\max[\underline{a},a]} x(v,b)\, dv$ for mechanisms enforcing the cutoff $\underline{a}$ with damages (left) and tolls (right).

Replacing damages with tolls is also beneficial for the designer, who does not benefit from damages but values agents' utility and toll revenue.

Remark 1. *The logic of Proposition 1 extends to a more general class of screening instruments. We could consider two wasteful screening instruments, where the cost of one is increasing more steeply in the value for good A than the cost of the other. An analogous result would then say that any mechanism using the former instrument is Pareto-dominated by one using only the latter instrument.*

While the result is simple, it provides economic insight. For example, it captures a key force in the model of congestion pricing of Vickrey (1973) and its subsequent generalization by Van Den Berg and Verhoef (2011). The model studies the use of congestion pricing to spread the amount of traffic passing through a capacity-constrained road. It shows that when drivers' values for time are identical, congestion pricing does not improve their utility. This is because the same allocation of driving times is implemented with or without tolls: the "prices" for traveling at specific times are then pinned down by market clearing, and it is irrelevant whether they are paid in money, through tolls, or in wasted time, through congestion. This conclusion is overturned, however, when agents' values for time are heterogeneous. Proposition 1 makes this clear, as waiting in traffic is a *damage*. When the road's capacity constraint at peak time is enforced through payments, everyone pays the same price as the marginal driver. When it is enforced through waiting in traffic, the marginal driver experiences the same disutility as she would from the payment, but the inframarginal drivers with the highest values for arriving early suffer strictly more.

5 Two-dimensional heterogeneity

I now turn to the case where both goods are scarce and values are heterogeneous in both dimensions. I therefore assume that $s_A, s_B > 0$, $s_A + s_B \leq 1$ and that the distribution of values F has full support on $[0,1]^2$. This two-dimensional heterogeneity makes the setting significantly richer. Indeed, tolls and damages will serve not only to exclude low-value agents, but also to direct recipients to choose goods in a socially efficient way. Before I discuss how screening devices "sort" agents into goods, however, I show that incentive-compatible mechanisms can be conveniently represented using a boundary in the type space.

Boundary structure of mechanisms. Let us first define good-specific indirect utility functions $U_A, U_B : [0,1] \to \mathbb{R}_+$, given by:

$$U_A(a) = \max_{\substack{(a',b') \\ y(a',b')=A}} \big(x(a',b')\, a - c(a',b')\big)_+, \qquad U_B(b) = \max_{\substack{(a',b') \\ y(a',b')=B}} \big(x(a',b')\, b - c(a',b')\big)_+.$$

Intuitively, $U_A(a)$ and $U_B(b)$ represent the highest utility type (a,b) could get from selecting some quality and toll option for the A- and the B-goods, respectively, or not participating. U_A and U_B are convex, increasing, and depend only on one dimension of the type—an agent's value for good B does not affect her choice of quality and toll option if she chooses good A. Notice also that agents for whom $U_A(a) > U_B(b)$ choose good A and those for whom $U_A(a) < U_B(b)$ choose good B. These observations let us define a boundary characterizing different types' good choices.

Definition 1. *Define a mechanism's* ***participation cutoffs*** *as follows:*

$$\underline{a} = \sup\{a : \ U_A(a) = 0\}, \quad \underline{b} = \sup\{b : \ U_B(b) = 0\}.$$

Let a ***boundary*** *be a strictly increasing, continuous function $z : [\underline{a}, \overline{a}] \to [\underline{b}, \overline{b}]$ such that $\overline{a} \leq 1$ and $\overline{b} \leq 1$, with one of them holding with equality.*

Proposition 2. *Suppose the mechanism allocates positive masses of both goods. Then agents' choices of goods are characterized by the mechanism's participation cutoffs $\underline{a}, \underline{b}$ and a boundary z:*

1. *Types $(a,b) < (\underline{a}, \underline{b})$ do not get either good.*
2. *Types for whom $a > \underline{a}$ and $b < \underline{b}$ get good A; types for whom $a < \underline{a}$ and $b > \underline{b}$ get good B.*
3. *Types $(a,b) > (\underline{a}, \underline{b})$ get good A if (a,b) is below the boundary z, that is, if $z(a) > b$, and get good B if (a,b) is above the boundary z, that is, if $z(a) < b$.*

Moreover, types on the boundary are indifferent between their favorite options for both goods, thus:

$$U_A(a) = U_B(z(a)) \quad \text{for all} \quad a \in [\underline{a}, \overline{a}]. \tag{1}$$

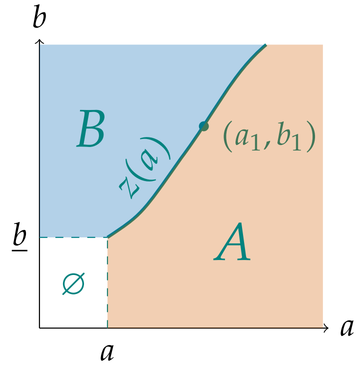


Figure 2: Types below the boundary (orange) choose good A and types above it (blue) choose good B.

When all offered A- and B-options come with tolls, types with sufficiently low values for both goods, i.e. $(a,b) < (\underline{a},\underline{b})$, do not participate. The choices of participating types are then described by an increasing boundary z along which the good-specific indirect utilities are equal. Types below this boundary choose good A, while types above it choose good B. Indeed, consider a type (a_1,b_1) on the boundary, so that $U_A(a_1) = U_B(b_1)$. Types lying below it satisfy $a > a_1$ and $b < b_1$, and so prefer good A to good B:

$$U_A(a) > U_A(a_1) = U_B(b_1) > U_B(b).$$

Analogously, types lying above (a_1,b_1) prefer good B.

I now discuss how different combinations of tolls and damages induce different boundaries z. This is in contrast with the one-dimensional setting studied in Section 4, where every pattern of sorting agents into goods could be achieved using either instrument.

Tolls and damages sort agents differently. For illustration, consider mechanisms that use only tolls and only damages. In the former case, we have $x(a,b) \equiv 1$, and so (IC) requires that each good be given with a single toll, c_A or c_B. Note that type-(a,b) agents will select good A if

$$a - c_A > b - c_B,$$

and select good B otherwise. This kind of screening can only lead to sorting patterns like the one illustrated in Figure 3a, where agents get good A if their types lie below a certain 45-degree line. It cannot create a sorting pattern like that in Figure 3b, where agents get good A if their types lie below a ray from the origin, that is, if

$$\frac{b}{a} < q,$$

for some q, and get good B otherwise. Such a sorting pattern can be achieved with damages, however. Consider a mechanism offering A with a damage, $x = q < 1$, and B without it, $x = 1$, with no tolls for either. The set of indifferent agents will then be given by:

$$a \cdot q = b \quad \Rightarrow \quad \frac{b}{a} = q.$$

Agents below and above this boundary will then choose goods A and B, respectively.

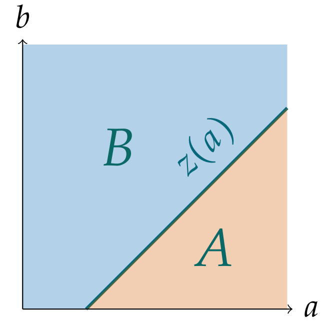


Figure 3a: Sorting pattern implementable with tolls alone.

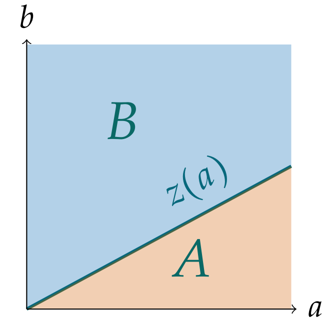
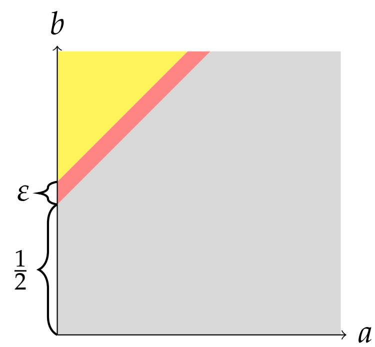


Figure 3b: Sorting pattern whose implementation requires damages.

Why damages can be optimal. As the following example shows, the optimal sorting pattern may sometimes require the use of damages:

Example 1. *Suppose the designer cares only about welfare, i.e. $\gamma = 0$, and consider the density*

$$f(a,b) = \begin{cases} \varepsilon \dfrac{2}{\left(\frac{1}{2} - \varepsilon\right)^2}, & \text{if } b - a \geq \frac{1}{2} + \varepsilon, \\ \dfrac{1}{3} \dfrac{2}{\varepsilon - \varepsilon^2}, & \text{if } b - a \in \left[\frac{1}{2}, \frac{1}{2} + \varepsilon\right), \\ \frac{8}{7}\left(\frac{2}{3} - \varepsilon\right), & \text{if } b - a < \frac{1}{2}. \end{cases}$$



and supplies given by $s_A = \frac{2}{3} - \varepsilon$ and $s_B = \frac{1}{3} + \varepsilon$. For $\varepsilon > 0$ sufficiently small, a mechanism using only tolls is not optimal.

The distribution from the example is illustrated in the figure to the right. The probability masses in all three shaded areas are constant in ε: they equal ε in the yellow area, $1/3$ in the red area, and $2/3 - \varepsilon$ in the gray one. The supply of good B is chosen to exactly match the total mass of the yellow and red areas, while the supply of good A matches the mass in the gray area.

Consider a mechanism for this distribution which uses only tolls but not damages. Discarding any supply of either good would not be helpful (this point is later shown formally in the proof of Theorem 1), so we can without loss consider only mechanisms where the whole supply is allocated. Without damages, this can only be achieved by giving out good A for free, and giving good B with a toll $c = 1/2$. Indeed, this mechanism induces a pattern of sorting illustrated in Figure 4a, with types shaded in orange getting good B and types shaded in blue getting good A. As discussed above, the boundary splitting the two regions is angled at 45 degrees.

Now, notice that for agents in the strip between the solid and dashed lines in Figure 4a, the surplus from getting good B over getting good A is at most ε. This is because most of their surplus is consumed by the toll $c = 1/2$. Consider then the case where $\varepsilon \to 0$. As ε falls, this surplus goes to zero for the whole 1/3-mass of agents in the aforementioned strip. Similarly, the mass of agents above the dashed line, equal to ε, also tends to zero. Consequently, total welfare then tends to that which would have resulted from all agents getting good A for free.

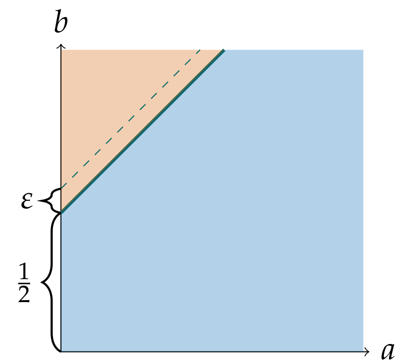


Figure 4a: The mechanism in Example 1 that gives good B with a toll.

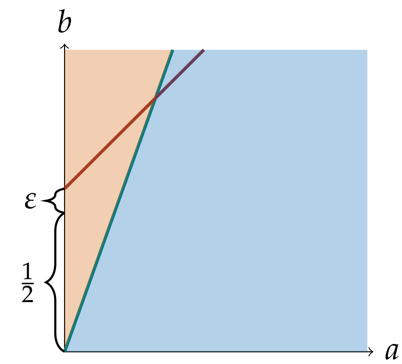


Figure 4b: The mechanism in Example 1 that damages good B.

Consider by contrast the mechanism which uses no tolls but damages good B to the point where the resulting boundary satisfies both goods' supply constraints (Figure 4b). Here too, total welfare of agents above the red line becomes negligible as $\varepsilon \to 0$. Note, however, that agents who are below the red line and far to the left of the green line now benefit substantially from getting good B over getting good A for free. These agents have a strong relative preference for B over A, and thus strongly prefer even damaged B to undamaged A. Since the mass of such agents does not go to zero as ε decreases, this mechanism creates substantial gains from allocating good B over A even in the limit case. This is in contrast to the toll-only mechanism, where all the gains from allocating B over A are "eaten away" by the toll used to screen high-b types, thus making the damage-based mechanism superior. Intuitively, this occurs because most of the consumers of good B under the toll-based mechanism have values very close to the "market-clearing toll".

When are damages not optimal? I now generalize the ideas from the previous section and provide sufficient conditions under which using damages is and is not optimal. I discuss the economic meaning of these conditions after stating both results. Throughout the rest of this section, I maintain the following assumption on the value distribution.

Assumption 1. *F has a strictly positive, Lipschitz continuous density f on $[0,1]^2$.*

For analytical convenience, I establish the first result under the following technical restriction. Note that it permits allocation rules to have finitely many discontinuities.

Assumption 2. *The designer is restricted to quality rules $x : [0,1]^2 \to [0,1]$ that are piecewise continuously differentiable in each dimension of the type.*

The final condition of the result is a shape restriction on the type distribution.

Assumption 3. *The inverse anti-hazard rates of the conditional distributions,*

$$\frac{F_{A|B}(a|b)}{f_{A|B}(a|b)}, \quad \frac{F_{B|A}(b|a)}{f_{B|A}(b|a)}, \tag{2}$$

are increasing in a and b, and for each ratio at least one of the two monotonicities is strict.

When the values for the two goods are independent, $F(a,b) = F_A(a)F_B(b)$, the condition then reduces to the requirement that

$$\frac{F_A(a)}{f_A(a)}, \quad \frac{F_B(b)}{f_B(b)}$$

be increasing, with at least one of them strictly so. This holds, for example, when the marginals are uniform, truncated normal, truncated decreasing exponential, or Beta(α,β) distributions with $\alpha,\beta \geq 1$.

Theorem 1. *Suppose Assumptions 1–3 hold. Then the optimal mechanism offers only two options: undamaged goods A and B with tolls of c_A^* and c_B^*, respectively. These tolls are chosen so that the whole supply of both goods is allocated.*

Thus, under the above conditions, the optimal mechanism replicates the competitive equilibrium allocation of goods: agents whose values for both goods are sufficiently low receive nothing, while all other agents get an undamaged version of one of the goods and pay the associated toll (Figure 5). While the optimality of this mechanism is immediate when tolls are welfare-neutral, the result asserts it remains optimal when the designer values a unit of toll revenue at any $\gamma \in [0,1]$ and so, in particular, when tolls are completely wasteful.

Throughout the rest of the paper, I refer to this mechanism as the *market-clearing toll mechanism* and write c_A^*, c_B^* for these market-clearing tolls. I also use

$$z_0(a) := a + c_B^* - c_A^*,$$

to denote the boundary generated by this mechanism and write $(\bar{a}^*, \bar{b}^*)$ for its end point.

Before providing an intuition for the conditions of Theorem 1, I state its partial converse which provides conditions under which the optimal mechanism does use damages.

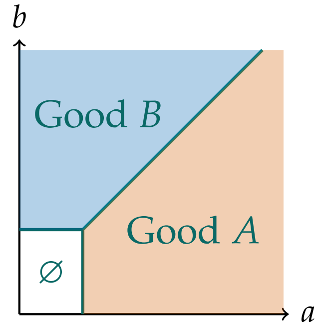


Figure 5: Optimal allocation under the conditions of Theorem 1.

When are damages optimal? Define the following objects for $a \in [c_A^*, 1]$ and $b \in [c_B^*, 1]$:

$$P_A^a := \int_0^{\min\{1, z_0(a)\}} f(a,t)\, dt, \quad P_B^b := \int_0^{\min\{1, z_0^{-1}(b)\}} f(t,b)\, dt, \quad P_{AB} := \int_{c_A^*}^{\bar{a}^*} f(a, z_0(a))\, da. \tag{3}$$

P_A^a is the density of agents who value good A at a and choose it under the market-clearing toll mechanism; P_B^b is defined analogously. P_{AB} measures the density of agents along the interior A-B boundary (Figure 6a).

Theorem 2. *Suppose Assumption 1 holds and the market-clearing toll for good B satisfies* $0 < c_B^* < 1$. *Suppose also that for some* $\tilde{b} \in (c_B^*, 1)$*, we have*

$$\int_{c_B^*}^{\tilde{b}} \big((1-\gamma)\tilde{b} - b\big) P_B^b\, db > (1-\gamma)\left(\alpha s_A + \beta s_B\right), \tag{4}$$

where

$$\alpha = \frac{P_B^{c_B^*}\left(P_{AB}(\tilde{b} - c_B^*) - Q\right)}{(P_A^{c_A^*} + P_{AB})(P_B^{c_B^*} + P_{AB}) - P_{AB}^2}, \qquad \beta = \frac{P_A^{c_A^*} Q + (P_A^{c_A^*} + P_{AB}) P_B^{c_B^*}(\tilde{b} - c_B^*)}{(P_A^{c_A^*} + P_{AB})(P_B^{c_B^*} + P_{AB}) - P_{AB}^2}, \tag{5}$$

and

$$Q := \int_{c_A^*}^{\bar{a}^*} \left(\tilde{b} - z_0(a)\right)_+ f(a, z_0(a))\, da.$$

Then the optimal mechanism uses damages.

Intuitively, Theorem 2 considers a local perturbation of the market-clearing mechanism from Theorem 1. Starting from the menu that offers good A at toll c_A^* and good B at toll c_B^*, the designer adds a slightly damaged option for good B. This new option gives B at quality $1-\varepsilon$ and offers a toll discount of $\varepsilon\tilde{b}$. An agent with value b for good B therefore trades off a quality loss worth εb against a toll reduction worth $\varepsilon\tilde{b}$.

This perturbation has two first-order effects. The first is the *inframarginal effect*: old B-choosers with $b < \tilde{b}$ benefit from the damaged option, because they receive a targeted toll reduction in exchange for a small quality loss. The second is the *toll-correction effect*: the damaged option also creates extra demand for good B, both from agents who previously consumed neither good and from agents who previously chose A. The latter group is shaded dark-blue in Figure 6b; to first order, its mass is given by Q. To restore market clearing, the designer must offset this extra demand by raising the original tolls on goods A and B, by $\varepsilon\alpha$ and $\varepsilon\beta$, respectively.

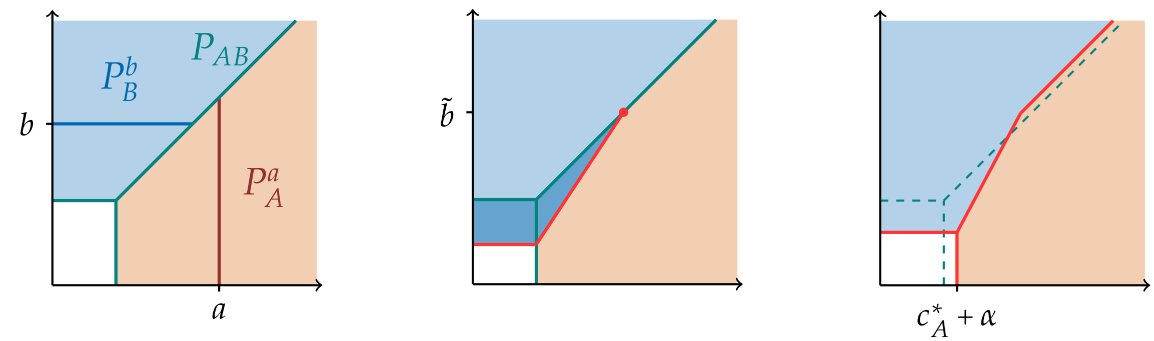


Figure 6a Figure 6b Figure 6c

Figure 6: The first subfigure illustrates the definitions in (3). The second shows the effect of introducing the damaged option to the menu from Theorem 1; the dark triangular region corresponds to Q. The last subfigure illustrates the toll adjustment that restores the supply constraints after the damaged option is added.

Condition (4) compares the inframarginal effect with the toll-correction effect. The left-hand side is the inframarginal effect: an old B-chooser with value $b < \tilde{b}$ loses quality worth εb, but receives a toll reduction worth $\varepsilon\tilde{b}$. Since the designer values a unit of toll revenue at γ, the net first-order gain from such a type is

$$(1-\gamma)\tilde{b} - b.$$

Aggregating this gain over old B-choosers gives the left-hand side of (4). The right-hand side is the toll-correction effect. Restoring market clearing requires raising the original tolls by α and β, and the net social cost of a one-unit toll increase is $1-\gamma$. Hence the total first-order cost of the toll correction is

$$(1-\gamma)\left(\alpha s_A + \beta s_B\right).$$

The condition says that the inframarginal effect dominates the toll-correction effect.

The requirement in Theorem 2 becomes especially simple when the total supply adds up to one and the designer does not value tolls:

Corollary 1. *Suppose Assumption 1 holds, $\gamma = 0$, $s_A + s_B = 1$, and the market-clearing toll for good B satisfies $0 < c_B^* < 1$. Then damages are optimal if for some $\tilde{a} \in (0, 1-c_B^*)$ we have*

$$\operatorname{Cov}\left(\frac{F_{A|B}(a \mid a + c_B^*)}{f_{A|B}(a \mid a + c_B^*)},\ (\tilde{a} - a)_+ \ \middle|\ b = a + c_B^*\right) > 0. \tag{6}$$

Formally, the conditional covariance is taken with respect to the probability measure on $[0, 1 - c_B^*]$ with density proportional to $f(a, a + c_B^*)$.

I now intuitively explain this covariance condition. First, note that since supplies add up to 1 and the market-clearing toll for good B is positive, the other toll is zero: $c_A^* = 0$. Indeed, if both tolls were positive, some agents would not choose either good and some supply would remain unallocated. Let us then consider a perturbation that introduces a slightly damaged option in this setting, analogous to the one used for Theorem 2. Figures 7a and 7b illustrate the effects of this perturbation, before and after the tolls are readjusted to preserve supply constraints.

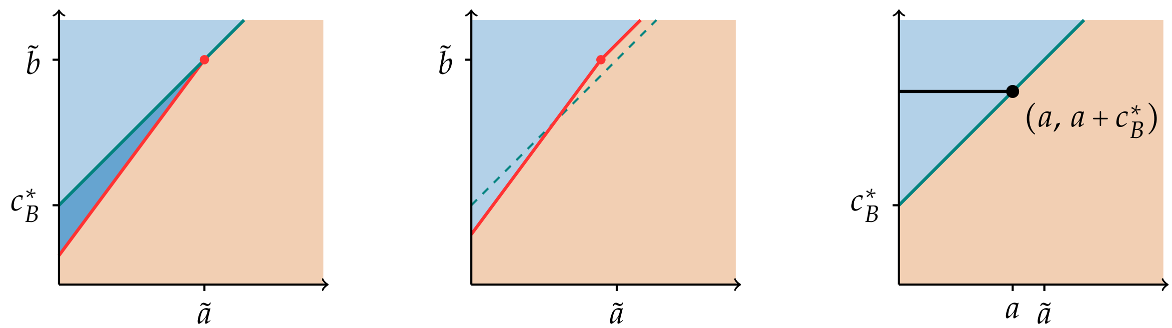


Figure 7a Figure 7b Figure 7c

Figure 7: The first subfigure shows how the type-split changes after the damaged B-option is added but before tolls are readjusted. The second subfigure shows how this split changes after this readjustment. The third subfigure illustrates the ratio between inframarginal old B-choosers and marginal A-B switchers among agents with a fixed value for b.

In this normalized case, the same two effects take a particularly simple form. The inframarginal effect benefits old B-choosers with $b < \tilde{b}$. The toll-correction effect comes from the induced movement of the old A-B boundary: some agents who previously chose A now switch into good B, so the designer must raise the B-toll to restore market clearing.

Now, fix a horizontal slice $b = a + c_B^*$. Along this slice, the inframarginal effect acts on old B-choosers on the black segment in Figure 7c. The toll-correction effect is generated by types on the old boundary z_0, marked by the black point. The relative density of these two groups is given by the inverse anti-hazard rate of the conditional distribution:

$$\frac{\int_0^a f(t, a + c_B^*)\, dt}{f(a, a + c_B^*)} = \frac{F_{A|B}(a \mid a + c_B^*)}{f_{A|B}(a \mid a + c_B^*)}. \tag{7}$$

Thus, the ratio in (7) measures the strength of the inframarginal effect relative to the toll-correction effect at the boundary point $(a, a + c_B^*)$. The covariance in (6) then aggregates this comparison across boundary points. Note that since $\tilde{b} = \tilde{a} + c_B^*$, we can write

$$(\tilde{b} - (a + c_B^*))_+ = (\tilde{a} - a)_+.$$

This term, which captures an agent's gain from taking the damaged option, scales the strength of both effects: where it is larger, inframarginal B-choosers enjoy larger welfare gains and more boundary types are induced to switch into B. Hence, the covariance in (6) asks whether the perturbation is strongest where the ratio of inframarginal beneficiaries to boundary switchers is high.

The covariance condition also clarifies the role of Assumption 3 in Theorem 1, which says that the inverse anti-hazard rate (7) is increasing in both arguments. This requirement ensures that

$$a \mapsto \frac{F_{A|B}(a \mid a + c_B^*)}{f_{A|B}(a \mid a + c_B^*)} \tag{8}$$

is increasing along the old boundary $b = a + c_B^*$. Since $a \mapsto (\tilde{a} - a)_+$ is decreasing, the covariance in (6) is then negative, and introducing a damaged option is not locally beneficial. This also suggests that Theorem 1 would continue to hold under weaker assumptions: monotonicity is a particularly strong way of ensuring the local optimality of not using damages.

Affiliated values and the optimality of damages. The monotonicity condition on the inverse anti-hazard ratio in (7) is related to affiliation. Following Milgrom and Weber (1982), values for goods A and B are affiliated when the joint density is log-supermodular. If f is strictly positive and twice continuously differentiable, this is equivalent to

$$\frac{\partial^2}{\partial a, \partial b} \log f(a, b) \geq 0. \tag{9}$$

The reverse inequality corresponds to log-submodularity, or negative affiliation.

Note that the conditions of Theorem 1, which require that damages are not optimal when the conditional inverse anti-hazard rates are increasing, rule out positive affiliation between the two values but is compatible with their negative dependence. To see this, suppose f is differentiable; then:

$$\frac{d}{db} \frac{F_{A|B}(a \mid b)}{f_{A|B}(a \mid b)} = \frac{1}{f(a,b)} \int_0^a f(t,b) \left(\frac{\partial}{\partial b} \log f(t,b) - \frac{\partial}{\partial b} \log f(a,b) \right) dt. \tag{10}$$

Since positive affiliation makes $\partial_b \log f(a,b)$ increasing in a, the integrand in (10) is negative, so affiliation pushes the inverse anti-hazard ratio downward as b increases. Negative affiliation reverses this inequality: the integrand is positive, and the inverse anti-hazard ratio is increasing in b, as required by Assumption 3.

Moreover, positive affiliation makes damages more likely to be optimal. Along the old boundary $b = a + c_B^*$, affiliation tends to make the ratio in (7) lower at high values of a, and therefore relatively higher at low values of a. Since $(\tilde{a} - a)_+$ is also highest at low values of a, this makes the covariance in (6) more likely to be positive.

The connection between affiliation and the optimality of damages has a simple economic interpretation. Affiliation affects the relative strength of the inframarginal effect and the toll-correction effect. When values for the two goods are positively related, there is more mass near the old A-B boundary, where agents are almost indifferent between the two goods, and less mass in the off-diagonal region where agents are inframarginal B-choosers. Thus affiliation strengthens the toll-correction effect relative to the inframarginal effect: the damaged option attracts more agents into B, requiring a larger toll increase to restore market clearing. By contrast, negative affiliation puts relatively more mass on inframarginal B-choosers and less mass near the boundary. The damaged option then delivers more targeted relief while creating a smaller toll correction, making damages more likely to be optimal.

I conclude with an example where sufficiently strong affiliation makes damages optimal.

Example 2. *Assume $\gamma = 0$ and fix $c_B \in (0,1)$. For each λ, consider the density*

$$f_\lambda(a,b) \propto e^{\lambda ab} \qquad (a,b) \in [0,1]^2.$$

Suppose supplies add up to 1 *and are such that the market-clearing tolls are c_A, c_B. Then, for all sufficiently large λ, the optimal mechanism uses damages.*

Indeed, λ directly indexes the strength of affiliation, since $\frac{\partial^2}{\partial a \, \partial b} \log f_\lambda(a,b) = \lambda$.

6 Proof of Theorem 1

This section gives the main steps of the proof of Theorem 1; I show the referenced lemmas in the appendix. The argument is based on the boundary representation from the previous section: any implementable mechanism is summarized by good-specific indirect utilities U_A and U_B, together with a boundary z that sorts types between the two goods. This lets us treat the two-dimensional problem as two endogenously connected one-dimensional screening problems.

We first make a useful normalization. Recall that the terminal point of the boundary lies on the boundary of the unit square: either $\overline{a} = 1$, so the boundary reaches the right edge first, or $\overline{b} = 1$, so it reaches the top edge first. Since the boundary can equivalently be parametrized by its inverse $z^{-1}(b)$ on $[\underline{b}, \overline{b}]$, we can break symmetry and without loss assume that $\overline{b} = 1$.

Reformulating welfare. Throughout most of the proof, I show the result for a purely welfarist objective, which is the special case for $\gamma = 0$; I then show that if the market-clearing toll mechanism is optimal when tolls have no social value, it remains optimal when the designer puts positive weight on toll revenue. I begin by rewriting aggregate welfare in terms of the boundary z and the A-indirect utility U_A. To that end, define the *extended boundary* $\hat{z}$ by

$$\hat{z}(a) = \begin{cases} 0, & \text{if } a < \underline{a}, \\ z(a), & \text{if } a \in [\underline{a}, \overline{a}], \\ 1, & \text{if } a > \overline{a}. \end{cases}$$

That is, $\hat{z}$ equals z on the latter's domain, takes value zero below it and takes value 1 above it (Figure 8a). Throughout, I use $\hat{z}^{-1}$ to denote its generalized inverse.

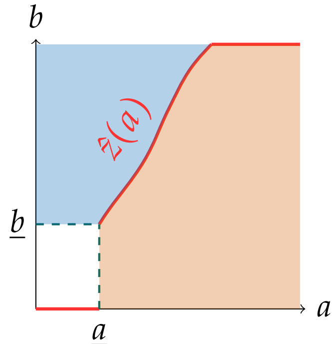


Figure 8a: Extended boundary $\hat{z}$.

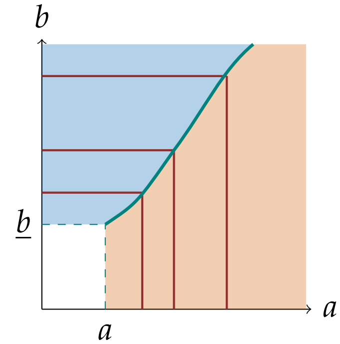


Figure 8b: Agents on the same L-shaped curves have equal utilities.

Lemma 1. *Consider a mechanism with a boundary z and A-indirect utility U_A. Total welfare under this mechanism is then given by:*

$$U_A(1) - \int_0^1 U_A'(a) \cdot F(a, \hat{z}(a))da. \tag{11}$$

To see why welfare can be expressed using z and U_A, note that the utility of an agent choosing good A depends only on a, while the utility of an agent choosing good B depends only on b. Since agents on the boundary are indifferent between the two goods, all types lying on the same inverted L-shaped curve in Figure 8b have the same utility. We can therefore calculate welfare by integrating over types who choose A while also taking into account the B-taking types on the same L-shaped curves. Such a calculation yields the expression (11).

This form of the objective also bears a resemblance to a Myersonian virtual value which would appear in a single-dimensional setting with a welfarist objective (see e.g. Condorelli (2012)). Moreover, by the envelope theorem, we can think of $U_A'(a)$ as the quality allocated to agents who receive good A and value it at a. Unfortunately, however, the setting does not lend itself to a Myersonian solution method. This is because the shape of the "virtual value" itself is endogenous to the choice of boundary. In fact, it is optimizing over the shape of the boundary which poses the greatest difficulty in proving the result.[7]

The proof of the $\gamma = 0$ case proceeds in three steps. First, I characterize the feasible pairs (z, U_A). Second, I fix a boundary z and find the welfare-maximizing U_A compatible with it. Third, I optimize over the boundary itself. The main step is the last one: using optimal control arguments,

[7]The difference between my approach and the Myersonian one is suggested by Assumption 3, which is imposed on the (conditional) *anti*-hazard rate, and not on the hazard rate, as it would be in a standard Myersonian setting with a welfarist objective. Indeed, the reason why the anti-hazard rate appears in my condition is logically distinct from the one for the presence of the hazard rate in the Myersonian problem.

I show that the welfare-maximizing boundary must be linear with slope 1. Such a boundary is implemented by undamaged goods and market-clearing tolls.

Characterizing feasible (z, U_A). We say (z, U_A) is *feasible* if there exists a mechanism (c, x, y) with A, B-indirect utilities U_A, U_B such that:

$$U_A(a) = U_B(z(a)) \quad \text{for all} \quad a \in [\underline{a}, \overline{a}]. \tag{12}$$

Lemma 2. *Under Assumption 2, the pair (z, U_A) is feasible if and only if:*

(i) *U'_A and U'_A/z' are non-decreasing and strictly positive above $\underline{a}$,*

(ii) *the boundary z has finite, strictly positive one-sided derivatives at every $a \in (\underline{a}, \overline{a})$, and a finite, strictly positive left derivative at $\overline{a}$,*

(iii) *the supply constraint (S′) holds:*

$$\int_{\underline{a}}^{1} \int_{0}^{\hat{z}(a)} f(a, v) dv \, da \le s_A, \quad \int_{\underline{b}}^{1} \int_{0}^{\hat{z}^{-1}(b)} f(v, b) dv \, db \le s_B, \tag{S′}$$

(iv) *U'_A and U'_A/z' are piecewise continuously differentiable, and z is piecewise twice continuously differentiable.*

To understand these conditions, note that the A-indirect utility U_A and the boundary z pin down U_B through (12). In particular, differentiating this condition gives

$$U'_A(a) = U'_B(z(a)) \cdot z'(a) \quad \Rightarrow \quad U'_B(z(a)) = \frac{U'_A(a)}{z'(a)}. \tag{13}$$

Since U_A and U_B must be convex, condition (i) requires the implied marginal utility profiles to be nondecreasing. Condition (ii) records the regularity of the boundary implied by (13). Condition (iii) rewrites the supply constraints in terms of the boundary: types below z receive good A, while types above z receive good B. Condition (iv) is the regularity imposed by Assumption 2: the induced marginal utilities and boundary must be piecewise smooth. Without this assumption, this condition would be omitted.

Welfare-maximizing U_A for a fixed boundary z. I now fix a boundary z and find the A-indirect utility U_A that maximizes total welfare (11) subject to (z, U_A) being feasible. This is the first step in the proof that needs Assumption 2, which rules out pathological boundaries and ensures that z, U'_A, and the implied B-marginal utility U'_A/z' are piecewise smooth.

Lemma 3. *Fix a piecewise twice continuously differentiable boundary z. Then (z, U_A), with U_A defined by (14), maximizes total welfare (11) among all feasible pairs $(z, \check{U}_A)$.*

$$U_A'(a) = \begin{cases} 0, & a \in (0, \underline{a}), \\ m(a) \cdot k, & a \in (\underline{a}, \overline{a}), \\ 1, & a \in (\overline{a}, 1). \end{cases} \tag{14}$$

where

$$m(a) = \exp\left(\int_{\underline{a}}^{a} \max\left[0, \frac{z''(s)}{z'(s)}\right] ds \right) \prod_{\substack{z'^{+}(t) > z'^{-}(t), \\ t \leq a}} \frac{z'^{+}(t)}{z'^{-}(t)}, \qquad k = \frac{1}{\max[m(\overline{a}),\ m(\overline{a})/z'^{-}(\overline{a})]}.$$

The main idea is that, for a fixed boundary z, the best implementation damages goods as little as possible while still satisfying incentive compatibility. Indeed, (14) has a simple implication: U_A' is constant on intervals where z is concave, while U_A' is proportional to z' on intervals where z is convex. Recall that

$$U_B'(z(a)) = \frac{U_A'(a)}{z'(a)}. \tag{15}$$

Thus, we can equivalently say that $U_A'(a)$ is constant on concave intervals of z while $U_B'(z(a))$ is constant on its convex intervals (Figure 9). Now, by Lemma 2, $U_A'(a)$ and $U_B'\big(z(a)\big)$ must

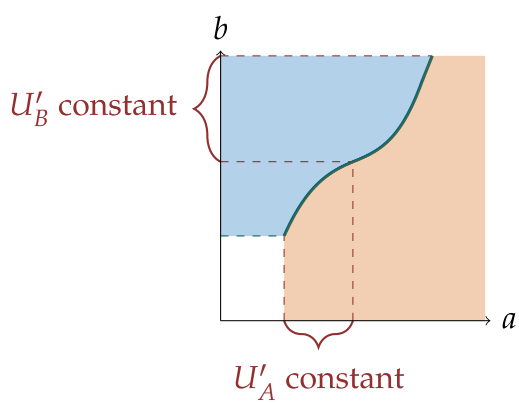


Figure 9: $U_A'(a)$ and $U_B'\big(z(a)\big)$ are constant where the boundary z is concave and convex, respectively.

be increasing. The above observation therefore means that at least one of these monotonicity constraints must always bind. Intuitively, were neither constraint to bind on some interval, we could increase $U_A'(a)$ and $U_B'\big(z(a)\big)$ pointwise in a (15)-preserving manner until one of them started binding. This would create pointwise higher utility profiles and thus produce superior welfare (11).

Showing the welfare-maximizing boundary is linear. Having pinned down the best U_A for each fixed boundary z, we can optimize over the boundary itself. This is where the inverse anti-hazard-rate condition in Assumption 3 enters.

Lemma 4. *Consider mechanisms that allocate positive masses of both goods. Then, under Assumptions 1–3, the welfare-maximizing mechanism features a linear boundary* $z^* : [\underline{a}^*, \overline{a}^*] \to [\underline{b}^*, \overline{b}^*]$.

Intuitively, I show that under Assumption 3, no boundary with strictly convex/concave parts or kinks can satisfy the necessary optimality conditions, and thus that the optimal boundary has to be linear. Consider some interval $[\underline{a}, \overline{a}]$ on which the boundary is concave and, for simplicity, assume it consists of multiple small, linear pieces (Figure 10a). We will consider the perturbations to these linear pieces on this part of the boundary. Notice, however, that such perturbations have to respect the supply constraint (S′), and thus must preserve the probability mass below and above the boundary. Still, we can construct perturbations preserving the supply constraint by perturbing one piece of the boundary upwards and another one downwards in a ratio that leaves the probability masses unchanged (Figure 10b). First-order optimality conditions then tell us that, when perturbing one such piece, we can capture the effect of this perturbation on the supply constraint by a Lagrange multiplier μ.

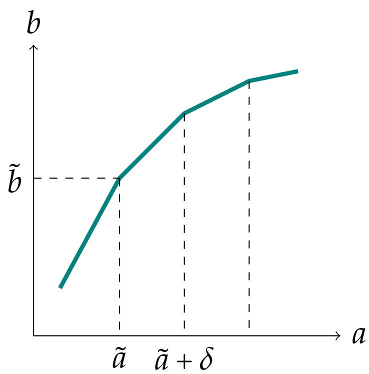


Figure 10a

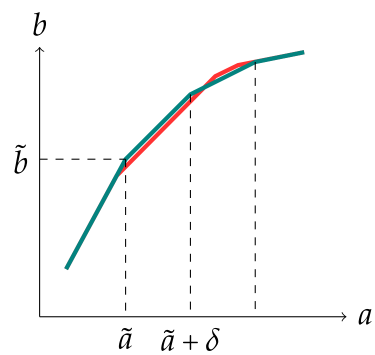


Figure 10b

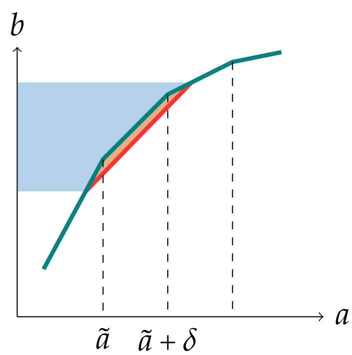


Figure 10c

Now, recall that by Lemma 1, total welfare is given by (11):

$$U_A(1) - \int_0^1 U_A'(a) \cdot F(a, \hat{z}(a))da.$$

Moreover, Lemma 3 tells us that U_A' is equal to some constant on the region where the boundary is concave. Thus, we can write the effect of the boundary in the region $[\tilde{a}, \tilde{a}+\delta]$, up to scaling, as follows:

$$-\int_{\tilde{a}}^{\tilde{a}+\delta} F\left(a, \tilde{b} + (a - \tilde{a}) \cdot s\right) da.$$

Now, consider the effect of a small downward perturbation to the height of the boundary on this interval, as illustrated in Figure 10c. The first-order effect of this perturbation is given by:

$$\frac{d}{d\tilde{b}} \int_{\tilde{a}}^{\tilde{a}+\delta} F\left(a, \tilde{b} + (a - \tilde{a}) \cdot s\right) da \; - \; \mu \cdot \frac{d}{d\tilde{b}} \int_{\tilde{a}}^{\tilde{a}+\delta} \int_{0}^{\tilde{b}+(a-\tilde{a})\cdot s} f(a,b) db da.$$

The latter term captures the effect of the perturbation on the probability mass under the boundary, which is valued according to the aforementioned multiplier μ on the supply constraint (S). Performing a small perturbation like this (in either direction) is not beneficial when:

$$0 \approx \int_{\mathcal{K}} f(a,b)\,\mathrm{d}(a,b) \; - \; \mu \int_{\mathcal{L}} f(a,b)\,\mathrm{d}(a,b)$$

where $\mathcal{K}$ and $\mathcal{L}$ are, respectively, the blue region in Figure 10c and the orange region in Figure 10c.When the length δ of the perturbed interval becomes small, this gives:

$$0 \; = \; \frac{F_{A|B}(\tilde{a}|z(\tilde{a}))}{f_{A|B}(\tilde{a}|z(\tilde{a}))} \; - \; \mu. \tag{16}$$

Consequently, when the boundary is strictly concave on some region, a profitable perturbation analogous to that in Figure 10b does not exist only if (16) holds at every point in that interval. However, this cannot be the case by Assumption 3, which guarantees that the inverse anti-hazard rate is strictly increasing there. Thus, if we were indifferent about perturbing the boundary slightly up or down at some level of a, we would strictly prefer to perturb it upwards for any higher a. The proof formalizes this reasoning in the continuous case using optimal control methods.

The welfare-maximizing linear boundary has unit slope. Lemma 4 tells us the welfare-maximizing boundary is linear and Lemma 3 pins down the best implementation of any linear boundary: in particular, if it has slope $s \geq 1$, then good A is undamaged and good B is delivered at quality $1/s$. It remains to show that the optimal slope is 1.

Lemma 5. *Consider mechanisms that allocate positive masses of both goods. Then, under Assumptions 1–3, the welfare-maximizing mechanism features a linear boundary with slope* 1.

Suppose the boundary z has slope $s > 1$ and reaches the ceiling of the unit square before reaching its right wall, as in Figure 11a. By Lemma 3, this boundary is implemented by offering good A undamaged at toll $\underline{a}_z$, and good B at quality $1/s$ and toll $\underline{b}_z/s$.

Now compare z to a slightly flatter boundary r that preserves the masses allocated to both goods, as in Figure 11b. The two boundaries cross once, at a^*, and the flatter boundary has a lower toll for good A: $\underline{a}_r < \underline{a}_z$. Let

$$\underline{\mathcal{D}} = \{(a,b) : 0 < a < a^*,\; \hat{z}(a) < b < \hat{r}(a)\}, \qquad \overline{\mathcal{D}} = \{(a,b) : a^* < a < 1,\; \hat{r}(a) < b < \hat{z}(a)\}.$$

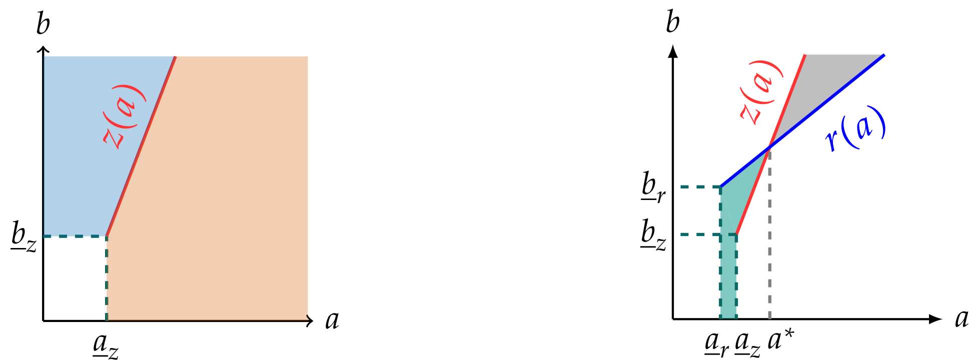


Figure 11a: s-sloped boundary z.

Figure 11b: Regions $\overline{\mathcal{D}}$ (gray) and $\underline{\mathcal{D}}$ (green).

These are the types switched from B to A, and from A to B, respectively, depicted by the green and gray regions in Figure 11b. Since r preserves the masses assigned to both goods,

$$\int_{\overline{\mathcal{D}}} f(a,b)\,\mathrm{d}(a,b) = \int_{\underline{\mathcal{D}}} f(a,b)\,\mathrm{d}(a,b). \tag{17}$$

Let Δ be the welfare difference between r and z. Since good A is undamaged under both boundaries, $U'_A \equiv 1$, and so

$$\begin{aligned}\Delta &= (1-\underline{a}_r) - \int_0^1 F\left(a,\hat{r}(a)\right)\,da - \left((1-\underline{a}_z) - \int_0^1 F\left(a,\hat{z}(a)\right)\,da\right)\\ &= (\underline{a}_z - \underline{a}_r) - \left(\int_{\underline{\mathcal{D}}} \frac{F_{A|B}(a\mid b)}{f_{A|B}(a\mid b)}\cdot f(a,b)\,\mathrm{d}(a,b) \;-\; \int_{\overline{\mathcal{D}}} \frac{F_{A|B}(a\mid b)}{f_{A|B}(a\mid b)}\cdot f(a,b)\,\mathrm{d}(a,b)\right).\end{aligned}$$

The first term is positive because $\underline{a}_r < \underline{a}_z$. The bracketed term is negative by Assumption 3, as the inverse anti-hazard rate is increasing in both arguments, while $\overline{\mathcal{D}}$ lies to the north-east of $\underline{\mathcal{D}}$ and, by (17), the two regions have equal probability mass.

Since $\Delta > 0$, any boundary with slope $s > 1$ is dominated by a flatter boundary, and so the optimal linear boundary has slope 1 which in turn is optimally implemented without damages.

The welfare-maximizing mechanism uses the whole supply. Finally, we need to show that the welfare-maximizing mechanism allocates all available goods. As shown above, this mechanism posts a toll for each good. Then, if this mechanism had some supply left over, we could improve welfare by simply lowering the associated tolls and allocating what remains.

Lemma 6. *The welfare-maximizing mechanism allocates the whole supply of both goods.*

We have therefore shown that the unique market-clearing toll mechanism is welfare-maximizing. It remains to show it is also optimal when the designer puts a weight of $\gamma \in [0,1]$ on toll revenue.

From welfare-maximization to the full objective. To extend this result to general objectives, note first that the same mechanism also maximizes allocative efficiency:

Proposition 3. *The market-clearing toll mechanism maximizes allocative efficiency:*

$$\int \left[\mathbb{1}_{y(a,b)=A}\, x(a,b)a + \mathbb{1}_{y(a,b)=B}\, x(a,b)b\right] dF(a,b), \tag{E}$$

subject to the supply constraint (S).

This is the standard efficiency property of competitive equilibrium: the tolls clear the two markets, and agents choose the good for which they have the highest net value.

Finally, observe that the designer's objective (O) can be written as a positive linear combination of welfare and allocative efficiency. Since the market-clearing toll mechanism separately maximizes both of them, it also maximizes their combination.

7 Extension: heterogeneous toll costs

The baseline model assumes that a given toll imposes the same cost on every agent. This is restrictive in some applications. For example, monetary payments may be more burdensome for poorer agents, who may also be the agents who benefit most from the program. I now allow the cost of completing a toll to vary across agents.

Suppose an agent's type is (a,b,r), where $r > 0$ is her unit cost of the toll. If she receives good A or B, her utility is respectively

$$\begin{aligned} &x \cdot a - r \cdot c \ \text{ if } \ y = A, \\ &x \cdot b - r \cdot c \ \text{ if } \ y = B. \end{aligned}$$

Assume that $r \in [\underline{r}, \overline{r}]$, with $\underline{r} > 0$, and that types are distributed according to $\tilde{F}$ with full support on $[0,1]^2 \times [\underline{r}, \overline{r}]$. This extension can be reduced to the baseline problem by a change of variables. Following Dworczak et al. (2021), observe that a type (a,b,r) is behaviorally equivalent to a type with unit toll cost and values

$$(\tilde{a}, \tilde{b}) = \left(\frac{a}{r}, \frac{b}{r}\right).$$

Let G denote the distribution of transformed values $(\tilde{a}, \tilde{b})$ induced by $\tilde{F}$. Its support is contained in $[0, 1/\underline{r}]^2$, and it has full support on this square. The transformation also changes the planner's objective. Since the utility of the original type (a,b,r) is equal to r times the utility of the transformed type $(a/r, b/r, 1)$, transformed types receive welfare weights

$$\lambda(\tilde{a}, \tilde{b}) = \mathbb{E}\left[r \;\middle|\; \frac{a}{r} = \tilde{a},\ \frac{b}{r} = \tilde{b}\right].$$

Thus, after renormalizing toll costs to one, the designer's objective becomes

$$\int \left[\lambda(\tilde{a},\tilde{b})u_{\tilde{a},\tilde{b}}(\tilde{a},\tilde{b}) + \gamma\tilde{c}(\tilde{a},\tilde{b})\right] dG(\tilde{a},\tilde{b}), \tag{18}$$

where $\gamma \in [0,1]$ and $\tilde{c}(\tilde{a},\tilde{b})$ denotes the toll assigned to transformed type $(\tilde{a},\tilde{b})$. Theorem 1 then has the following direct analogue.

Corollary 2. *Suppose Assumption 2 holds and that there are fewer goods than agents: $s_A + s_B \leq 1$. Let g denote the density of G, and suppose g is Lipschitz continuous. Moreover, assume that the weighted inverse anti-hazard ratios*

$$\frac{\int_0^a \lambda(v,b)g(v,b)\,dv}{g(a,b)}, \qquad \frac{\int_0^b \lambda(a,v)g(a,v)\,dv}{g(a,b)}$$

are increasing in both arguments, and that for each ratio at least one of the two monotonicities is strict. Then the market-clearing toll mechanism is optimal.

The proof is the same after replacing the density f in the welfare terms by the weighted density λg, while keeping the supply constraints expressed in terms of the unweighted density g. Thus the same argument goes through with the inverse anti-hazard ratios above in place of the corresponding ratios from the baseline model.

Theorem 2 can be modified analogously. The supply-preserving terms are still computed using the unweighted transformed density g, while utility changes are computed using the weighted density λg. The following is the normalized analogue of Corollary 1:

Corollary 3. *Let $\bar{v} := 1/\underline{r}$. Suppose g is strictly positive and Lipschitz continuous, the market-clearing toll for good B satisfies $0 < c_B^* < \bar{v}$, $s_A + s_B = 1$, and $\gamma = 0$. Then damages are optimal if for some $\tilde{a} \in (0, \bar{v} - c_B^*)$,*

$$\operatorname{Cov}\left(\frac{\int_0^a \lambda(v,a+c_B^*)g(v,a+c_B^*)\,dv}{g(a,a+c_B^*)}, (\tilde{a}-a)_+ \,\middle|\, b = a + c_B^*\right) > 0. \tag{19}$$

The conditional covariance is taken with respect to the probability measure on $[0,\bar{v} - c_B^*]$ with density proportional to $g(a, a + c_B^*)$.

8 Discussion

The main contribution of this paper is to distinguish between two kinds of screening instruments: tolls, whose costs are separable from agents' values for the allocated good, and damages, which are more burdensome for agents who value the good more. I study when a welfare-maximizing designer should use each class of instruments. When the allocated good is homogeneous, my results unambiguously predict that tolls are superior to damages. Consider, for instance, rationed access to specialist appointments or elective procedures. If access is allocated through a waitlist, the burden falls especially heavily on patients whose need for treatment is greatest. When such waitlists become long, introducing or raising a copayment is likely to be welfare-improving: in equilibrium, it screens out the lowest-need visits and reduces wait times for the sickest patients. Similarly, in food assistance programs such as SNAP, usage restrictions are disproportionately burdensome for households that rely on the program most. My model suggests they should be replaced by more stringent application or recertification requirements.

In settings where the allocated goods are heterogeneous, however, the prediction is more nuanced and depends on the statistical relationship between agents' values for the different goods. When these are strongly negatively related, the optimal mechanism is likely to use only tolls. Consider, for instance, the allocation of DMV appointments across offices dispersed throughout a city. Applicants may place high value on an appointment near where they live, while placing low value on appointments at more distant offices. In this case, the designer should use office-specific toll instruments—such as location-specific booking fees, application burdens, or walk-in queues—to clear excess demand and sort applicants across offices. Long waitlists at the most popular offices are likely to be suboptimal. When values for different goods are strongly positively related, the conclusion is likely to change. This can be the case for an affordable housing program managing projects located in the same area: the most needy households may value all units highly, while those with better outside options may not. When such correlation is strong, market-clearing tolls are more likely to absorb much of the surplus generated by the allocation, and damages may improve sorting across goods. Thus, in some housing settings, differentiated waitlists may be part of an optimal mechanism, possibly alongside project-specific toll-like instruments, such as higher rent contributions for overdemanded projects or rent subsidies for less demanded ones. Characterizing the exact optimal design of such mechanisms is an avenue for future research.

Finally, on a methodological level, the paper provides a tractable approach to two-dimensional screening problems under the restriction to deterministic mechanisms. The two-good case captures qualitatively important economic forces, such as screening agents into and out of different options, while remaining analytically simple. The approach could therefore be useful in other screening problems where deterministic mechanisms are economically plausible. For example, it may be applicable to two-good monopoly pricing with unit-demand consumers, or to optimal tax problems in which a couple chooses whether to file jointly or separately and faces different labor-supply incentives under the two filing regimes.

References

Akbarpour, M., P. Dworczak, and F. Yang (2023): "Comparison of Screening Devices," in *Proceedings of the 24th ACM Conference on Economics and Computation*, 60–60.

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, 371–427.

ARNOSTI, N. AND P. SHI (2020): "Design of lotteries and wait-lists for affordable housing allocation," *Management Science*, 66, 2291–2307.

BARZEL, Y. (1974): "A theory of rationing by waiting," *The Journal of Law and Economics*, 17, 73–95.

BESLEY, T. AND S. COATE (1991): "Public Provision of Private Goods and the Redistribution of Income," *The American Economic Review*, 81, 979–984.

——— (1992): "Workfare versus welfare: Incentive arguments for work requirements in poverty-alleviation programs," *The American Economic Review*, 82, 249–261.

BLOCH, F. AND D. CANTALA (2017): "Dynamic Assignment of Objects to Queuing Agents," *American Economic Journal: Microeconomics*, 9, 88–122.

BROT-GOLDBERG, Z. C., S. BURN, T. LAYTON, AND B. VABSON (2023): "Rationing medicine through bureaucracy: authorization restrictions in Medicare," Tech. rep., National Bureau of Economic Research.

BULOW, J. AND P. KLEMPERER (2012): "Regulated prices, rent seeking, and consumer surplus," *Journal of Political Economy*, 120, 160–186.

CITY OF VANCOUVER (2016): "Affordable Home Ownership Pilot Program," Policy Report to the Standing Committee on City Finance and Services, Vancouver City Council.

CONDORELLI, D. (2012): "What money can't buy: Efficient mechanism design with costly signals," *Games and Economic Behavior*, 75, 613–624.

CURRIE, J. AND F. GAHVARI (2008): "Transfers in Cash and In-Kind: Theory Meets the Data," *Journal of Economic Literature*, 46, 333–383.

DASKALAKIS, C., A. DECKELBAUM, AND C. TZAMOS (2017): "Strong duality for a multiple-good monopolist," *Econometrica*, 85, 735–767.

DENECKERE, R. J. AND P. R. MCAFEE (1996): "Damaged goods," *Journal of Economics & Management Strategy*, 5, 149–174.

DESHPANDE, M. AND Y. LI (2019): "Who is screened out? Application costs and the targeting of disability programs," *American Economic Journal: Economic Policy*, 11, 213–248.

DUPAS, P., V. HOFFMANN, M. KREMER, AND A. P. ZWANE (2016): "Targeting health subsidies through a nonprice mechanism: A randomized controlled trial in Kenya," *Science*, 353, 889–895.

DWORCZAK, P. (2026): "How to Allocate Money?" *American Economic Journal: Microeconomics*, 18, 312–339.

DWORCZAK, P., S. D. KOMINERS, AND M. AKBARPOUR (2021): "Redistribution Through Markets," *Econometrica*, 89, 1665–1698.

FEINBERG, E. A. AND A. B. PIUNOVSKIY (2006): "On the Dvoretzky–Wald–Wolfowitz theorem on nonrandomized statistical decisions," *Theory of Probability & Its Applications*, 50, 463–466.

FINKELSTEIN, A. AND M. J. NOTOWIDIGDO (2019): "Take-up and targeting: Experimental evidence from SNAP," *The Quarterly Journal of Economics*, 134, 1505–1556.

HARTLINE, J. D. AND T. ROUGHGARDEN (2008): "Optimal mechanism design and money burning," in *Proceedings of the fortieth annual ACM symposium on Theory of computing*, 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, 54–90.

LESHNO, J. D. (2022): "Dynamic matching in overloaded waiting lists," *American Economic Review*, 112, 3876–3910.

MANELLI, A. M. AND D. R. VINCENT (2006): "Bundling as an optimal selling mechanism for a multiple-good monopolist," *Journal of Economic Theory*, 127, 1–35.

MANNING, W. G., J. P. NEWHOUSE, N. DUAN, E. B. KEELER, AND A. LEIBOWITZ (1987): "Health insurance and the demand for medical care: evidence from a randomized experiment," *The American economic review*, 251–277.

MILGROM, P. R. AND R. J. WEBER (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50, 1089–1122.

NEUSTADT, L. W. (1976): *Optimization: A Theory of Necessary Conditions*, Princeton University Press.

NODA, S. AND G. OKADA (2024): "No Screening is More Efficient with Multiple Objects," *arXiv preprint arXiv:2408.10077*.

NICHOLS, A. L. AND R. J. ZECKHAUSER (1982): "Targeting Transfers through Restrictions on Recipients," *The American Economic Review*, 72, 372–377.

NICHOLS, D., E. SMOLENSKY, AND T. N. TIDEMAN (1971): "Discrimination by waiting time in merit goods," *The American Economic Review*, 61, 312–323.

ROCHET, J.-C. AND P. CHONÉ (1998): "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66, 783–826.

Seierstad, A. and K. Sydsaeter (1986): *Optimal control theory with economic applications*, Elsevier North-Holland, Inc.

Tokarski, F. (2026): "Targeting Without Transfers," *arXiv preprint arXiv:2602.00487*.

Van Den Berg, V. and E. T. Verhoef (2011): "Winning or losing from dynamic bottleneck congestion pricing?: The distributional effects of road pricing with heterogeneity in values of time and schedule delay," *Journal of Public Economics*, 95, 983–992.

Van Ommeren, J. N. and A. J. Van der Vlist (2016): "Households' willingness to pay for public housing," *Journal of Urban Economics*, 92, 91–105.

Vickrey, W. (1973): *Pricing, metering, and efficiently using urban transportation facilities*, 476.

Waldinger, D. (2021): "Targeting in-kind transfers through market design: A revealed preference analysis of public housing allocation," *American Economic Review*, 111, 2660–2696.

World Health Organization (2023): "High-value referrals: learning from challenges and opportunities of the COVID-19 pandemic: concept paper," *High-value referrals: learning from challenges and opportunities of the COVID-19 pandemic: concept paper*.

Yang, F. (2021): "Costly multidimensional screening," *arXiv preprint arXiv:2109.00487*.

A Justifying the direct-revelation formulation

In the main model, I restrict the designer to deterministic menus, meaning that each menu option allocates good A, good B, or nothing with certainty. In this environment, the standard revelation principle does not immediately apply, because there may exist equilibria that are more favorable for the designer in which agents randomize over menu options. I now show that this is not the case: for any feasible randomized selection among agents' favorite menu options, there exists a deterministic selection that allocates the same aggregate quantities of both goods, collects the same toll revenue, and leaves every agent's utility unchanged. Then, under such a selection rule, the usual revelation argument applies.

First, note that we can restrict attention to menus where $(x, c) \in [0,1]^2$ because any option charging a toll $c > 1$ would be dominated by non-participation. Let us then fix two menus $M_A, M_B \subset [0,1] \times [0,1]$ and denote the outside option by $\varnothing$. I use Ω to denote the set of options available to agents:

$$\Omega := \{\varnothing\} \cup \big(\{A\} \times M_A\big) \cup \big(\{B\} \times M_B\big).$$

I write the utility of type (a, b) from option $\omega \in \Omega$ as $u_{a,b}(\omega)$ and use $\kappa(\omega)$ to denote the toll paid under that option:

$$u_{a,b}(\omega) := \begin{cases} ax - c & \text{if } \omega = (A, x, c), \\ bx - c & \text{if } \omega = (B, x, c), \\ 0 & \text{if } \omega = \varnothing. \end{cases}, \qquad \kappa(\omega) := \begin{cases} c & \text{if } \omega = (A, x, c) \text{ or } \omega = (B, x, c), \\ 0 & \text{if } \omega = \varnothing. \end{cases}$$

Given the menus M_A and M_B, a *selection rule* is a measurable stochastic kernel $\sigma : [0,1]^2 \to \Delta(\Omega)$, where $\sigma_{a,b}(E)$ is the probability that type (a, b) selects an option in the Borel set $E \subseteq \Omega$. An extended mechanism consists of the menus together with such a selection rule.

The mechanism σ satisfies the supply constraints if

$$\int \sigma_{a,b}\big(\{A\} \times M_A\big)\, dF(a,b) \;\le\; s_A, \qquad \int \sigma_{a,b}\big(\{B\} \times M_B\big)\, dF(a,b) \;\le\; s_B.$$

For each type (a, b), let

$$G(a,b) := \arg\max_{\omega \in \Omega} u_{a,b}(\omega).$$

A mechanism σ is *agent-optimal* if $\sigma_{a,b}(G(a,b)) = 1$ for every (a, b). Note G is nonempty, compact-valued, and measurable because Ω is compact and $u_{a,b}(\omega)$ is continuous in ω.

The following result shows that it is without loss to use a deterministic selection rule.

Proposition 4. *Fix closed and bounded menus M_A, M_B, and let σ be an agent-optimal selection rule satisfying the supply constraints. Then there exists a deterministic selection rule $\tilde{\sigma}$, of the form $\tilde{\sigma}_{a,b} = \delta_{\tilde{m}(a,b)}$ for some measurable map $\tilde{m} : [0,1]^2 \to \Omega$, such that*

$$\int \tilde{\sigma}_{a,b}\big(\{A\} \times M_A\big)\, dF(a,b) = \int \sigma_{a,b}\big(\{A\} \times M_A\big)\, dF(a,b),$$

$$\int \tilde{\sigma}_{a,b}\big(\{B\} \times M_B\big)\, dF(a,b) = \int \sigma_{a,b}\big(\{B\} \times M_B\big)\, dF(a,b),$$

$$\int \kappa(\tilde{m}(a,b))\, dF(a,b) = \int \int_\Omega \kappa(\omega)\, d\sigma_{a,b}(\omega)\, dF(a,b),$$

and

$$u_{a,b}\big(\tilde{m}(a,b)\big) = \int_\Omega u_{a,b}(\omega)\, d\sigma_{a,b}(\omega) \quad \text{for every } (a,b).$$

Consequently, the deterministic selection rule has the same aggregate use of each good, the same aggregate toll, and the same value of the designer's objective as the randomized selection rule.

Proof. Define the finite-dimensional payoff vector

$$h(\omega) := \Big(\mathbb{1}_{\omega \in \{A\} \times M_A}, \mathbb{1}_{\omega \in \{B\} \times M_B}, \kappa(\omega)\Big).$$

Since the menus are bounded, h is bounded and measurable. Moreover, F is atomless and $\sigma_{a,b}(G(a,b)) = 1$ for every (a, b), so a standard purification theorem applies (see, for example, Feinberg and Piunovskiy (2006)). Hence, there exists a measurable selection $\tilde{m}(a,b) \in G(a,b)$, up to an F-null set, such that

$$\int h(\tilde{m}(a,b))\, dF(a,b) = \int \int_\Omega h(\omega)\, d\sigma_{a,b}(\omega)\, dF(a,b). \tag{20}$$

Changing $\tilde{m}$ on the null set if necessary, using any measurable selection from G, we may take

$\tilde{m}(a,b) \in G(a,b)$ for every (a,b). Define $\tilde{\sigma}_{a,b} := \delta_{\tilde{m}(a,b)}$. The first two coordinates of (20) give

$$\int \tilde{\sigma}_{a,b}\big(\{A\} \times M_A\big)\, dF(a,b) = \int \sigma_{a,b}\big(\{A\} \times M_A\big)\, dF(a,b),$$

$$\int \tilde{\sigma}_{a,b}\big(\{B\} \times M_B\big)\, dF(a,b) = \int \sigma_{a,b}\big(\{B\} \times M_B\big)\, dF(a,b).$$

Hence the deterministic selection satisfies the same supply constraints as σ. The third coordinate gives

$$\int \kappa(\tilde{m}(a,b))\, dF(a,b) = \int \int_\Omega \kappa(\omega)\, d\sigma_{a,b}(\omega)\, dF(a,b).$$

It remains only to note that every type's utility is unchanged. Recall $\sigma_{a,b}$ is supported on $G(a,b)$ and $\tilde{m}(a,b) \in G(a,b)$, so both selections give type (a,b) her maximal utility:

$$u_{a,b}\big(\tilde{m}(a,b)\big) = \int_\Omega u_{a,b}(\omega)\, d\sigma_{a,b}(\omega) = \max_{\omega \in \Omega} u_{a,b}(\omega).$$

Since the deterministic selection preserves each type's utility and the aggregate toll, it also preserves the value of the objective. □

B Waitlists

I now show that the baseline model can be interpreted as a reduced-form description of a steady-state waitlist environment with free re-entry.

Waitlist model. Consider a stationary waitlist environment in continuous time. There are two goods, A and B. Agents have values $(a,b) \in [0,1]^2$, where a is the value of receiving good A and b is the value of receiving good B. At each instant $\tau \in \mathbb{R}$, flow masses $s_A, s_B > 0$ of goods A and B arrive, with $s_A + s_B \leq 1$. At the same time, a unit flow mass of agents arrives, with types distributed according to F. Agents discount utility at rate $\rho > 0$. Receiving good A gives utility a, receiving good B gives utility b, and receiving no good gives utility zero. tolls enter utility additively. An agent who fails to receive a good remains eligible to participate again. Thus unsuccessful agents may freely re-enter the waitlist system.

Mechanisms. There is a separate waitlist for each good. For each waitlist, the designer chooses a menu of triples (c,t,p), where $c \in \mathbb{R}_+$ is a toll, $t \in \mathbb{R}_+$ is a wait time, and $p \in [0,1]$ is the probability of receiving the good at the end of the wait. An agent who chooses (c,t,p) incurs the toll immediately, waits t units of time, and then receives the good with probability p. If she does not receive the good, she may re-enter the mechanism.

I restrict attention to stationary mechanisms that admit a steady state. Thus, the same menus are offered at every date, and all agents of the same type choose the same good and the same menu option, independently of their arrival date and of how many times they have already participated. Steady-state feasibility then requires

$$\int \mathbb{1}_{(a,b)\text{ chooses }A}\, dF(a,b) \le s_A, \qquad \int \mathbb{1}_{(a,b)\text{ chooses }B}\, dF(a,b) \le s_B. \tag{21}$$

The constraint takes this form because agents can freely re-enter after an unsuccessful attempt. Hence, any agent who chooses a waitlist option with $p > 0$ eventually receives the corresponding good almost surely. Per-attempt success probabilities affect the timing of receipt, but not the eventual mass of goods consumed.

Reduction to the baseline model. Consider a type-(a,b) agent who chooses option (c,t,p) on the waitlist for good A, with $p > 0$. Let $V_A(a;c,t,p)$ be her expected discounted utility at the moment she chooses the option. After an unsuccessful attempt, she faces the same continuation problem, shifted forward by t. Hence

$$V_A(a;c,t,p) = -c + pe^{-\rho t}a + (1-p)e^{-\rho t}V_A(a;c,t,p).$$

Solving gives

$$V_A(a;c,t,p) = \frac{-c + pe^{-\rho t}a}{1-(1-p)e^{-\rho t}} = xa - \tilde{c}, \tag{22}$$

where

$$x := \frac{pe^{-\rho t}}{1-(1-p)e^{-\rho t}} \in [0,1], \qquad \tilde{c} := \frac{c}{1-(1-p)e^{-\rho t}} \in \mathbb{R}_+.$$

The case of good B is analogous. Thus, each waitlist option has a reduced-form representation $(x,\tilde{c})$. Conversely, any $(x,\tilde{c}) \in (0,1] \times \mathbb{R}_+$ can be implemented by a waitlist option: take $p = 1$, $t = -(1/\rho)\log x$, and $c = \tilde{c}$. The case $x = 0$ corresponds to the outside option.

Using this reduced form and the revelation principle established in Section A, the designer's problem reduces to choosing an allocation rule

$$(\tilde{c}, x, y) : [0,1]^2 \to \mathbb{R}_+ \times [0,1] \times \{A, B, \varnothing\},$$

where $y(a,b)$ is the waitlist chosen by type (a,b), $x(a,b)$ is the discounted allocation intensity, and $\tilde{c}(a,b)$ is the reduced-form toll cost. The utility of type (a,b) from reporting (a',b') is

$$u_{a,b}(a',b') := \begin{cases} a\,x(a',b') - \tilde{c}(a',b') & \text{if } y(a',b') = A, \\ b\,x(a',b') - \tilde{c}(a',b') & \text{if } y(a',b') = B, \\ 0 & \text{if } y(a',b') = \varnothing. \end{cases}$$

The direct mechanism must satisfy incentive compatibility, individual rationality, and the steady-state resource constraints

$$\int \mathbb{1}_{y(a,b)=A}\, dF(a,b) \le s_A, \qquad \int \mathbb{1}_{y(a,b)=B}\, dF(a,b) \le s_B.$$

This matches the form of the baseline problem.

C Omitted proofs

First, recall that U_A and U_B are convex and strictly increasing on $[\underline{a}, \overline{a}]$ and $[\underline{b}, \overline{b}]$, respectively, with $U_A(\underline{a}) = U_B(\underline{b}) = 0$. Hence each is differentiable except at countably many points, and is pinned down by its derivative wherever the derivative exists. At nondifferentiability points, the one-sided derivatives are obtained as limits of derivatives at nearby differentiability points. For example, if D_A is the set of points at which U_A' exists, then

$$U_A'^{-}(t') = \lim_{t \uparrow t',\, t \in D_A} U_A'(t), \qquad U_A'^{+}(t') = \lim_{t \downarrow t',\, t \in D_A} U_A'(t).$$

Analogous statements hold for U_B.

C.1 Proof of Proposition 1

By single crossing, any feasible mechanism is characterized by a cutoff $\underline{a}$ such that all types above the cutoff receive good A:

$$\underline{a} := \inf\{a \in [0,1] : y(a,b) = A\}, \qquad \inf \varnothing := 1.$$

Fix such a cutoff. I show that any mechanism implementing it using damages can be replaced with a toll mechanism without lowering any agent's utility or the designer's objective.

First, incentive compatibility implies that all types below the cutoff receive the same indirect utility $\underline{U}$. By the envelope theorem, the indirect utility of type a can then be written as

$$U(a,b) = \underline{U} + \int_{\underline{a}}^{\max\{\underline{a}, a\}} x(v,b)\, dv.$$

Since $x(v,b) \le 1$, we have

$$U(a,b) \le \underline{U} + \max\{\underline{a}, a\} - \underline{a}. \tag{23}$$

This upper bound is implementable without damages by offering good B with a toll of $b - \underline{U}$ and good A with a toll of $\underline{a} - \underline{U}$. These tolls are nonnegative: $\underline{U} \le b$, because the outside option B has value b, and $\underline{U} \le \underline{a}$, because no A-option can give utility above a to a type with value a. Note this menu implements the same cutoff and attains the utility bound in (23).

It remains to check the designer's objective. We can write type (a,b)'s contribution to (O) as:

$$u_{a,b} + \gamma c(a,b) = u_{a,b} + \gamma(xv - u_{a,b}) = (1-\gamma)u_{a,b} + \gamma xv,$$

where v denotes her value for the good she receives. Note the above replacement weakly raises $u_{a,b}$ and sets $x = 1$. Since $\gamma \in [0,1]$, this contribution weakly increases for all types.

C.2 Proof of Proposition 2

Agents choose among good A, good B, and non-participation by comparing $U_A(a)$, $U_B(b)$, and 0. By the definitions of $\underline{a}$ and $\underline{b}$, types with $a < \underline{a}$ get no positive utility from good A, and types with $b < \underline{b}$ get no positive utility from good B. Hence every type $(a,b) < (\underline{a},\underline{b})$ chooses non-participation. If $a > \underline{a}$ and $b < \underline{b}$, then good A gives positive utility while good B does not, so the type chooses good A. Symmetrically, if $a < \underline{a}$ and $b > \underline{b}$, the type chooses good B. Since positive masses of both goods are allocated, $\underline{a}, \underline{b} < 1$. It remains to describe choices on the upper rectangle. Suppose, without loss, that $U_B(1) \geq U_A(1)$. Since U_A and U_B are continuous and strictly increasing on their active intervals, for each $a \in [\underline{a}, 1]$ there is a unique $z(a) \in [\underline{b}, 1]$ satisfying $U_B(z(a)) = U_A(a)$. Equivalently, $z = U_B^{-1} \circ U_A$ on this interval, so z is continuous and strictly increasing, with $z(\underline{a}) = \underline{b}$ and $z(1) \leq 1$. For any type $(a,b) > (\underline{a},\underline{b})$, if $b < z(a)$, then $U_B(b) < U_B(z(a)) = U_A(a)$, so the type chooses good A. If $b > z(a)$, the reverse inequality holds, so the type chooses good B. Types with $b = z(a)$ are indifferent. This gives the desired boundary representation.

C.3 Proof of Theorem 2

Step 1: The perturbing menu. Consider the market-clearing toll menu consisting of the A-option $(1, c_A^*)$ and the B-option $(1, c_B^*)$. Fix $\tilde{b} \in (c_B^*, 1)$. For small $\varepsilon > 0$, take the alternative menu

$$\left\{(A, 1, c_A^* + \varepsilon\alpha),\ (B, 1, c_B^* + \varepsilon\beta),\ (B, 1-\varepsilon, c_B^* + \varepsilon\beta - \varepsilon\tilde{b})\right\}.$$

Let U_A and U_B denote the A- and B-indirect utilities induced by this perturbed menu. Then

$$U_A(a) = (a - c_A^* - \varepsilon\alpha)_+, \quad \text{and} \quad U_B(b) = \max\left\{0,\ b - c_B^* - \varepsilon\beta,\ (1-\varepsilon)b - c_B^* - \varepsilon\beta + \varepsilon\tilde{b}\right\}.$$

For any fixed $b > c_B^*$, this type continues to participate for all sufficiently small ε. Hence, away from the lower participation margin, we only need to compare the two B-options:

$$\max\left\{b - c_B^* - \varepsilon\beta,\ (1-\varepsilon)b - c_B^* - \varepsilon\beta + \varepsilon\tilde{b}\right\} = b - c_B^* + \varepsilon\left[(\tilde{b} - b)_+ - \beta\right].$$

Thus the damaged option raises the B-side payoff by $\varepsilon(\tilde{b} - b)$ for types with $b < \tilde{b}$, and has no direct effect on the B-side payoff of types with $b \geq \tilde{b}$, apart from the common toll increase $\varepsilon\beta$.

Step 2: Mass change from the interior A-B boundary. We now compute the first-order change in the masses assigned to the two goods. Define

$$D_\varepsilon(a,b) := U_A(a) - U_B(b),$$

and fix a compact interval $K \subset (c_A^*, \bar{a}^*)$. For $a \in K$ and b in a small neighborhood of $z_0(a)$, all agents strictly prefer either good to nothing. Hence

$$D_\varepsilon(a,b) = z_0(a) - b + \varepsilon\left[\beta - \alpha - \left(\tilde{b} - b\right)_+\right].$$

Let $z_\varepsilon(a)$ denote the perturbed A-B boundary over K; then $D_\varepsilon(a, z_\varepsilon(a)) = 0$. Since $\beta - \alpha - (\tilde{b} - b)_+$ is bounded uniformly on K, this equation implies $z_\varepsilon(a) - z_0(a) = O(\varepsilon)$, uniformly in $a \in K$. Since

$b \mapsto (\tilde{b} - b)_+$ is Lipschitz, $\left(\tilde{b} - z_\varepsilon(a)\right)_+ = \left(\tilde{b} - z_0(a)\right)_+ + O(\varepsilon)$, uniformly in $a \in K$. Substituting this into $D_\varepsilon(a, z_\varepsilon(a)) = 0$ gives

$$z_\varepsilon(a) = z_0(a) + \varepsilon\left[\beta - \alpha - \left(\tilde{b} - z_0(a)\right)_+\right] + o(\varepsilon), \tag{24}$$

uniformly in $a \in K$.

Now, fix $\eta > 0$ and set $K_\eta := [c_A^* + \eta, \bar{a}^* - \eta]$. Define

$$\Delta M_A^{AB}(\varepsilon; K_\eta) := \int_{K_\eta} \int_{z_0(a)}^{z_\varepsilon(a)} f(a,b)\, db\, da.$$

This is the signed change in the mass assigned to good A coming from the movement of the A-B boundary over K_η. Using (24) and the continuity of f, for each fixed $\eta > 0$,

$$\Delta M_A^{AB}(\varepsilon; K_\eta) = \varepsilon \int_{K_\eta} \left[\beta - \alpha - \left(\tilde{b} - z_0(a)\right)_+\right] f(a, z_0(a))\, da + o_\eta(\varepsilon).$$

We now pass from K_η to the full old interior boundary. Let $E_\eta := [c_A^*, \bar{a}^*] \setminus K_\eta$. This set has length at most 2η. In addition, all relevant cutoffs and the A-B boundary move by at most order ε, uniformly for small ε. Hence the region swept out over E_η has area at most $C_0\varepsilon\eta$, for some constant C_0 independent of η. Since f is bounded, there is a constant C_1, also independent of η, such that

$$\left|\Delta M_A^{AB}(\varepsilon; E_\eta)\right| \le C_1 \varepsilon \eta.$$

Therefore, for every fixed $\eta > 0$,

$$\limsup_{\varepsilon \downarrow 0} \left| \frac{\Delta M_A^{AB}(\varepsilon; [c_A^*, \bar{a}^*])}{\varepsilon} - \int_{K_\eta} \left[\beta - \alpha - \left(\tilde{b} - z_0(a)\right)_+\right] f(a, z_0(a))\, da \right| \le C_1 \eta.$$

Now, let $\eta \downarrow 0$. Since the integrand is bounded and continuous on $[c_A^*, \bar{a}^*]$, the integral over K_η converges to the integral over $[c_A^*, \bar{a}^*]$. Hence

$$\Delta M_A^{AB}(\varepsilon; [c_A^*, \bar{a}^*]) = \varepsilon \int_{c_A^*}^{\bar{a}^*} \left[\beta - \alpha - \left(\tilde{b} - z_0(a)\right)_+\right] f(a, z_0(a))\, da + o(\varepsilon).$$

Since good B lies on the other side of the same boundary, the corresponding contribution to the mass assigned to good B is given by $\Delta M_B^{AB} = -\Delta M_A^{AB}$.

Step 3: The participation margins. It remains to compute the participation-margin contributions. Along the lower A-participation margin, the cutoff changes from c_A^* to $c_A^* + \varepsilon\alpha$. Therefore the signed change in the mass assigned to good A along this margin is

$$-\int_0^{c_B^*} \int_{c_A^*}^{c_A^* + \varepsilon\alpha} f(a,b)\, da\, db = -\varepsilon\alpha \int_0^{c_B^*} f(c_A^*, b)\, db + o(\varepsilon).$$

By the definition of $P_A^{c_A^*}$, this equals $-\varepsilon\alpha P_A^{c_A^*} + o(\varepsilon)$. The possible overlap with the lower B-participation margin is confined to an $O(\varepsilon)\times O(\varepsilon)$ neighborhood of (c_A^*, c_B^*), and hence has mass $o(\varepsilon)$.

Along the lower B-participation margin, the undamaged B-option has cutoff $c_B^* + \varepsilon\beta$, while the damaged B-option has cutoff solving $(1-\varepsilon)b - c_B^* - \varepsilon\beta + \varepsilon\tilde{b} = 0$. Thus, the lower cutoff for participation in the B-menu is

$$b_\varepsilon = c_B^* + \varepsilon(\beta + c_B^* - \tilde{b}) + o(\varepsilon),$$

because $\tilde{b} > c_B^*$ makes the damaged option the lower-threshold B-option. Hence the signed change in the mass assigned to good B along this margin is

$$-\int_0^{c_A^*}\int_{c_B^*}^{b_\varepsilon} f(a,b)\,db\,da = -\varepsilon(\beta + c_B^* - \tilde{b})\int_0^{c_A^*} f(a, c_B^*)\,da + o(\varepsilon).$$

By the definition of $P_B^{c_B^*}$, this equals $\varepsilon P_B^{c_B^*}(\tilde{b} - c_B^* - \beta) + o(\varepsilon)$. Again, the possible overlap with the lower A-participation margin is confined to an $O(\varepsilon)\times O(\varepsilon)$ neighborhood of (c_A^*, c_B^*), and hence has mass $o(\varepsilon)$.

Step 4: Preserving supplies to first order. We now choose α, β so that this first-order perturbation leaves the total allocated amounts of both goods unchanged. Combining the lower participation-margin contributions with the interior A-B boundary contribution gives

$$\Delta M_A = \varepsilon\left\{-\alpha P_A^{c_A^*} + \int_{c_A^*}^{\bar{a}^*}\left[\beta - \alpha - \left(\tilde{b} - z_0(a)\right)_+\right] f(a, z_0(a))\,da\right\} + o(\varepsilon).$$

$$\Delta M_B = \varepsilon\left\{P_B^{c_B^*}(\tilde{b} - c_B^* - \beta) - \int_{c_A^*}^{\bar{a}^*}\left[\beta - \alpha - \left(\tilde{b} - z_0(a)\right)_+\right] f(a, z_0(a))\,da\right\} + o(\varepsilon).$$

Setting these first-order changes equal to zero gives

$$-\alpha P_A^{c_A^*} + \int_{c_A^*}^{\bar{a}^*}\left[\beta - \alpha - \left(\tilde{b} - z_0(a)\right)_+\right] f(a, z_0(a))\,da = 0. \tag{25}$$

$$P_B^{c_B^*}(\tilde{b} - c_B^* - \beta) + \int_{c_A^*}^{\bar{a}^*}\left[\left(\tilde{b} - z_0(a)\right)_+ + \alpha - \beta\right] f(a, z_0(a))\,da = 0. \tag{26}$$

Using the definitions of P_{AB} and Q, equations (25) and (26) can be written as

$$(P_A^{c_A^*} + P_{AB})\alpha - P_{AB}\beta = -Q, \quad \text{and} \quad -P_{AB}\alpha + (P_B^{c_B^*} + P_{AB})\beta = P_B^{c_B^*}(\tilde{b} - c_B^*) + Q.$$

The coefficient matrix is

$$\begin{pmatrix} P_A^{c_A^*} + P_{AB} & -P_{AB} \\ -P_{AB} & P_B^{c_B^*} + P_{AB} \end{pmatrix}; \tag{27}$$

it has the determinant

$$(P_A^{c_A^*} + P_{AB})(P_B^{c_B^*} + P_{AB}) - P_{AB}^2 = P_A^{c_A^*} P_B^{c_B^*} + P_A^{c_A^*} P_{AB} + P_B^{c_B^*} P_{AB} > 0, \tag{28}$$

which is positive because $P_A^{c_A^*} > 0$ and $P_{AB} > 0$. The latter follows from $c_A^* < 1$, $c_B^* < 1$, and the fact that the old interior boundary has positive length. The supply-preserving first-order corrections are thus unique and satisfy (5).

Moreover, these corrections are nonnegative. Indeed, for $a \geq c_A^*$, we have $\big(\tilde{b} - z_0(a)\big)_+ \leq \tilde{b} - c_B^*$, and hence $Q \leq P_{AB}(\tilde{b} - c_B^*)$. Therefore $\alpha \geq 0$. The formula for β immediately gives $\beta \geq 0$. Thus the old A- and B-option tolls remain nonnegative. Since $c_B^* > 0$, the damaged B-option also has a nonnegative toll for all sufficiently small $\varepsilon > 0$.

Step 5: The first-order objective gain. We now compute the first-order effect on the objective. The movements of the choice boundaries do not create additional first-order utility terms, because types on the old A-B boundary are indifferent between the two goods, and types on the old participation margins receive zero utility. Hence the first-order utility change is obtained by integrating the direct utility changes over the old assignment regions.

Old A-choosers receive the first-order utility change $-\alpha$. Since the undamaged menu clears the supply of good A, their utility contribution is $-\alpha s_A$. Old B-choosers with value b receive the first-order utility change $(\tilde{b} - b)_+ - \beta$. Since P_B^b is the density of old B-choosers with value b, their utility contribution is

$$\int_{c_B^*}^{1} \left[(\tilde{b} - b)_+ - \beta\right] P_B^b \, db.$$

Therefore the first-order change in the utility component is

$$\varepsilon \left\{ \int_{c_B^*}^{\tilde{b}} (\tilde{b} - b) P_B^b \, db - \alpha s_A - \beta s_B \right\} + o(\varepsilon).$$

Now consider the toll component. Since the perturbation preserves the masses assigned to both goods to first order, the zeroth-order toll terms $c_A^* s_A + c_B^* s_B$ do not change to first order. The first-order change in total tolls is

$$\varepsilon \left\{ \alpha s_A + \beta s_B - \tilde{b} \int_{c_B^*}^{\tilde{b}} P_B^b \, db \right\} + o(\varepsilon).$$

Indeed, old A-choosers face the toll increase $\varepsilon\alpha$, old B-choosers face the common toll increase $\varepsilon\beta$, and old B-choosers with $b < \tilde{b}$ select the damaged option, whose toll is lower by $\varepsilon\tilde{b}$.

Combining the utility and toll components gives

$$\varepsilon \left\{ \int_{c_B^*}^{\tilde{b}} \big((1-\gamma)\tilde{b} - b\big) P_B^b \, db - (1-\gamma)(\alpha s_A + \beta s_B) \right\} + o(\varepsilon).$$

Substituting from (5), condition (4) says that the coefficient of ε is strictly positive. Hence the objective strictly increases for all sufficiently small $\varepsilon > 0$.

Step 6: Exact feasibility. The previous steps choose α and β so that the supply constraints are preserved to first order. I now perturb these two numbers slightly so that the constraints hold exactly. For $i \in \{A, B\}$, let $M_i(\varepsilon, \alpha, \beta)$ denote the mass assigned to good i by the perturbed menu, and let M_i^0 denote the mass assigned to good i by the original menu. For $\varepsilon > 0$, define the normalized supply errors

$$H_i(\varepsilon, \alpha, \beta) := \frac{M_i(\varepsilon, \alpha, \beta) - M_i^0}{\varepsilon}.$$

The expansions above are uniform for (α, β) in a small neighborhood of (α, β). Therefore H_A and H_B extend continuously to $\varepsilon = 0$, with

$$H_A(0, \alpha, \beta) = -\alpha P_A^{c_A^*} + \int_{c_A^*}^{\bar{a}^*} \left[\beta - \alpha - \left(\tilde{b} - z_0(a)\right)_+\right] f(a, z_0(a))\, da,$$

$$H_B(0, \alpha, \beta) = P_B^{c_B^*}(\tilde{b} - c_B^* - \beta) - \int_{c_A^*}^{\bar{a}^*} \left[\beta - \alpha - \left(\tilde{b} - z_0(a)\right)_+\right] f(a, z_0(a))\, da.$$

By construction, $H_A(0, \alpha, \beta) = H_B(0, \alpha, \beta) = 0$. Moreover, the derivative of (H_A, H_B) with respect to (α, β) at $(0, \alpha, \beta)$ is the negative of (27), and so its determinant is given by (28). Hence, by the implicit function theorem, for all sufficiently small $\varepsilon > 0$, there exist

$$\alpha_\varepsilon = \alpha + o(1), \qquad \beta_\varepsilon = \beta + o(1),$$

such that $H_A(\varepsilon, \alpha_\varepsilon, \beta_\varepsilon) = H_B(\varepsilon, \alpha_\varepsilon, \beta_\varepsilon) = 0$, and so, equivalently,

$$M_A(\varepsilon, \alpha_\varepsilon, \beta_\varepsilon) = M_A^0, \qquad M_B(\varepsilon, \alpha_\varepsilon, \beta_\varepsilon) = M_B^0.$$

Thus the adjusted perturbed menu preserves both supplies exactly. Finally, since $\alpha_\varepsilon \to \alpha$ and $\beta_\varepsilon \to \beta$, these exact-feasibility adjustments do not change the first-order coefficient of the objective gain computed in Step 5. By condition (4), that coefficient is strictly positive. Therefore, for all sufficiently small $\varepsilon > 0$, the adjusted perturbed menu is feasible and strictly improves on the original undamaged menu.

C.4 Proof of Corollary 1

Since $s_A + s_B = 1$, the market-clearing toll mechanism must allocate some good to almost every type. If both tolls were strictly positive, all types with $a < c_A^*$ and $b < c_B^*$ would choose neither good, contradicting market clearing. Since $c_B^* > 0$, it follows that $c_A^* = 0$. Thus the old interior A-B boundary is $b = a + c_B^*$ for $a \in [0, 1 - c_B^*]$.

Set $\tilde{b} := \tilde{a} + c_B^*$. By Theorem 2 and the assumption that $\gamma = 0$, it is enough to show

$$\int_{c_B^*}^{\tilde{b}} (\tilde{b} - b) \left[\int_0^{b - c_B^*} f(t, b)\, dt \right] db > \alpha s_A + \beta s_B.$$

Because $c_A^* = 0$,

$$P_B^{c_B^*} = \int_0^{c_A^*} f(a, c_B^*)\, da = 0,$$

and hence $\alpha = 0$. Moreover,

$$P_{AB} = \int_0^{1 - c_B^*} f(a, a + c_B^*)\, da \quad \text{and} \quad Q = \int_0^{1 - c_B^*} (\tilde{b} - a - c_B^*)_+ f(a, a + c_B^*)\, da = \int_0^{\tilde{a}} (\tilde{a} - a) f(a, a + c_B^*)\, da,$$

so

$$\beta = \frac{\int_0^{\tilde{a}} (\tilde{a} - a) f(a, a + c_B^*)\, da}{\int_0^{1 - c_B^*} f(a, a + c_B^*)\, da}.$$

By changing variables, the left-hand side of the damage condition becomes

$$\int_{c_B^*}^{\tilde{b}} (\tilde{b} - b) \left[\int_0^{b - c_B^*} f(t, b)\, dt \right] db = \int_0^{\tilde{a}} (\tilde{a} - a) \left[\int_0^a f(t, a + c_B^*)\, dt \right] da.$$

Also, since the undamaged menu clears supplies and $s_A + s_B = 1$,

$$s_B = \int_{c_B^*}^1 \left[\int_0^{b - c_B^*} f(t, b)\, dt \right] db = \int_0^{1 - c_B^*} \left[\int_0^a f(t, a + c_B^*)\, dt \right] da.$$

Thus the sufficient condition from Theorem 2 becomes

$$\int_0^{\tilde{a}} (\tilde{a} - a) \left[\int_0^a f(t, a + c_B^*)\, dt \right] da > \frac{\int_0^{\tilde{a}} (\tilde{a} - a) f(a, a + c_B^*)\, da}{\int_0^{1 - c_B^*} f(a, a + c_B^*)\, da} \int_0^{1 - c_B^*} \left[\int_0^a f(t, a + c_B^*)\, dt \right] da.$$

Equivalently,

$$\frac{\int_0^{\tilde{a}} (\tilde{a} - a) \left[\int_0^a f(t, a + c_B^*)\, dt \right] da}{\int_0^{\tilde{a}} (\tilde{a} - a) f(a, a + c_B^*)\, da} > \frac{\int_0^{1 - c_B^*} \left[\int_0^a f(t, a + c_B^*)\, dt \right] da}{\int_0^{1 - c_B^*} f(a, a + c_B^*)\, da},$$

which is equivalent to (6).

C.5 Proof of Lemma 1

By Proposition 2, welfare can be written as

$$\int_{\underline{a}}^{\overline{a}} \int_0^{z(a)} U_A(a) f(a, b)\, db\, da + \int_{\underline{b}}^1 \int_0^{\hat{z}^{-1}(b)} U_B(b) f(a, b)\, da\, db + \int_{\overline{a}}^1 \int_0^1 U_A(a) f(a, b)\, db\, da.$$

Using the boundary relation $U_B(z(a)) = U_A(a)$, the middle term can be rewritten by the change of variables $b = z(a)$. Hence the utility part equals

$$\int_{\underline{a}}^{\overline{a}} U_A(a)\, dF(a, z(a)) + \int_{\overline{a}}^{1} U_A(a)\, dF(a, 1).$$

Equivalently, using the extended boundary $\hat{z}$, this is

$$\int_0^1 U_A(a)\, dF(a, \hat{z}(a)).$$

Integrating by parts gives

$$U_A(1)F(1, \hat{z}(1)) - U_A(0)F(0, \hat{z}(0)) - \int_0^1 U_A'(a)F(a, \hat{z}(a))\, da.$$

Since $F(1, \hat{z}(1)) = F(1, 1) = 1$ and $F(0, \hat{z}(0)) = 0$, this becomes

$$U_A(1) - \int_0^1 U_A'(a)F(a, \hat{z}(a))\, da.$$

C.6 Proof of Lemma 2

We first prove necessity. Fix a feasible mechanism implementing (z, U_A), and let U_B be the induced B-indirect utility. Since U_A and U_B are convex, their derivatives are non-decreasing wherever they exist. By Proposition 2, $U_A(a) = U_B(z(a))$ for all $a \in [\underline{a}, \overline{a}]$. Differentiating at points where the relevant derivatives exist gives

$$U_A'(a) = U_B'(z(a))z'(a), \qquad \text{so} \qquad \frac{U_A'(a)}{z'(a)} = U_B'(z(a)).$$

Since z is increasing and U_B' is non-decreasing, U_A'/z' is non-decreasing. This proves (i).

Next, Proposition 2 also gives $z = U_B^{-1} \circ U_A$ on the active interval. Hence the one-sided derivatives of z are determined by the corresponding one-sided derivatives of U_A and U_B, i.e.,

$$z'^{+}(a) = \frac{U_A'^{+}(a)}{(U_B'^{+} \circ z)(a)},$$

and the analogous formula holds for left derivatives. Since the relevant one-sided derivatives of U_A and U_B are finite and strictly positive on the active intervals, z has finite, strictly positive one-sided derivatives at every $a \in (\underline{a}, \overline{a})$, and a finite, strictly positive left derivative at $\overline{a}$. This proves (ii).

The supply constraints follow from Proposition 2. Up to null sets, the agents receiving good A are those below the extended boundary, and the agents receiving good B are those above it.

Therefore

$$\int \mathbb{1}_{y(a,b)=A}\, dF(a,b) = \int_{\underline{a}}^{1} \int_{0}^{\hat{z}(a)} f(a,v)\, dv\, da, \quad \int \mathbb{1}_{y(a,b)=B}\, dF(a,b) = \int_{\underline{b}}^{1} \int_{0}^{\hat{z}^{-1}(b)} f(v,b)\, dv\, db.$$

Feasibility of the original mechanism therefore implies (iii).

It remains to prove the regularity condition. Under Assumption 2, the quality rule is piecewise continuously differentiable. By Proposition 2, types $(a,0)$ with $a > \underline{a}$ choose good A, and types $(0,b)$ with $b > \underline{b}$ choose good B. The envelope theorem then implies that, wherever the relevant derivatives exist, $U_A'(a) = x(a,0)$ and $U_B'(b) = x(0,b)$. Thus U_A' and U_B' are piecewise continuously differentiable. Since $z = U_B^{-1} \circ U_A$, the boundary z is piecewise continuously differentiable. On each interval on which the relevant functions are differentiable,

$$z'(a) = \frac{U_A'(a)}{U_B'(z(a))}.$$

The right-hand side is piecewise continuously differentiable, so z is piecewise twice continuously differentiable. Finally,

$$\frac{U_A'(a)}{z'(a)} = U_B'(z(a)),$$

so U_A'/z' is piecewise continuously differentiable. This proves (iv).

We now prove sufficiency. Fix (z, U_A) satisfying (i)-(iv). Define U_B on the active interval by

$$U_B(b) := U_A(z^{-1}(b)) \quad \text{for } b \in [\underline{b}, \overline{b}].$$

Set $U_B(b) = 0$ for $b \le \underline{b}$. For $b \ge \overline{b}$, extend U_B linearly with slope

$$q := \lim_{a \uparrow \overline{a}} \frac{U_A'(a)}{z'(a)}, \quad \text{that is,} \quad U_B(b) = U_B(\overline{b}) + q(b - \overline{b}) \quad \text{for } b \ge \overline{b}.$$

By (i), this extension is convex and non-decreasing. Moreover, on the active interval,

$$U_B'(z(a)) = \frac{U_A'(a)}{z'(a)}$$

wherever the derivative exists, and hence $U_B(z(a)) = U_A(a)$ for all $a \in [\underline{a}, \overline{a}]$. We now construct a direct mechanism. If neither good gives positive indirect utility, assign non-participation and set $x(a,b) = c(a,b) = 0$. Otherwise assign the good that gives the larger indirect utility:

$$y(a,b) = \begin{cases} A, & \text{if } U_A(a) \ge U_B(b) \text{ and } U_A(a) > 0, \\ B, & \text{if } U_B(b) > U_A(a), \\ \varnothing, & \text{otherwise.} \end{cases}$$

For types assigned good A and B, respectively, set

$$x(a,b) = U_A'^{-}(a), \qquad c(a,b) = aU_A'^{-}(a) - U_A(a).$$

$$x(a,b) = U_B'^{-}(b), \qquad c(a,b) = bU_B'^{-}(b) - U_B(b).$$

For types assigned $\varnothing$, keep $x(a,b) = c(a,b) = 0$. Since U_A is an A-indirect utility and U_B was constructed with the corresponding slope U_A'/z', these qualities lie in $[0,1]$. Convexity and $U_A(\underline{a}) = U_B(\underline{b}) = 0$ also imply that the tolls are nonnegative.

It remains to verify incentive compatibility. Consider the A-options. Since U_A is convex, $U_A'^{-}(a)$ is a subgradient of U_A at a. Thus for every a',

$$U_A(a') \geq U_A(a) + U_A'^{-}(a)(a' - a).$$

Equivalently, type a' gets weakly higher utility from the A-option designed for a' than from the A-option designed for a. Therefore no type wants to misreport among A-options. The same argument applies to B-options.

The utility from the assigned A-option is $aU_A'^{-}(a) - \big(aU_A'^{-}(a) - U_A(a)\big) = U_A(a)$, and the utility from the assigned B-option is $U_B(b)$. Since the mechanism assigns each type the good that gives the larger of $U_A(a)$ and $U_B(b)$, no type wants to switch goods. Individual rationality follows because both indirect utilities are nonnegative and the outside option is available.

By construction $U_A(a) = U_B(z(a))$ for all $a \in [\underline{a}, \overline{a}]$, so the mechanism implements the boundary z. Finally, the supply constraints hold by *(iii)*, because Proposition 2 identifies the masses assigned to goods A and B with the two integrals in (S′). Hence the constructed mechanism is feasible and implements (z, U_A).

C.7 Proof of Lemma 3

Fix the boundary z. By Lemma 2, feasibility of $(z, \check{U}_A)$ is equivalent to $\check{U}_A'$ and $\check{U}_A'/z'$ being non-decreasing, together with the bounds $0 \leq \check{U}_A' \leq 1$, $0 \leq \check{U}_A'/z' \leq 1$, and the regularity requirements stated there. Since the boundary is fixed, the supply constraints do not depend on $\check{U}_A$.

Let $p(a) := m(a)k$ for $a \in (\underline{a}, \overline{a})$. We first show that the U_A defined in (14) is feasible. The function p is non-decreasing by construction. Moreover, on every interval on which z is twice continuously differentiable,

$$\big(\log(p/z')\big)' = \max\left\{0, \frac{z''}{z'}\right\} - \frac{z''}{z'} \geq 0.$$

At an upward jump of z', the product in the definition of m makes p/z' continuous. At a downward jump of z', the function p is continuous while z' falls, so p/z' jumps upward. Hence both p and p/z' are non-decreasing. Finally,

$$p^{-}(\overline{a}) = m(\overline{a})k = \frac{1}{\max\{1, 1/z'^{-}(\overline{a})\}} = \min\{1, z'^{-}(\overline{a})\}.$$

Since p and p/z' are non-decreasing, this implies $p \leq 1$ and $p/z' \leq 1$ on $(\underline{a}, \overline{a})$. Thus U_A satisfies the feasibility conditions of Lemma 2.

Now take any other feasible pair $(z, \check{U}_A)$, and write $\check{p} = \check{U}'_A$. We show that $\check{p} \leq p$ almost everywhere. On every smooth interval on which $\check{p} > 0$,

$$(\log \check{p})' \geq 0 \qquad \text{and} \qquad (\log(\check{p}/z'))' \geq 0; \quad \text{thus} \quad (\log \check{p})' \geq \max\left\{0, \frac{z''}{z'}\right\} = (\log p)'.$$

Therefore $\log(p/\check{p})$ is non-increasing on each such interval.

It remains only to check jumps. If z' jumps upward at t, then

$$\frac{p^+(t)}{p^-(t)} = \frac{z'^+(t)}{z'^-(t)}.$$

Since $\check{p}/z'$ is non-decreasing,

$$\frac{\check{p}^+(t)}{\check{p}^-(t)} \geq \frac{z'^+(t)}{z'^-(t)}.$$

Hence $(p/\check{p})^+(t) \leq (p/\check{p})^-(t)$. If z' jumps downward, then p is continuous and $\check{p}$ is non-decreasing, so again $(p/\check{p})^+(t) \leq (p/\check{p})^-(t)$. The same conclusion is immediate at any jump of $\check{p}$ at which z' is continuous. Hence $p/\check{p}$ is non-increasing on $(\underline{a}, \overline{a})$.

By feasibility, $\check{p}^-(\overline{a}) \leq 1$ and $\frac{\check{p}^-(\overline{a})}{z'^-(\overline{a})} \leq 1$, so

$$\check{p}^-(\overline{a}) \leq \min\{1, z'^-(\overline{a})\} = p^-(\overline{a}).$$

Since $p/\check{p}$ is non-increasing, it follows that $\check{p}(a) \leq p(a)$ for a.e. $a \in (\underline{a}, \overline{a})$. Also, both marginal utilities are zero on $(0, \underline{a})$, while on $(\overline{a}, 1)$ feasibility gives $\check{U}'_A \leq 1 = U'_A$. Thus $\check{U}'_A(a) \leq U'_A(a)$ for a.e. $a \in (0,1)$.

Finally, since $U_A(a) = 0$ below $\underline{a}$, we can write (11) as

$$\int_0^1 \big(1 - F(a, \hat{z}(a))\big) U'_A(a)\, da.$$

For the fixed boundary z, the weight $1 - F(a, \hat{z}(a))$ is non-negative and independent of U_A. Since the constructed U'_A pointwise dominates the derivative of every other feasible $\check{U}_A$, it maximizes (11) among all feasible pairs $(z, \check{U}_A)$.

C.8 Proof of Lemma 4

First, note that since z^* is strictly increasing, Assumption 3 gives:

$$\frac{F_{A|B}(a|z^*(a))}{f_{A|B}(a|z^*(a))} \text{ is strictly increasing in } a, \quad \frac{F_{B|A}(b|(z^*)^{-1}(b))}{f_{B|A}(b|(z^*)^{-1}(b))} \text{ is strictly increasing in } b. \tag{29}$$

The proof has two steps. First, I show that z^* is piecewise linear. Then I rule out kinks.

Step 1: z^* is piecewise linear. Suppose towards a contradiction that z^* is not affine on some concave C^2 piece. Then there is a closed interval $[\alpha, \beta] \subset (\underline{a}^*, \overline{a}^*)$ on which $z^{*\prime\prime}(a) < 0$ for all $a \in [\alpha, \beta]$.

By Lemma 3, the welfare-maximizing A-indirect utility for a fixed boundary satisfies $U_A'(a) = k$ on any concave interval of the boundary, for some $k > 0$. Consider local perturbations of z^* on $[\alpha, \beta]$ that keep fixed $z^*(\alpha), z^*(\beta)$, the one-sided slopes $z^{*\prime}(\alpha), z^{*\prime}(\beta)$, and the mass below the boundary on this interval:

$$\int_\alpha^\beta \int_0^{z(a)} f(a,b)\, db\, da = \int_\alpha^\beta \int_0^{z^*(a)} f(a,b)\, db\, da.$$

For all sufficiently small perturbations of this kind that preserve concavity and keep $z' > 0$, the same U_A remains feasible. Indeed, U_A' is still constant and U_A'/z' remains non-decreasing. Since the endpoint slopes are fixed, these monotonicity constraints also remain valid where the perturbed interval is pasted back into the unchanged boundary. The fixed-mass condition preserves the supply of good A on the interval, and therefore also preserves the supply of good B on the same vertical strip. By Lemma 1, such perturbations affect welfare only through

$$-k \int_\alpha^\beta F(a, z(a))\, da.$$

Thus $z^*|_{[\alpha,\beta]}$ must locally maximize

$$-\int_\alpha^\beta F(a, z(a))\, da$$

among concave paths with the same endpoint values, endpoint slopes, and mass below the boundary. Equivalently, writing $y = z'$, $u = y'$ and $q'(a) = \int_0^{z(a)} f(a,b)\, db$, the local maximization has fixed values of z and y at both endpoints and fixed $q(\beta) - q(\alpha)$.

Let ξ, ϕ, μ denote the costates associated with z, y, q, respectively. The Hamiltonian is

$$H(a, z, y, u, q, \xi, \phi, \mu) = -F(a,z) + \mu \int_0^z f(a,b)\, db + \xi y + \phi u.$$

The costate on q is constant, so μ is constant. The remaining costate equations are

$$\xi'(a) = \int_0^a f(v, z^*(a))\, dv - \mu f(a, z^*(a)), \qquad \phi'(a) = -\xi(a).$$

Hence

$$\phi''(a) = \mu f(a, z^*(a)) - \int_0^a f(v, z^*(a))\, dv = f(a, z^*(a)) \left(\mu - \frac{F_{A|B}(a|z^*(a))}{f_{A|B}(a|z^*(a))} \right). \tag{30}$$

On $[\alpha, \beta]$, the control $u^* = z^{*\prime\prime}$ is strictly below the constraint $u \le 0$. Therefore the maximum condition for the free control gives $\phi(a) = 0$ for all $a \in (\alpha, \beta)$. Thus, $\phi''(a) = 0$ on (α, β). By (30),

$$\frac{F_{A|B}(a|z^*(a))}{f_{A|B}(a|z^*(a))} = \mu \qquad \text{for all } a \in (\alpha, \beta).$$

This, however, contradicts the strict monotonicity in (29). Hence z^* cannot be strictly concave on any open interval. Applying a symmetric argument rules out strictly convex intervals. Therefore $z^{*\prime\prime} = 0$ wherever the second derivative exists. Since z^* is piecewise twice continuously differentiable by Lemma 2, it is piecewise linear.

Step 2: z^* has no kinks. We now know that z^* is piecewise linear. Pick $\overline{v}$ such that $[\underline{a}^*, \overline{v}]$ is the largest interval starting at $\underline{a}^*$ on which z^* is either convex or concave. Suppose first that z^* is concave on this interval. The convex case is symmetric after passing to the inverse boundary and using the B-side inverse anti-hazard rate.

On this initial concave interval, possible nonlinearity of z^* can only take the form of downward jumps in the slope. Consider local variations of z^* on $[\underline{a}^*, \overline{v}]$ that keep fixed the endpoint values $z^*(\underline{a}^*)$ and $z^*(\overline{v})$, keep fixed the terminal slope $z^{*\prime}_{-}(\overline{v})$, and preserve the mass below the boundary:

$$\int_{\underline{a}^*}^{\overline{v}} \left(\int_0^{z(a)} f(a,b)\, db \right) da = \int_{\underline{a}^*}^{\overline{v}} \left(\int_0^{z^*(a)} f(a,b)\, db \right) da.$$

The initial slope is free. The admissible paths are concave, so away from kinks we can write

$$z'(a) = y(a), \qquad y'(a) = u(a), \qquad u(a) \le 0,$$

and at each kink a_i, $y_+(a_i) - y_-(a_i) = v_i \le 0$, while z and the accumulated-mass state do not jump.

The restriction of z^* to $[\underline{a}^*, \overline{v}]$ must solve this local problem. Indeed, any admissible improvement can be pasted into the original boundary. The endpoint conditions preserve continuity of the boundary and the terminal slope, and the mass condition preserves the supply of good A on the interval, hence also the supply of good B on the same vertical strip. Since Lemma 3 gives $U_A' = k$ on a concave interval, for some constant $k > 0$, the same pasting argument implies that any improvement in

$$-\int_{\underline{a}^*}^{\overline{v}} F(a, z(a))\, da$$

would strictly improve total welfare, contradicting the fact that z^* maximizes it.

We now show that such a local optimum cannot have a downward jump in slope. Let (ξ, ϕ, μ) be the costates associated with z, y, and the accumulated-mass state. Away from jump points, the Hamiltonian is

$$\mathcal{H} = -F(a, z^*(a)) + \mu \int_0^{z^*(a)} f(a,b)\, db + \xi(a) y(a) + \phi(a) u(a). \tag{31}$$

The costate on the accumulated-mass state is constant, so write it as μ. The remaining costate equations are

$$\xi'(a) = \int_0^a f(v, z^*(a))\, dv - \mu f(a, z^*(a)), \tag{32}$$

and

$$\phi'(a) = -\xi(a). \tag{33}$$

Since the initial value of y is free, the transversality condition gives

$$\phi(\underline{a}^*) = 0; \tag{34}$$

see Neustadt (1976, p. 234).

Because $u \le 0$ and the Hamiltonian is linear in u, the maximum condition implies

$$\phi(a) \ge 0 \tag{35}$$

away from jump points. Moreover, by the maximum principle with jumps (Seierstad and Sydsaeter, 1986, Theorem 7, pp. 196–197), the adjoints are continuous across each jump in the present problem. Hence ξ and ϕ are continuous across every jump. Since $\phi' = -\xi$ on both sides of each jump, it follows that

$$\lim_{a \uparrow a_i} \phi'(a) = \lim_{a \downarrow a_i} \phi'(a). \tag{36}$$

At any genuine downward jump, the jump size satisfies $v_i < 0$. Since v_i enters the jump map only through $y_+ = y_- + v_i$, the first-order condition for the jump size gives

$$\phi(a_i) = 0. \tag{37}$$

Combining (35) and (37), every genuine interior jump $a_i \in (\underline{a}^*, \overline{v})$ must be a local minimum of ϕ. Together with (36), this implies

$$\phi'(a_i) = 0. \tag{38}$$

I now show that no such point can exist.

Away from jump points, differentiating (33) and using (32) gives

$$\phi''(a) = -\xi'(a) = f(a, z^*(a))\left(\mu - \frac{F_{A|B}(a|z^*(a))}{f_{A|B}(a|z^*(a))}\right). \tag{39}$$

Since $f > 0$, (29) implies that the sign of ϕ'' changes sign at most once. It is either always weakly positive, always weakly negative, or positive up to some cutoff and negative after it.

Because $\phi(\underline{a}^*) = 0$ and $\phi \ge 0$, we have $\phi'_+(\underline{a}^*) \ge 0$. If $\phi'' \ge 0$ throughout the interval, then ϕ' is non-decreasing between jumps and continuous across jumps. Hence ϕ' cannot vanish at an interior local minimum of ϕ, so no interior jump can occur.

If $\phi'' \le 0$ throughout the interval, then ϕ' is non-increasing. The case $\phi'_+(\underline{a}^*) = 0$ is impossible, because then ϕ' would be negative immediately to the right of $\underline{a}^*$, forcing $\phi < 0$ nearby. Thus $\phi'_+(\underline{a}^*) > 0$. Since ϕ' is non-increasing, it can cross zero at most once. Such a crossing is a local maximum of ϕ, not a local minimum. Hence no interior jump can occur.

Finally, suppose there is a cutoff $\tilde{a}$ such that $\phi'' > 0$ on $(\underline{a}^*, \tilde{a})$ and $\phi'' < 0$ on $(\tilde{a}, \overline{v})$. Then ϕ' is increasing up to $\tilde{a}$, so, since $\phi'_+(\underline{a}^*) \geq 0$, we have $\phi'(a) > 0$ on $(\underline{a}^*, \tilde{a}]$. Hence no jump can occur weakly before $\tilde{a}$. For $a > \tilde{a}$, ϕ' is strictly decreasing and can cross zero at most once. At such a crossing, ϕ is strictly positive, because ϕ has already been increasing on $(\underline{a}^*, \tilde{a}]$. After the crossing, $\phi' < 0$. Thus there is again no interior point at which both $\phi(a) = 0$ and $\phi'(a) = 0$.

We have thus shown that z^* cannot have a downward jump in slope on the initial concave interval $[\underline{a}^*, \overline{v}]$. Since z^* is piecewise linear, it must therefore consist of a single linear piece.

C.9 Proof of Lemma 5

By Lemma 4, the optimal boundary z^* is linear. Let $\sigma > 0$ denote its slope. We now show $\sigma = 1$. For each s in a neighborhood of σ, let z_s denote the linear boundary with slope s that allocates the same masses of goods A and B as z^*. Write $\hat{z}_s$ for its extended boundary, and write $\underline{a}_s, \underline{b}_s$ for its participation cutoffs. Since f is strictly positive and continuous, these endpoints are locally uniquely pinned down by the two mass constraints and vary continuously with s.

We first record a monotonicity fact. If $s_1 > s_2$, then the two extended boundaries $\hat{z}_{s_1}$ and $\hat{z}_{s_2}$ cross exactly once. Let (a^*, b^*) be their crossing point. The flatter boundary $\hat{z}_{s_2}$ is above $\hat{z}_{s_1}$ to the left of a^*, and below it to the right of a^*. In particular, $\underline{a}_s$ is strictly increasing in s. Define

$$I_A(s) := \int_0^1 F(a, \hat{z}_s(a))\, da.$$

I claim that $I_A(s)$ is strictly increasing in s. To see this, fix $s_1 > s_2$, and define

$$\mathcal{D}^- := \{(a,b) : a < a^*,\ \hat{z}_{s_1}(a) < b < \hat{z}_{s_2}(a)\}, \quad \text{and} \quad \mathcal{D}^+ := \{(a,b) : a > a^*,\ \hat{z}_{s_2}(a) < b < \hat{z}_{s_1}(a)\}.$$

Then

$$\begin{aligned} I_A(s_1) - I_A(s_2) &= \int_0^1 \int_{\hat{z}_{s_2}(a)}^{\hat{z}_{s_1}(a)} \int_0^a f(v,b)\, dv\, db\, da \\ &= \int_{\mathcal{D}^+} \frac{F_{A|B}(a|b)}{f_{A|B}(a|b)} f(a,b)\, da\, db - \int_{\mathcal{D}^-} \frac{F_{A|B}(a|b)}{f_{A|B}(a|b)} f(a,b)\, da\, db. \end{aligned}$$

By Assumption 3, we have:

$$\frac{F_{A|B}(a|b)}{f_{A|B}(a|b)} > \frac{F_{A|B}(a^*|b^*)}{f_{A|B}(a^*|b^*)} \quad \text{on } \mathcal{D}^+, \qquad \frac{F_{A|B}(a|b)}{f_{A|B}(a|b)} < \frac{F_{A|B}(a^*|b^*)}{f_{A|B}(a^*|b^*)} \quad \text{on } \mathcal{D}^-.$$

Moreover, because z_{s_1} and z_{s_2} allocate the same mass of good A,

$$\int_{\mathcal{D}^+} f(a,b)\, da\, db = \int_{\mathcal{D}^-} f(a,b)\, da\, db,$$

so $I_A(s_1) - I_A(s_2) > 0$.

Now suppose, toward a contradiction, that $\sigma > 1$. For any linear boundary with slope $s > 1$, Lemma 3 gives $m \equiv 1$ and

$$k = \frac{1}{\max\{1, 1/s\}} = 1.$$

Thus the optimal A-indirect utility satisfies

$$U'_{A,s}(a) = 0 \quad \text{for } a < \underline{a}_s, \qquad U'_{A,s}(a) = 1 \quad \text{for } a > \underline{a}_s.$$

Using Lemma 1, welfare under z_s is therefore

$$\begin{aligned} W[z_s] &= \int_0^1 \big(1 - F(a, \hat{z}_s(a))\big) U'_{A,s}(a)\, da \\ &= \int_{\underline{a}_s}^1 \big(1 - F(a, \hat{z}_s(a))\big)\, da \\ &= 1 - \underline{a}_s - \int_0^1 F(a, \hat{z}_s(a))\, da = 1 - \underline{a}_s - I_A(s), \end{aligned}$$

where the third equality uses $F(a, 0) = 0$. Since both $\underline{a}_s$ and $I_A(s)$ are strictly increasing in s, welfare is strictly decreasing in s on the region $s > 1$. Hence, if $\sigma > 1$, lowering the slope slightly while preserving the masses of both goods would strictly increase welfare. This contradicts the optimality of z^*.

It remains to rule out $\sigma < 1$. Let $\hat{w}_s(b) := \hat{z}_s^{-1}(b)$ denote the generalized inverse boundary. When $s < 1$, the inverse boundary has slope $1/s > 1$. Applying the preceding argument with the roles of A and B reversed, the expression

$$I_B(s) := \int_0^1 F(\hat{w}_s(b), b)\, db$$

is strictly decreasing in s on the region $s < 1$, and $\underline{b}_s$ is also strictly decreasing in s. Moreover, applying Lemma 3 to the inverse boundary gives

$$U'_{B,s}(b) = 0 \quad \text{for } b < \underline{b}_s, \qquad U'_{B,s}(b) = 1 \quad \text{for } b > \underline{b}_s.$$

Using the symmetric version of Lemma 1, welfare can be written as

$$W[z_s] = \int_0^1 \big(1 - F(\hat{w}_s(b), b)\big) U'_{B,s}(b)\, db = 1 - \underline{b}_s - I_B(s).$$

Since both $\underline{b}_s$ and $I_B(s)$ are strictly decreasing in s, welfare is strictly increasing in s on the region $s < 1$. Hence, if $\sigma < 1$, raising the slope slightly while preserving the masses of both goods would strictly increase welfare. This again contradicts the optimality of z^*.

C.10 Proof of Lemma 6

I first show that the optimal mechanism allocates positive masses of both goods. Indeed, consider a mechanism allocating only A. Now, augment it by adding to its menu of options one that allocates good B with $x = 1$ and a toll of $c = 1 - \varepsilon$. Since the distribution F of values (a, b) is full-support, for any $\varepsilon > 0$ there will be a mass of agents who prefer the B-good option to any A-good option offered; that is, introducing this option increases their welfare. Note that adding the B-option only relaxes the supply constraint on A, and that the mass of the agents taking the B-option converges to 0 as $\varepsilon \to 0$. Thus, the augmented mechanism satisfies the supply constraint for B for ε small enough.

Then, by Lemmas 5 and 3, the optimal mechanism can be written as two undamaged options, A at toll c_A and B at toll c_B, so each type chooses the best element of $\{0,\ a - c_A,\ b - c_B\}$.

Suppose, toward a contradiction, that the supply constraint for good A is slack. Since a positive mass of agents receive A, we have $c_A < 1$. Also $c_A > 0$; otherwise almost every agent would receive some good, contradicting slackness of A's constraint together with $s_A + s_B \leq 1$. Now lower c_A to $c_A - \varepsilon$, leaving c_B unchanged. For every $\varepsilon \in (0, c_A)$, all agents are weakly better off, and the agents who previously chose A are strictly better off. Since A was allocated to a positive mass of agents, welfare strictly increases.

It remains to check feasibility. Since F has a density and the indifference set is null, the mass choosing A converges to its original value as $\varepsilon \downarrow 0$. Since the original A-constraint was slack, it remains satisfied for all sufficiently small ε. The mass choosing B can only fall, since the A-option has become more attractive and the B-option is unchanged. Thus, the B-constraint remains satisfied, so the modified mechanism is feasible; contradiction.

C.11 Proof of Proposition 3

The allocation of the market-clearing toll mechanism is:

$$q_A^*(a,b) = \mathbb{1}\{a - c_A^* > \max\{0, b - c_B^*\}\}, \quad \text{and} \quad q_B^*(a,b) = \mathbb{1}\{b - c_B^* > \max\{0, a - c_A^*\}\}.$$

up to a measure-zero set of indifferent types. By Lemma 6, this mechanism allocates the whole supply of both goods, so $\int q_A^* \, dF = s_A$ and $\int q_B^* \, dF = s_B$. Now, consider any feasible mechanism. Define

$$q_A(a,b) := \mathbb{1}_{y(a,b)=A} x(a,b), \qquad q_B(a,b) := \mathbb{1}_{y(a,b)=B} x(a,b).$$

Then $q_A, q_B \in [0,1]$, $q_A + q_B \leq 1$, $\int q_A \, dF \leq s_A$ and $\int q_B \, dF \leq s_B$. Thus, it is enough to show that (q_A^*, q_B^*) maximizes the relaxed problem over all such q_A, q_B.

For every type (a,b), the rule (q_A^*, q_B^*) chooses a maximizer among 0, $a - c_A^*$ and $b - c_B^*$. Therefore, pointwise,

$$q_A^*(a,b)(a - c_A^*) + q_B^*(a,b)(b - c_B^*) \geq q_A(a,b)(a - c_A^*) + q_B(a,b)(b - c_B^*).$$

Integrating gives

$$\int q_A^* a + q_B^* b \, dF - c_A^* s_A - c_B^* s_B \geq \int q_A a + q_B b \, dF - c_A^* \int q_A \, dF - c_B^* \int q_B \, dF.$$

Since $c_A^*, c_B^* \geq 0$, $\int q_A \, dF \leq s_A$, and $\int q_B \, dF \leq s_B$, it follows that

$$\int q_A^* a + q_B^* b \, dF \geq \int q_A a + q_B b \, dF.$$

Hence the market-clearing toll mechanism 1 maximizes allocative efficiency.

D Verifying examples

D.1 Verifying Example 1

Consider mechanisms which do not use damages. By an argument analogous to the proof of Lemma 6, we can restrict attention to mechanisms allocating all the available supply. The only such mechanism takes the form

$$y(a,b) = B, \; c(a,b) = 1/2 \quad \text{when } b - a > 1/2, \qquad \text{and} \qquad y(a,b) = A, \; c(a,b) = 0 \quad \text{when } b - a < 1/2.$$

Thus, the total welfare from this mechanism is

$$\int_{\{b-a>1/2\}} \left(b - \tfrac{1}{2}\right) f(a,b) \, da \, db + \int_{\{b-a<1/2\}} a \, f(a,b) \, da \, db = \frac{14\varepsilon^2 - 9\varepsilon + 23}{42} - \frac{28\varepsilon^2 - 46\varepsilon + 25}{252(1-\varepsilon)}.$$

This converges to $\frac{113}{252}$ as $\varepsilon \to 0^+$.

Now consider an alternative mechanism which offers two options: good A with $x_A = 1$ and good B with $x_B^\varepsilon < 1$, with no tolls for either. For small enough ε, there exists x_B^ε for which both supply constraints hold with equality. We can verify that

$$x_B^\varepsilon = \frac{7}{16} - \frac{287}{1024}\varepsilon + O(\varepsilon^2),$$

and so $x_B^\varepsilon \to 7/16$ as $\varepsilon \to 0^+$. Calculation confirms that the limit welfare from this mechanism is

$$\lim_{\varepsilon \to 0^+} \left[\int_{\{x_B^\varepsilon b > a\}} x_B^\varepsilon b \, f(a,b) \, da \, db + \int_{\{x_B^\varepsilon b < a\}} a \, f(a,b) \, da \, db \right] = \frac{17497}{36288}.$$

$\frac{17497}{36288} > \frac{113}{252}$, so this mechanism dominates the no-damage mechanism for $\varepsilon > 0$ sufficiently small.

D.2 Verifying Example 2

Note Corollary 1 applies and

$$\frac{F_{A|B}(a \mid a+c_B^*)}{f_{A|B}(a \mid a+c_B^*)} = \frac{\int_0^a e^{\lambda t(a+c_B^*)}\,dt}{e^{\lambda a(a+c_B^*)}} = \frac{1-e^{-\lambda a(a+c_B^*)}}{\lambda(a+c_B^*)}.$$

Fix any $\tau > 0$, and set $\tilde{a}_\lambda := 1 - c_B^* - \tau/\lambda$. For all sufficiently large λ, we have $\tilde{a}_\lambda \in (0, 1-c_B^*)$. Let A_λ denote a random variable distributed according to the boundary measure. Define the random variable $Y_\lambda := \lambda(1 - c_B^* - A_\lambda)$. Its density is obtained from the change of variables $a = 1 - c_B^* - \frac{y}{\lambda}$, so for $y \in [0, \lambda(1-c_B^*)]$, the density of Y_λ is proportional to

$$\exp\left\{\lambda\left(1 - c_B^* - \frac{y}{\lambda}\right)\left(1 - \frac{y}{\lambda}\right)\right\} = \exp\left\{\lambda(1-c_B^*) - (2-c_B^*)y + \frac{y^2}{\lambda}\right\}.$$

After normalizing, the constant term $\lambda(1-c_B^*)$ drops out. Hence Y_λ converges in distribution to an exponential random variable Y with rate $2 - c_B^*$. Then

$$(\tilde{a}_\lambda - A_\lambda)_+ = \frac{(Y_\lambda - \tau)_+}{\lambda}.$$

Also,

$$\frac{1-e^{-\lambda A_\lambda(A_\lambda + c_B^*)}}{\lambda(A_\lambda + c_B^*)} = \frac{1}{\lambda(1 - Y_\lambda/\lambda)} + o(\lambda^{-2}) = \frac{1}{\lambda} + \frac{Y_\lambda}{\lambda^2} + o(\lambda^{-2}),$$

Therefore

$$\lambda^3 \operatorname{Cov}\left(\frac{F_{A|B}(a \mid a+c_B^*)}{f_{A|B}(a \mid a+c_B^*)}, (\tilde{a}_\lambda - a)_+ \;\middle|\; b = a + c_B^*\right) \to \operatorname{Cov}\left(Y, (Y-\tau)_+\right).$$

Since $Y \sim \operatorname{Exp}(2 - c_B^*)$,

$$\operatorname{Cov}\left(Y, (Y-\tau)_+\right) = \mathbb{E}\left[Y(Y-\tau)_+\right] - \mathbb{E}[Y]\mathbb{E}[(Y-\tau)_+] = e^{-(2-c_B^*)\tau}\left(\frac{\tau}{2-c_B^*} + \frac{1}{(2-c_B^*)^2}\right) > 0.$$

Hence, for all sufficiently large λ,

$$\operatorname{Cov}\left(\frac{F_{A|B}(a \mid a+c_B^*)}{f_{A|B}(a \mid a+c_B^*)}, (\tilde{a}_\lambda - a)_+ \;\middle|\; b = a + c_B^*\right) > 0.$$