Pricing priorities in waitlists*

Filip Tokarski
Stanford GSB

December 14, 2024

Abstract

I study the problem of introducing pricing into waitlists when extracting revenue from participants is undesirable. The designer hands out heterogeneous goods to arriving agents and aims to maximize allocative efficiency. She can incentivize agents to join socially optimal waitlists by charging joining fees and allowing agents to pay to reduce their wait-time. I show that screening through wait-times, while not wasteful, can only extract information about agents' relative values for the offered goods. In contrast, using payments incurs waste but allows for screening on absolute valuations, improving allocative efficiency. This trade-off has a relatively simple resolution: the optimal mechanism charges a price for joining only one of the waitlists and offers a discrete menu of pay-to-skip options.

1 Introduction

Waitlists are the most common alternative to market-based allocation mechanisms. Frequently employed by policymakers prioritizing recipients' welfare and unwilling to extract revenue from participants, they are used to allocate goods as diverse as affordable housing, daycare spots, and camping permits. However, unlike prices, waitlists do not allow participants to express *how much* they value the distributed good. For instance, whenever joining the waitlist is free, everyone with positive value for the good has an incentive to sign up, resulting in some of the supply being given to those with little need for it. This paper asks how such issues can be addressed by combining waitlists with pricing, while recognizing that extracting revenue from participants may in itself be undesirable.

I study a model where heterogeneous goods and agents arrive over time. Each kind of good is associated with a separate waitlist and arriving agents decide which waitlist to join. The designer aims to allocate goods to those who value them most, and thus to incentivize agents to join socially optimal waitlists. She can do so by adjusting the wait-times in the waitlists and by charging participants fees. Specifically, the designer can price joining waitlists and let agents pay to reduce their wait-time in a particular waitlist (which I call 'pay-to-skip

*I am grateful to Joey Feffer, Federico Llarena, Sam Wycherley, Rafael Berriel, Flint O'Neill, Andrzej Skrzypacz, Piotr Dworczak, Michael Ostrovsky, Ilya Segal and Rebecca Diamond their helpful comments and suggestions.

options'). However, the designer views extracting money from participants as wasteful. She can therefore combine two screening instruments: wait-times and payments. Wait-times are a non-wasteful instrument, but since the costs of waiting are associated with delaying the good's receipt, they can only extract information about agents' *relative* values for the offered goods, and not their *absolute* values for them. Payments, on the other hand, incur waste by extracting resources from participants, but allow the designer to screen on agents' absolute valuations, improving allocative efficiency. I describe the structure of the optimal mechanism: it charges a price for joining only one of the waitlists and offers finitely many pay-to-skip options.

From a technical perspective, my model is an instance of a tractable multidimensional screening problem. I characterize the design of pricing schemes for two waitlists as two 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 join each waitlist. 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 optimal boundary among all implementable ones.

The literature has not studied combining waitlists with payments in settings with heterogeneous goods. However, a substantial literature examines such waitlists where monetary transfers are infeasible. Ashlagi et al. (2024) demonstrate that allocative efficiency can be improved by coarsening agents' information about the qualities of allocated goods. Arnosti and Shi (2020) compare common non-monetary mechanisms in terms of targeting and match efficiency. 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 distinguishing between *waitlists*, where waiting is passive and its cost stems from delayed receipt, and *queues*, where agents actively waste time waiting. I show that in the former case, the ability of wait-times to 'act as prices' is impeded, as they can only screen on agents' relative values for the offered goods. My work also relates to the literature on allocation problems without money (e.g. Hylland and Zeckhauser (1979) and Budish (2011)). While my paper shares the motivation that using money can be costly or undesirable in some settings, it relaxes the extreme approach of not using it at all and investigates whether limited use of money can still be beneficial. Finally, the wasteful nature of payments in my setting relates to the literature on costly screening and money burning (e.g. Hartline and Roughgarden (2008), Condorelli (2012)). However, this literature focuses on cases where the designer only has access to the wasteful screening device. An exception is Yang (2021) who considers a monopolist with both wasteful and non-wasteful instruments and characterizes cases where the wasteful one should not be used. In contrast, this paper focuses on optimally combining such instruments.

2 Model

A designer distributes two types of goods, A and B, to agents with heterogeneous preferences. Agents' types are characterized by a pair $(a,b) \in [0,1]^2$, where a and b represent their values for goods A and B, respectively. The utility of a type-(a,b) agent who receives good

x, pays p, and waits for time t is given by:

$$\begin{cases} e^{-\rho t}(a-p) & \text{if } x = A, \\ e^{-\rho t}(b-p) & \text{if } x = B, \end{cases}$$

where $\rho > 0$ is the discount rate. Note the payment occurs at the end of the waiting period.

Goods and agents arrive continuously over time. At every instance $\tau \in \mathbb{R}$, flow masses $\mu_A, \mu_B > 0$ of goods A and B arrive, with $\mu_A + \mu_B = 1$. Concurrently, a unit flow mass of agents arrives, with types (a, b) distributed according to a joint distribution F. I assume that F has full support on $[0,1]^2$ and has a Lipschitz continuous and differentiable density f.

There are two separate first-come-first-serve waitlists, one for each good. The designer can choose prices for joining each waitlist, as well as menus of pay-to-skip options for them. That is, she can offer agents joining a particular waitlist the choice to pay extra to reduce their wait-time in it. Such pay-to-skip menus can involve a continuum of options and can differ between the two waitlists. Arriving agents must therefore choose which waitlist to join (if any) and whether to pay extra to reduce their wait-time.

I focus on mechanisms that admit a steady state of the system. This means that the distribution of types in each waitlist is stationary, as are the distributions of types joining and leaving each waitlist (which are also always equal to each other). Thus, under such mechanisms, almost all agents of the same type choose the same waitlist and pay-to-skip option, regardless of when they arrive. By the Revelation Principle, we can then think of the designer as choosing steady state allocation rules for payments $p : [0,1]^2 \to \mathbb{R}_+$, wait-times $t : [0,1]^2 \to \mathbb{R}_+$ and goods $x : [0,1]^2 \to \{A, B, \varnothing\}$. I call a triplet (p, t, x) a mechanism. The designer selects a mechanism subject to the following constraints:

$$\text{for all } (a,b),(a',b') \in [0,1]^2, \quad U[a,b,(p,t,x)(a,b)] \geq U[a,b,(p,t,x)(a',b')], \tag{IC}$$

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

$$\int \mathbb{1}_{x(a,b)=A} \, \mathrm{d}F(a,b) \leq \mu_A, \quad \int \mathbb{1}_{x(a,b)=B} \, \mathrm{d}F(a,b) \leq \mu_B, \tag{S}$$

where $U[a,b,(p,t,x)(a',b')]$ denotes the utility type (a,b) gets from reporting type (a',b'). The supply constraints (S) ensure that the mass of agents assigned to either waitlist in any period does not exceed the mass of the corresponding good that arrives in it. I also assume agents can choose not to participate and receive nothing at no cost. I call a mechanism (p,t,x) satisfying (IC),(IR) and (S) *feasible*.

The designer chooses a feasible mechanism to maximize the following objective combining revenue and allocative efficiency:

$$\gamma \cdot R + W,$$

where $\gamma \in [0,1]$, and:

$$R = \int p(a,b) \, \mathrm{d}F(a,b),$$

$$W = \int \underbrace{\mathbb{1}_{x(a,b)=A}(a - p(a,b))}_{\text{Agents getting good } A} + \underbrace{\mathbb{1}_{x(a,b)=B}(b - p(a,b))}_{\text{Agents getting good } B} \, \mathrm{d}F(a,b).$$

The parameter γ captures the value the designer puts on revenue. In one extreme, $\gamma = 1$, the designer is indifferent about payments made to her by the participants; this may correspond to the case where the designer can costlessly rebate the revenue to them. The other extreme, $\gamma = 0$, means that the designer considers all such payments completely wasted. This might correspond to an environment where payments are not monetary, but represent a wasteful ordeal such as form-filling or travelling to a distant office. Intermediate values of γ represent cases where the designer values money in agents' pockets more than in her own, but also has some use for the generated revenue. This case may capture the nature of social programs whose participants are less wealthy than the general taxpayer, so a redistributive designer would not want to use them to generate revenue.[1] For γ to be intermediate, rebating revenue to the participants has to be costly. In the case of private enterprises, this could be because monetary payments are taxed (while in-kind benefits are not). In the case of government programs, this could be because distributing cash lacks the screening benefits of in-kind transfers. For instance, when the designer hands out a free (or subsidized) inferior good, only relatively poor agents will want to participate as wealthier ones can afford higher-quality alternatives. Thus, the subsidy is automatically targeted to those who need it most (Besley and Coate, 1991). As soon as the designer hands out cash, however, such targeting disappears as money is desired by everyone, regardless of wealth.

The latter term W in the objective captures the value provided by the good to those who receive it. Observe, however, that wait-times do not explicitly appear in W. While this may seem counterintuitive, it follows naturally from the properties of steady states when the designer cares only about agents' values for the goods, and not their arrival times. That is, when the designer faces two agents who value the good identically but differ in how long they have been waiting already, she is indifferent about giving the good to either. Note, moreover, that as long as the designer hands out all goods as soon as they arrive, wait-times have a zero-sum nature—giving the good earlier to one agent requires pushing other agents back. Furthermore, in the steady-state, the types of agents pushed back this way are always identical to the agent who had been moved forward. The designer is therefore indifferent between any two steady states which charge the same types the same payments and give them the same goods.

A further consequence of this logic is that the designer is indifferent about certain types 'skipping ahead'. To see why, consider a particular steady state and compare it to another which differs only in that the wait-times of all agents with some particular type are reduced. Since the same types still join the same waitlists, the inflows and outflows of all types are unchanged relative to the old steady state. Moreover, since the wait-times were reduced for

[1]While the model does not explicitly account for wealth differences or heterogeneous welfare weights among agents, this can be viewed as an approximation of a scenario where such differences exist but are relatively small *between* participants compared to the gap between participants and the average taxpayer. This is especially likely when the designer allocates inferior goods, such as public housing in undesirable areas. The designer's welfare-weighted objective can then be approximated by a constant weight on all participants, which is distinct from that on revenue, representing her welfare weight for the average taxpayer.

all agents of a given type, this effectively corresponds to a 'reindexing' of such types—agents of that type arriving later now replace 'earlier copies of themselves' (Figure 1).

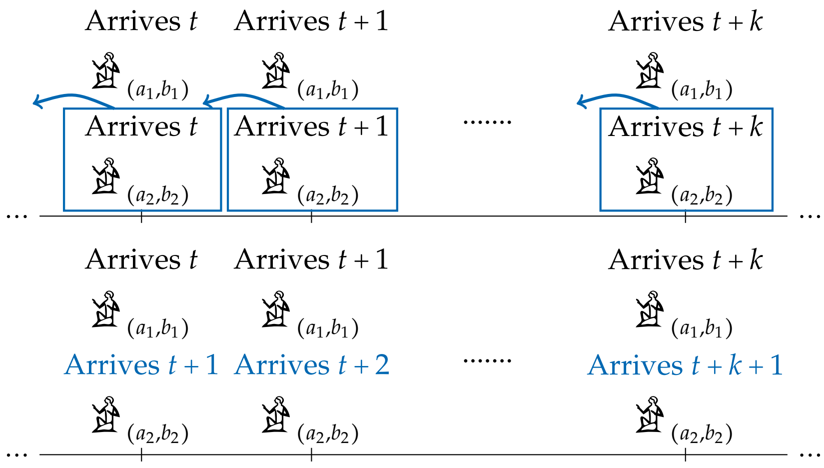

Figure 1: Skipping ahead only 'reindexes' agents of the same type.

Consequently, when the designer offers agents the options to pay to skip, she does not do so for the sake of affecting their wait-times. Instead, she uses such options as a means of attracting certain types to join a particular waitlist, and therefore to affect the allocation of goods between types.

3 Feasible mechanisms

Let us first describe the space of feasible mechanisms. It will be convenient to characterize them in terms of waitlist-specific indirect utility functions $U_A, U_B : [0,1] \to \mathbb{R}_+$, defined as:

$$U_A(a) = \max_{(a',b'):\ x(a',b') \in \{A, \varnothing\}} e^{-\rho \cdot t(a',b')}(a - p(a',b')), \tag{1}$$

$$U_B(b) = \max_{(a',b'):\ x(a',b') \in \{B, \varnothing\}} e^{-\rho \cdot t(a',b')}(b - p(a',b')). \tag{2}$$

Intuitively, $U_A(a)$ and $U_B(b)$ give the highest utility type (a,b) could get from selecting some pay-to-skip option from the A- and the B-waitlists, respectively (or not participating). Note that waitlist-specific indirect utilities depend only on one dimension of the type—an agent's value for good B does not affect her choice of pay-to-skip option in waitlist A.

We will use U_A and U_B to describe agents' choices of waitlist. Indeed, type (a,b) for whom $U_A(a) > U_B(b)$ will join the A-waitlist ($x(a,b) = A$) and vice versa. Let us also define the *lowest participating values* as follows:

$$\underline{a} = \sup\{a :\ U_A(a) = 0\}, \quad \underline{b} = \sup\{b :\ U_B(b) = 0\}.[2] \tag{3}$$

[2]Since the mechanism offers a non-participation option, we always have $U_B(0) = U_A(0) = 0$. Thus, these suprema are well-defined.

We can now equivalently characterize feasible mechanisms in terms of the waitlist-specific indirect utility functions U_A, U_B they induce. We will say that U_A, U_B are feasible if there exists a mechanism inducing them.

Lemma 1. *Waitlist-specific indirect utility functions U_A, U_B are feasible if and only if:*

1. *U_A, U_B are convex,*
2. *$U'_A(a), U'_B(b) \in [0,1]$ whenever they exist,*
3. *$U_A(0) = U_B(0) = 0$,*
4. *$\int \mathbb{1}_{U_A(a)>U_B(b)} \,\mathrm{d}F(a,b) \leq \mu_A$, $\int \mathbb{1}_{U_A(a)<U_B(b)} \,\mathrm{d}F(a,b) \leq \mu_B$.*

While proof is in the appendix, the result follows from Myerson's lemma (Myerson, 1981) and an envelope argument (Milgrom and Segal, 2002). Intuitively, point 1 corresponds to the mechanism satisfying (IC). Point 2 is due to discounting terms $e^{-\rho \cdot t}$ being bounded above by 1, and thus limiting how steeply agents' indirect utilities can change with type. Point 3 corresponds to (IR) and point 4 captures the supply constraint (S).

I later use the above characterization to rewrite the designer's problem as one of selecting waitlist-specific indirect utility functions U_A, U_B. The following result helps us by characterizing how agents self-select into the two waitlists given U_A and U_B:

Definition 1. *Let a **boundary** be a function $g : [\underline{a}, \overline{a}] \to \mathbb{R}$ that is continuous, strictly increasing and satisfies $\overline{a} \leq 1$ and $g(\overline{a}) \leq 1$, with one of them holding with equality.*

Lemma 2. *All types point-wise below the lowest participating values do not join either waitlist, that is $x(a,b) = \varnothing$ for all $(a,b) < (\underline{a}, \underline{b})$.*[3] *Suppose a positive mass of agents joins either waitlist. Then the waitlist choices of types $(a,b) > (\underline{a}, \underline{b})$ are characterized by a boundary $g : [\underline{a}, \overline{a}] \to \mathbb{R}$ which satisfies:*

$$\int_{\underline{a}}^{1} \Phi_A(a) da \leq \mu_A, \quad \int_{g(\underline{a})}^{1} \Phi_B(b) db \leq \mu_B, \tag{S'}$$

where:

$$\Phi_A(a) := \int_0^{g(\min[a,\overline{a}])} f(a,v) dv, \quad \Phi_B(b) := \int_0^{g^{-1}(\min[b, g(\overline{a})])} f(v,b) dv.$$

Then a type $(a,b) > (\underline{a}, \underline{b})$ joins waitlist A if it is below the boundary g, that is, if $g(a) > b$, and joins waitlist B if it is above the boundary g, that is, if $g(a) < b$. Moreover, types at the boundary are indifferent between joining either waitlist, thus:

$$U_A(a) = U_B(g(a)) \quad \textit{for all} \quad a \in [\underline{a}, \overline{a}]. \tag{I}$$

*I then say that U_A, U_B **implement** the boundary g.*

Lemma 2 describes how U_A and U_B pin down the waitlist choices of almost everyone (indifferent types constitute a measure-zero set). These choices are summarized in Figure 2a.

[3] When comparing vectors, I will use $\geq$ and $>$ for pointwise comparisons.

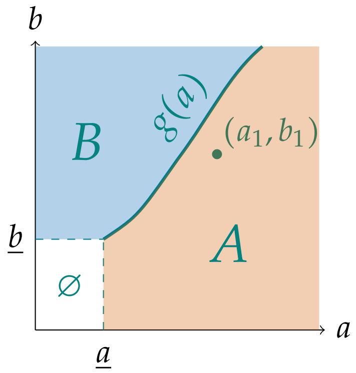

Figure 2a: Types below the boundary (orange) self-select into waitlist A and types above it (blue) self-select into waitlist B.

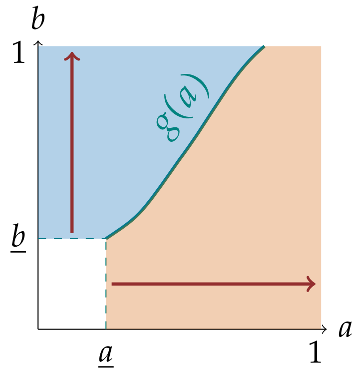

Figure 2b: The supply condition (S′) ensures that the probability masses below the boundary (orange) and above it (blue) are at most μ_A and μ_B, respectively. The red arrows mark the direction of integration in the left-hand-sides in (S′).

Types (a, b) for whom $(a, b) < (\underline{a}, \underline{b})$ do not participate. This happens when entering the waitlists is priced at $\underline{a}$ and $\underline{b}$, so such types never find joining worth it. On the contrary, anyone who values one of the goods above the corresponding price will join some waitlist—doing so guarantees them positive surplus, even if it is realized after a long wait.

Let us now give an intuition for the waitlist choices of participating types. Consider some type (a_1, b_1) choosing waitlist A (Figure 2a). Then any type (a, b) with $a > a_1$ and $b < b_1$ will also join waitlist A—since she values the A-good even more than (a_1, b_1) and values the B-good even less, all options in waitlist B are strictly less attractive to her than they are to (a_1, b_1). We can now notice that the types who are indifferent between their best options in both waitlists lie on an upwards-sloping curve g originating from $(\underline{a}, \underline{b})$. By the above logic, all types below this curve join waitlist A and pick some pay-to-skip option from its menu, while types above it join waitlist B.

We can thus think of our multidimensional mechanism design problems as two single-dimensional problems connected endogenously through the boundary g. While agents on its either side effectively face one-dimensional problems, making one of the waitlists more attractive invites more types to switch to it, effectively deforming the boundary.

Lemma 2 also changes the way we express the supply constraint. Rather than look at allocations $x(a, b)$, it takes advantage of the fact that types joining waitlist A (B) are below (above) the boundary. It then measures the masses of agents joining either waitlist by integrating over agents below and above g (Figure 2b).

4 Role of payments

I now explain why the designer might benefit from incorporating payments into the mechanism despite the fact that such payments generate waste. To that end, let us first consider two extreme cases: one where the designer does not use money, i.e. does not price joining waitlists or offer pay-to-skip options, and one where payments are not wasteful, so $\gamma = 1$.

4.1 Mechanisms without money

Proposition 1. *Consider a mechanism without money, so one where $p(a,b) \equiv 0$, that allocates the good to some agents. Then there exists $k \in (0,\infty)$ such that all types with $a/b > k$ join waitlist A and all types with $a/b < k$ join waitlist B.*

Proof. Note that unless the designer allocates nothing, she must allocate the whole available supply of both A and B. Suppose that some type (a,b) received one of the goods for zero payment while a positive mass of agents received nothing. Then almost all such agents would strictly prefer to report (a,b). (IC) thus requires that all agents receive either good, implying that $\underline{a} = 0$ or $\underline{b} = 0$. Since a unit mass of goods is allocated, the supply conditions (S′) for both goods have to hold with equality.

Now, when $p(a,b) \equiv 0$, all types joining either waitlist must have the same waiting time—otherwise all types joining a waitlist would report the type with the shortest wait-time in it. Let t_A and t_B be the wait-times in the two waitlists. The boundary g for this mechanism must then satisfy the boundary indifference condition:

$$a \cdot e^{-\rho \cdot t_A} = g(a) \cdot e^{-\rho \cdot t_B} \quad \text{for all } a \in [0,\bar{a}]. \tag{I}$$

Thus, by Lemma 2, all types for whom $a/b > k = e^{\rho(t_A - t_B)}$ join waitlist A, and all types for whom $a/b < k$ join waitlist B. □

Under such a mechanism joining either waitlist is free, and so all applicants with non-zero value for either good will join one of the waitlists. Moreover, in the absence of payments, the 'market will clear' based on wait-times—if good A is overdemanded, that is, preferred by more agents than the mass of this good—the wait-time for it will be longer. This will in turn deter some agents who prefer good A from joining this waitlist and encourage them to wait for good B instead. Importantly, however, wait-times can only screen agents based on their *relative values* for the two goods, that is, the ratio a/b. Graphically, this corresponds to the boundary g partitioning the type space along a ray originating from zero (Figure 3). The slope of the boundary reflects the ratio a/b for which agents are indifferent between the waitlists. This slope is pinned down by the relative supply of the two goods.

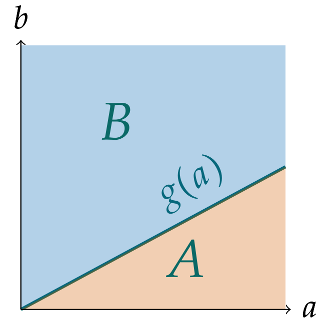

Figure 3: Without money, agents choose waitlists based on a/b.

Intuitively, waiting can only screen on relative values because its costs come from discounting, and thus are multiplicative with agents' values for the goods. This draws a contrast between *waitlists*, where waiting is passive, and its cost comes from delaying receipt, and *queues* where agents actively waste time waiting. In the latter case, the cost of waiting is the alternative cost of time, which is independent of one's value for the offered good.

4.2 Non-wasteful payments ($\gamma = 1$)

In the case where payments are not wasteful, the optimal mechanism takes a simple form:

Proposition 2. *Suppose payments are not wasteful, so $\gamma = 1$. Then the optimal mechanism does not offer any pay-to-skip options and prices only one waitlist. The joining price is chosen to equalize the wait-times in both waitlists.*

Proof. Since $\gamma = 1$, the designer's objective can be written as:

$$\int \mathbb{1}_{x(a,b)=A} \cdot a \; + \; \mathbb{1}_{x(a,b)=B} \cdot b \; \mathrm{d}F(a,b).$$

Consider first the relaxed problem where the designer chooses the mechanism subject only to supply constraints (S). The optimal mechanism would then allocate all of the available A-good to the μ_A agents with the highest difference $a - b$, and allocate the B-good to the μ_B remaining ones (Figure 4). Thus, such a mechanism would correspond to a boundary $g^* : [\underline{a}^*, \overline{a}^*] \to \mathbb{R}$ with a slope of 1 for which the supply conditions (S′) hold with equality (such a boundary g^* exists by the continuity of the density f). Suppose that g^* crosses the a axis first, that is, $\underline{a}^* \geq 0$ (the opposite case is symmetric). Then the relaxed-optimal boundary g^* can be implemented by the following waitlist-specific indirect utilities:

$$U_A(a) = \max[0,\; a - \underline{a}^*], \quad U_B(b) = b,$$

which are feasible by Lemma 1. Recall also that the indirect utilities of types (a,b) in waitlists A and B are equal to $U_A(a)$ and $U_B(b)$, respectively. Thus, by the envelope formula, $U'_A(a), U'_A(b) = e^{-\rho \cdot t(a,b)}$ for such types. Since $U'_A(a) = U'_B(b) = 1$ for all such types, all agents' wait-times must be equal. □

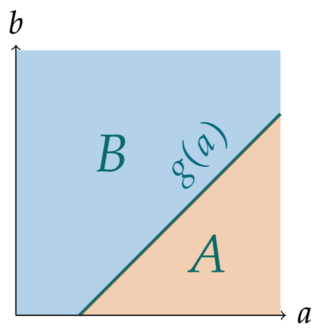

Figure 4: Optimal allocation subject only to supply constraints.

When transfers are not wasteful, the optimal mechanism simply prices entry to the waitlist for the overdemanded good. This simple mechanism not only solves the problem with private information about type, but also implements the first-best allocation subject only to supply constraints. This is unsurprising—when the designer is indifferent about transfers between herself and agents, prices costlessly elicit agents' valuations and assign the goods to whoever values them most. Given the efficiency properties of prices, screening through wait-times becomes redundant, and even harmful. This is because, as noted in the previous section, wait-times can only screen agents on their relative preferences. Since the designer's objective depends on agents' *absolute* values (which screening with prices can recover) altering allocations based on relative values could only worsen the outcome.

4.3 Trade-off: efficient assignment vs. minimizing payments

Juxtaposing the above examples illustrates the main trade-off associated with screening through prices. When the designer does not use payments in the mechanism, agents will still self-select into waitlists based on their values for the goods. However, they will do so only based on their *relative* values for them. Consequently, the designer will not be able to distinguish between two agents whose value ratios a/b are equal but whose *absolute* values differ. However, the designer cares about these two agents' allocations to different extents. If the former agents' values for both goods are higher, it is more important to give her the good she prefers. The first-best mechanism would therefore distort the assignments to agents whose values for both goods are low (by giving them the less demanded one) and leave the overdemanded good to those whose absolute value for it is large. This is exactly what the mechanism from Proposition 2 manages to accomplish. However, it does so through charging relatively high prices to agents joining the waitlist for the overdemanded good—this extreme solution might not be optimal when transfers are wasteful.

Introducing payments to the mechanism therefore presents a trade-off: on the one hand, it lets the designer screen on *absolute* rather than *relative* values, which lets her allocate goods more efficiently. Graphically, this lets her produce an 'area split' more like the right panel of Figure 5 than the left one. On the other, charging agents money generates waste. While Figure 5 illustrates the two extreme ends of this trade-off, the designer might in general prefer to compromise the 'area split' to reduce the payments that achieving it requires.

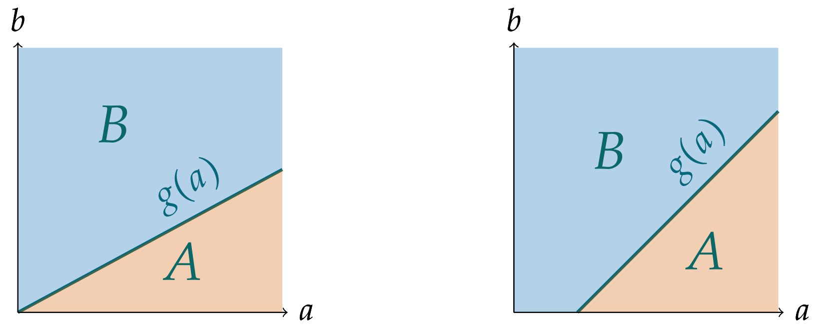

Figure 5: The designer trades off the 'area split' against the size of payments.

5 Optimal mechanism

We now turn to the case where payments are at least somewhat wasteful, that is, $\gamma < 1$. I impose the following technical assumption on the space of admissible mechanisms:

Assumption 1. *The designer is restricted to allocation rules for time, $t : [0,1]^2 \to \mathbb{R}_+$, that are piecewise continuously differentiable in each dimension of the type.*

Importantly, Assumption 1 does not require the allocations of wait-times to be continuous, but only piecewise continuous. The assumption guarantees that the boundary g separating types who join waitlists A and B is well-behaved.

Fact 1. *Let g be the boundary implemented by any mechanism satisfying Assumption 1. Then g is piecewise twice continuously differentiable.*

The proof is in the appendix. Under Assumption 1, the structure of the optimal mechanism is pinned down by the following result:

Theorem 1. *The optimal mechanism prices entry to only one waitlist and offers finitely many pay-to-skip options.*

While the model allows for offering a continuum of pay-to-skip options for each waitlist, Theorem 1 prescribes a relatively simple payment policy: a discrete menu of pay-to-skip options and a single price charged for joining only one of the waitlists; intuitively, the one for the overdemanded good.

The proof shows that, in optimum, the boundary cannot have strictly convex or concave regions, and so each smooth piece of the boundary has to be linear. Each kink in the boundary then corresponds to a different option in the pay-to-skip menu. The proof of Theorem 1 also yields the following corollary:

Corollary 1. *Suppose the boundary g^* associated with the optimal mechanism is differentiable. Then the optimal mechanism does not offer any pay-to-skip options and prices entry to only one waitlist.*[4]

Interestingly, simulations suggest that the optimal boundary may be smooth for a wide class of distributions. Indeed, the intuitive reason underlying the lack of strictly convex and concave parts of the boundary also applies in the case of 'kinked' convex and concave regions. However, formalizing this intuition for non-differentiable regions of the boundary presents technical difficulties. I conjecture that a reasonably permissive condition on the density f can be formulated under which the hypothesis of Corollary 1 is without loss.

6 Proof of Theorem 1

I break the proof up into steps that let me highlight the core intuitions, along with the features of the model that make it tractable despite multidimensional types. In the first step I

[4] Strictly speaking, pay-to-skip options can be assigned to a zero-measure set of types.

rewrite the planner's objective in terms of waitlist-specific indirect utilities U_A, U_B and the boundary g they implement. I then break the problem up into two stages: in the first stage, I consider all the U_A, U_B pairs that implement a particular boundary g and find the one that does so optimally. Then, in the second stage, I consider mechanisms optimally implementing different boundaries g and look for the optimal boundary g^*. I first demonstrate that the optimal boundary does not exclude any types, that is, almost every type gets either good A or good B in optimum. I then use optimal control tools to show that the optimal boundary g^* has to be piecewise linear.

Throughout the proof I will focus on piecewise twice continuously differentiable boundaries. Let $\mathcal{G}$ be the set of all boundaries g satisfying this additional requirement.

6.1 Objective in terms of indirect utilities

Let us first recast the planner's problem as one of choosing the optimal U_A and U_B. Recall, however, that the planner's objective did not feature agents' utilities per se, but rather the values (net of payments) recipients derived from the goods. Thus, the objective will feature U_A and U_B normalized by the agents' discounting terms $e^{-\rho \cdot t}$. We therefore need to re-express the discounting term in terms of U_A and U_B. Observe, however, that the discounting term is analogous to an allocation in a standard quasilinear screening problem, as it is the only quantity that multiplies the (relevant component) of the agent's type. Thus, by the envelope theorem, this discounting term is equal to U_A' or U_B'.

Lemma 3. *Suppose that under the optimal mechanism a positive mass of agents joins either waitlist. We can then characterize the designer's problem as follows. Choose convex, increasing and Lipschitz continuous $U_A, U_B : [0,1] \to \mathbb{R}_+$ satisfying $U_A(0) = U_B(0) = 0$ to maximize:*

$$\int_{\underline{a}}^{1} \left(a \cdot \gamma + (1-\gamma) \cdot \frac{U_A(a)}{U_A'(a)} \right) \cdot \Phi_A(a) da + \int_{g(\underline{a})}^{1} \left(b \cdot \gamma + (1-\gamma) \cdot \frac{U_B(b)}{U_B'(b)} \right) \cdot \Phi_B(b) db,$$

subject to the supply condition:

$$\int_{\underline{a}}^{1} \Phi_A(a) da \leq \mu_A, \quad \int_{g(\underline{a})}^{1} \Phi_B(b) db \leq \mu_B. \tag{S'}$$

*I will call pairs U_A, U_B satisfying the conditions of this problem **admissible**.*

Proof. We can rewrite the designer's objective as:

$$\int \mathbb{1}_{x(a,b)=A}(a - (1-\gamma) \cdot p(a,b)) \; + \; \mathbb{1}_{x(a,b)=B}(b - (1-\gamma) \cdot p(a,b)) \; \mathrm{d}F(a,b).$$

We aim to express $p(a,b)$ in terms of U_A, U_B. First, consider types with $x(a,b) = A$. For them:

$$U_A(a) = e^{-\rho \cdot t(a,b)}(a - p(a,b)).$$

Moreover, by the envelope theorem $U_A'(a) = e^{-\rho \cdot t(a,b)}$. Thus, for these types we have:

$$a - p(a,b) = \frac{U_A(a)}{U_A'(a)} \quad \Rightarrow \quad p(a,b) = a - \frac{U_A(a)}{U_A'(a)},$$

with an analogous equality holding for types taking B. Substituting in yields:

$$\int \mathbb{1}_{x(a,b)=A} \left(a \cdot \gamma + (1-\gamma) \cdot \frac{U_A(a)}{U_A'(a)} \right) + \mathbb{1}_{x(a,b)=B} \left(b \cdot \gamma + (1-\gamma) \cdot \frac{U_B(b)}{U_B'(b)} \right) \mathrm{d}F(a,b).$$

We can now use Lemma 2 to rewrite objective in terms of the implemented boundary g:

$$\int_{(a,b)>(\underline{a},\underline{b})} \mathbb{1}_{b<g(\min[a,\bar{a}])} \left(a \cdot \gamma + (1-\gamma) \cdot \frac{U_A(a)}{U_A'(a)} \right) + \mathbb{1}_{b>g(\min[a,\bar{a}])} \left(b \cdot \gamma + (1-\gamma) \cdot \frac{U_B(b)}{U_B'(b)} \right) \mathrm{d}F(a,b),$$

which we can further break up into two integrals in terms of a and b, respectively. Finally, we can rewrite these integrals in terms of the density f to obtain the form of the objective stated in the lemma.

Now, by Lemma 1, U_A, U_B are feasible if and only if they are convex, $U_A(0) = U_B(0) = 0$, $U_A'(a), U_B'(b) \in [0,1]$, and they satisfy the supply condition:

$$\int \mathbb{1}_{U_A(a)>U_B(b)} \mathrm{d}F(a,b) \leq \mu_A, \quad \int \mathbb{1}_{U_A(a)<U_B(b)} \mathrm{d}F(a,b) \leq \mu_B.$$

However, by Lemma 2, this last condition is equivalent to (S′). Moreover, the requirement that $U_A'(a), U_B'(b) \in [0,1]$ can without loss be replaced by U_A, U_B being increasing and Lipschitz continuous. The original requirement can then be ensured by rescaling U_A, U_B appropriately. Note that such rescaling does not affect boundary indifference (I) or $\frac{U_A}{U_A'}$ and $\frac{U_B}{U_B'}$, and thus does not affect the implemented boundary or the objective. □

6.2 Optimally implementing a fixed boundary

We now fix any piecewise twice continuously differentiable boundary $g \in \mathcal{G}$ and look for the waitlist-specific indirect utilities U_A, U_B that optimally implement it. Notice that since the supply constraint (S′) is expressed purely in terms of the boundary g, we need not worry about whether U_A and U_B implementing our fixed boundary satisfy it.

By examining the objective in Lemma 3, we can see that making U_A and U_B 'more convex' decreases the value of the objective. To see why, note that the objective includes the terms:

$$\frac{U_A(a)}{U_A'(a)}, \quad \frac{U_B(b)}{U_B'(b)},$$

with positive weights. When U_A and U_B are linear (as 'non-convex' as can be), these grow with a and b at the rate of 1. However, as soon as U_A and U_B become strictly convex on some interval, they are bound to grow slower: the value accrued on the interval is then divided by the function's highest slope on it. Thus, we will aim to find the 'least convex' U_A and U_B

that implement the fixed boundary g. Recall also that whenever U_A and U_B implement g, they satisfy the boundary indifference condition (I):

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

Differentiating it gives:

$$U_A'(a) = U_B'(g(a)) \cdot g'(a). \tag{DI}$$

Now, consider an interval on which the boundary g is convex. On this region, the 'least convex' we can possibly make $U_B(g(a))$ is to make it linear, and so make $U_B'(g(a))$ constant. This in turn means that $U_A'(a)$ has to be proportional to $g'(a)$ on that interval. Conversely, whenever g is concave on some interval, $U_A'(a)$ is constant on it and $U_B'(a)$ is inversely proportional to $g'(a)$ (Figure 6). This intuition underlies the following lemma:

Definition 2. *Let an open interval $\mathcal{I}$ be a **convex (concave) region** of g if g is convex (concave) on $\mathcal{I}$ but not on any larger open interval containing $\mathcal{I}$.*

Lemma 4. *Suppose payments are wasteful, so $\gamma < 1$. Fix any implementable boundary $g \in \mathcal{G}$. Then the U_A, U_B optimally implementing g are unique up to scaling and satisfy the following properties:*

1. *For every $a \in (\underline{a}, \overline{a})$, at least one of $U_A'(a)$ and $U_B'(g(a))$ exists.*
2. *On convex regions, $U_A'(a) \propto g'(a)$ and $U_B'(g(a))$ is constant.*
3. *On concave regions, $U_B'(g(a)) \propto 1/g'(a)$ and $U_A'(g(a))$ is constant.*
4. *Unless $\overline{a} = g(\overline{a}) = 1$, one of $U_A'(\overline{a})$ and $U_B'(g(\overline{a}))$ exists. If $\overline{a} < 1$, U_A' is constant on $(\overline{a}, 1]$. If $g(\overline{a}) < 1$, U_B' is constant on $(g(\overline{a}), 1]$.*

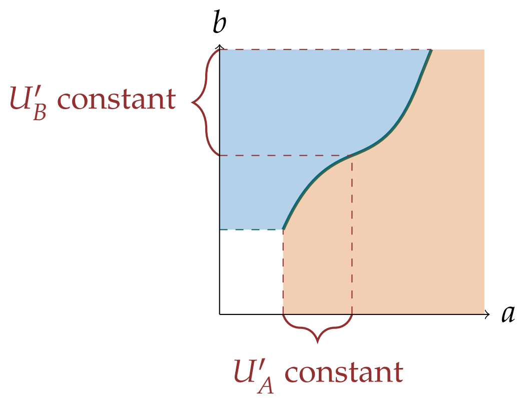

Figure 6: U_B' (U_A') is constant where the boundary g is convex (concave).

The motive for making U_A and U_B 'as non-convex as possible' has a corresponding economic intuition. Recall that in standard mechanism design problems, indirect utility is strictly convex on some region when agents with types in this region are separated, i.e. get distinct allocations. Recall also that in our problem, the 'allocation' corresponds to the discounting term $e^{-\rho \cdot t}$. Thus, whenever U_A is strictly convex on some region, agents who join waitlist A

and have values a in that region must all have different wait-times. Supporting this wait-time allocation, however, requires charging agents with lower wait-times more, which is wasteful. As mentioned, inducing such separation can still be useful to the extent that it serves to attract agents to join a specific waitlist (produce a 'better area split'). However, conditional on fixing the area split (which fixing the boundary effectively means) increasing separation, and thus payments, only generates waste.

Given that, consider the effects of creating separation for types on both sides of the same region of the boundary. Assigning different pay-to-skip options to agents on the A-side encourages additional high-a types to join this waitlist, as it now offers them larger information rents. Graphically, this bends the boundary upwards, making it convex there and expanding the region under it. Now, notice that having agents on the B-side of such a region also take different pay-to-skip options would generate the exact opposite effect and attract agents to waitlist B! Since offering separation requires increasing wasteful payments, this would never be optimal. The designer would then prefer to make the allocations on both sides 'less separating' in such a way that they would still implement the same boundary g.

Finally, note that the problem of finding the optimal U_A, U_B that implement a specific boundary is greatly simplified by the fact that the designer does not care about the assignment of wait-times. In a static model where the designer allocates amounts of the good (and has preferences over these amounts), she could find it beneficial to make the allocation on both sides of the boundary more separating, even if this had no effect on the implemented boundary. This is never the case in my model where, conditional on implementing the same boundary, 'more separating' allocations are always more wasteful.

6.3 Choosing the optimal boundary

Having pinned down the optimal way to implement a given boundary, we can turn to searching for the best boundary among all optimally implemented ones. Let g^* be optimal boundary and $\underline{a}^*, \underline{b}^*$ be lowest participating values associated with it. Suppose that g^* is piecewise twice continuously differentiable, that is, $g^* \in \mathcal{G}$. I first show that the optimal boundary extends all the way to the bottom of the type-space, and thus that the whole supply of both goods is allocated.

Lemma 5. *The optimal mechanism allocates the whole supply of both goods, that is, the supply constraint* (S) *binds for it.*

Equivalently, the result says that almost all types participate, that is, either $\underline{a}^* = 0$ or $\underline{b}^* = 0$. The intuition behind this is illustrated in Figure 7. Whenever a mechanism excludes some types, this is because it prices entry into both waitlists. But since the total supply of the good arriving each period is equal to the mass of arriving agents, we can always lower one of these prices to zero and allocate the unused supply to agents who would have otherwise not joined. Graphically, this corresponds to extending the boundary g downwards at an angle that makes supply constraints hold with equality. The proof is relegated to the appendix; it tackles the technical difficulties which arise when such a boundary extension is vertical.

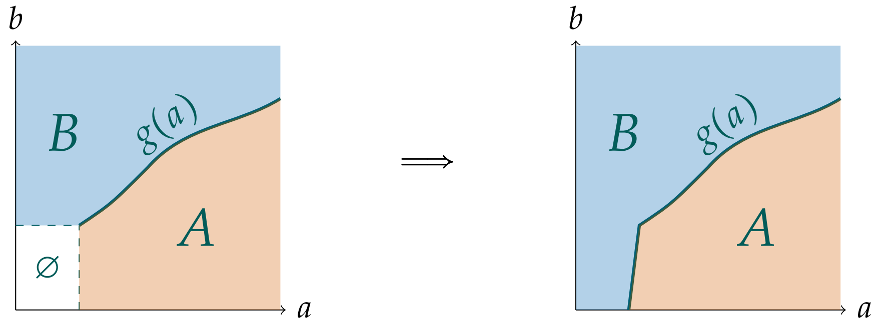

Figure 7: 'Extending the boundary' to allocate the unused supply.

Thus, the optimal mechanism does not price entry into both waitlists. Moreover, we can restrict our search for the optimal boundary to those that exclude no types. As it turns out, the optimal boundary among them has to be piecewise linear:

Proposition 3. *If the optimal boundary $g^* : [\underline{a}^*, \overline{a}^*] \to \mathbb{R}$ belongs to $\mathcal{G}$, it is piecewise linear.*

To prove the proposition, I set up an optimal control problem of selecting the optimal boundary on a part of a convex/concave region. I show that any boundary with strictly concave/convex parts cannot satisfy the necessary optimality conditions, and thus that the optimal boundary has to be piecewise linear. Intuitively, the simple structure of the solution stems from the normalization of waitlist-specific indirect utilities by their derivatives, U'_A and U'_B. This normalization introduces a convexity into the optimal control problem, which in turn leads to corner solutions.

It remains to show that a piecewise linear boundary g^* implies that finitely many pay-to-skip options are offered. Let U^*_A and U^*_B be the waitlist-specific indirect utilities that implement it. Then, by Lemma 4, U^*_A and U^*_B are also piecewise linear. Recall also that U_A and U_B are the indirect utilities of agents joining waitlists A and B, respectively. By the envelope theorem, we then have $e^{-\rho \cdot t(a,b)} = U'_A(a)$ for all types joining waitlist A, and $e^{-\rho \cdot t(a,b)} = U'_B(a)$ for all types joining waitlist B. Since U'_A and U'_B only take finitely many values, the set of assigned wait-times is also finite.

7 Conclusion

While payments are often deemed infeasible in contexts like transplant waitlists, programs allocating affordable housing, daycare slots, and medical procedures frequently involve some form of monetary transfers. This paper studies how to optimally integrate such transfers into waitlist mechanisms in order to enhance their allocative efficiency. It identifies a qualitative difference between screening with wait-times and prices: wait-times impose costs through delaying receipt, which scale with participants' valuation of the goods, eliciting only relative preferences. Prices, however, enable participants to express absolute valuations. Consequently, a designer who is indifferent about transfers between herself and participants would want to allocate the good through a pure price mechanism. However, such transfers reduce the recipients' surplus—this may conflict with the designers' motivations which led her to resort to using waitlists in the first place. I therefore solve for the optimal price-and-waitlist mechanism under the assumption that such transfers are undesirable. Despite considering a broad class of mechanisms allowing for continuously-changing price schemes, the structure of the optimal mechanism turns out to be much simpler: it involves a single fee for joining the oversubscribed waitlist while allocating the other good for free, and offers a discrete menu of pay-to-skip options. My model therefore suggests that limited pricing of waitlist sign-ups and wait reductions can help policymakers better target goods to high-value recipients while deterring low-value applicants.

8 Appendix: omitted proofs

Throughout, I will use $U_A'^-, U_B'^-$ and $U_A'^+, U_B'^+$ to denote the left- and right-derivatives of U_A and U_B. Since U_A, U_B are increasing and convex, their one-sided derivatives exist everywhere on $(0,1)$ with the right derivative being weakly greater than the left. Similarly, whenever $a > \underline{a}$ ($b > g(\underline{a})$), we have $U_A'^+(a) > 0$ ($U_B'^+(b) > 0$).

8.1 Proof of Lemma 1

($\Rightarrow$). Fix any indirect utilities U_A, U_B satisfying the lemma's conditions. I will show they are induced by the following mechanism, and that the mechanism is feasible:

$$e^{-\rho \cdot t(a,b)} = \begin{cases} 1, & \text{if } (a,b) \le (\underline{a}, \underline{b}), \\ U_A'^-(a), & \text{if } a > \underline{a} \text{ and } U_A(a) \ge U_B(b), \\ U_B'^-(b), & \text{if } b > g(\underline{a}) \text{ and } U_B(b) > U_A(a), \end{cases}$$

$$x(a,b) = \begin{cases} \varnothing, & \text{if } (a,b) \le (\underline{a}, \underline{b}), \\ A, & \text{if } a > \underline{a} \text{ and } U_A(a) \ge U_B(b), \\ B, & \text{if } b > g(\underline{a}) \text{ and } U_B(b) > U_A(a), \end{cases}$$

$$p(a,b) = \begin{cases} 0, & \text{if } (a,b) \le (\underline{a}, \underline{b}), \\ a - U_A(a) \cdot e^{\rho \cdot t(a,b)}, & \text{if } a > \underline{a} \text{ and } U_A(a) \ge U_B(b), \\ b - U_B(b) \cdot e^{\rho \cdot t(a,b)}, & \text{if } b > g(\underline{a}) \text{ and } U_B(b) > U_A(a). \end{cases}$$

A standard envelope and convexity argument verifies that no type (a,b) wants to misreport to (a',b') for which $x(a',b') = x(a,b)$, that is, conditional on joining the waitlist she was assigned to, (a,b) prefers her assigned allocation. This in turn implies that U_A and U_B indeed satisfy (1) and (2) for this mechanism. Thus, the mechanism indeed induces U_A and U_B.

It remains to show that the mechanism is feasible. Given the above, verifying that (IC) holds requires only checking that no (a,b) wants to misreport to (a',b') for which $x(a',b') \ne x(a,b)$. But since $U_A(a)$ and $U_B(b)$ are the best utilities she can get from either waitlist, this is true by the construction of $x(a,b)$.

Note also that (IR) must hold as type $(0,0)$ has indirect utility of 0 and U_A and U_B are increasing. It therefore remains to check the supply condition (S). However:

$$\int \mathbb{1}_{x(a,b)=A}\, \mathrm{d}F(a,b) = \int \mathbb{1}_{U_A(a)>U_A(b)}\, \mathrm{d}F(a,b) \le \mu_A,$$

where the inequality holds by condition 4 (an analogous expression holds for good B). Thus, the supply constraint (S) must hold.

($\Leftarrow$). Fix any feasible mechanism (p,t,x). Then U_A and U_B must be increasing by the standard single-crossing argument. Since any mechanism allows for non-participation, U_A, U_B must also be weakly positive. Furthermore, the envelope formula gives:

$$\text{for } (a,b):\ x(a,b) = A, \quad U_A'(a) = e^{-\rho \cdot t(a,b)},$$

$$\text{for } (a,b):\ x(a,b) = B, \quad U_B'(b) = e^{-\rho \cdot t(a,b)},$$

wherever these derivatives exist. Since $t(a,b) \ge 0$, it then follows that $U_A'(a), U_B'(0) \in [0,1]$. It therefore remains to show that 4. holds. However, $U_A(a) > U_B(b)$ for almost all agents for whom $x(a,b) = A$, so:

$$\int \mathbb{1}_{U_A(a)>U_A(b)}\, \mathrm{d}F(a,b) = \int \mathbb{1}_{x(a,b)=A}\, \mathrm{d}F(a,b) \le \mu_A,$$

where the inequality holds by (S). An analogous expression also holds for B.

8.2 Proof of Lemma 2

Consider first types $(a,b) < (\underline{a},\underline{b})$. By the definition of $(\underline{a},\underline{b})$, none of them can strictly benefit from joining either waitlist. Moreover, if some such (a,b) was indifferent about joining waitlist A or B, then some type $(a+\epsilon, b+\epsilon) < (\underline{a},\underline{b})$, for $\epsilon > 0$ small enough, would strictly prefer to join it. Thus, one of $U_A(a+\epsilon)$ and $U_B(b+\epsilon)$ would have to be strictly above zero. Since U_A, U_B are increasing by Lemma 1, this contradicts the definition of $(\underline{a},\underline{b})$. Thus, $x(a,b) = \varnothing$ for all $(a,b) < (\underline{a},\underline{b})$.

Analogously, all types for whom $a > \underline{a}$ or $b > \underline{b}$ strictly benefit from joining one of the waitlists. Moreover, note that $\underline{a},\underline{b} < 1$. This is because a positive mass of agents joins either waitlist and the set of agents for whom $a = \underline{a}$ or $b = \underline{b}$ is zero-mass.

Let us now identify the set of types $(a,b) \ge (\underline{a},\underline{b})$ who are indifferent between the two waitlists, that is, for whom $U_A(a) = U_B(b)$. This must be the case for $(\underline{a},\underline{b})$ by construction. Recall also that U_A, U_B are continuous and strictly increasing on $[\underline{a},1]$ and $[\underline{b},1]$, respectively. Therefore, all indifferent types must lie on a continuous and strictly increasing curve originating from $(\underline{a},\underline{b})$. Let $g(a) : [\underline{a},\overline{a}] \to \mathbb{R}$ describe this curve; notice that either $\overline{a} = 1$ or $g(\overline{a}) = 1$. Thus, any type $(a, g(a)) > (\underline{a},\underline{b})$ is indifferent between her best options in both waitlists but still strictly prefers to join either. Then, by the standard single-crossing argument, any type (a', b) with $a' > a$ strictly prefers to join waitlist A. Analogously, any type (a, b') with $b' > b$ strictly prefers to join waitlist B.

8.3 Proof of Fact 1

Let U_A and U_B be the waitlist-specific indirect utilities induced by the mechanism. By Lemma 2, types $(a, 0)$ where $a > \underline{a}$ join waitlist A and types $(0, b)$ where $b > g(\underline{a})$ join waitlist B. The envelope theorem then gives:

$$U_A(a) = \int_{\underline{a}}^{a} e^{-\rho \cdot t(v,0)} dv, \quad U_B(b) = \int_{g(\underline{a})}^{b} e^{-\rho \cdot t(0,v)} dv,$$

for $a \geq \underline{a}$ and $b \geq g(\underline{a})$. Since $t(a, b)$ is piecewise continuously differentiable in both variables, U_A and U_B are piecewise twice continuously differentiable.

Now, by Lemma 2, g satisfies the boundary indifference condition:

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

U_B is strictly increasing on $[g(\underline{a}), g(\overline{a})]$, so it is invertible there:

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

Since U_A and U_B were piecewise twice continuously differentiable on $[\underline{a}, \overline{a}]$ and $[g(\underline{a}), g(\overline{a})]$, respectively, g is piecewise twice continuously differentiable on $(\underline{a}, \overline{a})$.

8.4 Proof of Lemma 4

In what follows I prove properties 1,2 and 4; property 3 is symmetric to property 2. I then show that the U_A, U_B optimally implementing g are pinned down up to scaling.

Fact 2 (Property 1)**.** *For every $a \in (\underline{a}, \overline{a})$, at least one of $U'_A(a)$ and $U'_B(g(a))$ exists.*

Proof. Suppose there is $\hat{a} \in (\underline{a}, \overline{a})$ such that neither $U'_A(\hat{a})$ nor $U'_B(g(\hat{a}))$ exist; I will construct $\tilde{U}_A, \tilde{U}_B$ that implement g and are a strict improvement over U_A, U_B. Define:

$$\chi := \max\left[\frac{U'^{-}_A(\hat{a})}{U'^{+}_A(\hat{a})}, \frac{U'^{-}_B(g(\hat{a}))}{U'^{+}_B(g(\hat{a}))}\right]. \tag{4}$$

Since U_A and U_B are convex and not differentiable at $\hat{a}$ and $g(\hat{a})$, respectively, we have $\chi \in (0, 1)$. Assume without loss that $\frac{U'^{-}_B(g(\hat{a}))}{U'^{+}_B(g(\hat{a}))} \geq \frac{U'^{-}_A(\hat{a})}{U'^{+}_A(\hat{a})}$. Consider the proposed improvement:

$$\tilde{U}_A(a) = \begin{cases} U_A(a), & \text{if } a < \hat{a}, \\ U_A(\hat{a}) + \chi \cdot (U_A(a) - U_A(\hat{a})), & \text{if } a \geq \hat{a}, \end{cases}$$

$$\tilde{U}_B(b) = \begin{cases} U_B(b), & \text{if } b < g(\hat{a}), \\ U_B(g(\hat{a})) + \chi \cdot (U_B(b) - U_B(g(\hat{a}))), & \text{if } b \geq g(\hat{a}). \end{cases}$$

I will now verify that $\tilde{U}_A, \tilde{U}_B$ implement the boundary g, that they are admissible, and that they strictly improve upon U_A, U_B.

Implementing the boundary. It suffices to show that $\tilde{U}_A, \tilde{U}_B$ satisfy the boundary indifference condition:

$$\tilde{U}_A(a) = \tilde{U}_B(g(a)) \quad \text{for all} \quad a \in [\underline{a}, \overline{a}]. \tag{I'}$$

We know that U_A, U_B satisfy (I). Since $\tilde{U}_A, \tilde{U}_B$ coincide with U_A, U_B on $[\underline{a}, \hat{a}]$ and $[g(\underline{a}), g(\hat{a})]$, respectively, (I′) holds there too. For $a > \hat{a}$, we need to show:

$$U_A(\hat{a}) + \chi \cdot (U_A(a) - U_A(\hat{a})) = U_B(g(\hat{a})) + \chi \cdot (U_B(g(a)) - U_B(g(\hat{a}))), \tag{5}$$

but since $U_A(\hat{a}) = U_B(g(\hat{a}))$, (5) reduces to $U_A(a) = U_B(g(a))$, which holds by (I).

Admissibility. Since U_A, U_B implement boundary g, the supply condition (S′) must hold for this boundary, which means that it holds for $\tilde{U}_A, \tilde{U}_B$ too. $\tilde{U}_A$ and $\tilde{U}_B$ also trivially satisfy $\tilde{U}_A(0) = \tilde{U}_B(0) = 0$ as they agree with U_A, U_B in the neighborhood of 0. They also inherit the convexity of U_A, U_B on $[\underline{a}, \hat{a}]$, $(\hat{a}, 1]$ and $[g(\underline{a}), g(\hat{a})]$, $(g(\hat{a}), 1]$, respectively. We must therefore only check convexity where these intervals meet. First consider $\tilde{U}_B$; it suffices to show that:

$$\tilde{U}_B'^{+}(g(\hat{a})) \geq \tilde{U}_B'^{-}(g(\hat{a})).$$

Note, however, that χ removed the 'wedge' between the right and left derivatives of U_B at $g(\hat{a})$, and thus $\tilde{U}_B'^{-}(g(\hat{a})) = U_B'^{-}(g(\hat{a})) = \tilde{U}_B'^{+}(g(\hat{a})) = \tilde{U}_B'(g(\hat{a}))$.

Let us then consider the convexity of $\tilde{U}_A$ at $\hat{a}$. Similarly, it suffices to show that:

$$\tilde{U}_A'^{+}(\hat{a}) \geq \tilde{U}_A'^{-}(\hat{a}).$$

However:

$$\tilde{U}_A'^{+}(\hat{a}) = U_A'^{+}(\hat{a}) \cdot \frac{U_B'^{-}(g(\hat{a}))}{U_B'^{+}(g(\hat{a}))} \geq U_A'^{+}(\hat{a}) \cdot \frac{U_A'^{-}(\hat{a})}{U_A'^{+}(\hat{a})} = U_A'^{-}(\hat{a}) = \tilde{U}_A'^{-}(\hat{a}),$$

where the inequality holds by (4).

Recall also that U_A, U_B were increasing. Since $\tilde{U}_A, \tilde{U}_B$ agree with them on some neighborhood around 0 and are convex, they have to be increasing too. Finally, $\tilde{U}_A, \tilde{U}_B$ also inherit the Lipschitz continuity of U_A, U_B.

Improvement. I will now show that wherever U_A, U_B are differentiable, we have:

$$\frac{\tilde{U}_A(a)}{\tilde{U}_A'(a)} \geq \frac{U_A(a)}{U_A'(a)}, \quad \frac{\tilde{U}_B(g(a))}{\tilde{U}_B'(g(a))} \geq \frac{U_B(g(a))}{U_B'(g(a))}, \tag{6}$$

where the inequality is strict for $a > \hat{a}$. For $a \le \hat{a}$, (6) holds with equality. For $a > \hat{a}$:

$$\begin{aligned}
\frac{\tilde{U}_A(a)}{\tilde{U}'_A(a)} &= \frac{\int_{\underline{a}}^{\hat{a}} \tilde{U}'_A(v)\,dv}{\tilde{U}'_A(a)} + \frac{\int_{\hat{a}}^{a} \tilde{U}'_A(v)\,dv}{\tilde{U}'_A(a)} \\
&= \frac{\int_{\underline{a}}^{\hat{a}} U'_A(v)\,dv}{\chi \cdot U'_A(a)} + \frac{\int_{\hat{a}}^{a} U'_A(v)\,dv}{U'_A(a)} \\
&> \frac{\int_{\underline{a}}^{\hat{a}} U'_A(v)\,dv}{U'_A(a)} + \frac{\int_{\hat{a}}^{a} U'_A(v)\,dv}{U'_A(a)},
\end{aligned}$$

where the inequality holds since $\chi \in (0,1)$ and $U'_A(a) > 0$ for $a > \underline{a}$. An analogous inequality holds for U_B. Since $\frac{U_A}{U'_A}$ and $\frac{U_B}{U'_B}$ appear in the objective from Lemma 3 with strictly positive weights, $\tilde{U}_A, \tilde{U}_B$ are indeed a strict improvement over U_A, U_B. □

Fact 3 (Property 2)**.** *On convex regions, $U'_A(a) \propto g'(a)$ and $U'_B(g(a))$ is constant.*

Proof. Consider any interval $[\underline{v}, \overline{v}] \subset \mathcal{I}$ for some convex region $\mathcal{I}$ such that U_B is differentiable at $g(\underline{v})$ and $g(\overline{v})$. Define $\xi : [g(\underline{v}), g(\overline{v})] \to \mathbb{R}$:

$$\xi(v) = \frac{U'_B(g(\underline{v}))}{U'_B(v)}.$$

Recall that $U'_B(v) > 0$ for $v > g(\underline{a})$ and that U_B is convex on $[g(\underline{v}), g(\overline{v})]$. This tells us ξ is decreasing and takes values in $(0,1]$. I will now propose feasible $\tilde{U}_A, \tilde{U}_B$ that improve upon U_A, U_B. I will show that the improvement is strict whenever the statement of the fact does not hold for U_A and U_B. Let $\tilde{U}_A, \tilde{U}_B : [0,1] \to \mathbb{R}$ be continuous functions such that $\tilde{U}_A(0) = \tilde{U}_B(0) = 0$ and:

$$\tilde{U}'_A(a) = \begin{cases} U'_A(a) & \text{if } a \le \underline{v}, \\ U'_A(a) \cdot \xi(g(a)) & \text{if } \underline{v} < a \le \overline{v}, \\ U'_A(a) \cdot \xi(g(\overline{v})) & \text{if } a > \overline{v}, \end{cases} \quad \tilde{U}'_B(b) = \begin{cases} U'_B(b) & \text{if } b \le g(\underline{v}), \\ U'_B(b) \cdot \xi(b) & \text{if } g(\underline{v}) < b \le g(\overline{v}), \\ U'_B(b) \cdot \xi(g(\overline{v})) & \text{if } b > g(\overline{v}). \end{cases}$$

Since U'_A, U'_B exist at all but countably many points, these conditions pin down the values of $\tilde{U}_A$ and $\tilde{U}_B$ everywhere. The remainder of the proof consists of showing that $\tilde{U}_A, \tilde{U}_B$ implement the boundary g, that they are admissible, and that they improve upon U_A, U_B.

Implementing the boundary. It suffices to show that boundary indifference (I) holds for $\tilde{U}_A, \tilde{U}_B$. Recall that (I) holds for U_A, U_B. Since $\tilde{U}_A, \tilde{U}_B$ agree with U_A, U_B until $\underline{v}$, we have:

$$\tilde{U}_A(a) = \tilde{U}_B(g(a)) \quad \text{for all} \quad a \in [\underline{a}, \underline{v}]. \tag{7}$$

We thus only need to show (I) for $\tilde{U}_A, \tilde{U}_B$ on $(\underline{v}, \overline{a}]$. Note that differentiating (I) for U_A, U_B gives the following equation on $[\underline{v}, \overline{v}]$, except at countably many points:

$$U'_A(a) = U'_B(g(a)) \cdot g'(a). \tag{DI}$$

Multiplying both sides of (DI) for U_A, U_B by $\xi(g(a))$ gives:

$$\underbrace{U_A'(a)\cdot\xi(g(a))}_{=\tilde{U}_A'(a)} = \underbrace{U_B'(g(a))\cdot\xi(g(a))}_{=\tilde{U}_B'(g(a))}\cdot g'(a),$$

so (DI) holds for $\tilde{U}_A, \tilde{U}_B$ on $[\underline{v}, \overline{v}]$, except at countably many points. The argument for (DI) on $(\overline{v}, \overline{a}]$ is analogous. Together, (7) and (DI) for $\tilde{U}_A, \tilde{U}_B$ tell us that for any $a \in [\underline{v}, \overline{a}]$:

$$\tilde{U}_A(\underline{v}) + \int_{\underline{v}}^{a} \tilde{U}_A'(v)dv \;=\; \tilde{U}_B(g(\underline{v})) + \int_{\underline{v}}^{a} \tilde{U}_B'(g(v))\cdot g'(v)dv,$$

which is equivalent to $\tilde{U}_A(a) = \tilde{U}_B(g(a))$ on that interval.

Admissibility. The proof is analogous to the previous proof of admissibility except for a part of the argument for convexity. Here $\tilde{U}_A, \tilde{U}_B$ inherit convexity on $[0, \underline{v})$, $(\overline{v}, 1]$ and $[0, g(\underline{v}))$, $(g(\overline{v}), 1]$, respectively. I will show that $\tilde{U}_A, \tilde{U}_B$ are convex on $(\underline{v}, \overline{v})$ and $(g(\underline{v}), g(\overline{v}))$ and verify convexity at the points between these intervals. First, note that $\tilde{U}_B$ is trivially convex on $(g(\underline{v}), g(\overline{v}))$ because it is constant there:

$$\tilde{U}_B'(g(a)) = U_B'(g(a))\cdot\xi(g(a)) = U'(g(\underline{v})).$$

Turning to $\tilde{U}_A$, recall that (DI) holds for $\tilde{U}_B$ and $\tilde{U}_A$ on $(\underline{v}, \overline{v})$ except at countably many points. Thus:

$$\tilde{U}_A'(a) = \tilde{U}_B'(g(a))\cdot g'(a) = U_B'(g(a))\cdot\frac{U_B'(g(\underline{v}))}{U_B'(g(a))}\cdot g'(a) = U_B'(g(\underline{v}))\cdot g'(a), \tag{8}$$

on that interval except at countably many points. Since g is convex on $[\underline{v}, \overline{v}]$, g' is increasing there. Consequently, $U_A'(a)$ is increasing there too and so U_A is convex on $[\underline{v}, \overline{v}]$. The remainder of the proof—verifying convexity of $\tilde{U}_A, \tilde{U}_B$ at $\underline{v}, \overline{v}$ and $g(\underline{v}), g(\overline{v})$, respectively—is analogous to the previous argument.

Improvement. Suppose the statement of the fact fails on $[\underline{v}, \overline{v}]$. Then $U_B'(g(v))$ cannot be constant on $(\underline{v}, \overline{v})$: if it were, (DI) would ensure the other part of the fact's statement. Thus, there exists some $\hat{v} \in (\underline{v}, \overline{v})$ such that $U_B'(g(\hat{v})) < U_B'(g(\overline{v}))$. Since U_B is convex, we then have:

$$\frac{\xi(g(v))}{\xi(g(\overline{v}))} = \frac{U_B'^{+}(g(\overline{v}))}{U_B'^{+}(g(v))} > 1 \quad \text{for all } v < \hat{v}. \tag{9}$$

Now, like before, it suffices to show that wherever U_A, U_B are differentiable, we have:

$$\frac{\tilde{U}_A(a)}{\tilde{U}_A(a)} \geq \frac{U_A(a)}{U_A(a)}, \quad \frac{\tilde{U}_B(g(a))}{\tilde{U}_B'(g(a))} \geq \frac{U_B(g(a))}{U_B'(g(a))}, \tag{10}$$

with the inequality holding strictly on a positive-mass set. For $a \leq \underline{v}$, (10) holds with equality.

Look at $a > \overline{v}$ (the argument for $a \in [\underline{v}, \overline{v}]$ is analogous); there:

$$\begin{aligned}
\frac{\tilde{U}_A(a)}{\tilde{U}'_A(a)} &= \frac{\int_{\underline{a}}^{\underline{v}} \tilde{U}'_A(v)\,dv}{\tilde{U}'_A(a)} + \frac{\int_{\underline{v}}^{\overline{v}} \tilde{U}'_A(v)\,dv}{\tilde{U}'_A(a)} + \frac{\int_{\overline{v}}^{a} \tilde{U}'_A(v)\,dv}{\tilde{U}'_A(a)} \\
&= \frac{\int_{\underline{a}}^{\underline{v}} \xi(g(\underline{v})) \cdot U'_A(v)\,dv}{\xi(g(\overline{v})) \cdot U'_A(a)} + \frac{\int_{\underline{v}}^{\overline{v}} \xi(g(v)) \cdot U'_A(v)\,dv}{\xi(g(\overline{v})) \cdot U'_A(a)} + \frac{\int_{\overline{v}}^{a} \xi(g(\overline{v})) \cdot U'_A(v)\,dv}{\xi(g(\overline{v})) \cdot U'_A(a)} \\
&= \frac{\int_{\underline{a}}^{\underline{v}} \frac{\xi(g(\underline{v}))}{\xi(g(\overline{v}))} \cdot U'_A(v)\,dv}{U'_A(a)} + \frac{\int_{\underline{v}}^{\overline{v}} \frac{\xi(g(v))}{\xi(g(\overline{v}))} U'_A(v)\,dv}{U'_A(a)} + \frac{\int_{\overline{v}}^{a} U'_A(v)\,dv}{U'_A(a)}.
\end{aligned}$$

The final term is identical to the corresponding one for U_A, U_B. However, by (9), the former two are strictly larger, implying that (10) holds strictly for $a > \hat{v}$. The argument for the B-part of (10) is analogous. Thus, $\tilde{U}_A, \tilde{U}_B$ are indeed a strict improvement over U_A, U_B whenever the statement of the fact fails on some $[\underline{v}, \overline{v}] \subset \mathcal{I}$ where $U_B(g(v))$ is differentiable at the end-points. But since U_B is convex and $\mathcal{I}$ is an open interval, $\mathcal{I}$ can be covered arbitrarily well by such $[\underline{v}, \overline{v}]$. Consequently, the statement of the fact has to hold on its entirety. □

Fact 4 (Property 4)**.** *Unless $\overline{a} = g(\overline{a}) = 1$, one of $U'_A(\overline{a})$ and $U'_B(g(\overline{a}))$ exists. If $g(\overline{a}) < 1$, U'_B is constant on $(g(\overline{a}), 1]$. If $\overline{a} < 1$, U'_A is constant on $(\overline{a}, 1]$.*

Proof. Consider the case where $\overline{a} = 1$ and $g(\overline{a}) < 1$ (the opposite case is analogous). I first show that if U_B not is differentiable on $[g(\overline{a}), 1)$, it can be improved upon. Suppose U'_B is not differentiable for $\hat{v} \in [g(\overline{a}), 1)$. Then define:

$$\chi := \frac{U_B'^{-}(\hat{v})}{U_B'^{+}(\hat{v})}, \quad \tilde{U}_B(b) = \begin{cases} U_B(b), & \text{if } b < \hat{v}, \\ U_B(b) + \chi \cdot (U_B(b) - U_B(\hat{v})), & \text{if } b \geq \hat{v}. \end{cases}$$

The arguments for why $\tilde{U}_B$ is feasible and improves upon U_B are analogous to those in the proof on Fact 2. I now show that if U_B is not constant on $[g(\overline{a}), 1)$, we can improve upon it. Define $\xi : (g(\overline{a}), 1) \to \mathbb{R}$ such that:

$$\xi(v) = \frac{U_B'^{-}(g(\overline{a}))}{U_B'^{+}(v)},$$

and construct a continuous $\tilde{U}_B$ such that $\tilde{U}_B(0) = 0$ and:

$$\tilde{U}'_B(b) = \begin{cases} U'_B(b) & \text{if } b \leq g(\overline{a}), \\ U'_B(b) \cdot \xi(b) & \text{if } b \in (g(\overline{a}), 1). \end{cases}$$

The arguments for why $\tilde{U}_B$ is feasible and improves upon U_B are analogous to those in the proof on Fact 3. □

It remains to show that the U_A and U_B optimally implementing g are unique up to scale. I first inductively show this holds for them on $[\underline{a}, \overline{a}]$ and $[g(\underline{a}), g(\overline{a})]$, respectively. The argument extends to $[\overline{a}, 1]$ and $[g(\overline{a}), 1]$ analogously.

Lemma 6. *Let U_A, U_B be waitlist-specific indirect utilities optimally implementing g. Then U_A and U_B restricted to $[\underline{a}, \overline{a}]$ and $[g(\underline{a}), g(\overline{a})]$, respectively, are unique up to scaling.*

Proof. It suffices to show that $U_A(a)$ and $U_B(g(a))$ are pinned down (up to scale) on convex/concave regions. Their values at the ends of these intervals will then be pinned down by continuity.

Let $\mathcal{I}_1$ be the first convex/concave region. Suppose without loss that it is convex; then, by property 2, $U_B'(g(a)) = k_1$ and $U_A'(a) = k_1 \cdot g'(a)$ on this region for some $k_1 > 0$. Thus, the optimal U_A, U_B are pinned down up to scale on the first convex/concave region. I now show the inductive step. Let $\mathcal{I}_n$ and $\mathcal{I}_{n+1}$ be the nth and $n+1$st convex/concave regions, and $a_{n/n+1}$ be the point between them. It suffices to show that $U_A(a)$ and $U_B(g(a))$ for $a \in \mathcal{I}_n$ pin down those values for $a \in \mathcal{I}_{n+1}$. To prove this, I will use the following fact:

Fact 5. *If $g_-'(a_{n/n+1}) \geq g_+'(a_{n/n+1})$, U_A is continuously differentiable in some neighborhood of $a_{n/n+1}$. If $g_-'(a_{n/n+1}) \leq g_+'(a_{n/n+1})$, U_B is continuously differentiable in some neighborhood of $g(a_{n/n+1})$.*

Proof. Fix any $v \in (\underline{a}, \overline{a})$. Since g is piecewise twice continuously differentiable, it is continuously differentiable in $(v-\epsilon, v)$ and $(v, v+\epsilon)$ for a sufficiently small $\epsilon > 0$. By Fact 3, this implies that U_A' (U_B') is also continuous in these neighborhoods. Then if U_A (U_B) is differentiable at v, it is continuously differentiable in some neighborhood of v.[5] It therefore suffices to show that $U_A'(a_{n/n+1})$ and $U_B'(g(a_{n/n+1}))$ exist in the respective cases.

By Fact 2, at least one of $U_A'(a_{n/n+1})$ and $U_B'(g(a_{n/n+1}))$ exist. Recall also that (DI) holds everywhere on $\mathcal{I}_{n+1}$ and $\mathcal{I}_{n+1}$ (except at countably many points):

$$U_A'(a) = U_B'(g(a)) \cdot g'(a). \tag{DI}$$

Consider three cases.

Case 1: $g_-'(a_{n/n+1}) > g_+'(a_{n/n+1})$. Suppose that $U_B'(g(a_{n/n+1}))$ exists. Then U_B' has to be continuous in some neighborhood of $g(a_{n/n+1})$. But since the g' has a discontinuous downwards jump at $a_{n/n+1}$ and (DI) holds on both $\mathcal{I}_n$ and $\mathcal{I}_{n+1}$, this means that U_A' also has a discontinuous downwards jump at $a_{n/n+1}$; contradiction. Thus, $U_A'(a_{n/n+1})$ has to exist.

Case 2: $g_-'(a_{n/n+1}) < g_+'(a_{n/n+1})$. Suppose that $U_A'(a_{n/n+1})$ exists. Then U_A' has to be continuous in some neighborhood of $a_{n/n+1}$. But since the g' has a discontinuous upwards jump at $a_{n/n+1}$ and (DI) holds on both $\mathcal{I}_n$ and $\mathcal{I}_{n+1}$, this means that U_B' also has a discontinuous downwards jump at $a_{n/n+1}$; contradiction. Thus, $U_B'(g(a_{n/n+1}))$ has to exist.

Case 3: $g_-'(a_{n/n+1}) = g_+'(a_{n/n+1})$. In this case, $U_A'(a_{n/n+1})$ exists if and only if $U_B'(g(a_{n/n+1}))$ does. To see why, suppose $U_B'(g(a_{n/n+1}))$ exists. Then U_B' has to be continuous in some neighborhood of $g(a_{n/n+1})$. Since U_A and U_B are convex and U_B is differentiable at $g(a_{n/n+1})$, (DI) has to then hold at $a_{n/n+1}$, which means that $U_A'(a_{n/n+1})$ exists. The reverse direction is symmetric. □

[5]If a convex function f is continuously differentiable on (a, b) and (b, c) and $f'(b)$ exists, f is continuously differentiable on (a, c).

Note that we can exploit the symmetry of the setting and assume without loss that $\mathcal{I}_n$ is a convex region. Then, by Fact 3, $U_B'(g(a)) = k_n$ and $U_A'(a) = k_n \cdot g'(a)$ on this region for some $k_n > 0$. Consider first the case when $\mathcal{I}_{n+1}$ is a convex region. Then we similarly have $U_B'(g(a)) = k_{n+1}$ and $U_A'(a) = k_{n+1} \cdot g'(a)$ on this region for $k_{n+1} > 0$. If $g_-'(a_{n/n+1}) \geq g_+'(a_{n/n+1})$, Fact 5 tells us that U_A' exists and is continuous for a in some neighborhood of $a_{n/n+1}$. Thus:

$$U_A'(a_{n/n+1}) = k_n \cdot g_-'(a_{n/n+1}) = k_{n+1} \cdot g_+'(a_{n/n+1}) \quad \implies \quad k_{n+1} = k_n \cdot \frac{g_-'(a_{n/n+1})}{g_+'(a_{n/n+1})},$$

which means k_{n+1} is pinned down by k_n. If $g_-'(a_{n/n+1}) \leq g_+'(a_{n/n+1})$, Fact 5 tells us that U_B' exists and is continuous in some neighborhood of $g(a_{n/n+1})$. Thus, we have:

$$U_B'(a_{n/n+1}) = k_n = k_{n+1},$$

so k_{n+1} is pinned down by k_n. The case when $\mathcal{I}_{n+1}$ is concave is analogous. □

8.5 Proof of Lemma 5

I first show that the optimal mechanism allocates a positive mass of both goods. I then show that it does not exclude any agents.

Fact 6. *Under the optimal mechanism, a positive mass of agents joins either waitlist.*

Proof. First, I show that a mechanism allocating nothing cannot be optimal. By (IR) such a mechanism has to give zero utility to everyone, and the value of the objective for it is zero. Consider then a mechanism which offers good A at price $1-\epsilon$ with a wait-time of zero. All agents for whom $a > 1-\epsilon$ prefer to take good A at this price and almost all of them get strictly positive utility from it. All types (a, b) for whom $a < 1-\epsilon$ prefer to get nothing. For ϵ small enough, the mass of agents in the former group is strictly below μ_A, so this mechanism is feasible. Moreover, the value of the objective for this mechanism is strictly positive.

Now, consider a mechanism allocating only one of the goods; assume without loss it is good A. Then, by single-crossing, all agents with $a > \underline{a}^*$ receive good A and all agents with $a < \underline{a}^*$ receive nothing. Since $\mu_A < 1$, the set of agents receiving nothing has to have positive mass, and so $\underline{a}^* \in (0,1)$. The value of the objective for this mechanism is then at most:

$$\int_{\underline{a}^*}^{1} (a \cdot \gamma + 1 - \gamma) \cdot \left(\int_0^1 f(a,v) dv \right) da. \tag{11}$$

I will now propose a superior mechanism that allocates both goods. Consider a mechanism that offers good A at the price of $\underline{a}^*$ with a wait-time of 1 and offers good B at the price $1-\epsilon$ with a wait-time of δ. Fix ϵ so that:

$$\int_{1-\epsilon}^{1} \int_0^{\underline{a}^*} f(z,v) dz\, dv = \kappa,$$

for some $\kappa \in (0, \mu_B/2)$. The waitlist-specific indirect utilities for this mechanism are given by:

$$U_A(a) = \max[0, e^{-\rho}(a - \underline{a}^*)], \quad U_B(b) = \max[0, e^{-\rho \cdot \delta}(b - 1 + \epsilon)].$$

Now, for any δ, all agents for whom a is sufficiently close to 0 and b is sufficiently close to 1 will take good B. Similarly, all those for whom a is sufficiently close to 1 and b is sufficiently close to 0 will take good A. Thus, a positive mass of agents receives each good. We can therefore apply Lemma 2. The boundary g then satisfies:

$$e^{-\rho}(a - \underline{a}^*) = e^{-\rho \cdot \delta}(g(a) - 1 + \epsilon) \quad \Longrightarrow \quad g(a) = e^{\rho(\delta-1)}(a - \underline{a}^*) + 1 - \epsilon.$$

The boundary g thus has a constant slope of $e^{\rho(\delta-1)}$ which escapes to ∞ as $\delta \to \infty$. It follows that the value of the objective for this mechanism approaches the following as $\delta \to \infty$:

$$\int_{\underline{a}^*}^{1} (a \cdot \gamma + 1 - \gamma) \cdot \left(\int_0^1 f(a, v) dv \right) da + \int_{1-\epsilon}^{1} (b \cdot \gamma + 1 - \gamma) \cdot \left(\int_0^{\underline{a}^*} f(v, b) dv \right) db.$$

Note that the latter term is strictly positive and the first one equals to (11). Thus, for δ sufficiently large, the proposed mechanism dominates any mechanism offering only good A. Moreover, the mechanism satisfies the supply constraint (S) for sufficiently high δ. This is because the mass of agents taking good A is smaller than under the single-good mechanism and the mass of agents taking good B approaches κ. $\square$

The above fact ensures that the optimal mechanism has a 'boundary structure' and so that Lemma 2 applies to it. I will now prove that at least one of the lowest participating values for the optimal mechanism, $\underline{a}^*$ and $\underline{b}^*$, has to be zero. This will in turn imply that almost all agents join either waitlist, and so that the mass of allocated goods has to be equal to 1. Thus, the supply constraint (S) has to hold with equality for both goods.

Fact 7. *Under the optimal mechanism, either $\underline{a}^* = 0$ or $\underline{b}^* = 0$.*

Proof. Consider any mechanism implementing a boundary $g \in \mathcal{G}$ with lowest participating values $\underline{a}, \underline{b} \in (0, 1)$. I will construct a mechanism that strictly improves upon it. Intuitively, the new mechanism will allocate the unused supply of the good by 'extending the boundary' of the previous one.

For $\epsilon > 0$ small enough, g is twice continuously differentiable on $(\underline{a}, \underline{a} + \epsilon]$. Moreover, by the boundedness and continuity of f, and the fact that the supply condition (S′) held for g, for $\epsilon > 0$ small enough we have a continuous $\delta(\epsilon) \in (0, \infty)$ such that:

$$\int_{\max[0,\ \underline{a} - g(\underline{a}+\epsilon) \cdot \delta(\epsilon)]}^{\underline{a}+\epsilon} \int_0^{g(\underline{a}+\epsilon) - \delta(\epsilon) \cdot (\underline{a}+\epsilon-v)} f(v, z) dz\, dv + \int_{\underline{a}+\epsilon}^{1} \Phi_A(v) dv = \mu_A.$$

The right-hand-side equals to the probability mass under the curve produced by modifying g through extending it downwards from $g(\underline{a} + \epsilon)$ at the slope of $\delta(\epsilon)$. The slope is chosen so that the area underneath the modified boundary matches μ_A.

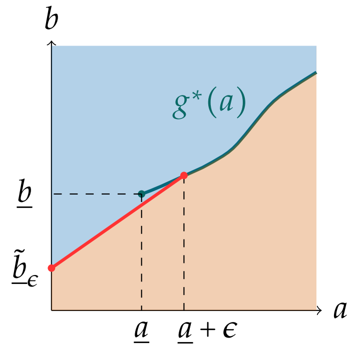

Figure 8: The slope $\delta(\epsilon)$ of the extension (red) was chosen so that the probability mass underneath the extended boundary (orange) equals to μ_A.

For such small enough ϵs, define the extended boundary as follows:

$$\tilde{g}_\epsilon(a) = \begin{cases} g(\underline{a}+\epsilon) - \delta(\epsilon)\cdot(\underline{a}+\epsilon-a) & \text{if } \max\left[0,\ \underline{a} - g(\underline{a}+\epsilon)\cdot\delta(\epsilon)\right] \le a < \underline{a}, \\ g(a) & \text{if } a \ge \underline{a}+\epsilon. \end{cases}$$

Let $\tilde{\underline{a}}_\epsilon, \tilde{\underline{b}}_\epsilon$ be the lowest participating values associated with the extended boundary. Note that for any ϵ, either $\tilde{\underline{a}}_\epsilon = 0$ or $\tilde{\underline{b}}_\epsilon = 0$, depending on whether the extension crosses the a or the b axis first. Moreover, since $\delta(\epsilon)$ is continuous, these lowest participating values will change continuously with ϵ.

Consider the case where $\lim_{\epsilon\to 0^*} \tilde{\underline{b}}_\epsilon > 0$.[6] Then $\tilde{\underline{a}}_\epsilon = 0$ and $\tilde{\underline{b}}_\epsilon = g(\underline{a}+\epsilon) - (\underline{a}+\epsilon)\cdot\delta(\epsilon)$ for all $\epsilon > 0$ small enough; restrict attention to such ϵs. Now, let U_A, U_B be the waitlist-specific indirect utility profiles optimally implementing g and define:

$$\tilde{U}_A(a) = \begin{cases} a\cdot\delta(\epsilon) & \text{if } a \le \underline{a}+\epsilon, \\ (\underline{a}+\epsilon)\cdot\delta(\epsilon) + U_A(a)\cdot c & \text{if } a > \underline{a}+\epsilon. \end{cases}, \quad \tilde{U}_B(b) = \begin{cases} 0 & \text{if } b < \tilde{\underline{b}}_\epsilon, \\ b - \tilde{\underline{b}}_\epsilon & \text{if } \tilde{\underline{b}}_\epsilon \le b \le \underline{b}, \\ \underline{b} - \tilde{\underline{b}}_\epsilon + U_B(b)\cdot c & \text{if } b > \underline{b}, \end{cases}$$

where $c \in \mathbb{R}_{++}$. I will show that for any c and $\epsilon > 0$ sufficiently small, $\tilde{U}_A, \tilde{U}_B$ are a strict improvement over U_A, U_B. I will later show that $\tilde{U}_A, \tilde{U}_B$ implement $\tilde{g}_\epsilon$ and that they are admissible for c large enough.

Improvement. for $a \in (0, \underline{a}+\epsilon]$ and $b \in (\tilde{\underline{b}}_\epsilon, g(\underline{a}+\epsilon)]$, we have:

$$\frac{\tilde{U}_A(a)}{\tilde{U}'_A(a)} = a, \quad \frac{\tilde{U}_B(b)}{\tilde{U}'_B(b)} = b - \tilde{\underline{b}}_\epsilon.$$

For any $a > \underline{a}+\epsilon$ and $b > g(\underline{a}+\epsilon)$:

$$\frac{\tilde{U}_A(a)}{\tilde{U}'_A(a)} = \frac{(\underline{a}+\epsilon)\cdot\delta(\epsilon)}{c\cdot U'_A(a)} + \frac{c\cdot U_A(a)}{c\cdot U'_A(a)} > \frac{U_A(a)}{U'_A(a)}, \tag{12}$$

[6] The case of $\lim_{\epsilon\to 0^*} \tilde{\underline{a}}_\epsilon > 0$ is symmetric; the case of $\lim_{\epsilon\to 0^*} \tilde{\underline{a}}_\epsilon = \lim_{\epsilon\to 0^*} \tilde{\underline{b}}_\epsilon = 0$ requires $\delta(0) \in (0,\infty)$ and can be tackled with a simpler version of the argument to follow in which $\tilde{g}_0$ is directly compared to g.

$$\frac{\tilde{U}_B(b)}{\tilde{U}_B'(b)} = \frac{\underline{b} - \tilde{\underline{b}}_\epsilon}{c \cdot U_B'(b)} + \frac{c \cdot U_B(b)}{c \cdot U_B'(b)} > \frac{U_B(b)}{U_B'(b)}. \tag{13}$$

where the inequalities hold because $\delta(\epsilon) \in (0, \infty)$ and $\underline{b} > \tilde{\underline{b}}_\epsilon$.[7] Let us now compare the values of the objective for U_A, U_B and $\tilde{U}_A, \tilde{U}_B$. For this purpose, define:

$$V_A^{[a_1,a_2]} := \int_{a_1}^{a_2} \left(a \cdot \gamma + (1-\gamma) \cdot \frac{U_A(a)}{U_A'(a)} \right) \cdot \Phi_A(a) da,$$

$$V_B^{[b_1,b_2]} := \int_{b_1}^{b_2} \left(b \cdot \gamma + (1-\gamma) \cdot \frac{U_B(b)}{U_B'(b)} \right) \cdot \Phi_B(b) db,$$

and let $\tilde{V}_A^{[a_1,a_2]}, \tilde{V}_B^{[b_1,b_2]}$ be the corresponding expressions for $\tilde{U}_A, \tilde{U}_B$. Notice that the difference between the objective values for $\tilde{U}_A, \tilde{U}_B$ and U_A, U_B is:

$$\tilde{V}_A^{[0,\bar{a}+\epsilon]} + \left(\tilde{V}_A^{[\bar{a}+\epsilon,1]} - V_A^{[\bar{a},1]} \right) + \tilde{V}_B^{[\tilde{\underline{b}}_\epsilon, g(\bar{a}+\epsilon)]} + \left(\tilde{V}_B^{[g(\bar{a}+\epsilon),1]} - V_B^{[g(\bar{a}),1]} \right).$$

We can use (12) and (13) to show that:

$$\lim_{\epsilon \to 0^+} \tilde{V}_A^{[\bar{a}+\epsilon,1]} \geq V_A^{[\bar{a},1]}, \quad \lim_{\epsilon \to 0^+} \tilde{V}_B^{[g(\bar{a}+\epsilon),1]} \geq V_A^{[\bar{a},1]}.$$

Therefore, to show that $\tilde{U}_A, \tilde{U}_B$ produces a strict improvement for some $\epsilon > 0$, it suffices to argue that $\tilde{V}_A^{[0,\bar{a}+\epsilon]} + \tilde{V}_B^{[\tilde{\underline{b}}_\epsilon, g(\bar{a}+\epsilon)]}$ are bounded away from zero as $\epsilon \to 0^+$. Indeed:

$$\lim_{\epsilon \to 0^+} \tilde{V}_A^{[0,\bar{a}+\epsilon]} \geq \int_0^{\underline{a}} (a \cdot \gamma + (1-\gamma) \cdot a) \cdot \left(\int_0^{\lim_{\epsilon \to 0^+} \tilde{\underline{b}}_\epsilon} f(a,v) dv \right) > 0.$$

Implementing the boundary. To show $\tilde{U}_A, \tilde{U}_B$ implement $\tilde{g}_\epsilon$ we must show that boundary indifference (I) holds for them. It holds on $[\underline{a}+\epsilon, 1]$ since $\tilde{U}_A(a), \tilde{U}_B(g(a))$ agree with $U_A(a), U_B(g(a))$ there. It thus remains to show it on $[0, \underline{a}+\epsilon)$. There:

$$\tilde{U}_A(a) = \tilde{U}_B(g(a)) \Leftrightarrow a \cdot \delta(\epsilon) = g(\underline{a}+\epsilon) - (\underline{a}+\epsilon-a) \cdot \delta(\epsilon) - \tilde{\underline{b}}_\epsilon,$$

which holds since $\tilde{\underline{b}}_\epsilon = g(\underline{a}+\epsilon) - (\underline{a}+\epsilon) \cdot \delta(\epsilon)$.

Admissibility. For any c, $\tilde{U}_A$ and $\tilde{U}_B$ are Lipschitz continuous. They also satisfy the first part of the supply constraint (S′) which holds with equality for $\tilde{g}_\epsilon$ by construction. Since $\mu_B = 1 - \mu_A$ and no type is excluded, the second part of the supply condition must hold with equality too. We must therefore only verify that for any sufficiently small $\epsilon > 0$, there exists $c \in \mathbb{R}_{++}$ such that $\tilde{U}_A, \tilde{U}_B$ are increasing and convex. As noted before, we focus on ϵ sufficiently small that $\delta(\epsilon) \in (0, \infty)$ and the case where $\underline{\tilde{a}} = 0$.

Note that $\tilde{U}_A, \tilde{U}_B$ are increasing and convex on $[0, \underline{a}+\epsilon]$, $(\underline{a}+\epsilon, 1]$ and $[0, g(\underline{a}+\epsilon)]$, $(g(\underline{a}+\epsilon), 1]$, respectively. It therefore suffices to check the point where these intervals meet. Since

[7] Lemma 4 ensures that $U_A'(\underline{a}+\epsilon), U_B'(g(\underline{a}+\epsilon))$ exist for $\epsilon > 0$ sufficiently small and are strictly positive.

$U_A'(\underline{a}+\epsilon), U_B'(g(\underline{a}+\epsilon)) > 0$, we can select $c > 0$ sufficiently large that:

$$\tilde{U}_A'^{-}(\underline{a}+\epsilon) = \delta(\epsilon) < c \cdot U_A'(\underline{a}+\epsilon), \quad \text{and} \quad \tilde{U}_B'^{-}(g(\underline{a}+\epsilon)) = 1 < c \cdot U_B'(g(\underline{a}+\epsilon)),$$

which guarantees $\tilde{U}_A, \tilde{U}_B$ to be convex and increasing on $[0,1]$. □

9 Appendix: proof of Proposition 3

Proposition 2 gives the result for the case of non-wasteful payments, $\gamma = 1$. I will therefore consider the case of $\gamma < 1$. I begin my showing that g^* has to solve a particular optimal control problem on some subsets of its domain.

9.1 Optimal control problem

Let $\mathcal{I}$ be a convex region of g^* and consider any $[\underline{v}, \overline{v}] \subset \mathcal{I}$ such that g^* differentiable in $[\underline{v}, \overline{v}]$. I will show that g^* restricted to $[\underline{v}, \overline{v}]$ must be the optimal g in the following control problem:

Problem 1. *Choose the control function $u : [\underline{v}, \overline{v}] \to \mathbb{R}_+$ and state functions $g, y, q : [\underline{v}, \overline{v}] \to \mathbb{R}$ to maximize:*

$$\int_{\underline{v}}^{\overline{v}} J(v, g(v), y(v)) dv, \tag{14}$$

where:

$$\begin{aligned} J(v, g(v), y(v)) := &\left(v \cdot \gamma + (1-\gamma) \cdot \frac{c_1 + g(v)}{y(v)}\right) \cdot \left(\int_0^{g(v)} f(v,z) dz\right) \\ &+ \left(g(v) + (1-\gamma) \cdot c_2\right) \cdot y(v) \cdot \left(\int_0^{v} f(z, g(v)) dz\right), \end{aligned}$$

where $c_1 + g^*(\underline{v}) \geq 0$, subject to the following laws of motion:

$$g'(v) = y(v), \quad y'(v) = u(v), \quad q'(v) = \int_0^{g(v)} f(v,z) dz,$$

and the following end-point constraints:

$$g(\underline{v}) = g^*(\underline{v}), \quad g(\overline{v}) = g^*(\overline{v}), \tag{15}$$

$$y(\underline{v}) = g_+^{*\prime}(\underline{v}), \quad y(\overline{v}) = g_-^{*\prime}(\overline{v}), \tag{16}$$

$$q(\underline{v}) = 0, \quad q(\overline{v}) = \int_{\underline{v}}^{\overline{v}} \int_0^{g^*(v)} f(v,z)\, dz\, dv. \tag{17}$$

The states g and y correspond to the boundary and its derivative, the control u corresponds to its second derivative, and q is introduced to capture the supply condition. The corresponding optimal control problem for concave regions is analogous.

Note that g^* restricted to $[\underline{v}, \overline{v}]$ is admissible in Problem 1. The proof thus consists of two steps. First, I show that for any g that is admissible in Problem 1, the following boundary is implementable in the original problem:

$$\tilde{g}(v) = \begin{cases} g(v) & \text{if } v \in [\underline{v}, \overline{v}], \\ g^*(v) & \text{elsewhere.} \end{cases}$$

Second, I show that the value of the objective from optimally implementing this boundary coincides with (14) up to a constant. Together, these two points imply that g^* on $[\underline{v}, \overline{v}]$ indeed has to be the optimal g in 1, as otherwise some superior boundary $\tilde{g}$ would exist.

Both steps of the proof rely on the following corollary which follows from a construction analogous to those in the proof of Lemma 4.

Corollary 2. *Let $g_1, g_2 : [\underline{a}, \overline{a}] \to \mathbb{R}$ be boundaries that satisfy the supply condition (S′) and belong to $\mathcal{G}$. Suppose $\mathcal{I}$ is a convex region for both g_1 and g_2 and let $[\underline{v}, \overline{v}]$ be its subset. Suppose further that g_1 and g_2 agree everywhere except $(\underline{v}, \overline{v})$, that is:*

$$g_1(a) = g_2(a) \quad \textit{for all } a \notin (\underline{v}, \overline{v}).$$

Then if there exist $U_{A,1}, U_{B,1}$ optimally implementing g_1, there also exist $U_{A,2}, U_{B,2}$ optimally implementing g_2. Moreover, $U_{A,1}$ and $U_{A,2}$ agree everywhere except $(\underline{v}, \overline{v})$ and $U_{B,1}$ and $U_{B,2}$ agree everywhere except $(g_1(\underline{v}), g_1(\overline{v}))$.

Intuitively, Corollary 2 says that perturbing a boundary inside an interval where it is convex does not affect the indirect utilities optimally implementing it outside this interval.

Consider any $g : [\underline{v}, \overline{v}] \to \mathbb{R}$ that is admissible in Problem 1 and let $\tilde{g}$ be its associated boundary. I will show Corollary 2 applies to $\tilde{g}$. This requires proving that:

1. g^* and $\tilde{g}$ agree outside of $(\underline{v}, \overline{v})$,
2. $\tilde{g}$ is a boundary,
3. $\tilde{g} \in \mathcal{G}$, that is, it is piecewise twice continuously differentiable,
4. g^* and $\tilde{g}$ are convex on $\mathcal{I}$,
5. the supply condition (S′) holds for $\tilde{g}$.

1. holds by the definition of $\tilde{g}$ and (15). Since the values of g^* at $\underline{a}$ and $\overline{a}$ are fixed, 2. requires showing that $\tilde{g}$ is continuous and strictly increasing. Note that Problem 1 does not allow for jumps in the state g, so it must be continuous on $(\underline{v}, \overline{v})$. It is also continuous with g^* at the end-points of this interval by (15). Now, note that $y(\underline{v})$ and the control u have to be positive; this ensures that any g is strictly increasing. 3. holds because the optimal control problem only admits absolutely continuous y. To see why 4. is satisfied, note first that g is convex on $[\underline{v}, \overline{v}]$ because the control, which corresponds to its second derivative, is positive on this interval. Moreover, (16) ensures that the 'pasting' of g and g^* at the ends of that interval preserve convexity on $\mathcal{I}$. Finally, to see why 5. holds, note that by Lemma 5, the supply

condition (S′) has to hold with equality for the optimal boundary g^*. Thus, we can without loss replace (S′) with:

$$\int_{\bar{a}}^{1} \Phi_A(v)dv = \mu_A. \tag{SE}$$

Since g^* is implementable, (SE) has to hold for it; consequently, (17) ensures that it holds for $\tilde{g}$ too.

Thus, Corollary 2 applies to g^* and $\tilde{g}$. Define U_A^*, U_B^* as the indirect utilities optimally implementing g^*. The corollary then tells us there exist some U_A, U_B that optimally implement $\tilde{g}$ and agree with U_A^*, U_B^* everywhere outside $(\underline{v}, \overline{v})$ and $(g^*(\underline{v}), g^*(\overline{v}))$, respectively.

Consequently, the objective depends on the state g only through two terms:

$$V_A[g] := \int_{\underline{v}}^{\overline{v}} \left(a \cdot \gamma + (1-\gamma) \cdot \frac{U_A(a)}{U_A'(a)}\right) \cdot \left(\int_0^{g(a)} f(a,z)dz\right) da,$$

$$V_B[g] := \int_{g^*(\underline{v})}^{g^*(\overline{v})} \left(b \cdot \gamma + (1-\gamma) \cdot \frac{U_B(b)}{U_B'(b)}\right) \cdot \left(\int_0^{g^{-1}(b)} f(z,b)dz\right) db.$$

Let us now pin down how $U_B(b)$ depends on $\tilde{g}$ for $b \in (g^*(\underline{v}), g^*(\overline{v}))$. Note that $(\underline{v}, \overline{v})$ is a subset of a convex region, so by Lemma 4 we know that $U_B'(b)$ is constant on $(g^*(\underline{v}), g^*(\overline{v}))$ and thus equals to $U_B^{*\prime+}(g^*(\underline{v}))$ on it. Since $U_B(g^*(\underline{v})) = U_B^*(g^*(\underline{v}))$, this means that $U_B(b) = U_B^*(b)$ for b in that interval. Hence, for any such b:

$$\frac{U_B(b)}{U_B'(b)} = \frac{U_B^*(g^*(\underline{v})) + U_B^{*\prime+}(b)(b - g^*(\underline{v}))}{U_B^{*\prime+}(b)} = c_2 + b,$$

where $c_2 = \frac{U_B^*(g(\underline{v}))}{U_B^{*\prime+}(b)} - g^*(\underline{v})$. Substituting into $V_B[g]$ yields:

$$\begin{aligned} V_B[g] &= \int_{g(\underline{v})}^{g(\overline{v})} \left(b \cdot \gamma + (1-\gamma) \cdot (c_2 + b)\right) \cdot \left(\int_0^{g^{-1}(v)} f(z,b)dz\right) db, \\ &= \int_{g^*(\underline{v})}^{g^*(\overline{v})} \left(b + (1-\gamma) \cdot c_2\right) \cdot \left(b \cdot \gamma + (1-\gamma) \cdot (c_2 + b)\right) \cdot \left(\int_0^{g^{-1}(v)} f(z,b)dz\right) db. \end{aligned}$$

By changing variable from b to $g(v)$ we obtain:

$$V_B[g] = \int_{\underline{v}}^{\overline{v}} \left(g(v) + (1-\gamma) \cdot c_2\right) g'(v) \cdot \left(\int_0^{v} f(z, g(v))dz\right) dv.$$

Let us now turn to the dependence of $U_A(a)$ on g for $a \in (\underline{v}, \overline{v})$. We know it satisfies (DI) at all but countably many points:

$$U_A'(a) = U_B'(g(a)) \cdot g'(a) = U_B^{*\prime+}(g(\underline{v})) \cdot g'(a).$$

We therefore have:

$$\begin{aligned} U_A(a) &= U_A^*(\underline{v}) + \int_{\underline{v}}^{a} U_A'(v)\,dv \\ &= U_A^*(\underline{v}) + \int_{\underline{v}}^{a} U_B^{*\prime+}(g^*(\underline{v}))\cdot g'(v)\,dv \\ &= U_A^*(\underline{v}) + U_B^{*\prime+}(g^*(\underline{v}))\cdot\big(g(a)-g^*(\underline{v})\big). \end{aligned}$$

Moreover:

$$\frac{U_A(a)}{U_A'(a)} = \frac{U_A^*(\underline{v}) + U_B^{*\prime+}(g^*(\underline{v}))\cdot(g(a)-g^*(\underline{v}))}{U_B^{*\prime+}(g^*(\underline{v}))\cdot g'(a)} = \frac{c_1+g(a)}{g'(a)},$$

where $c_1 = \frac{U_A^*(\underline{v})}{U_B^{*\prime+}(g^*(\underline{v}))} - g^*(\underline{v}) \geq -g^*(\underline{v})$. Substituting into $V_A[g]$ yields:

$$V_A[g] = \int_{\underline{v}}^{\overline{v}} \left(a\cdot\gamma + (1-\gamma)\cdot\frac{c_1+g(a)}{g'(a)}\right)\cdot\left(\int_0^{g(a)} f(a,z)dz\right)da.$$

This completes the proof.

9.2 Strict convexity/concavity of boundary

I now use the optimal control problem introduced above to show that the second derivative of the optimal boundary, $g^{*\prime\prime}(v)$, is zero wherever it exists. Since the optimal boundary g^* was piece-wise twice continuously differentiable, this will imply that it is piecewise linear.

Suppose towards a contradiction that $g^{*\prime\prime}(\hat{v}) > 0$ for some $\hat{v}$ (the case where $g^{*\prime\prime}(\hat{v}) < 0$ is analogous). Since g'' is piecewise continuous, there must be some interval $[\underline{v},\overline{v}]$ around $\hat{v}$ such that $g^{*\prime\prime}(v) > 0$ on it. Consider therefore Problem 1 for that interval. As shown, g^* restricted to $[\underline{v},\overline{v}]$ must be the optimal g for that problem. Let $(g^*, y^*, q^*, u^*, \xi, \phi, \eta)$ be optimal the collection of states, controls and costates associated with g^*. By the Maximum principle, $u^*(v) > 0$ must then maximize the following Hamiltonian for all $v \in (\underline{v},\overline{v})$:

$$\begin{aligned} \mathcal{H} = &\left(v\cdot\gamma + (1-\gamma)\cdot\frac{c_1+g^*(v)}{y^*(v)}\right)\cdot\left(\int_0^{g^*(v)} f(v,z)\,dz\right) \\ &+ \big(g^*(v) + (1-\gamma)\cdot c_2\big)\cdot y^*(v)\cdot\left(\int_0^{v} f(z,g^*(v))\,dz\right) \\ &+ \xi(v)\cdot y^*(v) + \phi(v)\cdot u + \eta(v)\cdot\left(\int_0^{g^*(v)} f(v,z)\,dz\right). \end{aligned}$$

Note that the Hamiltonian depends on the control linearly and so the optimal control can be interior only if $\phi(v) = 0$ on $(\underline{v},\overline{v})$. Since the control is singular on this interval, the optimal collection must satisfy the Kelley condition, which is a necessary condition for maximization (Robbins, 1967). It requires the following inequality to hold at the optimal collection:

$$\frac{\partial}{\partial u}\left(\frac{d^2}{dv^2}\mathcal{H}_u\right) \geq 0. \tag{18}$$

I will show the condition fails at the conjectured optimum. Notice that:

$$\frac{d^2}{dv^2}\mathcal{H}_u = \phi''(v).$$

Moreover, the Maximum Principle tells us that at the optimal state vector:

$$\phi'(v) = -\mathcal{H}_y = \frac{A(v)}{y^*(v)^2} - B(v) - \xi(v),$$

where:

$$A(v) := (1-\gamma)(c_1 + g^*(v))\left(\int_0^{g^*(v)} f(v,z)\,dz\right),$$
$$B(v) := \left(g^*(v) + (1-\gamma)\cdot c_2\right)\cdot\left(\int_0^{v} f(z,g^*(v))\,dz\right).$$

Consequently:

$$\phi''(v) = \frac{A'(v)}{y^*(v)^2} - 2\frac{A(v)}{y^*(v)^3}u - B'(v) - \xi'(v).$$

Direct computation confirms that $A'(v)$ and $B'(v)$ do not depend on u. Moreover, the Maximum Principle also tells us that at the optimal state vector:

$$\begin{aligned}\xi'(v) &= -\mathcal{H}_g \\ &= -\frac{1-\gamma}{y^*(v)}\cdot\left(\int_0^{g^*(v)} f(v,z)\,dz\right) - \left(\frac{c_1+g^*(v)}{y^*(v)} + \eta(v)\right)\cdot f(v,g^*(v)) \\ &\quad - y^*(v)\cdot\left(\int_0^{v} f(z,g^*(v))\,dz\right) - (g^*(v)+(1-\gamma)c_2)\cdot y^*(v)\cdot\left(\int_0^{v} f_2(z,g^*(v))\,dz\right),\end{aligned}$$

which does not depend on u either. Thus:

$$\frac{\partial}{\partial u}\left(\frac{d^2}{dv^2}\mathcal{H}_u\right) = \frac{\partial}{\partial u}\phi''(v) = -2\frac{A(v)}{y^*(v)^3} < 0,$$

which contradicts (18). To see why the last inequality holds, recall that g^* was strictly increasing on $(\underline{v},\overline{v})$, so $y^*(v) = g^{*\prime}(v) > 0$ and $g^*(v) > g^*(\underline{v})$. Recall also that $c_1 + g^*(\underline{v}) \geq 0$, so $c_1 + g^*(v) > 0$, giving $A(v) > 0$.

References

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

ASHLAGI, I., F. MONACHOU, AND A. NIKZAD (2024): "Optimal Allocation via Waitlists: Simplicity Through Information Design," *The Review of Economic Studies*, rdae013.

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.

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

BUDISH, E. (2011): "The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes," *Journal of Political Economy*, 119, 1061–1103.

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

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.

HYLLAND, A. AND R. ZECKHAUSER (1979): "The efficient allocation of individuals to positions," *Journal of Political Economy*, 87, 293–314.

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

MILGROM, P. AND I. SEGAL (2002): "Envelope theorems for arbitrary choice sets," *Econometrica*, 70, 583–601.

MYERSON, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

ROBBINS, H. M. (1967): "A generalized Legendre-Clebsch condition for the singular cases of optimal control," *IBM Journal of Research and Development*, 11, 361–372.

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