REDISTRIBUTION WITH PRICES, RATIONING, AND QUEUING

ANDREW CONKEY
Department of Computer Science, Stanford University

BENJAMIN WITTENBRINK
Department of Computer Science, Stanford University

## 1. INTRODUCTION

How does one optimally allocate a scarce resource? Many would answer to those who can best use it, or by need, but how to evaluate either of these? These fundamental questions have concerned humans for centuries and are at the heart of the political and the ethical. One of the crucial insights in economics is that the price system provides a simple and automatic mechanism to aggregate preferences and efficiently distribute a good based on who values it most as measured by willingness-to-pay. Moreover, for any different allocation, there always exists a mechanism combining the price system with a lump-sum transfer that can make everyone at least as well off. Nonetheless, many important societal resources, including but not limited to housing, health care, sustenance, and transportation, are often provided to the public at below-market rates and, occasionally, for free. Additionally, some items are distributed without a price system entirely, e.g., COVID-19 vaccines, but instead through a "queuing" procedure that prioritizes those deemed the most vulnerable and forces others to wait. Given the potential allocative distortions and inefficiencies that such policies have, why are they so commonplace?

Many policy-makers defend these policies on the basis of equity. Specifically, they argue that if vital societal resources were distributed solely based on market prices, individuals with the least ability to pay would be left out from benefiting from them. This poses an important ethical problem, particularly as those with the lowest willingness-to-pay are generally of lower socio-economic status and might actually be those in most need. In other words, monetary screening devices may not be efficient mechanisms in settings with sufficient inequality, as willingness-to-pay may not accurately capture true need.

In this paper, we study how to model this tradeoff between equity and efficiency. We study a setting with both market and non-market mechanisms at the disposal of the designer. Specifically, we wish to characterize the benefits in welfare and targeting associated with the introduction of queuing as a potential mechanism in addition to prices and rationing. While the tradeoffs associated with these non-market mechanisms have been studied extensively within the academic literature, this has always been done in isolation. Indeed, there is very little academic work focused on the multi-dimensional case with prices, queuing, and rationing.

To gain some intuition about the potential welfare benefits for introducing queuing as a mechanism to a setting with prices and a rationing options, we propose a simple example with two agents, 1 and 2, and only one good to allocate amongst them. We report their values for the good, money, and time in Table I. In particular, suppose that Agent 1 values both the good and money at 100 (we can think of the value for money as the utility cost of spending a dollar); Agent 2 has values equal to 1 for both money and the good. Clearly, their willingness to pay is identical (the ratio of their two values) and consequently the optimal mechanism would be to

Andrew Conkey: aconkey@stanford.edu
Benjamin Wittenbrink: witten@stanford.edu

ration between the two agents. In other words, both Agent 1 and 2 would individually receive 0.5 of the good. Thus, the overall welfare from this allocation would be 50.5.

Now suppose the agents also have a value for time, which is equal to one for both. Then, in the presence of queuing, we have the ability to offer the agents the following menu: burn $1+\epsilon$ units of time to receive 1 good. Note that agent 2 would not select this menu as it would give her negative utility $(1-(1+\epsilon)<0)$. In contrast, agent 1 would be delighted to select this option, as it yields a welfare equal to $100-(1+\epsilon)=99-\epsilon$. Moreover, this is a tremendous improvement over the case above without queuing, where we only able to achieve a total welfare of 50.5. While this is a simple example, it should provide the necessary intuition to understand the potential welfare benefits of queuing.

TABLE I

MOTIVATING EXAMPLE

| | Value for good | Value for money | Value for time |
|---|---|---|---|
| Agent 1 | 100 | 100 | 1 |
| Agent 2 | 1 | 1 | 1 |

This paper contributes to an extensive literature discussing tradeoffs between market and non-market mechanisms. In particular, previous work has studied the welfare implications of rationing and price controls (Weitzman, 1977, Condorelli, 2013, Reuter and Groh, 2020, Dworczak et al., 2021, Akbarpour et al., 2023a, Kang and Zheng, 2023) and of queuing and costly signals (Nichols et al., 1971, Nichols and Zeckhauser, 1982, McAfee and McMillan, 1992, Hartline and Roughgarden, 2008, Chakravarty and Kaplan, 2013, Rose, 2021, Zeckhauser, 2021, Dworczak, 2023). However, almost the entirety of this work has focused on these non-market mechanisms in isolation, whereas our focus is to understand the welfare impacts of employing both rationing and queuing.

We proceed as follows. Section 2 provides an overview of the related literature and Section 3 introduces a general model. We discuss our theoretical results in Section 4, specifically focusing on optimal mechanisms for a setting with two agent types. Finally, we discuss computational approaches and report these results in Section 5.

## 2. RELATED LITERATURE

### 2.1. *Rationing and redistributive markets*

How to appropriately allocate a commodity in the presence of scarcity is a fundamental question in economics. Economists generally favor the price system as the means for allocation, as it provides a simple and automatic mechanism to elicit preferences in order to match a scarce commodity with those who value it most. Moreover, for any other allocation, there always exists a combination of the price system and a lump-sum transfer that can make everyone at least as well off. However, as Weitzman (1977) notes

> That is true enough in principle, but not typically very useful for policy prescriptions, because the necessary compensation is practically never paid.

Indeed, Weitzman argues that this willingness-to-pay elicitation does not always accurately capture people's true needs, especially in the presence of broad inequality. In such settings, he shows that a free, random allocation can perform better than competitive pricing when considering overall welfare.

Condorelli (2013) characterized the optimality of market and non-market mechanisms for allocating identical goods in a setting where agent's willingness-to-pay does not perfectly capture the value they place on the item. Akbarpour et al. (2023a) extend this model by allowing the designer to have preferences over revenue to maximize a redistributive objective. Relatedly, Dworczak et al. (2021) study a more general market setting with buyers and sellers and show that in markets with significant same-side inequality, price controls may be optimal even though they induce rationing.

### 2.2. *Queuing / Ordeals / Costly Signals*

The potential utility of queuing as a tool to screen for targeted buyers has been discussed in theory work since Nichols et al. (1971) and Nichols and Zeckhauser (1982). Hartline and Roughgarden (2008) consider what they call "money burning" mechanisms, but analyze costly signaling in isolation, without also allowing for monetary transfers or considering redistributive concerns. Chakravarty and Kaplan (2013) also investigate optimal allocation with costly signals, but again do not concurrently allow for monetary transfers. Recent work, such as Dworczak (2023), has made some progress in characterizing optimal costly-signaling allocation mechanisms, but generally continues to limit the tools available to the mechanism designer to reduce the dimensionality of the problem. Others, like Rose (2021) and Zeckhauser (2021) have instead focused on equity and other ethical concerns in using queuing to allocate critical goods and services such as healthcare. The intention of this work is not to suggest that abstracting away from the more complex burdens of queuing, such as those discussed in Bertrand et al. (2004), is appropriate for all contexts.

On the empirical side, a small body of work has established the usefulness of queuing and ordeals for targeting in practical contexts. Alatas et al. (2016), for example, demonstrate that requiring beneficiaries of a transfer program in Indonesia to apply for benefit rather than enrolling them automatically results in poorer beneficiaries enrolling on average. Similarly, Dupas et al. (2016) observe substantially higher use rates of chlorine solution among recipients in Kenya when distributed via vouchers that must be redeemed monthly rather than through a free distribution program. Finkelstein and Notowidigdo (2019), meanwhile, observe that the SNAP-eligible elders most responsive to a set of interventions to reduce enrollment frictions are on average higher-income and less sick than those that enroll in the absence of intervention. In contrast to these works, Deshpande and Li (2019) observe a decrease in the targeting efficiency of disability benefits following the closure of Social Security Administration field offices (thus increasing the non-monetary costs of application via travel costs). However, this result is likely largely explained by the intuition (which we partially formalize) that non-monetary ordeals which are disproportionately burdensome on target recipients often cannot be used to improve targeting.

### 2.3. *Multidimensional mechanisms*

Recent work has begun to make progress on costly screening in multidimensional settings. Yang (2022) provides a necessary condition for the use of costly screening in a principal-agent setting where the principal has access to both productive (including monetary) and costly screening tools. Patel and Urgun (2022) characterize the optimal mechanism in a principal-agent setting without monetary transfers, but where the principal can make use of both costly screening and costly verification. We seek to differentiate from this work primarily by focusing on mechanisms that are optimal from a welfare-maximizing perspective in a setting with many agents, and by linking costly signaling to monetary inequality and other explicit redistributive preferences. Most recently, Akbarpour et al. (2023b) make progress in this area by providing conditions for the dominance of one costly signaling device over another, also in a setting involving the welfare-maximizing allocation of a set of goods.

## 3. MODEL

*Framework.* We study a setting where a designer allocates a mass of homogeneous goods of size $Q$ to a unit mass of agents distributed according to density function $f(v)$ characterized by a type vector $(v_x, v_m, v_t)$. We denote the overall type space by $V$. The three dimensions of agents' type vectors have a joint distribution in the population that is known to the designer. The first component of an agent's type is $v_x$, which denotes their value for receiving a good. Similarly, $v_m$ and $v_t$ represent the agent's value for money and for time, respectively. The utility an agent of type $(v_x, v_m, v_t)$ receives from an allocation of $x$ at a price of $p$ and a queuing cost $t$ is given additively as $v_x x - v_m p - v_t t$.

As a brief remark, this framework can be adapted to a willingness-to-pay and welfare weight formulation (as discussed in Dworczak et al. (2021) and Akbarpour et al. (2023a)). To do so, we simply re-parameterize our type space as $(r, \lambda, c)$ where $r \equiv v_x \big/ v_m$ denotes the agent's willingness-to-pay, $c \equiv v_t \big/ v_m$ represents the agent's willingness-to-queue, and $\lambda \equiv v_m$ is the social welfare weight. In this setting, the contribution of the agent's utility to the social welfare function is $\lambda(rx - p - ct)$. We apply this framework in some later sections of this paper.

*Mechanism.* A mechanism $m$ consists of an allocation function $x : V \to [0,1]$, a price function $p : V \to \mathbb{R}_+$, and a non-pecuniary time cost function $t : V \to \mathbb{R}_+$. Specifically, we restrict agents to at most unit demand, and impose the restriction that prices and time-costs be non-negative. The allocation function can be interpreted as yielding either fractional allocations of a divisible good or lottery probabilities for an indivisible good.

*Objective.* In our setting, the mechanism designer wishes to maximize social welfare. Additionally, the designer values revenue with weight $\alpha$, and any time cost incurred is wasted. Thus they maximize

$$\int \left(v_x x(v) - v_m p(v) - v_t t(v)\right) f(v) dv + \alpha \int p(v) f(v) dv \tag{1}$$

subject to individual rationality, incentive compatibility, and feasibility constraints. Some possible values of $\alpha$ have natural interpretations. For example, we can capture a setting in which the designer distributes all revenue raised using equal lump sum transfers for each agent with $\alpha = \int v_m f(v) dv$. We might also take $\alpha$ as the maximum value of the $v_m$, to reflect a setting in which the designer can target all revenue to the "poorest" agent. An $\alpha$ greater than all $v_m$ might indicate that the mechanism designer instead has some preferable outside use in mind for revenue.

## 4. THEORETICAL RESULTS

We begin our analysis of the model by restricting to a setting with only two types of agents. We assume that these agents are present in equal proportions (a unit mass of each) and that the total quantity of the good is $Q = 1$. Let $\theta_0$ represent the first type and $\theta_1$ the second. We will use the re-paramaterized type space discussed above, such that agents of type $\theta_i$ receive utility $\lambda_i(r_i x - p - c_i t)$ from burning $t$, spending $p$, and obtaining allocation $x$. To simplify further, we normalize the parameters with respect to type $\theta_0$'s values:

$$u_0(x_0, p_0, t_0) = x_0 - p_0 - t_0;$$
$$u_1(x_1, p_1, t_1) = \lambda(r x_1 - p_1 - c t_1).$$

### 4.1. *Optimal price mechanism without queuing*

We begin by examining the form of the optimal mechanism without queuing (that is, the optimal mechanism when we add an additional constraint that $t_0 = t_1 = 0$).

REMARK 1: Provided that $\frac{1+\lambda}{2} \leq \alpha \leq \lambda$ and $r \neq 1$, it is optimal when queuing is not permitted to sell at $p = 1$ to the type with the higher willingness-to-pay.

Intuitively, this result says that on this range $\alpha$ is small enough such that designer prefers not to extract maximum revenue from type 1 when $r > 1$ but is large enough that the revenue loss from rationing outweighs the targeting gain when $r < 1$. In this setting, the $r = 1$ case must lead to rationing for both types at the same price, since then they have an identical willingness to pay and it is not possible to ration to one type but not the other (when $Q = 1$). Given the constraint on $\alpha$, this rationing occurs at $p = 1/2$.

PROOF: Our objective function is now given by $x_0 - p_0 + \lambda(rx_1 - p_1) + \alpha(p_o + p_1)$. There are two cases we need to analyze: that where $r > 1$ and where $r < 1$.

First, suppose that $r > 1$. That is, $\theta_1$ has a higher willingness-to-pay in monetary terms than $\theta_0$. From the linearity of the objective function, there are three relevant options: we can sell the good to $\theta_1$ at $p = 1$, sell the good to $\theta_1$ at $p = r$, or we can ration the good between both types for some $p \leq 1/2$. Plugging in the values, we see that the first option generates a welfare of $\lambda(r-1) + \alpha$ and the second of $\alpha r$. Because $\alpha \leq \lambda$, the former always exceeds the latter. For the rationing case, we must solve the maximization

$$\max_{p \in [0,1/2]} \frac{\lambda r + 1}{2} + p(2\alpha - \lambda - 1).$$

We have $\alpha \geq \frac{1+\lambda}{2}$, and therefore $2\alpha - \lambda - 1 \geq 0$, so this quantity is maximized at $p = 1/2$. Thus the maximum welfare that can be obtained from rationing is $\frac{\lambda(r-1)}{2} + \alpha$. This is always strictly outperformed by selling at $p = 1$, which we demonstrated obtains welfare of $\lambda(r-1) + \alpha$. Consistent with the desired result, selling at $p = 1$ to $\theta_1$ is optimal.

Now consider the case when $r < 1$, that is that $\theta_0$ has a higher willingness-to-pay than $\theta_1$. Once again, from the linearity of the objective function, there are three relevant options: we can sell the good to $\theta_0$ at $p = r$, sell the good to $\theta_0$ at $p = 1$, or we can ration the good between both at $p \leq r/2$. Plugging in these values, we see that the first option generates welfare of $1 - r + \alpha r$ and the second of $\alpha$. Since $r > 1$ and $\alpha > 1$, the second option is preferable to the first. For the rationing case, we maximize the same objective function as in the first rationing case, this time over $p \in [0, r/2]$. By the same logic using our bound on $\alpha$, the maximum is found at $p = r/2$, yielding welfare of $\frac{1-r}{2} + \alpha r$. But this is strictly worse that selling the good at $p = r$, to obtain welfare of $1 - r + \alpha r$. Overall, then, selling at $p = 1$ to $\theta_0$ is optimal, again consistent with the desired result.

*Q.E.D.*

### 4.2. *Optimal mechanism with queuing*

We now move to analyzing the form of the optimal mechanism when queuing is permitted.

PROPOSITION 1: *Provided $\frac{1+\lambda}{2} \leq \alpha \leq \lambda$ and $r \neq 1$, the optimal mechanism uses queuing at a price of 0 to allocate the good to the target type when $c < \min\{1, r\} - \frac{\alpha}{\lambda}$. Otherwise the good is sold to the type with higher willingness-to-pay at $p = 1$ (without queuing).*

PROOF: Here we return to examining the full objective function,

$$x_0 - p_0 - t_0 + \lambda(rx_1 - p_1 - ct_1) + \alpha(p_0 + p_1).$$

We begin by considering optimal pure allocation mechanisms. Since $Q = 1$, this implies only a single non-zero quantity allocation. Denote by $t$ the queuing time associated with this allocation. We can conceptualize this as a transformation of $\theta_1$'s willingness-to-pay from $r$ to $r - ct$ and of $\theta_0$'s willingness-to-pay from 1 to $1 - t$, allowing us to apply much of our logic from the previous proof. We consider the optimal mechanism fixing a given $t$ and then optimize over values of $t$ to find the overall optimal pure-allocation mechanism.

Parallel to the first of our cases from the previous proof, we first consider when $r - ct > 1 - t$. Appealing to our previous logic and result, when this condition holds the optimal price is $p = 1 - t$, yielding welfare of $\lambda(r - ct - 1 + t) + \alpha(1 - t)$. Similarly, when $r - ct < 1 - t$, the optimal price is again $p = 1 - t$, yielding welfare of $\alpha(1 - t)$.

Given this, we move now to look at the optimal choice of $t$. Suppose that $r > 1$. If $c \leq 1$, we always have $r - ct > 1 - t$ for $t > 0$. Then we optimize $\lambda(r - ct - 1 + t) + \alpha(1 - t)$, so the optimal pure-allocation mechanism uses queuing with $t = 1$ (and $p = 0$) if $c < 1 - \frac{\alpha}{\lambda}$ and sells at $p = 1$ (with $t = 0$) otherwise. Alternatively, if $c > 1$, then minimizing $t$ is optimal both when $r - ct > 1 - t$ and when $r - ct < 1 - t$. Therefore, the optimal pure-allocation mechanism sells at $p = 1$ (with $t = 0$). We can combine these two cases to state that that when $r > 1$, the optimal pure-allocation mechanism uses queuing at a price of 0 to allocate the good to the target type when $c < 1 - \frac{\alpha}{\lambda}$, consistent with the desired result. For other values of $c$ when $r > 1$, the good is sold to $\theta_1$ at $p = 1$, also consistent with the desired result.

Now instead suppose that $r < 1$. If $c \geq 1$, we always have that $r - ct < 1 - t$, so $t = 0$ and $p = 1$ is optimal. If $c < 1$, with $t = 0$ and $p = 1$ we obtain welfare of $\alpha$, while with $t = 1$ and $p = 0$ we obtain welfare of $\lambda(r - c) + \alpha$. The optimal pure-allocation mechanism thus uses queuing with $t = 1$ (and $p = 0$) if $c < r - \frac{\alpha}{\lambda}$ and sells at $p = 1$ (with $t = 0$) otherwise, also consistent with our desired result.

What remains is to consider allocations involving rationing. Applying our previous reasoning, we can rule out rationing allocations where rationing is accomplished through a single queue and some combination of prices. To see that rationing with differing queue options for each type is also non-optimal, we offer a simple argument. Consider all allocations where one type queues for time $t$ and the other queues for at least $t$. We can again view the resulting queuing costs as shifting the agents' willingness-to-pay. But this returns us to the case of single-queue allocations, with valuations shifted to $r - ct$ and $1 - t$, where rationing is always non-optimal. We thus rule out all rationing options and complete the proof. *Q.E.D.*

This result matches our intuition that queuing is only used for targeting if it is sufficiently cheap for the target type relative to for the non-target type. For it to be optimal to use queuing for targeting, we must at least have $c < 1$, so queuing is not optimal when equally or more costly for the target type. We see also that the maximum $c$ that justifies queuing is decreasing in $\frac{\alpha}{\lambda}$. This reflects the fact that the more the mechanism designer cares about raising revenue relative to benefiting the target type, the stronger the screening effect of the queue must be to justify the revenue loss resulting from screening.

This result demonstrates also that queuing can be a useful tool even when the target type already has a higher willingness-to-pay than the non-target type, since queuing can allow the designer to still sell to target agents while screening non-target agents. That is, queuing can reduce the price (here to 0) at which the designer can sell exclusively to target agents.

### 4.3. *Single-bundle mechanism*

In this subsection, we consider a slightly different formulation of our original problem. We assume that the designer is only able to impose a fixed time cost $T$ and thus the non-pecuniary time cost function $t \in \{0, T\}$. Furthermore, we modify our utility function slightly, such that agents only incur this fixed time cost proportional to their allocation of the good. In this case, utility is given as

$$\int (x(v)(v_x - t(v)) - v_m p(v)) f(v) dv + \alpha \int p(v) f(v) dv$$

where we set $v_t \equiv 1$. Then, for each agent, we may calculate a modified version of willingness-to-pay $\hat{r} \equiv (v_x - T)/v_m$. Note that we have decomposed the problem into a unidimensional one; that is, an agent's preferences are fully captured by $\hat{r}$ given a time cost $T$. This maps directly to Dworczak et al. (2021)'s setting and thus we may apply their result (Proposition 2) that no matter the state of buyer-side inequality, the competitive price is the optimal single-price mechanism. Hence, the optimal price is determined by $T$, and so we denote this as $p_T^*$. Finally, to find the optimal $T$ it suffices to solve

$$T^* = \arg\max_T \left[ \int (x(v)(v_x - T) - v_m p_T^*) f(v) dv + \alpha \int p_T^* f(v) dv \right].$$

Thus, the optimal single-bundle menu imposes a queuing cost $T^*$ and lists the item for the resulting competitive price $p_T^*$.

## 5. COMPUTATIONAL APPROACHES

In this section, we consider computational approaches to our allocation problem. In particular, we will consider a discrete distribution of agents and formulate the model as a linear program. We will then present results for various distributions of agents.

### 5.1. *Linear program*

As discussed above, we restrict our framework to a discrete distribution of agents with $N$ types. As before, each agent has unit demand and is characterized by $(v_x, v_m, v_t)$. Thus, the utility of receiving a bundle of $(x, p, t)$ is given as $v_x x - v_m p - v_t t$. We adapt the objective function to the discrete case, but it is otherwise identical. Formally, we solve the following linear program

$$\begin{aligned}
\max_{(x_i, p_i, t_i)} \; & \sum_i v_{x,i} x_i - v_{p,i} p_i - v_{t,i} t_i + \alpha \sum_i p_i \\
\text{s.t.} \quad & x_i \geq 0,\; p_i \geq 0,\; t_i \geq 0, \forall i && \text{(Non-negativity)} \\
& x_i \leq 1 \,\forall i, \quad \sum_i x_i \leq Q && \text{(Feasibility)} \\
& v_{x,i} x_i - v_{m,i} p_i - v_{t,i} t_i \geq 0, \forall i && \text{(Individual Rationality)} \\
& v_{x,i} x_i - v_{m,i} p_i - v_{t,i} t_i \geq v_{x,i} x_j - v_{m,i} p_j - v_{t,i} t_j, \forall i \neq j && \text{(Incentive Compatibility)}
\end{aligned}$$

As discussed in Section 3, the first and second summation terms in the objective represent the total welfare and revenue, respectively. The feasibility constraints enforce unit demand for each agent and the overall scarcity of the good. The individual rationality constraint ensures that agents receive nonnegative utility from the allocation, i.e., that they wouldn't be better off not participating. Finally, the incentive compatibility enforces that each agent weakly prefer their bundle to any other offered bundle. Thus, we need one constriant for every pair of agents, which implies that the linear program has $O(n^2)$ constraints.

Using this approach, we now present three different simulations, each one computed using 100 agents, 50 goods, and $\alpha = \frac{1}{N}\sum_i v_i^m$, the average value for money. We solve the linear program using an open-source Python library, `PuLP`.

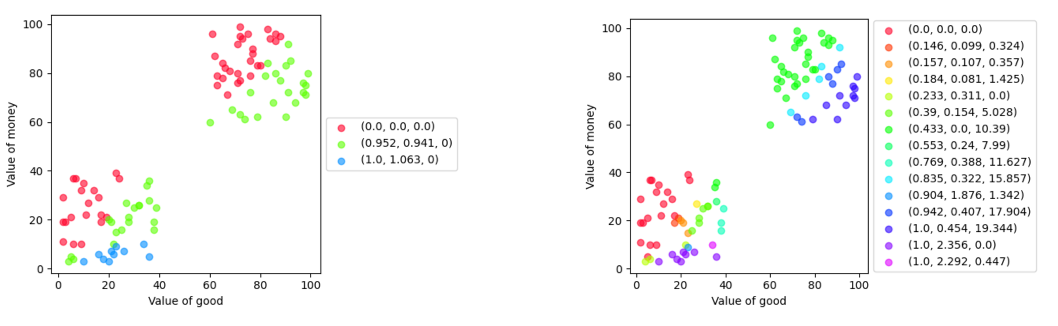
### 5.2. *Two types: "rich" and "poor"*

For the first simulation, we consider two types of agents: "rich" agents, which we will denote as $\theta_R$, and "poor" agents, denoted with $\theta_P$. This simulation is meant to capture the original intuition behind the motivating example from the introduction; specifically, both $\theta_R$ and $\theta_P$ have similar willingnesses-to-pay but $\theta_P$ has much greater values for the good and for money. Concretely, for $\theta_R$ we draw $v_i^x, v_i^m \sim U([1, 40])$ and for $\theta_P$ from $U([60, 100])$. These distributions were chosen to have sufficient differentiation in values within type while still ensuring separation from the other type; however, the exact values were chosen arbitrarily.

Figure 1 illustrates the results of solving the linear program to obtain an optimal mechanism without queuing (1a) and with queuing (1b). Each point in the graph depicts an agent in its type space (value of money by value of good). There are two distinct clusters: in the bottom left, $\theta_R$ type agents, and in the top right, $\theta_P$ type agents. The coloring of each agent in the plot denotes the bundle they receive in the optimal mechanism obtained by the linear program. The legend presents the menus as an ordered tuple $(x, p, t)$, such that a bundle $(1, 2, 0.5)$ would be interpreted as spending 2 and burning 0.5 to obtain 1 good. Red (the topmost entry in the legend) always denotes the zero menu.

FIGURE 1.—Optimal mechanism with $\theta_R$ and $\theta_P$ types

(a) *without* queuing (b) *with* queuing




We make the following observations about these results.

1. **Number of menu items.** Without queuing, there are two non-zero menu items: one rationing option and a full option. With queuing, however, we obtain fourteen distinct bundles. It is possible that over arbitrary distributions the optimal mechanism may even consist of an infinite menu of lottery options, as has been proven for other multidimensional mechanism design contexts (Daskalakis et al., 2015).

2. **Welfare improvement among $\theta_P$.** Without queuing, the $\theta_P$ agents with higher values for money did not receive any allocation because their willingness-to-pay was too low, despite having high valuations of the good. However, with queuing, every agent of type $\theta_P$ receives at least some amount of the good – at a minimum, 0.4 of the good. Moreover, the majority of the $\theta_P$ agents that were being rationed to in the without queuing case, now receive the good with certainty. These results are achieved by adding a sizable time burning component to the bundle, which screens out the agents of type $\theta_R$.
3. **Reduction in revenue.** We observe that the case without queuing raises substantially more revenue (0.502 per agent) than the with queuing case (0.355 per agent). This is because we are substituting some of the screening done by prices with time in the queuing case. For example, looking at the $\theta_P$ agents with the highest valuation of the good, without queuing they spent 0.941 to receive 0.952 of the good, whereas in the queuing case they have been reduced to spending 0.454 to obtain 1 good as they also burn 19.3 units of time. Nonetheless, even if we were able to perfectly redistribute the revenue raised (i.e., with ex-post lump sum transfers), the total welfare among $\theta_P$ would still be higher in the queuing case.

### 5.3. *Four types: all quadrants*

Next, we extend the type space to include the other two quadrants of the type space as well. Specifically, we now consider four types:

- $\theta_{BL}$: low $v_x$, low $v_m$ ("bottom left");
- $\theta_{BR}$: high $v_x$, low $v_m$ ("bottom right");
- $\theta_{TL}$: low $v_x$, high $v_m$ ("top left");
- $\theta_{TR}$: high $v_x$, high $v_m$ ("top right").

We define "low" and "high" as in Section 5.2 with the two types: that is low values are drawn from $U([1, 40])$ and high values from $U([60, 100])$. Note that $\theta_{BL} = \theta_R$ and $\theta_{TR} = \theta_P$ from above. These combinations were chosen in order to obtain coverage the majority of the type space while still maintaining some distinction between the agents. This initialization also seems like a more realistic setting: in the world, we are likely to have agents that are rich and poor, and that value and don't value the good.

FIGURE 2.—Optimal mechanism with four types ($\theta_{BL}, \theta_{BR}, \theta_{TL}, \theta_{TR}$)

(a) *without* queuing (b) *with* queuing

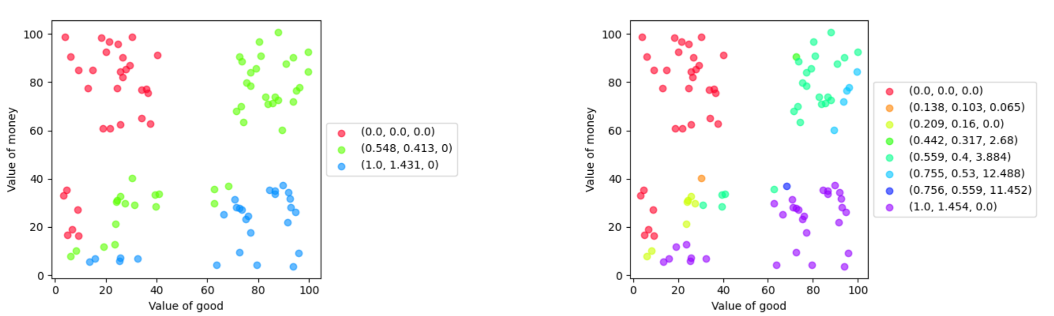


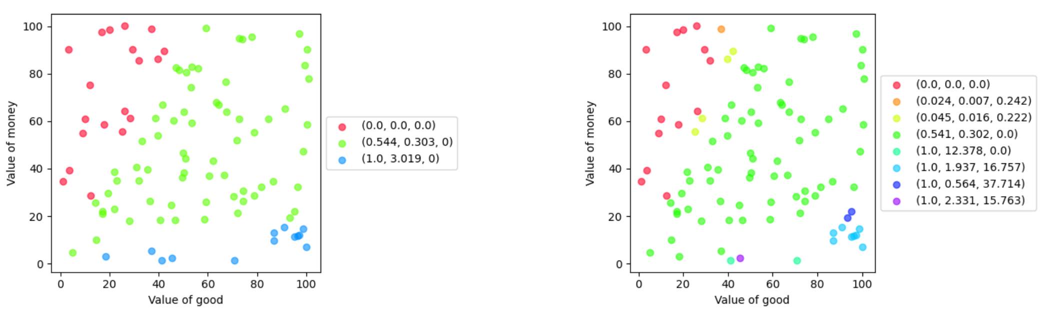

Figure 2 illustrates the results of solving the linear program for this type initialization without queuing (2a) and with queuing (2b). Similar to the results in Section 5.2, introducing queuing leads to an increase in menu items (though fewer than before), allows for better targeting of “rich” ($\theta_{BL}$) agents, and delivers most “poor” agents ($\theta_{TR}$) more of the allocation. However, in this case none of the $\theta_{TR}$ agents get the full good – instead, these are sold to agents with the greatest willingness-to-pay ($\theta_{BR}$ and some in $\theta_{BL}$). Interestingly, the overall extent of queuing in this setting is substantially reduced relative to before, likely because more agents are receiving the good from the direct price mechanism. Both with and without queuing, $\theta_{TL}$ agents, i.e., those with a low value for the good and with a high value for money, do not receive any allocation, which is intuitive as the marginal gain to our objective is always greater for the other types and given the scarcity of the good.

### 5.4. *Uniform and independent*

Finally, we draw agents' types from the same distribution, using $v_i^x, v_i^m \sim u([1, 100])$. This is similar to the case above, though there is no longer any enforced separation between agents. That is, the agent's values now cover the whole range.

FIGURE 3.—Optimal mechanism with uniform distribution

(a) *without* queuing (b) *with* queuing




The results of this simulation are displayed in Figure 3. Interestingly, the results for the majority of agents are approximately identical with and without queuing. In particular, there is a large green region that receives a bundle of (0.544, 0.303, 0) in the no queuing case and approximately the same region in the queuing case that receives a bundle of (0.541, 0.302, 0). This region accounts for the majority of agents in both plots. Moreover, a large amount of the revenue generated from the allocation in the queuing case is obtained from two agents, each who pay a price of 12.378 to receive one good. In addition, as observed previously, adding queuing reduces the fraction of agents receiving a zero menu.

## 6. CONCLUSION

A long tradition of economic thought has identified the usefulness of costly screening mechanisms, both in theory and in practice. Generally, however, analytical explorations of their use have kept their distance from multidimensional settings, with the exception of some recent work.

In this work, we first characterize the optimal use of queuing devices in a simple multidimensional context with monetary transfers where agents have varying values for money, or equivalently where the mechanism designer has redistributive preferences in the form of welfare weights. Our results align with simple intuitions about queuing devices. In particular, we establish that queuing devices which are relatively more costly (in monetary terms) for wealthier agents can improve targeting efficiency, and improve the ability of the designer to subsidize poorer agents. The "value-shifting" interpretation of costly signaling used in our analysis may potentially prove useful for further work, and we provide an additional example (the optimal single price/queue mechanism) further demonstrating its utility.

To build on this theory and provide additional intuition, we produce a computational approach for solving discrete distribution instances of our general model. We parameterize the mechanism designer's optimization problem as a linear program, and use a freely-available LP solver to generate numerical results. Here we again provide results that match intuition, in particular with respect to the mechanism designer's need to balance revenue generation, targeting, and deadweight loss from queuing. These results appear to validate the concern that allowing for queuing may greatly increase the complexity of the optimal mechanism, but future work may be able to address this issue by focusing on continuous distributions with full support or restricting the set of allowable distributions in other ways.

The overall problem of multidimensional mechanism design, whether involving costly signaling or other tools, remains challenging. Given the ubiquity of costly screens (such as queues and forms) in allocating critical goods, however, we believe it is crucial to better understand when and how they can be optimally used.

Even in our simple model, there remain avenues for further development, including by removing restrictions on $\alpha$ and $Q$ to allow for optimal mechanisms that involve rationing. Simple extensions to a small number of agents greater than two could also prove useful, as the complexity of the problem grows quickly with the complexity of the type space. One productive avenue for future work might lie in exploring approximately optimal mechanisms, in particular by limiting the permitted complexity of the menu used. In our computational work we frequently observe in the optimal mechanism that many agents are close to indifferent between several menu options, suggesting that approximations limiting the number of menu options might be able to approximate the optimal menu very closely (alternatively, this might suggest that an analysis of continuous distributions with full support would yield a simpler result). We are excited to see this area of research continue to develop, and perhaps continue to contribute to it ourselves.

## REFERENCES

AKBARPOUR, MOHAMMAD, PIOTR DWORCZAK, AND SCOTT DUKE KOMINERS (2023a): "Redistributive allocation mechanisms," *Forthcoming at Journal of Political Economy*. [2, 4]

AKBARPOUR, MOHAMMAD, PIOTR DWORCZAK, AND FRANK YANG (2023b): "Comparison of Screening Devices," *Available at SSRN 4456198*. [3]

ALATAS, VIVI, RIRIN PURNAMASARI, MATTHEW WAI-POI, ABHIJIT BANERJEE, BENJAMIN A OLKEN, AND REMA HANNA (2016): "Self-targeting: Evidence from a field experiment in Indonesia," *Journal of Political Economy*, 124 (2), 371–427. [3]

BERTRAND, MARIANNE, SENDHIL MULLAINATHAN, AND ELDAR SHAFIR (2004): "A behavioral-economics view of poverty," *American Economic Review*, 94 (2), 419–423. [3]

CHAKRAVARTY, SURAJEET AND TODD R KAPLAN (2013): "Optimal allocation without transfer payments," *Games and Economic Behavior*, 77 (1), 1–20. [2, 3]

CONDORELLI, DANIELE (2013): "Market and non-market mechanisms for the optimal allocation of scarce resources," *Games and Economic Behavior*, 82, 582–591. [2]

DASKALAKIS, CONSTANTINOS, ALAN DECKELBAUM, AND CHRISTOS TZAMOS (2015): "Mechanism Design via Optimal Transport," *CoRR*, abs/1503.01958. [8]

DESHPANDE, MANASI AND YUE LI (2019): "Who is screened out? Application costs and the targeting of disability programs," *American Economic Journal: Economic Policy*, 11 (4), 213–248. [3]

DUPAS, PASCALINE, VIVIAN HOFFMANN, MICHAEL KREMER, AND ALIX PETERSON ZWANE (2016): "Targeting health subsidies through a nonprice mechanism: A randomized controlled trial in Kenya," *Science*, 353 (6302), 889–895. [3]

DWORCZAK, PIOTR (2023): "Equity-efficiency trade-off in quasi-linear environments," Tech. rep., Technical Report, GRAPE Group for Research in Applied Economics 2022. [2, 3]

DWORCZAK, PIOTR, SCOTT DUKE KOMINERS, AND MOHAMMAD AKBARPOUR (2021): "Redistribution through markets," *Econometrica*, 89 (4), 1665–1698. [2, 3, 4, 7]

FINKELSTEIN, AMY AND MATTHEW J NOTOWIDIGDO (2019): "Take-up and targeting: Experimental evidence from SNAP," *The Quarterly Journal of Economics*, 134 (3), 1505–1556. [3]

HARTLINE, JASON D AND TIM ROUGHGARDEN (2008): "Optimal mechanism design and money burning," in *Proceedings of the fortieth annual ACM symposium on Theory of computing*, 75–84. [2, 3]

KANG, MINGSHI AND CHARLES Z ZHENG (2023): "Optimal design for redistributions among endogenous buyers and sellers," *Economic Theory*, 75 (4), 1141–1180. [2]

MCAFEE, R PRESTON AND JOHN MCMILLAN (1992): "Bidding rings," *The American Economic Review*, 579–599. [2]

NICHOLS, ALBERT L AND RICHARD J ZECKHAUSER (1982): "Targeting transfers through restrictions on recipients," *The American Economic Review*, 72 (2), 372–377. [2, 3]

NICHOLS, DONALD, EUGENE SMOLENSKY, AND T NICOLAUS TIDEMAN (1971): "Discrimination by waiting time in merit goods," *The American Economic Review*, 61 (3), 312–323. [2, 3]

PATEL, ROHIT AND CAN URGUN (2022): "Costly Verification and Money Burning," . [3]

REUTER, MARCO AND CARL-CHRISTIAN GROH (2020): "Mechanism design for unequal societies," *Available at SSRN 3688376*. [2]

ROSE, JULIE L (2021): "Rationing with time: time-cost ordeals' burdens and distributive effects," *Economics & Philosophy*, 37 (1), 50–63. [2, 3]

WEITZMAN, MARTIN L (1977): "Is the price system or rationing more effective in getting a commodity to those who need it most?" *The Bell Journal of Economics*, 517–524. [2]

YANG, FRANK (2022): "Costly Multidimensional Screening," . [3]

ZECKHAUSER, RICHARD (2021): "Strategic sorting: the role of ordeals in health care," *Economics & Philosophy*, 37 (1), 64–81. [2, 3]