The Structure of the Social Organization

Canice Prendergast*

University of Chicago

August 3, 2025

Abstract

We address how social organizations allocate/match services to clients. The market design literature on matching without prices prioritizes client preferences through mechanisms like Deferred Acceptance and Serial Dictatorship. But the enormous private social sector is primarily funded by fees to clients and external donors. We consider matching with *fair* prices, and show that it eliminates that priority. When average client income is low, services are assigned. With low total funding, it delegates to social service providers or uses bureaucratic rules. Neither incorporates client preferences and the empirical research highlights these mechanisms. Deferred Acceptance only arises with low client income but high external funding. When client income is sufficiently high and clients exhibit Decreasing Risk Aversion, the firm implements client choice at subsidized prices. Limited client choice is possible at lower client income, including one analogous to markets with artificial currency. In this way, we offer a pecking order of mechanisms for the non-profit and social enterprise.

*Thanks to Jacob Leshno, Eduardo Acevedo, and seminar participants at the NBER, University of Mannheim and Northeastern University for helpful comments.
canice.prendergast@chicagobooth.edu.

"You have not told me all concerning yourself;
and how shall I choose better than you?",
J.R.R. Tolkien,
The Fellowship of the Rings.

The private social sector is enormous. Non-profits alone employ more people than public education or manufacturing, accounting for 10% of private sector U.S. employment and expenditures of $3 trillion.[1] There are also ten million for-profit social enterprises worldwide. Social organizations play a fundamental role in society - they provide medical care to the sick, house those in need of shelter, and on and on. The majority of social organizations distribute services to clients in areas such as social assistance, housing, education and health care.[2] Our focus is on how best to do so.

Perhaps the central theme of modern microeconomics has been the design of mechanisms to transfer goods to consumers. For the for-profit sector, this has resulted in voluminous fields like optimal auctions, non-linear pricing and bundling. The market design literature on matching without prices assigns students to schools, patients to kidneys, and so on. Our starting point is that neither represents the reality of the private social organization. The literature on profit-seeking firms typically designs mechanisms that best sell to high willingness-to-pay clients. For social organizations, who typically exist to serve those in financial need, this is anathema to their mission. On the other hand, the market design literature does not address the firm's need to fund itself. Put more simply, while the for-profit literature is largely only about money, the market design literature typically ignores it. Social organizations sit squarely in the middle. Non-profits consistently report lack of funding as their greatest concern. Furthermore, they receive 49% of funding from fees from clients, with an additional 22% from donors.[3] The organization we model reflects that reality of needing funds to function, but not using willingness-to-pay to get them.

Since Gale and Shapley, the gold standard for understanding non-market allocation has been through matching without prices.[4] We instead study the canonical non-market setting of a social organization providing services to clients in need but where funding by clients and donors limits its coverage.[5] The firm chooses an allo-

[1] 12.8m people work in non-profits, compared to 12.7m in manufacturing, roughly 1.5m in public high schools and 8.8m overall in public schools and universities. Its importance has grown, with non-profit employment increasing by 33% over the last fifteen years compared to 9% in the for-profit sector (Bureau of Labor Statistics, 2024). The data on the revenue of the non-profit sector are from USA Facts, 2023, and employment from BLS, 2023.

[2] 83% of non-profits are in the social assistance, health care and educational sectors, which have obvious clients (BLS, 2023).

[3] Social enterprises receive almost all their funding from clients.

[4] See Roth, 2008, Pathak and Sonmez, 2008, Abdulkadiroglu et al., 2005, etc.

[5] For most social organizations, client need looms large in their allocation decisions and is often part of their legal mandate. For instance, non-profit hospitals are required to provide charitable

cation mechanism to distribute its services. To do so, it hires an agent motivated by his impact on client welfare. The central distinction is between assigning services on behalf of clients or allowing those clients to choose.[6] The market design literature typically prioritizes integrating client preferences in either case though mechanisms like Deferred Acceptance or Serial Dictatorship. This is not the case here. Clients privately know their need and services are vertically differentiated. As a result, Choice - the analog of Serial Dictatorship here - has an obvious appeal in overcoming the "how shall I choose better than you?" component of the opening quote. However need and willingness to pay differ here because clients vary in their income. The problem with pricing is that poorer clients may suffer compared to wealthier ones: the "you have not told me all concerning yourself" of the opening quote. To address this tradeoff, the firm must take a position on pricing. We assume that it is governed by a *Fairness Constraint*, where no client can be denied for lack of income. For many service providers, this defines their mission of serving those in need. As a result, we address matching with fair prices designed to avoid inequality.

Unrestricted Choice allows a client to choose any service provided by the firm. It can only be used if it satisfies the Fairness Condition that chosen outcomes reflect client need rather than ability to pay. When Choice fails that test, Assigned Consumption arises. Yet the preferred form of Assigned Consumption also does not prioritize client preferences. We consider three forms of Assigned Consumption. *Deferred Acceptance* integrates firm preferences with clients' strategy proof rankings. *Bureaucratic Assignment* assigns treatments based on the firm's information, where client rankings play no role. Such rigid practices are characteristic of bureaucracies. Finally, *Agent Assignment* allows autonomy to a social service provider to choose services, again without input from clients. While Deferred Acceptance has the highest allocative efficiency - matching services to need - among Assigned Consumption mechanisms for familiar reasons, it has the lowest level of client coverage. More generally, we show a tradeoff between allocative efficiency and client coverage.

Funding mitigates the familiar priority to Deferred Acceptance in favor of more rigid rules for two reasons. First, Deferred Acceptance has two foundational components: strategy proof rankings by clients, and a desired ex post criterion such as stability or Pareto efficiency. Both are costly here. This is because the feature that limits client choice of services - that fairness may be violated by benefiting higher income clients - also constrains clients from ranking them. We show that this problem can be overcome only by reducing prices paid by clients, thereby reducing funding. This implies that Deferred Acceptance raises the least revenue of any mechanism we consider.

The second reason relates to the importance of agent autonomy, another central concern of the literature reviewed below. Disagreement among stakeholders is a char-

---

care.

[6]This is a major theme in the empirical literature. See, for example, Chaumba and Locklear, 2021, and Cohen et al., 2016.

acteristic of many mission-based organizations. We assume that agents disagree both with the firm and each other over client needs. From the firm's perspective, this implies that Agent Assignment reduces allocative efficiency. However, it also reduces wages. This is because agents value how much they "make a difference" to client welfare. Agent Assignment offers higher marginal impact on client welfare and hence a lower wage. As a dollar spent on wages is a dollar less for service provision, this increases client coverage. This outcome is not generated by simply adding agency issues to the firm's problem: everyone is motivated by client welfare. Instead, it is that algorithms like Deferred Acceptance or Bureaucratic Allocation have the feature that the identify of the agent has no impact on outcomes. As wages are determined by marginal impact, Agent Assignment recruits at lower cost. Similar logic implies that a typical desideratum of the market design literature - Pareto efficiency - also increases wages, even though client welfare is commonly known to be higher.

These two effects mean that Deferred Acceptance has the lowest level of client coverage. Across the three Assigned Consumption mechanisms, client coverage and allocative efficiency are substitutes: Agent Assignment has the worst allocative efficiency but greatest client coverage, while Bureaucratic Assignment is intermediate on each dimension. Given these problems, the appeal of client choice is obvious. Choice is the analog to Serial Dictatorship in the market design literature, where it is typically chosen when the firm is indifferent about services received by clients. Here the alignment of client and firm interests is endogenous. Specifically, if clients exhibit Decreasing Absolute Risk Aversion, we show that (subsidized) Unrestricted Choice satisfies Fairness if and only if average client income is sufficiently high. If this is the case, it raises the most revenue from clients.

These outcomes provide a pecking order for assigning services based on the level and source of funding. Social organizations have many constituencies and objectives to satisfy: those of employees, a desire for consistency in outcomes, and the preferences of clients. When the firm is poorly funded, it prioritizes employee interests by using Agent Assignment. With somewhat higher funding, it transitions to allocating by rule, using Bureaucratic Assignment. Only when funding reaches a sufficient level are client preferences incorporated by transitioning to Deferred Acceptance. Better firm information makes both Agent Assignment and Deferred Acceptance less likely. Finally, when average client income is high enough such that Unrestricted Choice is possible, the firm never uses Deferred Acceptance or Bureaucratic Assignment. In this way, poor clients always receive services through a process that is out of their hands, while wealthier clients can choose. To illustrate the importance of these alternative mechanisms, we describe an extensive empirical literature on how the key transition for social service providers is between Bureaucratic and Agent Assignment. If funding doesn't matter and hence client coverage is exogenous, the firm uses either Deferred Acceptance or Unconstrained Choice. This replicates the outcome of the literature but here the preferred choice depends only on average client income.

These outcomes assume indivisible services and strategy proof mechanisms. We

show that other forms of Choice arise when we relax either assumption, even at lower client income levels. First, when services are divisible, *Restricted Choice* gives clients a choice between a larger quantity of a less desired good and a smaller quantity of a more desired one. With appropriate quantity restrictions and prices, this mechanism overcomes the Fairness problem for any level of client income. Restricted Choice can be implemented by a common market design, namely allowing clients access to artificial currency. Artificial currencies tend to be used when fiat money outcomes would result in unfair outcomes. Here quantity distortions decline with average client income and converge to Unrestricted Choice. The central problem with Restricted Choice is poor targeting of high priority clients. We use two recent market design settings, Feeding America and Covax, to illustrate this targeting issue.

Second, many social organizations implement a mixture between assignment and choice, where clients can choose from undersubscribed services but oversubscribed ones are assigned. For example, a client for a housing association may be able to choose a house in a less desired area, but if she turns down that offer, her next offer arises from an assignment process with a random outcome. We show that such *Partial Choice* can be implemented at lower levels of client income than Unrestricted Choice through a (non-strategy proof) First Choice mechanism. Both of these other Choice mechanisms raise more revenue than Assigned Consumption.

After describing the outcomes of the basic model, we consider a series of extensions. Most important among these is addressing ways in which the Fairness Constraint can be relaxed, allowing us to consider outcomes where clients are excluded from service.

So far, we have described strategies for the non-profit. Yet 10 million for-profit social enterprises account for 3% of global employment (World Economic Forum, 2024). The choice of profit status is a central concern in the literature on the social organization. The key characteristic of the non-profit is the Non-Distribution Constraint (Hansmann, 1980). The economics literature has focused on harmful firm incentives that arises from the for-profit. Yet much of the concern in the sector lies elsewhere, namely, how profit status affects funding. In particular, the concern is that donors will not give to a for-profit firm, even one with a clear social mission. To address this, consider donors in a setting where there is uncertainty about donations. If the firm exhibits diminishing returns to serving clients, as some clients are higher priority than others, there is a benefit to the firm acting as residual claimant in order to offer insurance to the firm's mission. This increases allocative efficiency. However, donors give holding the donations of others fixed. As a result, he evaluates his gift at the use of the marginal dollar received. If the marginal donation goes to profit, he will not give. Yet by acting as residual claimant, all marginal dollars are profit. This outcome of eliminating donations arises even if most infra-marginal funds are used on the mission. As a result, an organizational form that increases allocative efficiency once again reduces client coverage, such that the social enterprise can only exist if it is fully client-funded (Gertner, 2022, Brumme and Trelsted, 2024).

We begin in Section 1 by describing a model of the non-profit, while we address the for-profit Social Enterprise in Section 2. Section 3 concludes.

1 The Non-Profit Social Organization

A firm provides treatments (services) $t_i \in \{0,1\}$ with intensity $I_i \in \{0,1\}$ to clients i, where $i = 1,2,3,.....\infty$. The return to treatment depends on a client i's privately known type $\tau_i \in \{0,1\}$. The surplus from treatment for client i is

$$T(\tau_i, t_i) = t_i\mu + \phi\mathbb{1}(t_i = \tau_i). \tag{1}$$

The parameter $\mu > 0$ scales vertical differentiation and $\phi > 0$ reflects horizontal differentiation or the return to matching τ_i to t_i. Our objective is to characterize outcomes in terms of these two parameters.

We are interested in settings where the firm wishes to match t_i to τ_i. We assume that the supply of types $\tau \in \{0,1\}$ and the distribution of treatments $t = \{0,1\}$ are balanced and uniform, where there are $\frac{1}{2}$ of each type and treatment. (Below we show how the qualitative results extend to both unbalanced demand and supply and a more general technology on T.) We also assume that the number of clients is large to avoid sampling error. Surplus to the firm is

$$S = \sum_{i=1}^{\infty}(v_i + T(\tau_i, t_i))I_i \tag{2}$$

where v_i represents a client-specific value that the firm gets from treating client i. A central feature of the market design literature is firm priority over clients, here represented by v_i. Clients are ordered to reflect diminishing returns, where $v_i = v - \delta i$. Clients are served in order, $i = 1,2,.....$ To allocate services to clients, the firm chooses an assignment mechanism A described below. The firm charges prices $P = \{p_0, p_1\}$, where $p_j \geq 0$ is the unit charge for treatment t_j.[7]

Clients: The utility of client i with type τ_i and marginal utility of income m_i, receiving treatment t_j for a unit price p_j, is given by

$$U_i^c = (T(\tau_i, t_j) - m_i p_j)I_i. \tag{3}$$

We assume that m_i is privately known and defined over a support $[\overline{m}, \underline{m}]$. The parameter m_i derives from a client's indirect utility function $V(y)$, where $y \geq 0$ is her income, where her utility is $V(y) + T$. (Below we interpret this.) We assume that

[7] We discuss the restriction on non-negative prices below.

$V'(0) = \infty, V' > 0, V'' < 0$, so $m_i = V'(y_i)$.[8] Let the poorest clients have marginal utility $\underline{m}$ and the richest $\overline{m} < \underline{m}$. Income and type τ are independent. (We relax this below.)

There are many settings in the social sector that fulfill this combination of vertical and horizontal differentiation.[9] As a leading example, imagine that a housing association is assigning clients to two cities, *Chicago* ($t = 0$) and *NewYork* ($t = 1$). Then $T(k, j)$ for clients is

$$\begin{array}{cc} \tau = & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} Chicago \\ NewYork \end{array} & \begin{bmatrix} \phi & 0 \\ \mu & \mu + \phi \end{bmatrix} \end{array}$$

The problem is only of interest if client preferences do not lead to trivial outcomes. The technology $T(\tau_i, t_i)$ and supply of types implies that τ_i should receive treatment $t_i = \tau_i$. If clients always preferred that allocation, the firm would always allow choice. We choose preferences such that client ranking in the absence of prices is partially informative. We assume that ϕ is random, equal to $\overline{\phi}$ with probability h (h for "horizontal" preferences) and $\underline{\phi} < \overline{\phi}$ with probability $1 - h$. ϕ is private information for the client. To render the problem interesting, we make Assumption 1.

Assumption 1: $\overline{\phi} > \mu > \underline{\phi}$.

Absent pricing, those with $\phi = \underline{\phi}$ prefer $t = 1$, while those with $\overline{\phi}$ prefer τ_i. In the example above, a majority $1 - \frac{h}{2}$ prefer *New York* to *Chicago*. Let $\hat{t}_i = (t_i^1 \succ t_i^2)$ be the preference ranking of client i, where t_i^j is her jth preferred treatment.[10] Let $\hat{\phi} = h\overline{\phi} + (1 - h)\underline{\phi}$. Efficient assignment entails $\tau_i = t_i$, with expected surplus $T^* = \frac{\mu}{2} + \hat{\phi}$.

Agent: The firm hires an agent to implement t_i. Any agent a observes a (contractible) signal τ_i^a on each served client's type where $\tau_i^a = \tau_i$ with probability q, where $\frac{1}{2} \leq q < 1$. The signal τ_i^a represents the kind of observable data - family size,

[8] The marginal rate of substitution between services and the outside good is then $\frac{T}{m}$. We model m_i as variation in client income. This could alternatively be interpreted as savings, liquidity, an absence of debt, or access to alternative support. Clients do not care about v_i. This is to offer preferences for the firm that differ from those of the client. Allowing these as part of preferences only adds person-specific variation to prices.

[9] In ascending desirability for the typical client: (i) Covid-19 vaccines: Sinopharm, Astra Zeneca, Pfizer, (ii) food banks: produce, dairy, frozen chicken, (iii) housing: larger units, or units in more desirable locations, and (iv) mental health care: drugs, drugs plus therapist check-in monthly, intensive therapy.

[10] As we assume that all clients must be served, we do not need to include the null set in the preference ranking.

employment history, age, residential location etc. - used to allocate social services. The agent has preferences over wages and client welfare:

$$U^a = \sum_{i=1}^{\infty} (w + \alpha \hat{U}_i^c) I_i. \tag{4}$$

where α reflects altruism towards the client and $\hat{U}_j^c$ is his belief about client i's expected utility. The agent has a reservation utility $\overline{w}$ per unit of I.[11] A random agent is offered the position at a wage of w per unit of I. If he accepts the position, his utility per client receiving service is $U^a = w + \alpha \hat{U}^{ca}$. If he does not, let $U^a = \overline{w} + \alpha \overline{U}^{ca}$ be his utility per client, where $\overline{U}^{ca}$ is agent a's belief of the client's utility if he is not employed. Then the agent accepts any offer where

$$w \geq \overline{w} + \alpha(\overline{U}^{ca} - \hat{U}^{ca}). \tag{5}$$

This is a formalization of a commonly stated preference by workers in the social sector, namely the desire to "make a difference".

$\overline{U}^{ca} - \hat{U}^{ca}$ derives from an assumption that any agent can disagree with others on appropriate actions. We model this as differences in priors over client need. For the firm, τ_i^a is pivotal for choosing t_i. The agent holds a similar prior with probability λ, but with probability $1 - \lambda$, has a prior that a client is type $\tilde{\tau}$ with probability ≈ 1, where $\tilde{\tau}$ is uniformly distributed. We assume λ is independently distributed across agents. Conditional on a given agent holding a belief $\tilde{\tau}$, there is a probability $\rho \geq \frac{1}{2}$ of two randomly chosen agents agreeing. Let $\tilde{\tau}_i^a$ be the agent's posterior belief over τ_i after observing τ_i^a.

Client Coverage: External donations are D.[12] Client coverage depends on two cases. When $w > p$, donor-funding limits client coverage as

$$I = \frac{D}{w - p}. \tag{6}$$

The first best level of treatment for the firm is $\frac{v+T^A}{\delta}$. To avoid trivial outcomes where the firm discards donations, we assume $D < \frac{v+T^A}{\delta}$. When $p \geq w$, the firm is self-funded without donors.[13]

[11] First, $\overline{U}$ need not be the compensation that the agent could receive elsewhere, but instead could include any "warm glow" discount for working for the social organization. Second, the model has constant returns to scale with respect to $I = \sum I_i$. This is assumed to avoid wired-in benefits from the firm becoming large and to address the allocation of fixed costs.

[12] In this section, D refers to any source of funding aside from clients, such as endowments, government or foundation support.

[13] This characterization rules out the use of Rotating Savings mechanisms, where clients pool income and randomize who receives treatment. See Besley et al., 1993.

Pricing: Each client has two attributes that affect her willingness to pay in (3), τ_i and m_i. A defining characteristic of many social organizations is that no client should be denied treatment for lack of money. To reflect this, we assume that treatment is independent of m_i:

Fairness Constraint: $Et_i(\tau_i|m_i) = Et_i(\tau_i), EI_i(\tau_i|m_i) = EI_i(\tau_i)$, for all m_i.

This assumption implies that the firm cannot trade off who is served in order to gain more revenue. We also assume that all services are assigned to a client.[14] We restrict attention to pure strategies by the client, nor we do not allow clients to purchase lotteries over services.[15] Below we discuss outcomes that allow such lotteries.

1.1 Assignment Mechanisms

We now address the set of assignment mechanisms considered. There are three actors: clients, the firm, and the agent. Allowing for input from each of these, and for now considering only strategy proof mechanisms with indivisible services, results in the following matrix of mechanisms:

Client	**Input**	**Firm**	**Input**	**Agent**	**Input**	**None**
Yes		*Deferred*	*Acceptance*	*Agent*	*DA*	*Choice*
No		*Bureaucratic*	*Assignment*	*Agent*	*Assignment*	.

Assigned Consumption arises when the firm or agent has input to outcomes, the four items on the left of the matrix. Choice allows the client to select her preferred outcome independent of the firm or client.[16]

1. Assigned Consumption:

   - *Deferred Acceptance*: Assignment is based on τ_i^a and the client's ordinal ranking $\hat{t}_i$ in a standard DA mechanism.
   - *Bureaucratic Assignment*: Assignment is based on τ_i^a
   - *Agent Assignment*: Assignment is based on $\tilde{\tau}_i^a$.[17]

2. Choice:

[14] This allows us to ignore what we call Cash Cow mechanisms below.

[15] We do not consider mixed strategies as one of the central concerns of the market design literature is the need to offer simple choices to clients, with its focus on strategy proof mechanisms.

[16] Random Assignment has neither, but is not explored here as it is always dominated by Bureaucratic Assignment.

[17] We discuss Agent Deferred Acceptance below, where assignment is based on $\tilde{\tau}_i^a$ and the client's ordinal ranking $\hat{t}_i$ in a standard DA mechanism.

- *Unrestricted Choice*: The client chooses either treatment at price p_i.

All the elements in the matrix above are strategy proof, and arise when services are indivisible. Below, we consider two additional choice outcomes when one of these fails:

- *Restricted Choice*: The client chooses either treatment at prices p_i and receives I_i of her highest ranked preference, where $I_i < 1$ for at least one i.
- *Partial Choice:* The outcome is partially chosen and partially assigned. It is implemented by a First Choice mechanism described below.

Strategies and timing: The timing is as follows:

1. The firm chooses A, p_j^A, and w^A, which is observed by clients and potential agents.
2. The firm offers the position to one agent subject to (5).
3. ϵ is revealed.
4. All clients $i = 1, 2, ...I^A$ report $\hat{t}_i$.
5. The firm and agent observes τ_i^a for all $i = 1, 2, ..., I^A$.
6. A is implemented, p_i^A paid by clients, and w^A is paid, where
   - If $A =$ Bureaucratic Assignment, $t_i = \tau_i^a$,
   - If $A =$ Agent Assignment, $t_i = \tilde{\tau}_i^a$,
   - If $A =$ Unrestricted Choice, $t_i = \hat{t}^1$ if services are not oversubscribed, and
   - If $A =$ Deferred Acceptance, t_i is allocated through an algorithm described below.

Given this timing, the objective of the firm is to choose A, w^A and P^A to maximize surplus subject to (3), (4), (5), (6), the Fairness Constraint, and Assumption 1. The equilibrium concept used is Bayesian Nash. The program is formally defined in the Appendix.

A Benchmark: Let T^A be average per capita treatment surplus from assignment mechanism A, let p^A be average revenue per client, and let w^A be the wage paid per unit of I. Then $I^A = \frac{D}{w^A - p^A}$. As a result, surplus is

$$S^A = \frac{D}{w^A - p^A}\left(v + T^A - \frac{\delta D}{2(w^A - p^A)}\right). \tag{7}$$

Surplus depends on three endogenous variables: (i) wages w^A, (ii) client payments p^A, and (iii) assignment surplus T^A.

Note one feature before describing outcomes. This is not a setting where the client cannot identify the appropriate treatment. In such instances, the optimal outcome often entails the firm assigning consumption to the client.[18] Our interest is not in such cases, but rather why Assigned Consumption would arise with informed clients.

1.2 Assigned Consumption

Assigned Consumption arises when the firm or agent at least partially determines services. We begin by considering Deferred Acceptance.

1.2.1 Deferred Acceptance

Here the firm uses client strategy proof rankings $\hat{t}_i = (t_i^1 \succ t_i^2)$ to match services. The Deferred Acceptance mechanism is as follows. From the timing assumptions, the firm I^{da} when allocating t_i. It offers $I_i = 1$ to all $i \leq I^{da}$ and 0 to all other clients.

1. The firm chooses P^{da} and all clients $i < I^{da}$ report $\hat{t}_i$.
2. The firm allocates $t = 1$ to all clients $i < I^{da}$ for which $\tau_i^a = 1$ and $\hat{t}_i = (1 \succ 0)$. If there are remaining $t = 1$ slots, it randomly assigns them to clients for which $\tau_i^a = 1$ and $\hat{t}_i = (1 \succ 0)$ until there are no more $t = 1$ slots.
3. The firm allocates $t = 0$ to all other clients.

To understand outcomes, we first address how the assignment mechanism affects revenue. First, all clients must be willing to pay for assigned treatment. In any equilibrium where type $\tau = k$ receives treatment $t = j$ with positive probability, Individual Rationality implies

$$p_j \leq min_k \frac{T(k,j)}{\underline{m}}. \tag{8}$$

When both types of clients receive all services, $p_j \leq min_k \frac{T(k,j)}{\underline{m}}$ over all j. As $q < 1$, any Assigned Consumption outcome implies that each type receives both services in equilibrium, so Individual Rationality requires that $p_0 = 0$, and $p_1 \leq \frac{\mu}{\underline{m}}$. In addition, in any equilibrium where client ranking affects outcomes, her Incentive Compatibility constraint requires that $\hat{t}_i$ must be independent of m_i. Lemma 1 simplifies.

[18] For instance, Bansak et al, 2025, design a mechanism for allocating asylum seekers to regions based on employment prospects in those regions, a factor that is unlikely to be known by the clients.

Lemma 1:

- *Unrestricted Choice satisfies the Fairness Constraint if and only*

$$\frac{\overline{m}}{\underline{m}} \geq \frac{\mu - \underline{\phi}}{\mu + \underline{\phi}}. \tag{9}$$

  In that case, maximum prices are $P^{uc} = \{p_0^{uc}, \frac{\underline{\phi}+\mu}{\underline{m}}\}$, *where* $p_0^{uc} = \min\{\frac{\underline{\phi}-\mu}{\overline{m}} + \frac{\mu+\underline{\phi}}{\underline{m}}, \frac{\underline{\phi}}{\underline{m}}\}$. For $\frac{\overline{m}}{\underline{m}} \geq 1 - \frac{\underline{\phi}}{\mu}$, $p_0^{uc} = \frac{\underline{\phi}}{\underline{m}}$.

- *Agent Assignment and Bureaucratic Assignment can be implemented for any income level with maximum prices* $P^{ac} = \{0, \frac{\mu}{\underline{m}}\}$ *with revenue per client* $p^{ac} = \frac{\mu}{2\underline{m}}$.

- *Deferred Acceptance can be implemented for any income level. For* $\frac{\overline{m}}{\underline{m}} < \frac{\mu-\underline{\phi}}{\mu+\underline{\phi}}$, *maximum prices* $P^{da} = \{0, \frac{\mu-\underline{\phi}}{\underline{m}}\}$ *with revenue per client* $p^{da} = \frac{\mu-\underline{\phi}}{2\underline{m}} < p^{ac}$.

First, note the importance of uncertainty over m_i from (9). Without uncertainty, $\frac{\overline{m}}{\underline{m}} = 1$ and the firm can always achieve the efficient outcome though Choice.[19] Second, (9) implies that Unrestricted Choice can only arise if clients are sufficiently similar in their willingness to pay. Intuitively, this is more likely when there is more horizontal differentiation and less likely with more vertical differentiation. If this is the case, it raises more revenue than Assigned Consumption.[20]

Third, Deferred Acceptance raises less revenue than any of the other mechanisms. To understand why, consider the housing example with $\mu = 5$ and $\underline{\phi} = 2$. Then $T(k, j)$ for clients with $\phi = \underline{\phi}$ is

$$\begin{array}{cc} \tau = & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} Chicago \\ NewYork \end{array} & \begin{bmatrix} \mathbf{2} & \mathbf{0} \\ \mathbf{5} & \mathbf{7} \end{bmatrix} \end{array}$$

As $q < 1$, mistakes are made under Assigned Consumption, so both τ types receive each city. As no one is excluded from service, Chicago is priced at 0 and New York at $\frac{5}{\underline{m}}$. Revenue is $\frac{2.5}{\underline{m}}$. From (9), the firm can use Unconstrained Choice if $\frac{\overline{m}}{\underline{m}} \geq \frac{3}{7}$.[21] If $\frac{\overline{m}}{\underline{m}} < \frac{3}{7}$, information on τ_i cannot be perfectly revealed. Instead, the only client

[19] With common marginal utility m, the firm generates efficient outcomes with prices $p_0 \in [0, \frac{\underline{\phi}}{\underline{m}}]$ and $p_1 \in (\frac{\mu}{\underline{m}}, \frac{\mu+\underline{\phi}}{\underline{m}}]$.

[20] Unconstrained choice requires that all services are assigned. This requires that $p_0^{uc} \leq \frac{\underline{\phi}}{\underline{m}}$ and $p_1^{uc} \leq \frac{\mu+\underline{\phi}}{\underline{m}}$.

[21] It can charge prices of $p_0 = \frac{2}{\underline{m}}, p_1 = \frac{7}{\underline{m}}$ when $\frac{\overline{m}}{\underline{m}} \geq \frac{3}{5}$. This yields revenue of $\frac{\frac{\mu}{2}+\underline{\phi}}{\underline{m}} = \frac{4.5}{\underline{m}}$.

information that can be revealed through Deferred Acceptance is client ranking of $T(k,j)$. Fairness requires that all clients with type $\underline{\phi}$ report *New York* as their first choice. This implies that the price difference between *New York* and *Chicago* cannot be larger than $\frac{5}{\underline{m}} - \frac{2}{\underline{m}} = \frac{3}{\underline{m}}$. As a result, maximum DA prices are *Chicago* $= 0$ and *New York* $= \frac{3}{\underline{m}}$, with expected revenue of $\frac{\mu - \underline{\phi}}{2\underline{m}} = \frac{1.5}{\underline{m}}$.

In words, Fairness implies that outcomes cannot distinguish between richer and poorer clients. If both services were free, client rankings would not depend on income. However, poorer clients may be excluded from services as prices rise in a way not true for wealthier clients. As a result, prices need to be limited to preserve client rankings being fair.

Proposition 1 *Surplus from Deferred Acceptance is*

$$S^{da} = \frac{D}{\overline{w} - p^{da}} \left(v + T^{da} - \frac{\delta D}{2(\overline{w} - p^{da})} \right). \tag{10}$$

where

$$T^{da} = \frac{\mu}{2} + q\hat{\phi} + \frac{h(1-q)\overline{\phi}}{2} + \frac{r}{2}((1-q)\hat{\phi} - q(1-h)\underline{\phi}), \tag{11}$$

where $r = \frac{h(1-q)}{1-hq}$.

Consider each component of surplus, w^{da}, p^{da}, and T^{da}. Outcomes are independent of agent beliefs, so $\hat{U}^{ca} = \overline{U}^{ca}$ and hence $w^{da} = \overline{w}$ from (5). This implies that expected client coverage is $I^{da} = \frac{D}{\overline{w} - p^{da}}$. T^{da} reflects the benefits of Deferred Acceptance beyond assigning only on the firm's signal of quality q.[22] Note that r is decreasing in q: the input of clients is less important when the firm is better informed. Taken together, the final two terms in (11) reflect the utilitarian benefits of Pareto efficiency.[23] Following Balinski and Sonmez, 1999, this outcome is equivalent to the one-sided matching outcome where each service is a player with a priority list given by τ_i^a.[24]

[22]This is through the $\frac{h}{2}$ clients that do not report $t = 1$ as their first choice, and through r, the lottery over assignments to $t = 1$ among those who rank $t = 1$ highest but whose signal to the firm suggests assignment to $t = 0$.

[23]Deferred Acceptance is Pareto efficient here as preferences are acyclic (Ergin, 2002). This has identical outcomes to the more traditional way in which DA is implemented. To see this, let the firm offer $t_i = 1$ to all clients for which $\tau_i^a = 1$. All provisionally accept. The firm then offers $t_i = 0$ to all clients for which $\tau_i^a = 1$. A fraction $\frac{h}{2}$ turn down the initial offer of $t_i = 1$ and take $t = 0$ instead. The firm then randomly makes offers for the $(1-q)\frac{h}{2}$ clients who turned down $t = 1$ to clients who have $\tau_i^a = 0$ until all $t = 1$ is exhausted. All other clients receive $t = 0$.

[24]Those who receive a signal $\tau_i^a = 1$ are given a choice over both services. A fraction $\frac{h(1-q)}{2}$ will choose $t = 0$ and the remainder choose $t = 1$. The firm then randomly chooses from those with $\tau_i^a = 0$ the remaining $\frac{h(1-q)}{2}$ slots for $t = 1$ and gives all other clients $t = 0$.

1.2.2 Bureaucratic Assignment

Bureaucratic Assignment arises when outcomes depend only on the firm's signal. The firm knows I when allocating t_i. It allocates $t_i = \tau_i^a$ for each client $i < I^b = \frac{D}{\overline{w} - p^{ac}}$.

Proposition 2 *Surplus from Bureaucratic Assignment is*

$$S^b = \frac{D}{\overline{w} - p^{ac}} \left(v + T^b - \frac{\delta D}{2(\overline{w} - p^{ac})} \right). \tag{12}$$

where $T^b = \frac{\mu}{2} + q\hat{\phi}$.

This outcome follows automatically because the allocations do not depend on information provided by the client. The Incentive Compatibility condition on ranking outcomes can be dropped. As a result, the firm need only satisfy (8) for all k, yielding revenue per client of $p^{ac} = \frac{\mu}{2\underline{m}}$. Bureaucratic Assignment has worse allocative efficiency than Deferred Acceptance but higher client coverage.

1.2.3 Agent Assignment

Agent Assignment offers autonomy to the agent by allowing him to assign t with no constraints. The firm chooses P^a. The firm knows I^a when allocating t. The agent observes τ_i^a and allocates $I_i = 1$ of $t_i = \tilde{\tau}_i^a$ to each client $i < I^a = \frac{D}{w^* - p^{ac}}$.

Proposition 3 *Surplus from Agent Assignment is*

$$S^a = \frac{D}{w^* - p^{ac}} \left(v + T^a - \frac{\delta D}{2(w^* - p^{ac})} \right). \tag{13}$$

where

$$T^a = \mu + (\lambda q + \frac{1 - \lambda}{2})\hat{\phi}, \tag{14}$$

and $w^* = \overline{w} - \alpha(1 - \lambda)(1 - \frac{\lambda}{2} - (1 - \lambda)\rho)\hat{\phi}$.

Agent Assignment has lower allocative efficiency than Bureaucratic Assignment, where $T^b - T^a = (1-\lambda)\hat{\phi}(q - \frac{1}{2})$. Yet it has higher client coverage. This is because $\hat{U}^{ca} > \overline{U}^{ca}$, as she now "makes a difference". As a result, $w^* < \overline{w}$, where $\overline{w} - w^* \equiv \Delta w^* = \alpha(1 - \lambda)(1 - \frac{\lambda}{2} - (1 - \lambda)\rho)\hat{\phi} > 0$. Intuitively, the wage discount is higher when the extent of disagreement $(1 - \lambda)$ is greater, when the agent cares more about the client (α), and when the agent believes that an alternative agent is less likely to share his views (ρ lower). Agent Assignment has the worst allocative efficiency but the highest client coverage.

Empirical Research: The modeling above emphasizes three issues: that firms are concerned about funding, that clients are an important source of revenue, and that worker autonomy is important. We provide evidence on each.

First, empirical evidence consistently points to financing concerns as predominant for non-profits: 42% report "lack of adequate finances/resources" as their biggest challenge, with 57% reporting "generating income and achieving financial sustainability" as their greatest concern (Statista, 2024, Small Charities Data, 2023).[25] Yet this is only relevant if the allocation mechanism affects revenue. As non-profits receive 49% of their revenues from fees to clients, it clearly does, with another 22% coming from donors (National Council of Nonprofits, 2024). Because of this, we focus on p^A and D as the primary sources of funds for the social organization.

Finally, worker autonomy is an important component in the provision of social services.[26] This is typically empirically identified when firms bureaucratize, where "lack of trust in social workers competence and ability to exercise discretion well has led municipalities to increasingly rely on manuals in an attempt to regulate their discretionary freedom" (Petersen et al, 2020). Pascoe et al, 2023, survey almost forty academic publications on the impact of the transition to Bureaucratic Assignment on social workers. The response is overwhelmingly negative. Employees feel they have become "basically glorified data-inputters" (Dustin, 2006), spend their days "checking whether you have completed your paperwork correctly" (Carey, 2009), and suffer from "mission drift from providing personalized care to cost-cutting and efficiency" (Groves et al, 2016).[27] Many of these studies also find higher intension to quit by social service providers after such transitions. As a result, 29% of non-profits report "difficulty recruiting and retaining quality staff" as their biggest challenge, likely as workers are paid below market wages in the sector.[28] This transition is now typically part of the curriculum for non-profit leaders addressing the life cycle of the non-profit (Kenny Stevens, 2002). Note for this large empirical literature, the key choice for social service providers is between Agent Assignment and Bureaucratic Assignment.[29]

[25]Concern about funding is even higher for small non-profits, and the size distribution of non-profits heavily skews towards small firms, with 67% of non-profits having annual expenditures less than $500,000 (Faulk et al, 2021).

[26]For example, Petersen et al, 2020, describe how "social workers have traditionally been given a great deal of autonomy to use discretion, but their decisions have increasingly been accused of being subjective and random". The classic reference is Lipsky, 1980.

[27]For more studies corroborating this outcome, see Adler and Borys, 1996, Cantarelli et al, 2016, Langer et al, 2019, and Wright and Davis, 2003.

[28]See Statista, 2024, Leete, 2001, Ruhm and Borkowski, 2003, Salamon and Sokolowski, 2005, and Hirsch et al., 2017.

[29]We assume that agents know the assignment mechanism and their impact on client welfare when accepting a position at the firm. Realistically, the margin on which this matters is more likely retention of workers. With respect to this, there is a significant literature in the public administration literature correlating a service provider's intention to quit and the degree of autonomy on the job (for examples, see Liu at al., 2011, and Zhang et al., 2024).

Agent Deferred Acceptance and the Cost of Pareto Efficiency: Agent Assignment is not Pareto efficient: there are $I^a(1-q)\frac{h}{2}$ clients assigned $t=1$ who would swap allocations for $t=0$ and increase allocative efficiency. This could be implemented here by Agent Deferred Acceptance, where the DA algorithm is implemented using the agent's posterior. In addition to the cost of Deferred Acceptance above, this comes with a novel cost, as the agent demands a higher wage. This arise even though client welfare is commonly known be higher under Agent Deferred Acceptance.

Agent Deferred Acceptance allows clients to have more impact on outcomes than Agent Assignment. The $\frac{h}{2}$ clients who in the example above prefer Chicago to New York can be assured of receiving that by reporting $\hat{t} = \{Chicago \succ NewYork\}$. This increases welfare. However it reduces the impact of the agent, causing wages to rise. We show in the Appendix that the wage increase is proportional to $\frac{h}{2}$. This shows the tradeoffs with using the combination of Deferred Acceptance while allowing the agent control otherwise. It has the revenue cost of Deferred Acceptance above, while not receiving the benefit of the same wage reductions under Agent Assignment. Yet is has the allocative benefit of Pareto efficiency.

The agent is motivated only by altruism here. Despite this, Agent Assignment has a flavor of an agency problem. A common outcome in the agency literature on social organizations is the need to offer more autonomy to agents than in a standard profit-seeking firm.[30] A relevant antecedent here is Aghion and Tirole, 1997, where control is delegated to an agent to induce more effort. However, the results here are not the result of a simple agency extension. Instead, it is the implication of any algorithm that does not depend on the person implementing it. Here an algorithm that is commonly known to increase client welfare increases wages if it reduces the marginal impact of the agent. Yet there is one feature common to agency settings: that ex post outcomes are distorted to improve ex ante incentives.

1.3 Unrestricted Choice

Many social organizations offer subsidized choice to their clients.[31] Any model of the social sector has to reflect this. Unrestricted Choice delegates the allocation of the service to the client. In this sense, it is the analog to Serial Dictatorship in the market

[30] This can take the form of excessive delegation of control (Aghion and Tirole, 1997), hiring agents whose preferences do not coincide with that of the firm (Prendergast, 2007, Che and Kartik, 2009), or ignoring likely legitimate complaints against bureaucrats (Prendergast, 2003).

[31] There are many examples. First, housing associations often offer vouchers to qualified clients who choose from available housing units. Second, food pantries now typically allow clients to choose food (rather than being offered a pre-packed basket). Third, food pantries typically receive their food from the regional food bank, and choose food from a menu. Some foods are free, but the pantries must purchase other food at subsidized prices. Fourth, job seekers who receive training can tailor that retraining to their preferences (MacEachan et al., 2013). Finally, many mental health providers now offer patients choice over therapy options (Cook et al., 2019).

design literature. We characterize Unrestricted Choice satisfying fairness when each client's highest ranked service is $\hat{t}_i = \tau_i$ for any m_i, and all receive that choice. Unrestricted choice arises when the equilibrium entails all clients of type τ receiving $I_i = 1$ of $t_i = \tau_i$. After choosing $P^{uc} = \{p_0^{uc}, p_1^{uc}\}$, the mechanism to implement this is as follows:

1. Each client $i \leq I^{uc}$ reports $\hat{t}_i$.
2. The firm allocates all clients their first choice if i is undersubscribed.
3. For all $t = i'$ that are oversubscribed, the firm rations randomly.

So far, we have treated m_i as parametric. To offer conditions for when (9) arises, let $y_i = \hat{y} + \eta_i \geq 0$, $E\eta = 0$ and $y \sim [\hat{y} + \underline{\eta}, \hat{y} + \overline{\eta}]$. Average income is $\hat{y}$. Finally, assume Non-Increasing Absolute Risk Aversion for clients, where $-\frac{V''(y)}{V'(y)}$ is non-increasing in y.[32]

Proposition 4 *With Decreasing Absolute Risk Aversion, (9) holds for all income levels above $\hat{y}^{uc}$, where*

$$\frac{V'(\hat{y}^{uc} + \overline{\eta})}{V'(\hat{y}^{uc} + \underline{\eta})} \equiv \frac{\overline{m}^{uc}}{\underline{m}^{uc}} = \frac{\mu - \underline{\phi}}{\mu + \underline{\phi}}. \tag{15}$$

With Constant Absolute Risk Aversion, (9) either holds for all income levels or for no income levels. Surplus from Unrestricted Choice is

$$S^{uc} = \frac{D}{\overline{w} - p^{uc}} \left(v + T^* - \frac{\delta D}{2(\overline{w} - p^{uc})} \right), \tag{16}$$

where $T^* = \frac{\mu}{2} + \hat{\phi}$.

Wage are given by $\overline{w}$ as $\hat{U}^{ca} = \overline{U}^{ca}$. Prices are as in Lemma 1. The firm can induce efficient choice at $\hat{y}^{uc}$ with $p_0^{uc} = 0$. As average income rises above $\hat{y}^{uc}$, the firm can increase p_0^{uc} until it reaches its maximum of $\frac{\underline{\phi}}{\underline{m}}$ above average income of $\tilde{y}^{uc} > \hat{y}^{uc}$, where $\frac{V'(\tilde{y}^{uc} + \overline{\eta})}{V'(\tilde{y}^{uc} + \underline{\eta})} = 1 - \frac{\underline{\phi}}{\mu}$. Proposition 4 illustrates how Choice not only yields efficient assignment but also raises more revenue than either Bureaucratic Assignment or Deferred Acceptance.

The ability for clients to have agency through choice is a central theme of the literature on social service provision. See for example, Chaumba and Locklear, 2021, and Cohen et al., 2016. From a practitioners perspective, see Food Bank News, 2024, for empirical evidence on how food bank clients value choice. Social organizations cover a wide range of needy clients, from the poorest members of society to those who

[32] See Chiappori and Paiella, 2011, Guiso and Paiella, 2008, and Huber et al., 2023, for empirical evidence supporting DARA.

have assets and income. Decreasing Absolute Risk Aversion allows a stark difference between non-profits that serve poorer clients and those serving wealthier ones. Those serving wealthier clients not only get more revenue through reduced m_i, but it allows the firm to transition to the higher efficiency of Unrestricted Choice, which in turn raises more revenue. By contrast, the revenue from serving the poor is so low that the firm further compromises allocative efficiency below that in Deferred Acceptance to raise needed funds. In this sense, a model that has no technological increasing returns to scale endogenously creates them.

1.4 The Non-Profit Outcome

We now aggregate our results. As the paper addresses the impact of funding needs, we begin with a benchmark where client coverage is exogenous. This is not an equilibrium here but would arise if D changes to ensure a fixed level of client coverage. (This would arise for instance in a public sector setting where government revenue makes up the difference. D refers to any other source of funds.) If I is independent of w^A and p^A, the firm uses Deferred Acceptance if $\hat{y} < \hat{y}^{uc}$ and Unconstrained Choice otherwise. This illustrates the existing market design outcomes in the absence of funding concerns. Yet here the choice between these two mechanism is uniquely determined by average client income.

When funding does matter, outcomes change significantly. They vary in three ways: allocative efficiency, wages and prices. We now provide a pecking order of assignment mechanisms based on the level and source of firm funding. First consider the comparison between Deferred Acceptance and Bureaucratic Assignment. Let $\Delta p = p^{ac} - p^{da} = \frac{\phi}{2\underline{m}}$ be the additional per client revenue from Bureaucratic Assignment. The firm chooses Bureaucratic Assignment for all $D \leq D_b^*$, where D_b^* is uniquely defined by

$$(T^{da} - T^b) = \frac{\Delta p}{(\overline{w} - p^{da})(\overline{w} - p^{ac})} \left(v + T^b - \frac{\delta(D_b^* + \frac{\Delta p}{2})}{2(\overline{w} - p^{ac})} \right). \tag{17}$$

Similarly, the firm chooses Agent Assignment over Bureaucratic Assignment for all $D \leq D_a^*$, where D_a^* is uniquely defined by

$$(T^b - T^a) = \frac{\Delta w}{(w^* - p^{ac})(\overline{w} - p^{ac})} \left(v + T^a - \frac{\delta(D_a^* + \frac{\Delta w}{2})}{2(w^* - p^{ac})} \right). \tag{18}$$

Proposition 5 follows.

Proposition 5 *For $\hat{y} < \hat{y}^{uc}$, the firm uses Agent Assignment for $D \in (0, D_a^*)$, Bureaucratic Assignment for $D \in [D_a^*, D_b^*)$, and Deferred Acceptance for $D \geq D_b^*$.*

This offers a pecking order for Assigned Consumption. At the lowest level of funding, Agent Assignment is used. The firm then transitions to Bureaucratic Assignment,

and only with higher funding again uses Deferred Acceptance.[33] Not all mechanisms need be used. For instance, if $\alpha = 0$ wage savings from Agent Assignment are zero and it is never preferred over either alternative.

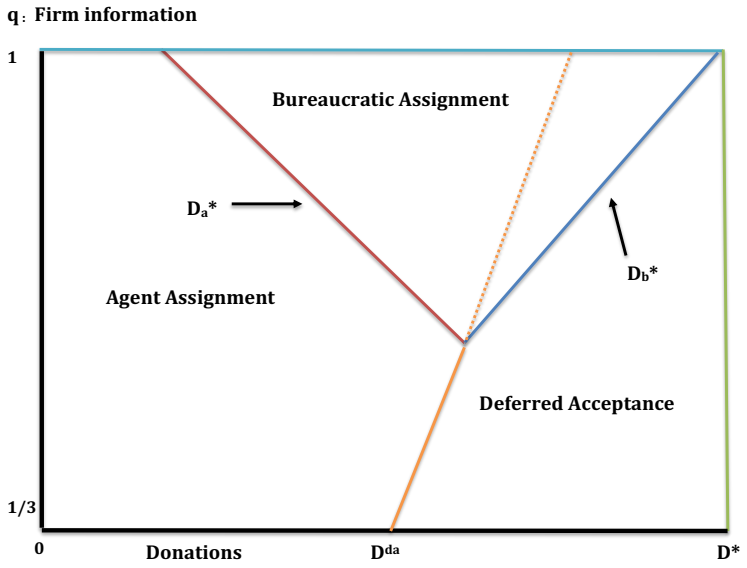


Figure 1: **Optimal Assigned Consumption.**

Figure 1 offers a taxonomy of Assigned Consumption when all three mechanisms are used. This figure holds $\hat{y}$ constant. D^* defines the maximum amount of donations that the firm would ever use ($D^* = \frac{v+T^{da}}{\delta}$). Two features beyond (17) and (18) are shown. First, the firm always prefers Agent Assignment over Bureaucratic Assignment for q low enough. Agent Assignment raises more funds, but at the cost of using worse information to assign treatments. As q falls, the value of better information declines, and $T^{da} - T^a \to 0$ as $q \to \frac{1}{3}$. Second, the firm always prefers Bureaucratic Assignment over Deferred Acceptance for q high enough. Deferred Acceptance incorporates client preferences not reflected in the firm's signal. As q increases, the value of the additional information provided by client rankings falls and $T^{da} - T^b \to 0$ as $q \to 1$ (the D_b^* line in Figure 1).[34] Deferred Acceptance arises among Assigned Consumption mechanisms only when funding is not a significant concern ($D \geq D_b^*$) and the firm is poorly

[33]The functional form of firm preferences is not innocuous. The separability between v_i and T has been chosen to reflect a reality of the non-profit, namely that its initial concern is to cover as many clients as possible, and only after that do issues of allocative efficiency become pre-eminent. With other functional forms, such as when the relationship between v_i and T is multiplicative, we would not get the same pecking order on D but the conceptual tradeoffs would remain.

[34]An even starker outcome arises as h varies, measuring the extent to which client preferences potentially offer valuable information for allocations. Specifically, as h gets small, the region of Figure 1 where Deferred Acceptance is used gets smaller, vanishing as $h \to 0$.

informed (q low). Random Serial Dictatorship is a common outcome in one-sided matching problems. This arises here only when $q = \frac{1}{2}$.

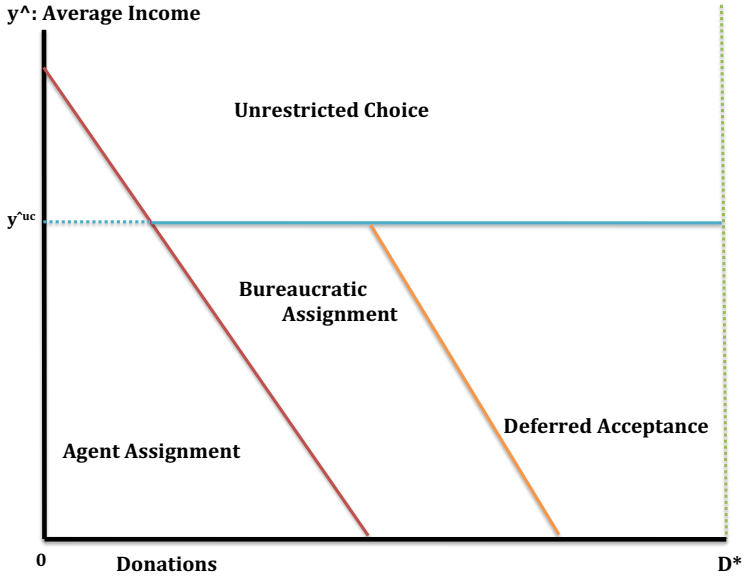


Figure 2: **The Source of Funding.**

The outcome depends not only on total funding for the firm but also its source. Figure 2 illustrates the outcome holding q fixed and varies each source of funding. When $\hat{y} < \hat{y}^{uc}$, the firm uses Assigned Consumption, in which case only total funding matters. The downward sloping lines below $\hat{y}^{uc}$ are combinations of donations and client revenue that hold total coverage constant. Deferred Acceptance is only used when donations are high, but client income is low. The source of funding matters when $\hat{y} \geq \hat{y}^{uc}$. Then the firm chooses Unrestricted Choice over either Bureaucratic Assignment or Deferred Acceptance. This does not arises as funds have marginally increased at this point but that Fairness has been satisfied. The firm may choose Agent Assignment even when Unconstrained Choice is possible. To see this, consider a case where D is low and $\hat{y}^{uc}$ is low so little revenue can be raised. Then if $w^* = 0$ but $\overline{w} > 0$ agents are willing to volunteer, but only when they are given autonomy. If there is an elastic supply of volunteers (or more generally low wage workers), it is simple to show cases where the firm uses Agent Assignment even when Unconstrained Choice is possible. (28% of Americans volunteered in 2022-23 (US Census Bureau, 2024)).

Worst Case Scenario: The market design literature typically uses an ordinal representation of utility. We use a cardinal measure where client choices are determined by marginal rates of substitution. We do so by necessity, as the marginal impact

of price changes on behavior is key. The results here rely on the firm having information about the preferences of potential clients. Yet one argument in favor of the ordinal representation of preferences is that it does not impose such constraints on the firm. Assume instead that all that the firm knows is that the marginal utility of clients is positive. Furthermore, many social organizations offer services to clients who cannot pay. The worst case scenario then is that $\underline{m} = \infty$, in which case the only price that satisfies Fairness is 0. Assume the firm prices for this worst case scenario. Then the firm never uses Bureaucratic Assignment, whose only benefit over Deferred Acceptance is more client revenue. The firm also can never use Unrestricted Choice as (9) always fails. The firm then chooses between Agent Assignment and Deferred Acceptance, with Agent Assignment for $D < D_a^{**}$, where

$$(T^{da} - T^a) = \frac{\Delta w}{w^* \overline{w}} \left(v + T^a - \frac{\delta(D_a^{**} + \frac{\Delta w}{2})}{2w^*} \right). \tag{19}$$

This is given by the hashed line in Figure 1. This extension at first glance may appear to offer a greater likelihood of Deferred Acceptance compared to the cases with positive prices. However, Deferred Acceptance is only chosen when the firm has enough donations that the highest priority clients are covered. When $p^A = 0$, fewer clients are covered for any level of donations. As a result, the required level of donations needed to transition to Deferred Acceptance is higher than above. Instead, it makes Agent Assignment more likely.

Central to outcomes above is that prices can vary with the service provided. Social organizations routinely vary prices for their services.[35] Zero prices is a special case of uniform pricing, a common practice among some social organizations.[36] Other than in the case where $\underline{m} = \infty$, uniform pricing is not an equilibrium strategy, but is a useful thought experiment. Under uniform pricing, the firm never uses Bureaucratic Assignment over Deferred Acceptance.[37]

The Alignment of Interests: One-sided matching is typically defined as settings where one side (the firm here) is indifferent to assignments (services here). This is what yields Serial Dictatorship, while two sided matching - where the firm cares -

[35] A mental health provider charges differently for therapy rather than medication. Larger houses, and those in better neighborhoods, have higher rents. Food pantries are typically offers some foods for free (produce, canned goods) but must pay for more desired goods (meat or fish).

[36] As an example from the public sector, the recipients of Section 8 housing vouchers pay approximately 30% of their income as rent independent of the housing unit they obtain.

[37] Uniform pricing raises another relevant conceptual assumption. The commonly used modeling device of liquidity constraints would not have a similar implication that clients pay more for more desired services. For instance, consider a case where all clients can only pay $1. Then the revenue of the firm per client is bounded by that dollar, independent of the service provided by the firm. This would yield a similar outcome to that under uniform pricing.

yields outcomes like Deferred Acceptance.[38] Our setting by that definition is two-sided, yet the outcome in many cases is the analog of Serial Dictatorship, namely Choice. Here the source of its use is not indifference but alignment of interests between the firm and client. When average income is high enough though (9), there is complete alignment through choice of prices, while when it lies below that the only possibility is partial alignment via Deferred Acceptance, but at the cost of lower prices and revenue. For low enough donations, there is no alignment.

A characteristic of social organizations is many competing interests and objectives: those of clients, a desire that outcomes should be consistent, and those of the organization's employees. The analysis above offers a priority ranking for these based both on the level of funding, where less funding first prioritizes employees and bureaucratic consistency, while either high level of donations or client payment lead to client preferences becoming pivotal.

The social sector is complex and varied, ranging from homeless shelters to elite private universities and museums, and no single paper can entirely capture that complexity. We have addressed a canonical setting where a firm allocates services to needy clients, in a setting where some services are more desirable. For instance, Covid-19 patients preferred Pfizer to Astra Zeneca, food pantries typically choose meat over canned goods for their clients, and houses in better neighborhoods are preferred to those in worse neighborhoods. While prices could be used to screen for familiar reasons, the firm is concerned that wealthier clients should not disproportionately benefit.[39]

Prices play three roles in the canonical economic transaction. First, efficient pricing allows the firm to delegate choice to clients rather than being assigned by a third party. Second, prices serve as information to coordinate the actions of suppliers. In the famous words of Hayek, they "act to co-ordinate the separate actions of different people". Finally, they pay the firm's bills. All three roles for prices fail in different ways above. The failure of the Fairness Condition implies the need for Assigned Consumption. Second, the reason why Agent Assignment arises above is because of disagreement between the firm and agents over the appropriate treatment of clients. Yet this arises because clients do not reveal their preferences through choice. Finally, the reason that the firm choose Agent Assignment or Bureaucratic Assignment over Deferred Acceptance is because prices don't pay the bills. In this sense, the paper

[38]The balance of demand and supply implies that there is no need for the common outcome of Random Serial Dictatorship in allocating service. Serial Dictatorship may also be preferred due to its ex ante combination of fairness and efficiency.

[39]This was the central reason why choice was not used in two market design settings that involved the author. Feeding America was aware that choice with pricing in fiat money could avoid many of the inefficiencies of assigning food to food banks. Yet it did not do so as some food banks had more monetary resources than others. Similarly, Covax was also aware that pricing could reduce the frequency of countries taking vaccines that they would not subsequently distribute. Yet if they were to price vaccines, higher income countries would disproportionately benefit.

offers a theory of the social organization based on the failure of the three roles of pricing.

1.5 Other Forms of Choice

When $I_i \in \{0,1\}$ and $\hat{y} < \hat{y}^{uc}$, the firm uses Assigned Consumption. We now show how Choice can arise for lower income levels in two settings. First, when services are divisible, such as where a student takes many courses, the frequency of therapy sessions can be chosen, or where a food bank receives many pounds of food. We show that Restricted Choice can occur for any income level. Second, we consider a case where there are more than two goods, where Partial Choice allows the client to choose some services but not others. This is implemented by a non-strategy proof First Choice mechanism.

1.5.1 Restricted Choice

As an initial thought experiment, consider a case where the firm can degrade services, where instead of offering $I_j = 1$, it can choose a worse version $I_j = \kappa_j \in [0,1]$ at no cost. Client utility becomes $U_i^c = (T(\tau_i, t_j) - m_i p_j)\kappa_j$. Consider the κ_j that are consistent with Choice for any given income level, where $I_0 = 1$, $I_1 = \kappa_1 \leq 1$. We show in the Appendix that with per unit prices P^*, Choice satisfies the Fairness Constraint for any $\kappa_1 \leq \kappa_1^*$ where

$$\kappa_1^* = \frac{\underline{\phi}}{\mu - \frac{\overline{m}\underline{\phi}}{\underline{m} - \overline{m}}}. \tag{20}$$

This shows how choice can be implemented for any income level. However, the distortion $1 - \kappa_1^*$ is higher at lower levels: it is decreasing in $\frac{\overline{m}}{\underline{m}}$ and is 0 at $\hat{y} = \hat{y}^{uc}$. Expected treatment surplus is $\tilde{T}^{rc} = \frac{\hat{\phi} + \kappa_1^*(\hat{y})(\hat{\phi}+\mu)}{2}$. This shows how the firm can implement $\tau_i = t_i$ but at the cost of distorting services.

Now consider a setting where instead of degrading services, they are divisible $I_i \in [0,1]$. The client then chooses between *quantities* $I_0 = 1$, and $I_1 = \kappa_1^*$. (20) still applies. The only difference is that there is now residual supply of service $t = 1$ beyond client $I^{rc} = \frac{D}{\overline{w} - p^{rc}}$. After client $i = I^{rc}$ there is no more remaining supply of $t = 0$, but there is $I^{rc}\frac{1-\kappa_1}{2}$ of $t = 1$. The firm allocates $t = 1$ to all clients from I^{rc} to $I^{rc}(1 + \frac{1-\kappa_1}{2})$.

Proposition 6 *Let $\hat{y} < y^{uc}$ and $I^{rc} = \frac{D}{\overline{w} - p^{rc}}$. Assume that $P = P^*$ for all clients*

$i \leq I^{rc}$. *Surplus from Restricted Choice is*

$$S^{rc} = I^{rc}\left(v + \mu + \hat{\phi}(1 - \frac{1-\kappa_1^*}{2}) + \frac{1-\kappa_1^*}{2}\frac{\hat{\phi}}{2}\right) - \delta\frac{I^{rc2}}{2}\left((1 - \frac{1-\kappa_1^*}{2}) + (1-\kappa_1^*)(1 + \frac{1-\kappa_1^*}{4})\right). \quad (21)$$

where $p^{rc} = (1 - \frac{1-\kappa_1^*}{2})p^{uc} + \frac{1-\kappa_1^*}{2}p^{ac}$.

The outcome has three notable components. First, clients $i \leq I^{rc}$ choose efficiently $t_i = \tau_i$ and pay unit prices P^*. Second, all covered clients $i > I^{rc}$ receive $t = 1$ so expected surplus is lower. Revenue is a weighted combination of that from Unrestricted Choice and Bureaucratic Assignment. A new issue is *client targeting*. This is given by the second bracketed term in (21) and reflects the fact that services are now spread beyond the highest priority clients. This cost is decreasing in κ_1 and hence in $\hat{y}$. This is what we mean below by poor targeting of clients. We give empirical examples of such targeting issues below.[40]

Poor targeting leads to a more ambiguous ranking of this choice outcome to Assigned Consumption. When Unrestricted Choice satisfies fairness it always is preferred to Deferred Acceptance or Bureaucratic Assignment, this is not the case for Restricted Choice. It is both intuitive and straightforward to show that for δ low enough, the firm may prefer either of these Assigned Consumption mechanisms to Restricted Choice as it ensures complete coverage of the highest priority clients.

Artificial Currency: Restricted Choice screens on marginal rates of substitution. Given this, it should not be surprising that there is a relationship to competitive equilibrium. The outcome above implements competitive equilibrium with equal incomes (Foley, 1967). We show this in the Appendix, where the firm offers supply of $t_0 = \frac{I^{rc}}{2}$ and $t_1 = \frac{\kappa_1^* I^{rc}}{2}$ and gives all clients $i = 1, 2, ...I^{rc}$ one unit of artificial currency to bid on services. The outcome above is an equilibrium where each client chooses between $I_0 = 1$ and $I_1 = \kappa_1^*$. Competitive equilibrium with artificial currency is the most common market design alternative to matching without prices (Budish, 2011, Budish and Cantillon, 2012, Prendergast, 2017, 2022, Goedde, 2024, and Wong, 2024).

In many of the applications where artificial currency has been used, it arises as outcomes would be seen as unfair if fiat money were used instead. For instance, colleges would worry about wealthier students getting better classes, and Feeding America was concerned about better resourced food banks receiving higher quality

[40]Proposition 6 assumes that the firm charges P^*. The issue with fairness is to induce clients with more income to choose $t = 0$ over the other services. There are two ways to do this: reducing quantities (κ_i) and reducing p_0. This implies that there is one other possible equilibrium outcome, where $p_0 = 0$. This is done to increase κ_i above the levels in (20). This does not change the tradeoffs, and is described in the Appendix.

food. Artificial currencies are used implement a fair playing field. Such inequalities from using fiat money underlie much of the analysis here. This section offers a link between the two existing mechanisms typically used. Yet the link is continuous in client income, where lower client income limits fiat money payments, resulting in more distorted service assignment. Furthermore, it shows how the outcome converges to Unrestricted Choice as $\hat{y} \to \hat{y}^{uc}$.[41]

We give two examples of the client targeting problem. First, Feeding America allocates food using an artificial currency, where food banks choose between higher quantities of a less desired good and lower quantities of a more desired one. The outcome with bidding with artificial currency is a Restricted Choice outcome, with prices (the analog to $\frac{1}{\kappa_i^*}$) given in Figure 2, taken from Prendergast, 2022. For instance, for every pound of produce ($I_i = 1$), a food bank can get $\frac{1}{7}$ pounds of processed fruit or $\frac{1}{50}$ pounds of cereal. Before introducing this system, the firm assigned the analog to $I_i = 1$ randomly. Under that system, a food bank received 1 pound of cereal. Now it gets $\frac{1}{50}$ of a pound. Where does the other $\frac{49}{50}$ go, or more generally how does the Restricted Choice mechanism reallocate food across the national network? Feeding America had a desire that the amount of food would be reoriented towards smaller, poorer food banks.[42] The distributional impact of the Feeding America design reinforced its allocative benefits. Small food banks increased their food consumption by 72% compared to the ex ante allocation, at very little cost to nutritional value (Prendergast, 2022). In the language of the model, targeting improved though the move to Restricted Choice.

A different outcome arose for vaccine distribution during the Covid-19 pandemic. The international agency Covax allocated over a billion vaccines to 125 countries after early 2021. Here $S = \sum_{i=1}^{N}(v_i + T(\tau_i, t_i))I_i$ represents different characteristics of vaccines t_i to be allocated to the countries (i) needing those attributes most τ_i, where v_i represents Covax view of country priority, and I_i is coverage of vaccines in country i. Covax chose to allocate vaccines in a manner that largely ignored vaccine attributes or priority. Instead, it offered all vaccines to all countries in a lock-step fashion (Budish et al., 2022). A problem that arose was that many vaccines were close to their expiry date. In the language of the model, $t = 0$ vaccines were close to expiry while $t = 1$ vaccines were far from their expiry date. Countries varied in their capacity to quickly distribute vaccines. As an example, Turkey's health care system was far more capable of responding than that of the typical West African nation. The

[41] Restricted Choice arises above when services are divisible. We do not allow clients to purchase lotteries. If we did, an alternative would allow the κ_i^* to be probabilities of receiving the service, whether divisible or not. Screening can then arise through single crossing in probabilities. This approach would mimic Hylland and Zeckhauser, 1979. In reality most social organizations simply will not fail to offer service to its most needy clients *ex post*, even if it has appealing *ex ante* properties.

[42] The democratic nature of Feeding America restricted them beforehand to treat all areas equally (conditional on statistics measures of need), but Feeding America's v_i would have focused more on smaller food banks.

logic of economic efficiency suggests distributing more "close to expiry" vaccines to Turkey than to West Africa. The allocation system did not use expiry date in the allocation mechanism. This led to remarkable waste, with estimates of almost 40% of all vaccines sent to sub-Saharan Africa not being used.[43]

Covax did not use expiry date because of a political issue deriving from countries not wanting an allocation mechanism that disproportionately gave them soon-to-expire vaccines. However, Restricted Choice mechanisms are designed to solve such problems, by offering countries a choice between a larger number of soon-to-expire vaccines and a smaller number of vaccines far from expiry. With an appropriately chosen rate of exchange (the κ_i^* of the model above), both sets of countries benefit. Covax did not pursue this. A likely concern was the perception that Covax was failing to focus on the most-needy countries, by giving higher income members *more* vaccines than the lower income countries of sub-Saharan Africa. Once again, this example focuses on how Restricted Choice mechanisms rest partially on how they affect the perception of targeting.[44]

1.6 Partial Choice and More Services

So far, we have assumed two services, $t_i \in \{0, 1\}$. This allows to describe all mechanisms above except Partial Choice. To address this and the generality of the insights, now assume that the firm provides n treatments $j \in \{0, 1, ..., n-1\}$ to clients $i \in \{0, 1, ..., n-1\}$, both uniform and balanced, with returns given by (1). Further assume that t_i^a is correct with probability q, and equal to $t \neq t_i$ with probability $\frac{1-q}{n-1}$.[45] The analog to Assumption 1 is now given by:

Assumption 1a: $\overline{\phi} > n\mu$ and $\underline{\phi} < \mu$.

It is simple with this extension to generate the analog to Lemma 1 below:

Lemma 1a:

- *Unrestricted Choice satisfies the Fairness Constraint if and only* $\frac{\overline{m}}{\underline{m}} \geq \frac{(n-1)\mu - \underline{\phi}}{(n-1)\mu + \underline{\phi}}$. *When* $\frac{\overline{m}}{\underline{m}} \geq 1 - \frac{\underline{\phi}}{(n-1)\mu}$, *maximum prices are* $p_j^{uc} = \frac{j\mu + \underline{\phi}}{\underline{m}}$.

[43] This led to well-publicized complaints from the African Union that "the majority of the donations have been ad hoc, provided with little notice and short shelf lives" (African Union, 2021).

[44] There is an important difference between both of these applications and the model above. The model assumes that the firm allocates services in the order of *its* priority ranking. Political considerations implied that both Feeding America and Covax had to offer equal access. Yet Feeding America's priorities were to reallocate food to poorer food banks, and Covax wished to vaccinate the poorest countries.

[45] Here the analogous condition for ρ is $\rho \geq \frac{1}{n}$.

- *Deferred Acceptance can be implemented for any income level. For* $\frac{\overline{m}}{\underline{m}} < \frac{n\mu-\underline{\phi}}{n\mu+\underline{\phi}}$, *maximum revenue is* $p_j^{da} = \frac{j(\mu-\underline{\phi})}{\underline{m}}$.
- *Bureaucratic Assignment and Agent Assignment can be implemented for any income level. Maximum revenue is* $p_j^{ac} = \frac{j\mu}{\underline{m}}$.

This shows how the tradeoffs above generalize to more treatments. We also show in the Appendix that the Deferred Acceptance outcome has the same structure as above. Expected revenue is now given by

$$p^A = \begin{cases} \frac{(n-1)\mu}{2} + \underline{\phi} & \text{if } A = uc \\ \frac{(n-1)\mu}{2} & \text{if } A = b, a \\ \frac{(n-1)(\mu-\underline{\phi})}{2} & \text{if } A = da \end{cases} \tag{22}$$

This has the same features as the tradeoff above. It is both intuitive and simple to show that the allocative efficiency retains the ranking above, so the tradeoff between the mechanisms retains the features of the two state example above. However, it adds one potential additional optimal mechanism, Partial Choice.

Partial Choice: With Unrestricted Choice, if the client does not choose $t = 1$, she can choose $t = 0$. This is not the case with Assigned Consumption, where if the client does not consume the treatment assigned, she consumes nothing. However, in many social settings this is not the case. One case is social housing when a client is matched to a housing unit. The unit may not be the client's most desired type or in her preferred location. If she turns the offer down, she is not excluded from consuming a house. Instead, the offer typically comes with a cost: if the client turns down a "reasonable" offer, she is placed lower on waiting list going forward. With repeated turn-downs, she may be be taken off the list (see for example *citizensadvice.org.uk*). As a result, the client needs to strategize about whether to accept. Partial Choice arises when some services are chosen and some are assigned. This cannot happen when there are two treatments, because by definition if one is chosen, the remaining one must be the alternative. To address the possibility of Partial Choice, let $n = 3$.[46] Furthermore, let $\hat{y} < \hat{y}^{uc}$ and $I_i \in \{0, 1\}$, so neither Unconstrained nor Restricted Choice can be used.

The novel mechanism here is where the client can choose one t but where the alternative is an assigned mechanism over the other two services: Partial Choice. In our example above, assume that in addition to Chicago ($t = 0$) and New York ($t = 1$), we add Boston ($t = 2$).[47] Purely for simplicity, we assume that the assignment component of the mechanism is Bureaucratic Assignment. Partial Choice can then be

[46] In this case $\hat{y}^{uc} = \frac{\overline{m}}{\underline{m}} \geq \frac{2\mu-\underline{\phi}}{2\mu+\underline{\phi}}$ and Assumption 1 is $\overline{\phi} > \mu$ and $\underline{\phi} > 2\mu$.

[47] In any of the assignment mechanisms that allowed client input, the client always received Chicago ($t = 0$) in equilibrium if it was ranked first.

implemented by a First Choice mechanism, which we now describe.

The firm knows I^{1st} when allocating t_i. It offers $I_i = 1$ to all $i \leq I^{1st}$. The First Choice Mechanism is:

1. The firm chooses $P^{1st} = \{p_0^{1st}, p_1^{1st}, p_2^{1st}\}$.
2. Each client $i \leq I^{1st}$ reports $\hat{t}_i$.
3. The firm allocates $t = i$ to all clients k for which $\hat{t}^k = i \succ ...$ if i is undersubscribed.
4. For all $t = i'$ that are oversubscribed, the firm implements Bureaucratic Assignment among those where $\hat{t}^k = i' \succ ...$

In the Appendix, we show that there is only one equilibrium of this form, where any client can choose $t = 1$ (New York) with the other two assigned. A Partial Choice outcome arises when only type $\tau = 1$ does so. Proposition 7 shows the outcome.

Proposition 7 *A Partial Choice equilibrium exists with $T^{1st} > T^b$ and $p^{1st} > p^{ac}$ if $\hat{y} \geq \hat{y}^{1st}$, where $\hat{y}^{1st} < \hat{y}^{uc}$. $\hat{y}^{1st}$ is increasing in q.*

The value of Partial Choice is that it reduces assignment error from Assigned Consumption, as some clients choose efficiently. Partial Choice is not strategy proof. Non-strategy proof First Choice mechanisms are well known in the market design literature. See Abdulkadiroğlu et al., 2011, Agarwal and Somaini, 2018, Calsamiglia et al., 2019, and Kapor et al., 2020, for differing views on the merits of this over Deferred Acceptance. The central outcome here is not that its analog, Partial Choice, can result in higher utilitarian allocative efficiency. Instead, there are two contributions. First, a Partial Choice equilibrium exists for lower client income than is needed for Unconstrained Choice.[48] For $q < q^* = \frac{2\mu+\underline{\phi}}{3(2\mu-\underline{\phi})}$, the Partial Choice outcome exists for any income level. Second, Partial Choice can raise more revenue than Bureaucratic Assignment. This arises if

$$\frac{\overline{m}}{\underline{m}} > 1 - \frac{(3q+1)\underline{\phi}}{(3q-1)2\mu}. \tag{23}$$

When $q = 1$ this is identical to the definition of $\tilde{y}^{uc}$ above.

Social organizations routinely offer non-strategy proof choices akin to this. A case familiar to readers may be students applying to universities via Early Decision. If the student is accepted by that school, she is obliged to accept the offer.[49] While this is

[48] The reason is simple: with Unconstrained Choice, the wealthier client has to be deterred from choosing $t = 2$ for sure when $\tau \neq 2$. This is not the case here - if the client does not choose $t = 1$, the alternative is random assignment, which is less valuable than the preferred one for sure.

[49] This is not a legally binding commitment, but is treated as binding by most students. They can also apply through Regular Decision, where the student is not committed to a school if accepted.

not Partial *Choice* in the sense above, probabilities of acceptance are often four times higher than though the regular process. The use of ED has increased significantly over the last decade (Murphy, 2022).[50] The conceptual novelty here is that the firm uses Partial Choice to generate more funding than Assigned Consumption. It is well known that Early Decision students pay higher average tuition (Antecol and Smith, 2012, Murphy, 2023). Also consistent with the model is White, 2024, who provides data that colleges with lower endowments (the analog to D in the model) fill a larger fraction of their class through Early Decision.

To summarize, we have offered a theory of the non-profit built on four components. First, for the last four decades the gold standard has been the market design literature on matching without prices. Yet one worker in ten works in a sector that ranks funding as their greatest concern and raises half its revenue from clients. Adding the idea of fair pricing, we have proposed a larger set of optimal assignment mechanisms. Second, we have assumed that the motives of the firm are "pure", only interested in client welfare. This assumption has been made to understand the complexity of the firm's tradeoffs even in the absence of agency concerns or profit motives. Third, we have assumed that agents are motivated by their impact on client welfare and that the incidence of any benefits is on them. This can easily be relaxed by assuming that $\alpha = 0$, in which case Agent Assignment would never be used. Finally, we have assumed fair pricing. We relax this below.

The key conceptual issue that generates the pecking order above is the relationship between funding and allocative efficiency. A striking difference arises between Assigned Consumption and Choice. For assigned consumption, coverage is ranked as Agent Assignment $\succ$ Bureaucratic Assignment $\succ$ Deferred Acceptance, so the two objectives are substitutes. By contrast, in the Choice domain, Unrestricted Choice $\succ$ Restricted Choice, and Unrestricted Choice $\succ$ Partial Choice, so the objectives are complementary.

1.7 Relaxing the Fairness Constraint

The assumption of Fairness assumes that no client can be denied any coverage for lack of income. It plays an important role in tying down prices and the ability to use Choice. We now discuss its importance and how to relax it. To address this, it is necessary to assume a distribution over m_i. Assume that a fraction $G(m)$ of clients have marginal utility below m, where $G(\underline{m}) = 1$, $G(\overline{m}) = 0$ and $G' \geq 0$.

Non-Negative Prices: The outcomes here arise from an inequality-averse firm assigning services to a privately informed client when funding is valuable. Dworczak, 2024, uses similar assumptions to address when rationing is needed rather than using

[50] Avery and Levin, 2011, show that it can be used to screen students on their enthusiasm for a school in a conceptually similar way to that described above.

choice through the price system. We discuss this here as his work addresses the tradeoff between pricing, revenue and coverage. He also shows that low inequality (reflected in less variation in willingness to pay) allows the use of prices while rationing arises with more inequality. The alignment of choice and low inequality arises for a different reason than here. In his work, the "firm" can compensate the poor (those with high social welfare weights) with *money* from the revenue obtained from wealthy clients purchasing. The price system optimally arises when the preferences of the principal are for wealthier clients (high willingness to pay) to purchase with higher probability. In itself, this harms inequality but this is resolved by the firm distributing its profit as a lump sum monetary transfer to those who did not purchase. This is equivalent to allowing negative prices above, which we have ruled out.

Our interest so far is in non-profits. Money transfers from non-profits to clients are very rare. Per IRS rules, they can legally only be used when they align with the charitable function of the non-profit. Examples of such cases would be disaster relief funds, payment of scholarships or emergency financial assistance where a non-profit pays bills like rent or utilities.[51] Outside of these kinds of settings, they cannot be used. In addition, even if such transfers had no legal problems, many non-profits are paternalistic and would be reluctant to give money to clients for fear it would be used in ways that do not align with their interests. For this reason, we have assumed that the only transfer that can be made by the firm to clients is providing services, and the cost of pricing high, which we now turn to, is clients who do not benefit.[52]

Pricing to exclude low income clients: The Fairness Constraint implies that the firm cannot sacrifice serving a high priority client to serve more lower priority clients. Some social organizations have no choice but to exclude some low income clients.[53] To address this, consider a relaxation of the Fairness Constraint such that the firm excludes some low income clients from service but treats all those served fairly. (A familiar example from our own profession is where admission rates to elite universities are higher for those who pay full tuition or whose families are donors, but access to classes is equal across all students.) We now assume a relaxed Fairness Constraint that the fraction of clients receiving no service cannot exceed $\Gamma > 0$. We retain the assumption that conditional on $I_i = 1$, t_i is independent of m_i.

Previously the firm normalized prices by $\underline{m}$. Let it now do so by $M \leq \underline{m}$. This does not change the conceptual tradeoffs above but instead renormalizes the support of clients. To see this, first consider Unrestricted Choice, where prices are $p_0 = \frac{\phi}{M}$

[51] See for example wscadv.org/wp-content/uploads/2017/08/Grants-to-Individuals.pdf.

[52] Another distinction from Drowczak is that a central point here is not simply that rationing arises when equity concerns rule out Choice, but rather *how* services are rationed, providing an avenue to the literature on matching without prices.

[53] As an example, private housing associations will not rent units to those who cannot pay at all, but instead have a minimum rent.

and $p_1 = \frac{\underline{\phi}+\mu}{M}$.[54] The fraction of clients not served is Γ. This is defined by $\Gamma = (1-h)(1-G(M)) + h(\frac{(1-G(\tilde{m}_0))+(1-G(\tilde{m}_1))}{2})$, where $\tilde{m}_0 = \frac{M\overline{\phi}}{\underline{\phi}}$ and $\tilde{m}_1 = \frac{M(\overline{\phi}+\mu)}{\underline{\phi}+\mu}$.[55] This implies that the support of marginal utilities is changed to $[\overline{m}, \tilde{m}_0]$. The qualitative results remain unchanged. Revenue is now $\tilde{p}^{uc} = \frac{\frac{\mu}{2}+\underline{\phi}}{M}$. The expected number of clients before one is served is $\frac{1}{1-\Gamma}$, so this changes the discount factor over clients from δ to $\tilde{\delta} = \frac{\delta}{1-\Gamma}$ and surplus is

$$\tilde{S}^{uc} = \frac{D}{\overline{w} - \tilde{p}^{uc}} \left(v + T^* - \frac{\tilde{\delta} D}{2(\overline{w} - \tilde{p}^{uc})} \right). \tag{24}$$

Compared to (16), prices are higher but the discount factor increases. The same logic arises with the other assignment mechanisms. For example, Deferred Acceptance prices are now $\tilde{P}^{da} = \{0, \frac{\mu-\underline{\phi}}{M}\}$ where here $1-\Gamma = \frac{1}{2} + \frac{(1-h)(1-G(M))+h(1-G(\tilde{m}_{da}))}{2}$, where $\tilde{m}_{da} = \frac{M\overline{\phi}}{\underline{\phi}-\mu}$. The conceptual tradeoffs above remain: instead, the outcome simply truncates low income clients.[56]

When wealthy clients get better services: Now consider the inverse violation of Fairness to that above where all clients are served, but better service ($t = 1$) is given to to wealthier clients. Here the Fairness Constraint is simply that I_i is independent of m_i but t_i can depend on m_i. The client plays no role in both Agent Assignment and Bureaucratic Assignment, so this makes no difference to outcomes there. In each case, the market clears and Individual Rationality binds so no changes arise. With Unconstrained Choice and Deferred Acceptance, the firm can now charge more. It can charge up to p_1^* defined by $1 = h(G(m_{00*}) + G(m_{11*})) + (1-h)(G(m_{01*}) + G(m_{10*}))$, where $m_{00*} = \frac{\mu-\overline{\phi}}{p_1^*}$, $m_{11*} = \frac{\mu+\overline{\phi}}{p_1^*}$, $m_{01*} = \frac{\mu-\underline{\phi}}{p_1^*}$ and $m_{11*} = \frac{\mu+\underline{\phi}}{p_1^*}$. At this price, demand is exactly $\frac{1}{2}$ per client. The cost of this is that the price of $t = 0$ is $p = 0$.

The impact of this change is ambiguous for Unrestricted Choice: the firm can charge more for $t = 1$ but must reduce the price for $t = 0$. This is not the case

[54] We also need to extend the equilibrium strategy of the client to allow $\hat{t} = \emptyset$ where the client choose no service.

[55] All covered clients with $m_i \leq M$ and $\phi = \underline{\phi}$ choose $\hat{t} = 1$ and all covered clients with $m_i \leq \tilde{m}_j$ and type $\phi = \overline{\phi}$ choose $\hat{t} = 1$, where $\tilde{m}_j = \frac{M(\overline{\phi}+j\mu)}{\underline{\phi}+j\mu}$. The necessary condition for Unrestricted Choice is now $\frac{\overline{m}}{\tilde{m}_0} \geq \frac{\mu-\underline{\phi}}{\mu+\underline{\phi}}$.

[56] This simply adds an additional choice for the firm. When δ is high, the firm will gravitate towards the outcomes described in the body of the paper. The opposite extreme is Milton Friedman's famous claim that the firm should maximize profits, and use those profits for socially responsible actions if desired. Consider the case where $\Gamma = 1$ and $\delta \to 0$. Then the optimal strategy for the firm is to maximize revenue by charging $p_0 = \frac{\overline{\phi}}{\overline{m}}$ and $p_1 = \frac{\overline{\phi}+\mu}{\overline{m}}$. These prices are only paid by the wealthiest clients ($\overline{m}$) with the highest valuations ($\overline{\phi}$). This yields the highest revenue and covers the most clients, though only the wealthiest ones.

for Deferred Acceptance. Because we no longer require Fairness, the revenue from Deferred Acceptance increases. More importantly, there is a range of prices under which both revenue and allocative efficiency rise, but at the cost of unfairness. To see this, consider the firm increasing the price from the price above $p_1^{da} = \frac{\mu - \underline{\phi}}{\underline{m}}$ to a price $\tilde{p}$ where $\frac{\mu - \underline{\phi}}{\underline{m}} < \tilde{p} \leq \frac{\mu}{\underline{m}}$. Then revenue rises, but so also does allocative efficiency as the only clients who no longer report $\hat{t} = \{1 \succ 0\}$ are clients of type $\tau = 0$. As a result, pricing above p_1^{da} improves both revenue and allocative efficiency. The problem is that the outcome is unfair, as high income $\tau = 0$ clients are more likely to consume $t = 1$ than low income clients. As a result, this section points to a tradeoff between revenue, allocative efficiency and fairness.[57]

1.8 Observations on the Non-Profit

Here we discuss some modeling choices and examples.

More General Technology: The ability of the client to reveal information was characterized by two results above: (i) Unrestricted Choice can only be used for income high enough, and (ii) Deferred Acceptance reduces revenue. This was shown using preferences in (1). We now show necessary and sufficient conditions for these outcomes with a more general technology. Assume $T(k, j)$ for the horizontal ($\overline{\phi}$) types is $\overline{\mu}_{kj}$, and $\underline{\mu}_{kj}$ for the vertical ($\underline{\phi}$) types. The analog to Assumption 1 is (i) $\overline{\mu}_{kk} \geq \overline{\mu}_{kj}$, and (ii) $\underline{\mu}_{k1} \geq \underline{\mu}_{k0}$. Screening satisfies the Fairness Constraint in the analog above for Unrestricted Choice with prices $P^* = \{\frac{\underline{\mu}_{00}}{\underline{m}}, \frac{\underline{\mu}_{11}}{\underline{m}}\}$ if $\frac{\overline{m}}{\underline{m}} \geq \frac{\underline{\mu}_{01} - \underline{\mu}_{00}}{\underline{\mu}_{11} - \underline{\mu}_{00}}$.

Now consider revenue from Deferred Acceptance. Individual rationality with Assigned Consumption implies that maximum revenue is $p_j = min_i \underline{\mu}_{ij}$ so $p^{ac} = \sum_j min_i \underline{\mu}_{ij}$. Let $p_0 = min_i \underline{\mu}_{0i}$. Then incentive compatibility under Deferred Acceptance implies that for a vertical ($\underline{\phi}$) client to rank $1 \succ 0$, the maximum price of $t = 1$ is $p_1 = min_i(\underline{\mu}_{i1} - \underline{\mu}_{i0}) - p_0$. Deferred Acceptance does not reduce revenue relative to the other Assigned Consumption mechanisms if

$$min_k \underline{\mu}_{k,j} - min_k \underline{\mu}_{k,j'} = min_k(\underline{\mu}_{k,j} - \underline{\mu}_{k,j'}). \tag{25}$$

This arises if (i) $\underline{\mu}_{ij}$ is non-decreasing in j for all i, j, and (ii) $\underline{\mu}_{(i+1)j} - \underline{\mu}_{ij}$ is non-decreasing in j. Lemma 2 follows immediately. Then Deferred Acceptance does not reduce client revenue if and only if type (τ) and services (t) are weak complements

[57] Fairness also matters for the priority ranking above. For instance, consider a social organization with low levels of donations. Above, the Fairness Constraint implied that the only avenue to getting greater client coverage was to delegate control to the agent, giving rise to Agent Assignment as the preferred mechanism. However, in a setting where the firm can offer better treatment to wealthier clients, this becomes particularly appealing. As a result, the assumption of fairness is central to the pecking order above.

for the $\underline{\phi}$ client. If horizontal differentiation is limited to this extent, the firm never chooses Bureaucratic Assignment over Deferred Acceptance. Yet note that if weak complements holds for all agents, there is no value to ordinal rankings as all agents prefer $t = 1$.

Unbalanced Demand and Supply: We have assumed that the supply of services is uniform and matches the distribution of types. This assumption eliminates the need for rationing other than with Deferred Acceptance. Second, Unrestricted Choice implies that all clients receive their preferred service. The qualitative results extend to unbalanced demand and supply. In the Appendix, we show this when there an excess of $t = 1$ slots relative to $t = 0$ but the analysis is identical for the other case. Assume that the distribution of types $\{\tau_0, \tau_1\}$ is not uniform $(\frac{1}{2}, \frac{1}{2})$ but instead $\{\frac{1-\zeta}{2}, \frac{1+\zeta}{2}\}$ for $\zeta > 0$. The qualitative outcomes remain unchanged in the sense that (i) Unrestricted Choice is feasible if (9) holds, (ii) the revenue from Deferred Acceptance is p^{da} and the revenue from Bureaucratic and Agent Assignment is p^b, and (ii) there is still a wage reduction from Agent Assignment. This leads to the same qualitative outcomes. The only substantive differences are (i) Unrestricted Choice involves rationing as there is excess demand for $t = 1$ at the price of $\frac{\mu+\underline{\phi}}{\underline{m}}$ (ii), $p_0 = 0$ under Unrestricted Choice as some $\tau = 1$ clients are mismatched, and (iii) surplus is lower as the first best entails $\frac{\zeta}{2}$ type $\tau = 1$ clients being assigned to $t = 0$. With unbalanced demand and supply, choice becomes analogous to the familiar Random Serial Dictator mechanism.

Where m_i and τ_i are not Independent: Consider now a case where τ_i and m_i are not independent. Specifically, let a client with type τ_i have marginal utility that is distributed over a support $[\overline{m}_i, \underline{m}_i]$. Then prices for Bureaucratic and Agent Assignment are $p_0 = 0, p_1 = \frac{\mu}{max\{\underline{m}_0, \underline{m}_1\}}$, prices under Deferred Acceptance are $p_0 = 0, p_1 = \frac{\mu-\underline{\phi}}{max\{\underline{m}_0, \underline{m}_1\}}$, and the analog to (9) is $\frac{\overline{m}_0}{\underline{m}_1} \geq \frac{\mu-\underline{\phi}}{\mu+\underline{\phi}}$. This changes qualitative outcomes only if $\frac{\overline{m}_0}{\underline{m}_1} > 1$ for all income levels, in which case Unrestricted Choice is always possible. This is because the type $\tau = 0$ client is always too poor to mimic the $\tau = 1$ allocation as the utility cost of the price to be paid $\frac{\mu+\underline{\phi}}{\underline{m}_1}$ is too high. Otherwise the results remain unchanged.

This extension also allows us to simply consider cases where v_i and y_i are correlated. For instance, the firm may have objectives to prioritize the poorest clients. We have not addressed this because there is now a correlation between funding and the average income of clients served. However, this link is mechanical and simply means that the analysis would need to condition average income for each level of donation and prices paid, but does not change the analysis.

Excluded Mechanisms: We have excluded two classes of mechanisms: those where clients purchase lotteries, and when the firm does not distribute some services. Both allow us to ignore what we call *Cash Cow* mechanisms.

The price under Unconstrained Choice above was bounded above by $p_1 = \frac{\underline{\phi}+\mu}{\underline{m}}$. This is because the firm needed to fill half of its slots so that even under Unconstrained Choice, they could not price beyond $p_1 = \frac{\underline{\phi}+\mu}{\underline{m}}$. Yet there are $\frac{h}{2}$ clients who have valuations given by $\overline{\phi}$ rather than $\underline{\phi}$. Choosing a price of $p_1 = \frac{\overline{\phi}+\mu}{\underline{m}}$ is akin to choosing a monopoly price which raises more revenue but at the cost of excluding some from service. We call these Cash Cow mechanisms as the highest valuation clients are being used to cover more clients. We have assumed that purchasing lotteries is not possible and that services must be distributed to ignore this outcome.

To see this, first consider now a case where the firm prices at $p_1 = \frac{\overline{\phi}+\mu}{\underline{m}}$. Only the highest valuation client (the Cash Cow) with $\tau = 1$ and $\phi = \overline{\phi}$ will purchase at that price. The firm now has excess supply of $t = 1$ of $\frac{1-h}{2}$ per client. One possibility is to not assign that, but instead to assign $t = 0$ to every other client. For $\overline{\phi}$ high enough, this can cover enough type $\tau = 0$ clients that it is the preferred mechanism, and does not violate Fairness. However, it relies on having a set of high "willingness to pay" clients that allows the firm to fund all other clients. This is far from the setting that most social organizations find themselves in. Furthermore, allowing this mechanism does not add anything conceptually. The assumption that all services must be allocated allows us to ignore the monopoly solution of pricing high and denying some services.

Second, consider an alternative where we allow lotteries to be chosen. The second version of the Cash Cow mechanism is where the firm charges a price $\overline{p}_1$, where $\frac{\overline{\phi}+\mu}{\underline{m}} > \overline{p}_1 > \frac{\underline{\phi}+\mu}{\underline{m}}$. This will only potentially be purchased by those clients with $\phi = \overline{\phi}$. As there are only a fraction $\frac{h}{2}$, this implies that there remains $\frac{1-h}{2}$ residual supply of $t = 2$. The firm then runs a lottery for those who have not purchased at price $\overline{p}_1$ to purchase $\tau = 1$ at a price $\underline{p}_1$ or $\tau = 0$ at a price p_0. A client is then choosing between $t = 1$ for sure at a price $\overline{p}_1$ or a lottery where with probability $\frac{1-h}{2-h}$ she receives $t = 1$ at a price $\underline{p}_1$ or with probability $\frac{1}{2-h}$ she receives $t = 0$ at a price p_0. Incentive compatibility then requires that the client with type $\tau = 1$ and $\overline{\phi}$ purchases for $\overline{p}_1$, while those with $\tau = 0$, or $\tau = 1$ and $\underline{\phi}$ choose the lottery. It is straightforward to show that there exist parameter values such this this form of price discrimination can yield higher revenue. Such price discrimination based on lotteries is not used in the social sector, largely as clients are unlikely to be well enough informed to compute such probabilities, and is ignored for that reason.[58]

[58]We have assumed that the firm can only use a single mechanism for all clients. This is for the sake of realism where it is likely to be difficult to communicate to both clients and agents if combinations of mechanisms are used. However, if such communication is possible, it maybe be desirable to use a distribution of mechanisms. Let z^j be the fraction of clients for which mechanism $j \in \{a, da, 1st, uc, rc\}$ is used, where $\sum_j z^j = 1$. Then the surplus obtained from the non-profit is adapted from (30) to $ES^* = \frac{D}{w^A - \sum_j z^j p^j} \left(v + \sum_j z^j T^j - \frac{\delta D}{2(w^A - \sum_j z^j p^j)} \right) - \frac{\delta \sigma_\epsilon^2}{2(w^A - \sum_j z^j p^j)^2}$. The reason for combinations is due to the $\frac{\delta D}{2(w^A - \sum_j z^j r^j)}$ term. As an example, imagine that w^* is far

Interpreting the Client's Preferences: Above, we have interpreted the client's preferences as $V(y) + T$, where y purchases a composite outside good. The central feature that causes the Fairness Constraint to fail is that higher income consumers are willing to pay more for services than those with lower incomes. To understand settings where this holds in the social sector, assume that client i consumes two goods z_1 and z_2, where the goods are weak substitutes and have diminishing returns, and has income y_i. The client faces market prices $\overline{p}_{z1}$ and $\overline{p}_{z2}$. In the absence of the firm, the client choose z_1^* and z_2^* to maximize $U(z_1^*, z_2^*)$ and $\overline{p}_{z1}z_1 + \overline{p}_{z2}z_2 \leq y_i$. If the solution involves $z_i^* > 0$ for both i, the solution is characterized by the familiar

$$\frac{U_{z1}}{\overline{p}_{z1}} = \frac{U_{z2}}{\overline{p}_{z2}} \tag{26}$$

and $\overline{p}_{z1}z_1^* + \overline{p}_{z2}z_2^* = y_i$. Assume that the non-profit gives covered clients 1 unit of z_2. Then the utility of the covered client becomes $U(z_1^{**}, z_2^{**} + 1)$, where z_i^{**} is the equilibrium choice given the firm's presence. To understand willingness to pay for the firm's contribution to z_2, there are three cases:

1. When $z_2^* \geq 1$ for all clients. Here $\frac{U_{z1}(z_1^{**}, z_2^{**})}{\overline{p}_{z1}} = \frac{U_{z2}(z_1^{**}, z_2^{**})}{\overline{p}_{z2}}$ and there is no difference in willingness to trade off good z_1 for good z_2 for richer and poorer clients. Then giving the client the service at the prices p_j above where $p_j \equiv p_{z2}$ is equivalent to simply increasing the client's income by $\overline{p}_{z2} - p_j$. If this is the case, then the solution to the client's problem is (26) subject to the clients income being $y_i + \overline{p}_{z2} - p_j$.[59]

2. When $z_2^* < 1$ for all clients. This is when the client consumes less of z_2 without the firm. A leading case is where $z_2^* = 0$, which we describe here. For many goods provides by the social sector, they are not chosen by clients in the absence of the firm. Then the price a client is willing to pay for the firm's service p_{z2}^* for a given y_i is given by

$$U(\frac{y_i - p_{z2}^*}{\overline{p}_{z1}}, 1) = U(\frac{y_i}{\overline{p}_{z1}}, 0). \tag{27}$$

   There are two reasons why willingness to pay varies by income. First, as all income is spent on z_1 higher income client consume more, so diminishing marginal

below $\underline{U}$. Then Agent Assignment has far higher client coverage than say Deferred Acceptance. But as there are diminishing returns, at some point the additional client coverage is of little value. As a result, the firm would prefer at some point to transfer a client from DA to Agent Assignment, and using a combination of mechanisms becomes optimal.

[59]Note however that even in this case, Hastings and Shapiro, 2018, show that Food Stamps increases food consumption relative to propensity to consume out of cash. Also note that Feeding America provides food support to the poor. In their view, much of the need for food arises as families have unexpected expenses that cannot be avoided - such as medical costs - or families fail to budget to ensure nutrition until their next pay check arrives.

utility increases the marginal rate of substitution towards good z_2. Second, through income effects. Most services provided by non-profits for the poor are those typically consumed by those with higher incomes: high quality housing in safe areas, therapy for mental health illness, tutoring for disadvantaged high school students, etc.. If these are normal goods, then the willingness to pay by a high income client will be higher than for a lower income client.

3. When $z_2^* > 1$ for wealthier clients, but $z_2^* < 1$ for poorer clients. Consider a case where clients are receiving food from a food pantry. Wealthier clients consume more food, and importantly add to their food consumption by purchasing on the private market. Poorer clients do not. Then the marginal price paid by the wealthier clients is the market price, while the marginal price for the poorer clients is that provided by the firm. In this instance, even if the canonical marginal rate of substitution equals relative price condition continues to hold, it does so at different prices, leading to different willingness to pay.

In the second and third case, the willingness to pay by higher income clients is higher, as required above. It also helps to explain why paternalism is a strong component of the non-profit mission. Almost no non-profits simply give money to their clients, but instead typically target services that would not otherwise be consumed.

Diminishing Returns: We have assumed that diminishing returns are linear: $v_i = v - \delta i$ to allow the impact of uncertainty in funding to be separable from funding levels. If instead we do not make this assumption, the analog to (7) is $S = \frac{D}{w^A - p^A} T^A + \sum_{i=1}^{\frac{D}{w^A - p^A}} v_i - \int_{\underline{\epsilon}}^{\overline{\epsilon}} \frac{x(v_{D+x} - v_D)}{w^A - p^A} f(x)dx$. For instance, Bureaucratic Assignment is now preferred to Agent Assignment if $\frac{D}{\overline{U} - p^{ac}}(T^b - T^a) - \sum_{i=\frac{D}{w^* - p^{ac}}}^{\frac{D}{\overline{U} - p^{ac}}} (v_i + T^a) + \int_{\underline{\epsilon}}^{\overline{\epsilon}} \frac{x(v_{D+x} - v_D)}{\overline{U} - p^{ac}} - \frac{x(v_{D+x} - v_D)}{w^* - p^{ac}} f(x)dx > 0$. This does not change any qualitative results above.

Unrestricted Choice and Income Differences: We have assumed above that the difference in income between the wealthiest and poorest client does not change with average income. Below we discuss the case where the support itself does not change when we address tolerance for unfairness. More generally, $\overline{\eta}$ and $\underline{\eta}$ may depend on $\hat{y}$. Let η be indexed by $\hat{y}$. The result that Choice can be used when average client income is high enough holds if $\frac{V'(\hat{y} + \underline{\eta}(\hat{y}))}{V'(\hat{y} + \overline{\eta}(\hat{y}))}$ does not increase too rapidly with $\hat{y}$. First, empirical evidence suggests that among low earners, dispersion of earnings is decreasing in average income (Farahan, 2015). If this translates to the support of outcomes, Decreasing Absolute Risk Aversion is sufficient but not necessary. Second, when randomness scales in a linear way with $\hat{y}$, $\eta(\hat{y}) = \eta\hat{y}$, the outcomes is analogous to finance settings that scale with investor wealth. Then our outcome arises if

$\frac{V''(\hat{y}+\overline{\eta}(\hat{y}))}{V'(\hat{y}+\overline{\eta}(\hat{y}))}\frac{1+\overline{\eta}}{1+\underline{\eta}} < \frac{V''(\hat{y}+\underline{\eta}(\hat{y}))}{V'(\hat{y}+\underline{\eta}(\hat{y}))}$. This is the analog to Decreasing *Relative* Risk Aversion in our setting.[60]

Unrestricted Choice and Fairness Tolerance: Above we showed how Unrestricted Choice arose for $\hat{y} \geq \hat{y}^{uc}$. This assumed that average income increases also increased the incomes of the highest and lowest clients in a similar way. Above we extended this to the multiplicative case, where an analog to Decreasing Relative Risk Aversion was key.

Assume now that the income of the wealthiest and poorest client do not change as average income changes. Under the Fairness Constraint above, Unrestricted Choice becomes no more likely with higher income. However, now consider the case where there is a tolerance for unfairness, and let income be generated by a Binomial distribution where all clients have n trials to generate income, with success rate ν. Each success gives the client income of Y. Then increasing ν is equivalent to increasing average income. In this case, the highest earning client has income of $y_i = nY$ and the lowest earning client has $y_i = 0$. Under the strict Fairness Constraint, the firm cannot charge a positive price as some clients have no income. Assume now that the firm has a relaxed Fairness Constraint with tolerance Γ. Unrestricted Choice arises if the analog of (9) holds for the m^* for which a fraction Γ are excluded from service. The CDF of the binomial is $G(k) = \sum_{i=0}^{k} \frac{n!}{(n-i)!}\nu^i(1-\nu)^{n-i}$. Let m^{**} be the marginal utility for the lowest earning covered client at which Γ are excluded. Note that $G(k)$ is decreasing in ν. As a result, $G(m^{**})$ is decreasing in ν, so that there exists a ν^* such that the firm can use Unrestricted Choice as $G(m^{**}) < \Gamma$. Hence the result that higher income allows Unrestricted Choice generalizes to this case where the support of income does not change with average income.[61]

Ability to Pay: We have interpreted ability to pay in terms of income, but it could alternatively represent variation in client assets, access to credit card debt, or capacity to receive a loan from friends or family. A small number of large non-profits have access to the financial position of clients. For instance, a major part of the social sector is health care facilities like hospitals. For many patients, their ability to pay is known simply because they have insurance. Our analysis does not apply to them. Instead, it relates to the uninsured population, which has declined from 16.8% of the US population in 2013 to 9.5% in 2023. Approximately 10% of non-profit

[60] More generally, when $\underline{\eta}'(\hat{y}) \geq 0$ the analysis holds if $\overline{\eta}' < \frac{\left(\frac{V''(\hat{y}+\underline{\eta}(\hat{y}))}{V'(\hat{y}+\underline{\eta}(\hat{y}))}\right)}{\left(\frac{V''(\hat{y}+\overline{\eta}(\hat{y}))}{V'(\hat{y}+\overline{\eta}(\hat{y}))}\right)} - 1$. This characterization holds for any parallel shift in supports, where $\underline{\eta}' = \overline{\eta}'$.

[61] Note that this result does not rely on DARA. This is because the difference in income between the m^{**} percentile of income and the highest earner is decreasing in ν. This shows how the result on DARA arises when we allow neutrality in income differences as average income increases.

hospital patients were uninsured over the 2015 to 2019 period (Wen et al., 2022). Consistent with the premise here, prices for that population are lower than for the insured population (Wang et al., 2023). Yet few non-profits have the capacity to both receive and process such information into price gradients by income, as they are typically too small. For example, 88% of non-profits have annual expenses of less than $500,000. Even those that do means testing for clients typically only use if as a maximum income exclusion for high earning potential clients.[62]

Donor Information and Preferences: We have assumed that donations are exogenously given by D. In reality, donors likely know the kind of clients who will benefit, so D may be indexed by $\hat{y}$. All results continue to hold if $D(\hat{y}) + p^A$ is decreasing in $\hat{y}$. This simply means that donations only partially make up for lower revenue from clients. There is considerable evidence that donations in charitable settings partially offset funds raised elsewhere (Andreoni, 1993, Andreoni and Payne, 2009, Bolton and Katok, 1998, and Robert, 1984).

2 The For-Profit Social Enterprise

Many social organizations operate as for-profit social enterprises with objectives beyond profit maximization.[63] As an example, for-profit hospitals offer similar charitable care to their non-profit counterparts (Bruch and Bellamy, 2021). The World Economic Forum has estimated 10 million such enterprises worldwide, accounting for 3% of world employment (World Economic Forum, 2024). The key component of the for-profit firm is the absence of the Non-Distribution Constraint. The literature on non-profit status has focused on agency problems (Hansmann, 1980, Glaeser and Schleifer, 2001). Yet a central concern in the charitable sector has instead focused on funding issues, and in particular the role of donors. Above we noted that 22% of non-profit funds derive from donors. Donors overwhelmingly report that their motivation is the impact of their gift.[64] We argue donors may be deterred by a for-profit firm as they correctly infer that "their" gift will not be used for the mission.

Any realistic model of donors must take account of their limited information. For instance, the literature on whether donations are responsive to firm efficiency show at best a weak relationship (Buchheit and Parsons, 2006, Parsons, 2007, and Yoruk, 2013). We assume that potential donors do not observe allocation mechanisms or

[62]For example, Gai et al., 2021, document how non-profit hospitals use income data only to determine eligibility caps for discounted care.

[63]For instance, there are 1,300 for-profit hospitals and 1,100 for-profit charter schools in the United States (American Hospital Association, 2024, Network for Public Education, 2023).

[64]This is called pure altruism in the literature. Ottoni-Wilhelm, et al., 2017, argue for the importance of such altruism for charitable giving. Survey evidence suggests that "97% of donors cite the impact of their gift as their main reason for giving" (CCS, 2024). This section extends the logic of Prendergast, 2023.

contracts, but instead only observe the profit status of the firm. They condition their donations on that feature alone. Our objective is to show that if donor equilibrium beliefs about the use of their gift are correct, the firm can only become a social enterprise if completely funded by clients (Gertner, 2022, Brumme and Trelsted, 2024).

To address this issue, three changes are made to the model.

Donations: Donations are given by $\mathbb{D}+\epsilon$, where ϵ is a mean zero error term defined over a support of $[\underline{\epsilon}, \overline{\epsilon}]$ according to a distribution function F, with density $f > 0$.

Profit Status: We now allow the firm to operate as a non-profit or social enterprise. A social enterprise can take profit π, where $E\pi = 0$ (this is simply to distinguish the firm from being a donor). The for-profit chooses $\pi \leq \mathbb{D} + \epsilon$ while the non-profit must spend all revenue on the mission. We assume that π can be contingent on $\mathbb{D} + \epsilon$. Then

$$I = \frac{\mathbb{D} - \pi(\epsilon) + \epsilon}{w^A - p^A}. \tag{28}$$

Donors: There are N potential donors to the firm. Each donor k gives $d_k \in \{0, 1\}$ to the firm. If a donation is used on a client, the donor believes that client utility increases by Δ. If it is used as profit, the donation has no value. The utility of the donor is $U^d = \mathbb{I}^c \alpha_k \Delta - d_k$, where $\mathbb{I}^c = 1$ iff the donation is used on a client and 0 otherwise, and α_k is the donor's altruism distributed according to a distribution over $(0, \infty)$ according to a CDF L. Let all donors assume that the probability that the donation is used on clients is b. Let donations be $\mathbb{D}(b)$, where $\mathbb{D}(0) = 0$, $\mathbb{D}'(b) \geq 0$, $\mathbb{D}(1) \equiv D$. Then

$$\mathbb{D}(b) = N(1 - L(\frac{1}{b\Delta})). \tag{29}$$

We assume that $b = 1$ for the non-profit, as all funds are spend on the mission. Then donations above, $D = N(1 - L(\frac{1}{\Delta}))$.

Strategies and timing: The timing is now as follows:

1. The firm chooses A, P^A, w^A, π^A and Π, which is observed by clients and agents. Π is observed by donors, who are endowed with a belief b.
2. The firm offers the position to one agent subject to (5).
3. The firm and agent observes τ_i^a for all i.
4. ϵ is revealed.
5. A is implemented, p_i^A paid by clients, and w^A and π^A are paid.

The model changes in the following way. First, the firm chooses its profit status. If it is non-profit, donations are D above. If a social enterprise, donations are $\mathbb{D}(b)$. If a social enterprise, it writes an enforceable contracts on how much profit it can take, π, as a function of donations $\mathbb{D}+\epsilon$. This is not observed by donors, who only observe the firm's profit status. Otherwise the model is as above.

Expected surplus is now given by $\int_{\underline{\epsilon}}^{\overline{\epsilon}} \frac{(\mathbb{D}(b|\Pi)-\pi^A(x)+x)}{w^A-p^A}(v+T^A-\frac{\delta(\mathbb{D}(b|\Pi)-\pi^A(x)+x)}{2(w^A-p^A)})f(x)dx$, or

$$ES = \frac{\mathbb{D}(b|\Pi)}{w^A-p^A}(v+T^A-\frac{\delta\mathbb{D}(b|\Pi)}{2(w^A-p^A)}) - \int_{\underline{\epsilon}}^{\overline{\epsilon}} \frac{\delta(-\pi(x)+x)^2}{w^A-p^A}f(x)dx. \qquad (30)$$

Let b^* be the fraction of the marginal donation that is spent on the mission in equilibrium. Let $\pi^*(\epsilon)$ be the equilibrium contract, where $\int_{\underline{\epsilon}}^{\overline{\epsilon}} \pi^*(x)f(x)dx = 0$. Then for a realization ϵ, a fraction $1-\pi^{*'}(\epsilon)$ of the marginal dollar is spend on I rather than on π. Then $b^* = \int_{\underline{\epsilon}}^{\overline{\epsilon}}(1-\hat{\pi}^{*'}(x))f(x)dx$.

Now consider the outcome. From (30), if $\Pi = 1$, the firm chooses $\pi^*(\epsilon) = \epsilon$, as donors cannot observe contracts and hence donations are independent of the π contract. Proposition 8 follows.

Proposition 8 *The social enterprise chooses $\pi^{*'}(\epsilon) = 1$. If $b = b^*$, then $\mathbb{D}(0) = 0$.*

Donors rarely give to for-profit firms. Proposition 8 offers a simple reason why, even for a firm which has a clear social mission. First, the firm values efficient insurance as it guarantees client coverage. Second, donations are fungible so all donors are marginal. Finally, donors do not observe the π function. The central conclusion of Proposition 8 is then the impossibility of a for-profit charity.[65] Note that this arises even though *in expectation* all revenue is spent on the mission.

We are not assuming that $b = b^*$. Instead, Proposition 8 simply says that if donors have correct beliefs about the firm's action, any marginal dollar received by the firm goes into the pocket of its owners. At the outset of the paper, we described a complex relationship between funding and efficiency in the social sector. Providing financial certainty to the firm increases expected surplus, but eliminates the funding that comes from donors. Once again, efficiency and funding are substitutes.[66] Note that viewing the non-profit through this lens offers an additional implication, namely, that it has to live "hand to mouth", never sure of its ability to cover its mission. Being "hand to mouth" is necessary to encourage donations, as a donor must know that her donation affects the mission.

The central message of this section is that donors are unlikely to give if their gift is used for purposes other than the mission of the firm. This has long been a concern among *non-profits*. We have assumed that $\pi' = 0$ in order to qualify as a non-profit. In reality, the legal standard is not so strict, where some degree of incentive pay is

[65] Posner and Malani, 2006, have argued the merits of the for-profit charity.

[66] Rose Ackerman, 1982, addresses how non-profits allocate excessive resources to fund-raising, as a way of business stealing from other non-profit organizations.

allowed.[67] Compensation among fund raisers is governed by ethical standards set down by the Association of Fundraising Professionals. That ethical standard "prohibits members from working for percentage based compensation" as it "undermines the integrity of the voluntary sector". Specifically, "donor attitudes can be unalterably damaged in reaction to... the awareness that a commission will be paid for a fundraiser from his or her gift, thus compromising the trust on which charity relies". (American Association of Fundraising Professionals, 2016). In a similar vein, Sandberg, 2006, proposes appropriate legal transparency rules for pay-for-performance for charities. His concern is "that most donors would see a high fundraising commission as unreasonable and illogical irrespective of its economic rationale. If fully informed, many donors would probably refuse to contribute". As a result, it is widely believed in the sector that using donations as marginal compensation deters donors.[68]

Firms choose whether to incorporate as non-profit or for-profit. Assume that the firm can self-fund with mechanism A. Then the firm chooses a non-profit with mechanism A' over a self-funded one with A if $ES^{A'} \geq \frac{(v+T^A)^2}{2\delta}$. As $I^{A*} = \frac{v+T^A}{\delta}$ surplus from a self-funded firm is $\frac{(v+T^A)^2}{2\delta}$. Yet it may be that the only way to self-fund is with an inefficient assignment mechanism A. For example, it could be that only with Agent Assignment is $w^* < p^{ac}$. We now address differences in how non-profits and social enterprises allocate services.

Proposition 9 *Assume that $w^* > p^{ac}$. Then if*

$$\frac{\mu}{V'(\hat{y}^{uc} + \underline{\eta})} < \overline{U} < \frac{\mu + \frac{p_0^* + \underline{\phi}}{2}}{V'(\hat{y}^{uc} + \underline{\eta})} \tag{31}$$

all social enterprises use Unrestricted Choice and no Assigned Consumption firm does.

This follows trivially from the definition of $\hat{y}^{uc}$, when the firm can transition to Unrestricted Choice. Client payments increase discretely from $\frac{\mu}{V'(\hat{y}^{uc}+\underline{\eta})}$ to $\frac{\mu+\frac{p_0^*+\underline{\phi}}{2}}{V'(\hat{y}^{uc}+\underline{\eta})}$. Then if the extra revenue allows the firm to cover its marginal cost $\overline{U}$, the social enterprise allows Choice but as $w^* > p^{ac}$, no Assigned Choice non-profit does.

People often complain about bureaucracies being unresponsive to their needs. This is contrasted with the "customer is always right" ethos of the typical for-profit setting. Proposition 9 provides a condition where this arises. More generally, the ability to self-fund is a comparison between the cost of the agent w^A and the ability to raise

[67] The law on the non-profit allows "reasonable" pay-for-performance for their employees (Founding Church of Scientology v. United States).

[68] Here the donor makes an undirected donation. It is notable that social enterprises often attempt to attract donors through crowdfunding campaigns for a specific initiative. These serve to both target a mission-related initiative and to offer the prospect of the donor seeing herself as marginal. Though success rates are low, see Farhoud et al., 2021, for examples of some successes.

funds p^A. In Proposition 10 in the Appendix we show that while for-profit firms can exist with any allocation method, it more likely when client choice is involved. In this sense, it is a generalization of (31) and the view that non-profit firms are inherently more likely to make allocative errors.

The literature on the for-profit choice has studied a setting where a firm controls unobserved quality, and the concern is that a for-profit firm cuts corners on unobservable quality. There are many examples of this behavior (Gertner, 2022). However, evidence for the most important non-profit sectors - schooling and medical care - find no evidence of lower quality among for-profits (Joynt et al, 2014, McClellan and Staiger, 1999, Dynarski et al., 2018, Cohodes and Parham, 2021, Taylor et al., 2023). As a result, we take a more positive approach to incentive contracting.

In the Appendix, we provide an alternative interpretation of the for-profit decision. The firm invests in unobserved quality of its services. Unlike the existing literature, the firm's objective is to maximize client welfare. Because quality is costly, there is a tradeoff between quality and clients covered. This changes the firm's incentives, in the opposite direction of the existing literature. Here the non-profit firm covers more clients by reducing quality below the efficient level. As clients infer quality in equilibrium, this backfires. The value of for-profit status is that by suitable contracting, this incentive can be corrected. However, the optimal contract renders the firm residual claimant on the firm's revenues. As above, donors realize that the firm gets at least some of the donor's marginal contribution, and donations fall or are eliminated.

This offers two alternative outcomes to that above. First, it places the analysis in the context of the existing literature, but relating it to a moral hazard setting with quality choices. Second, the extension allows for the possibility that all donations are not crowded out from the profit decision.

3 Conclusion

Social organizations do not use market clearing prices to allocate their services. The academic literature has championed Deferred Acceptance or Serial Dictatorship as a preferred alternative to choice with prices. Its appeal is couched in terms of strategy proofness and stability. The central premise here is that many social organizations are likely more concerned with raising funds to expand their mission, retaining and attracting staff, and exploring the possibility of client choice to avoid assignment error. It interprets why a firm would allow a bureaucrat with the "wrong" preferences unfettered control when something as apparently appealing as Deferred Acceptance is available. We also showed how the cost for an organization serving very low income clients is not just that their funding is low, but also that allocations entail error, because choice can only offered to clients with more income.

By now, the canonical applications of matching without prices - school choice, kidney exchange, etc.- have been extensively mined. One purpose of this work is to offer a bridge to a sector that employs roughly one in ten. The insights of the

market design literature - ex ante fairness, ex post efficiency, and strategy proofness - remain. Yet because of a link between funding and ex post efficiency, they may not be pivotal in how best to distributes services. There are a number of key inputs. First, that understanding the social organization requires addressing that funding primarily derives from clients, but that existing pricing models do not relate well to their constraints. It is for this reason that we have proposed the Fairness Constraint as the relevant obstacle for social organizations. Second, that in the absence of allocation mechanisms that reveal client preferences, stakeholders can reasonably disagree. As an example, is it more important for a food bank that everyone be fed a sufficient amount, or that they only be fed nutritious food? Third, that the logic of the agency literature for distorting ex post allocations to improve ex ante incentives also applies to *ex post* Pareto efficiency. Finally, that donors who care about the impact of their gift are (rightly) suspicious if the firm can take their donation as profit.

The paper is not intended to explain the actions of all social sector firms. Most importantly, we do not address cases in the public sector such as school choice, where funding occurs through tax revenue. Given this, it is striking that funding concerns for the core market design applications are minimal, occurring through taxation (school choice) or where the potential funds received are trivial compared to overall funding for the institutions (medical resident matching). In addition, many social organizations do not used the price system because there are externalities to client choice. For example, schools care about the composition of their student body. Firms in the sector vary in both their objectives and willingness to exclude clients. For instance, some want profit in addition to doing good. The objective of the firm here is "pure", only interested in the welfare of its clients. We have assume this to show the complexity of tradeoffs even with such pristine objectives. Similarly, the firm here is unwilling to deny service to potential clients if they cannot pay. While this may be a good reflection of a social service provider, it may resonate less well for museums or elite universities. We have offered a view of one slice of that sector, one focused on serving those in need.

For social organizations, pricing fails to fulfill their usual role in many ways. It stops the firm from trusting clients to choose, leading to Assigned Consumption. Efficient prices also identify client preferences, and disagreements about what client truly need are more common in their absence. Finally, prices pay the bills. Much of the paper has focused on how firms respond to these price failures, often by trading off the intensive margin of allocative efficiency for the extensive margin of expanding access to care. In this sense, one contribution of the paper may be to embed our understanding of social organizations in a price (failure)-theoretic framework.

References

Abdulkadiroglu, Atila, Parag A. Pathak, Jonathan Schellenberg,

and Christopher R. Walters. "Do Parents Value School Effectiveness?", *American Economic Review*, 2020, 110(5): 1502–1539.

Abdulkadiroglu, Atila, Yeon-Koo Che and Yosuke Yasuda, "Resolving Conflicting Preferences in School Choice: The "Boston Mechanism" Reconsidered", *American Economic Review*, 2011, Vol. 101, No. 1, 399-410.

Adler P. S., and B. Borys, "Two types of bureaucracy: Enabling and coercive", *Administrative Science Quarterly*, 1996, 41(1), 61-89.

African Union, Africa CDC, et al., "Joint Statement on Dose Donations of COVID-19 Vaccines to African Countries", 2021. https://www.who.int/news/item/29-11-2021-joint-statement- on-dose-donations-of-covid-19-vaccines-to-african-countries.

Agarwal, Nikhil, and Paulo Somaini, "Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism", *Econometrica*, 2018, 86 (2), 391-444.

Aghion, Philippe, and Jean Tirole, "Formal and Real Authority in Organizations", *Journal of Political Economy*, 1997, 105 (1), 1-29.

American Association of Fundraising Professionals, "Professional Compensation: A Position Paper", 2016. https://afpglobal.org/sites/default/files/attachments/2018-10/2016ProfCompensationPositionPaper.pdf.

American Hospital Association, "Fast Facts on U.S. Hospitals, 2024", 2024. https://www.aha.org/statistics/fast-facts-us-hospitals.

Andreoni, J., "An Experimental Test of the Public Goods Crowding-Out Hypothesis", *American Economic Review*, 1993, 83, 1317-27.

Andreoni, J. and A. Payne, "Is crowding out due entirely to fundraising? Evidence from a panel of charities", *Journal of Public Economics*, 2009, 95, 334-43.

Antecol, Heater and Janet Smith, "The Early Decision Option in College Admission and Its Impact on Student Diversity", *Journal of Law and Economics*, 2012, vol. 51, no. 1, 217–249.

Avery C., and J. Levin, "Early Admissions at Selective Colleges", *American Economic Review*, 2010, vol. 100, 2125-2156.

Balinski, M. and T. Sonmez, "A Tale of Two Mechanisms: Student Placement", *Journal of Economic Theory*, 1999, 84, 73-94.

Bansak, K., S. Lee, V. Manshadi, R. Niazadeh, and E. Paulson, "Dynamic Matching with Post-allocation Service and its Application to Refugee Resettlement", 2025, working paper.

Besley, Timothy, Stephen Coate, and Glenn Loury, "The Economics of Rotating Savings and Credit Associations", *American Economic Review*, 1993, 83, 792-810.

Bolton, G. and E. Katok, "An Experimental Test of the Crowding Out Hypothesis: The Nature of Beneficent Behavior", *Journal of Economic Behavior and Organization*, 1998, 37, 315-31.

Bruch, J.D., and Bellamy, D., "Charity Care: Do Nonprofit Hospitals Give More than For-Profit Hospitals?", *Journal of General Internal Medicine*, 2021, 36, 3279–3280.

Brumme, Cait, and Brian Trelstad, "Should Your Start-up Be For-Profit or Nonprofit?", *Harvard Business Review*, May-June, 2023. https://hbr.org/2023/05/should-your-start-up-be-for-profit-or-nonprofit

Buchheit, S., and L. Parsons, "An experimental investigation of accounting information's influence on the individual giving process", *Journal of Accounting and Public Policy*, 2006, 25(6):666-686.

Budish, Eric, "The Combinatorial Assignment Problem: Approximate Competitive Equilibrium from Equal Incomes", *Journal of Political Economy*, 2011, 119 (December): 1061-1103.

Budish, Eric, and and Estelle Cantillon, "The Multi-Unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard", *American Economic Review*, 2012, 102 (August): 2237-71.

Budish, Eric, Hannah Kettler, Scott Duke Kominers, Erik Osland,

Canice Prendergast and Andrew A Torkelson, "Distributing a billion vaccines: COVAX successes, challenges, and opportunities", 2022, *Oxford Review of Economic Policy*, 2022, 38(4), 941-74.

Bureau of Labor Statistics, "Celebrating National Nonprofit Day with BLS Data", August 15, 2023. https://www.bls.gov/blog/2023/celebrating-national-nonprofit-day-with-bls-data.htm.

Bureau of Labor Statistics, "Nonprofits: a look at national trends in establishment size and employment", January, 2024.

Calsamiglia, Caterian, Chao Fu, and Maia Guell, "Structural Estimation of a Model of School Choices: The Boston Mechanism Versus its Alternatives", *Journal of Political Economy*, 2019, vol. 128, issue 2, pages 642-80.

Cantarelli P., Belardinelli P., and Belle N., "A meta-analysis of job satisfaction correlates in the public administration literature", *Review of Public Personnel Administration,* 2016, 36, 115-144.

Carey, M., "It's a bit like being a robot or working in a factory": Does Braverman help explain the experiences of State Social Workers in Britain since 1971?", 2009, *Organization*, 16(4), pp. 505–27.

CCS Fundraising, "2024 Philanthropic Landscape, 13th Edition", 2024.

Che, Yeon-Koo, and Navin Kartik, "Opinions as Incentives", *Journal of Political Economy*, 2009, vol. 117, issue 5, pages 815-60.

Chaumba, J. and A. Locklear, "Strategies for Promoting Client Self-Determination: A Review of the Literature", *Journal of Social Work and Ethics*, 2021, 18(1), 60-71.

Chiappori, P.-A., and Paiella, M., "Relative risk aversion is constant: Evidence from panel data", *Journal of the European Economic Association*, 2011, 9(6), 1021–1052.

Cohen, N., Benish, A., and A. Shamriz-Ilous, "When the Clients Can Choose", *Social Service Review,* 2016, 90(4), 620-646.

Cohodes, Sarah R., and Katharine S. Parham, "Charter Schools' Effectiveness, Mechanisms, and Competitive Influence", 2018, NBER Working Paper No. 28477.

Cook, J. , S. Shore, J. Burke-Miller , J. Jonikas, M. Hamilton, B Ruckdeschel,

W. Norris, A. Markowitz, M. Ferrara, and D. Bhaumik., "Mental Health Self-Directed Care Financing: Efficacy in Improving Outcomes and Controlling Costs for Adults With Serious Mental Illness", *Psychiatric Services*, 2019, 70(3), 191-201.

Dustin, D., "Skills and knowledge needed to practise as a care manager: Continuity and change", *Journal of Social Work*, 2006, 6(3), pp. 293–313.

Dworczak, P., "Inequality and Market Design", 2024, forthcoming, *ACM SIGecom Exchanges*.

Dynarski, Susan, Daniel Hubbard, Brian Jacob and Silvia Robles, "Estimating the Effects of a Large For-Profit Charter School Operator", 2018, NBER Working Paper No. 24428.

Ergin, H., "Efficient Resource Allocation on the Basis of Priorities", *Econometrica*, 2002, 70, 2489–2497.

Farahan, Fatih, "Understanding Earnings Dispersion", 2015, Federal Reserve Bank of New York. https://libertystreeteconomics.newyorkfed.org/2015/11/understanding-earnings-dispersion/.

Farhoud, M., S. Shah, P. Stenholm, E. Kibler, M. Renko, and S. Terjesen, "Social enterprise crowdfunding in an acute crisis", *Journal of Business Venturing Insights*, 2021, e00211.

Faulk, Lewis, Mirae Kim, Teresa Derrick-Mills, Elizabeth T. Boris,

Laura Tomasko, Nora Hakizimana, Tianyu Chen, Minjung Kim, Layla Nath "Nonprofit Trends and Impacts 2021", 2021, *Urban Institute*.

Food Bank News, "Study Quantifies the Benefits of Client Choice", 2024. foodbanknews.org/study-quantifies-the-benefits-of-client-choice/.

Gai G, Zare H, Eisenberg MD, Polsky D, Anderson GF, "Comparison of Trends in Nonprofit Hospitals' Charity Care Eligibility Policies Between Medicaid Expansion States and Medicaid Nonexpansion States", *Medical Care Research and Review*, 2021, 79(3):458-468.

Gertner, Robert, "The Organization of Social Enterprises", *Annual Review of Economics*, 2023, 15:41–62.

Glaeser, E., and A. Shleifer, "Not-for-profit entrepreneurs", *Journal of Public Economics*, 2011, 81 (1), 99-116.

Goedde, J., "Pricing in Markets without Money: Theory and Evidence from Home Exchanges", 2024, Paris School of Economics.

Groves, J., Ho, W. and Siu, K., "Exploring the sacred-secular dialect in everyday social work practice: An analysis of religious responses to managerialism among outreach social workers in Hong Kong", 2016, *British Journal of Social Work*, 46(5), pp. 1411–28.

Guiso, L., and Paiella, M., "Risk aversion, wealth, and background risk", *Journal of the European Economic Association*, 2008, 108(12), 1109–1150.

Hastings, J., and J. Shapiro, "How Are SNAP Benefits Spent? Evidence from a Retail Panel", *American Economic Review*, 2018, 6(6), 3493–3540.

Hansmann, Henry, "The Role of Nonprofit Enterprise", *The Yale Law Journal*, 1980, 89(5), 835-901.

Hirsch, B,, D. McPherson, and A. Preston, "Nonprofit Wages: Theory and Evidence", 2017, IZA Discussion Paper.

Huber, T., Jaspersen, J.G., Richter, A., and D. Strumpel , "On the change of risk aversion in wealth: a field experiment in a closed economic system, *Experimental Economics*, 26, 1–26 (2023). https://doi.org/10.1007/s10683-022-09762-x

Hylland, Aanund, and Richard Zeckhauser, "The Efficient Allocation of Individuals to Positions", *Journal of Political Economy*, 1979, Vol. 87, No. 2, 293-314.

Joynt, Karen E, E John Orav, and Ashish K Jha, "Association between hospital conversions to for-profit status and clinical and economic outcomes", *Journal of the American Medical Association*, 2014, 16, 1644-52.

Kapor, A., C. Neilson and S. Zimmerman, "Heterogeneous Beliefs and School Choice Mechanisms", *American Economic Review*, 2020, vol. 110, no. 5, 1274–1315.

Kenny Stevens, Susan, *Nonprofit Lifecycles: Stage-Based Wisdom for Nonprofit Capacity*, 2002, Stagewise Enterprises.

Langer, Julie, Mary K. Feeney, and Sang Eun Lee, "Employee Fit and Job Satisfaction in Bureaucratic and Entrepreneurial Work Environments," *Review of Public Personnel Administration*, 2019, 39 (1), 135-155.

Leete, L., "Whither the nonprofit wage differential? Estimates from the 1990 Census", *Journal of Labor Economics*, 2001, 19 (1), 136-70.

Lipsky, Michael, *Street Level Bureaucracy: Dilemmas of the Individual in Public Services*, 1980, New York, Russell Sage Foundation.

Liu, Dong, Shu Zhang, T. W. Lee, "The effects of autonomy and empowerment on employee turnover: test of a multilevel model in teams.", *Journal of Applied Psychology*, 2011, 96(6), 1305-15.

MacEachan, E., Kosny, A., Ferrier, S., Lippel, K., Neilson, C.,

Franche, R. L., and Pugliese, D. "The ideal of consumer choice in social services: challenges with implementation in an Ontario injured worker vocational retraining programme.", *Disability and Rehabilitation*, 2013, 35(25), 2171–2179.

McClellan, Mark and Doug Staiger, "The Quality of Health Care Providers", 1999, NBER Working Paper 7327.

Murphy, James, "The Future of Fair Admissions", *Education Reform Now*, 2022.

Murphy, James, "Early Decision Benefits Rich Kids. That Doesn't Necessarily Mean That It's Evil", *Slate*, 2023, December 14.

National Council of Nonprofits, "Downloadable Charts and Figures", 2024.

https://www.nonprofitimpactmatters.org/data/downloadable-charts/.

Network for Public Education, "Chartered for Profit", 2023. https://networkforpubliceducation.org/chartered-for-profit/.

Nichols, Albert L., and Richard J. Zeckhauser, "Targeting Transfers through Restrictions on Recipients", *American Economic Review*, 1982, vol. 72(2), pages 372-377, May.

Ottoni-Wilhelm, Mark, Lise Vesterlund, and Huan Xie, "Why Do People Give? Testing Pure and Impure Altruism", *American Economic Review*, 2017, 107(11), 3617-3633.

Pathak, Parag, and and Tayfun Sonmez, "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism", *American Economic Review*, 2008, Vol 98(4), 1636-52.

Parsons, L. M., "The impact of financial information and voluntary disclosures on contributions to not-for-profit organizations", *Behavioral Research in Accounting*, 2007, 19(1):179-196.

Pascoe, Katheryn, Bethany Waterhouse-Bradley, and Tony McGinn , "Social Workers' Experiences of Bureaucracy: A Systematic Synthesis of Qualitative Studies", *The British Journal of Social Work*, 2023, Volume 53, Issue 1, 513–533, https://doi.org/10.1093/bjsw/bcac106 .

Petersen, Anette, Lars Rune Christensen, and Thomas T. Hildebrandt , "The Role of Discretion in the Age of Automation", *Computer Supported Cooperative Work*, 2020, https://doi.org/10.1007/s10606-020-09371-3.

Posner, Eric, and Anup Malani, "The Case for For-Profit Charities", 2006, John M. Olin Program in Law and Economics Working Paper No. 304.

Prendergast, Canice, "The limits of bureaucratic efficiency", *Journal of Political Economy*, 2003, 111, no. 5: 929-958.

Prendergast, Canice, "The motivation and bias of bureaucrats", *American Economic Review*, 2007, 97, no. 1: 180-96.

Prendergast, Canice, "How Food Banks Use Markets to Feed the Poor", *Journal of Economic Perspectives,* 2017, 31 (4), 145-162.

Prendergast, Canice, "The Allocation of Food to Food Banks", *Journal of Political Economy*, 2022, 130 (8), 1993-2017.

Prendergast, Canice, "Organizational Design for Making a Difference", *Journal of Public Economics*, 2023, vol. 218, 104817.

Roberts, R., "A Positive Model of Private Charity and Public Transfers", *Journal of Political Economy*, 1984, 92, 138-48.

Roth, Al, "What Have We Learned From Market Design?", *Economic Journal*, 2008, Vol 118, 285-310.

Rose-Ackerman, Susan, "Charitable Giving and 'Excessive' Fundraising.", *Quarterly Journal of Economics*, 1982, 97, 193-212.

Ruhm, C., and C. Borkowski, "Compensation in the nonprofit sector", *Journal of Human Resources*, 2003, 38 (4), 992-1021.

Salamon, L., and S. W. Sokolowski, "Nonprofit organizations: New insights from QCEW data", *Monthly Labor Review*, 2005, September, 19-26.

Sandberg, Haim, "Where Do Your Pennies Go? Disclosing Commissions for Charitable Fundraising", *The International Journal of Not-for-Profit Law*, 2006, 8(4).

Small Charities Data, "Trading and enterprise: how small charities earn their income", 2024. https://smallcharitiesdata.org/topic/trading-and-enterprise-how-small-charities-earn-their-income/

Statista, "Expenses of reporting 501(c)(3) nonprofit organizations in the United States from 1998 to 2020", 2024. https://www.statista.com/statistics/189257/expenses-of-reporting-non-profit-organizations-in-the-united-states-since-1998/.

Taylor, Lauren A., Berkeley Franz, Amanda Zink, Andrew Fair, and

Cory E. Cronin, "Public perceptions of US for-profit, nonprofit, and public hospitals", *Health Affairs Scholar,* 2023, 1(4), 1-9.

US Census Bureau, "U.S. Volunteerism Rebounding After COVID-19 Pandemic", November 2024.
https://www.census.gov/library/stories/2024/11/civic-engagement-and-volunteerism.html

USA Facts, "How many nonprofits are there in the US?", November 2023. https://usafacts.org/articles/how-many-nonprofits-are-there-in-the-us/

Wang, Y., M. Katz Meiselbach, Cox., J., G. Anderson and G. Bai, "The Relationships Among Cash Prices, Negotiated Rates, And Chargemaster Prices For Shoppable Hospital Services", *Health Affairs*, 2023, 42(4).

Wen, G., H. Zare, M. Eisenberg, and G. Anderson, "Association between nonprofit hospital community benefit spending and health outcomes", *Health Services Review*, 2022, 58(1):107–115.

White. S., "Early Decision Provides A Big Bang For Your Buck At These Colleges". *Forbes*, July 15, 2024.

Wong, M.B., "Money and Barter in the Field: Evidence from the Life and Death of a Digital Currency", 2024, University of Hong Kong.

World Economic Forum, "The State of Social Enterprise: A Review of Global Data 2013–2023", 2024, Schwab Foundation for Social Entrepreneurship.

Wright B. E. and Davis B. S., "Job satisfaction in the public sector the role of the work environment". *The American Review of Public Administration*, 2003, 33, 70-90.

Yoruk, Baris, "Charity Ratings", *Journal of Economics and Management Strategy*, 2016, 25(1), 195-219.

Zhang, Yiran, Wen Xu, Susan Yoon, Weizhi Chen, and Sarah Parmenter, "Workplace support, job autonomy, and turnover intention among child welfare workers in China: The mediating role of job satisfaction", *Child Protection and Practice*, 2024, Volume 2, August 2024, 100027.

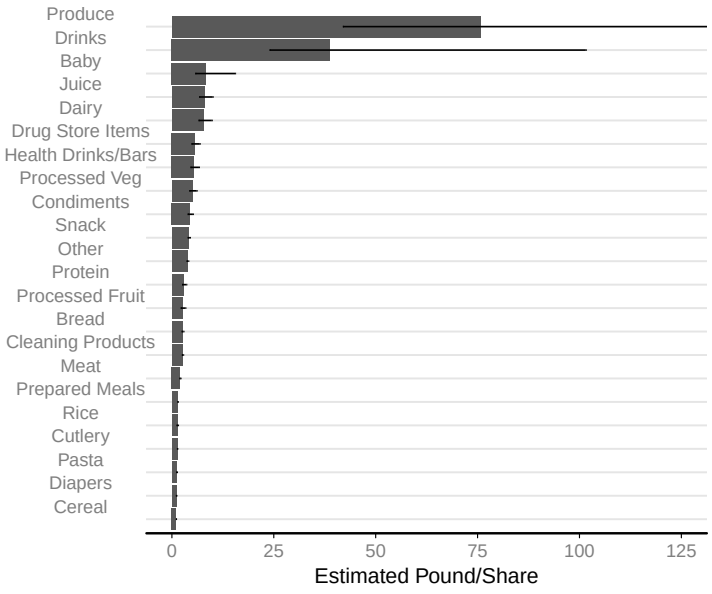


Figure 3: **Feeding America Price of Different Goods.**

4 Appendix

The Program: The outcome here is shown for both the for-profit where π need not be zero and for Restricted Choice where I_i is divisible. For the non-profit, $\pi = \Pi = 0$. The firm's objective given A is to choose $I_i^{*A}, P^A, \pi^{*A}, \Pi^{*A}$ and w^{*A} maximize

$$ES = \int_{\underline{\epsilon}}^{\bar{\epsilon}} \sum_{i=1}^{\infty} (v_i + T^A(t_i, \tau_i)) I_i(x) f(x) dx, \tag{32}$$

where $v_i = v - \delta i$ and $I_i \in [0, 1]$. If $w^A < p^A$, then $I^A = \sum_{i=1}^{\infty} I_i^A$ is defined by

$$I^A \leq \frac{\mathbb{D} - \pi^A(\tilde{I}^A(\epsilon)) + \epsilon}{w^A - r^A}, \tag{33}$$

where $\tilde{I}^A(\epsilon) = \frac{\mathbb{D}^A + \epsilon}{w^A - r^A}$. ES is maximized subject to

1. $E\pi^A = 0$. $\pi^A(\tilde{I}^A) = 0$ if $\Pi = 0$ and $\pi^A \leq \mathbb{D} + \epsilon$ if $\Pi = 1$.
2. Wages are given by (5).
3. $T(i, j|\underline{\phi}) \geq \underline{mp}_j^A$, in any equilibrium where type τ_i receives treatment $t = j$ with positive probability.
4. Client Reporting in A:

- Let $H_{ij}^A(\hat{t}_l|\hat{t}_{-l})$ be the probability that client l of type $\tau = i$ receives $t = j$ after reporting a ranking $\hat{t}_l$ and others report $\hat{t}_{-l}$. In a Bayesian Nash equilibrium, client l reports $\hat{t}$ when others report $\hat{t}_{-l}$ if

$$\sum_j H_{ij}^A(\hat{t}_l|\hat{t}_{-l})(T(i,j|\phi)-m_i p_j^A) \geq \sum_j H_{ij}^A(\hat{t'}_l|\hat{t}_{-l})(T(i,j|\phi)-m_i p_j^A), \forall m \in [\underline{m}, \overline{m}], \tag{34}$$

  where H_{ij}^A are generated by Bayes rule.

- If A involves Assigned Consumption and the allocation mechanism is strategy proof, this simplifies to

$$T(i,1|\underline{\phi}) - m_i p_1^A \geq T(i,0|\underline{\phi}) - m_i p_0^A, \forall m_i \in [\underline{m}, \overline{m}], \tag{35}$$

  and

$$T(i,i|\overline{\phi}) - m_i p_j^A \geq T(i,j|\overline{\phi}) - m_i p_k^A, \forall m \in [\underline{m}, \overline{m}]. \tag{36}$$

- If A involves Choice where the client chooses $I_i^A = k_i^A \leq 1$ at unit price p_i^A

$$k_i^A(T(i,i|\underline{\phi}) - m_i p_i^A) \geq p_j^A(T(i,j|\underline{\phi}) - m_i p_j^A), \forall m \in [\underline{m}, \overline{m}]. \tag{37}$$

5. In Section 2, donations are $d(b)$ where $b = \int_{\underline{\epsilon}}^{\overline{\epsilon}} \pi^{*\prime A}(\tilde{I}(x)) f(x) dx$ in equilibrium.

Proof of Lemma 1: Unrestricted Choice requires that for a client i of type $\tau_i = a$, $a^2 + \phi - m_i p_a \geq ab - m_i p_b$, for all a, b, ϕ and m_i. Under Client Choice, let $\hat{t}(\tau, \underline{\phi}, m)$ be the treatment chosen (meaning their first choice) by a client of type $\tau, \underline{\phi}$ with marginal utility m. Consider all cases where a client misreports $\hat{t}^1 > \tau$:

- $\hat{t}(0, \underline{\phi}, m) = 0 \succ \hat{t}(0, \underline{\phi}, m) = 1$ if $\underline{\phi} - mp_0 \geq \mu - mp_1, \forall m$, or $p_1 - p_0 \geq \frac{\mu - \underline{\phi}}{\overline{m}}$, for $\frac{\underline{\phi}}{\underline{m}} \geq p_0 \geq 0$ and $p_1 \leq \frac{\mu + \underline{\phi}}{\underline{m}}$. This can be satisfied if

$$\overline{m} p_0^* \leq \mu - \underline{\phi} + \overline{m} \frac{\mu + \underline{\phi}}{\underline{m}}, \tag{38}$$

  for $p_0 \geq 0$. As $p_0 \geq 0$, the Fairness Constraint can only be satisfied if

$$\frac{\overline{m}}{\underline{m}} \geq \frac{\mu - \underline{\phi}}{\mu + \underline{\phi}}. \tag{39}$$

It also implies that if (39) holds, $p_0^* = \min\{\frac{\underline{\phi} - \mu}{\overline{m}} + \frac{\mu + \underline{\phi}}{\underline{m}}, \frac{\underline{\phi}}{\underline{m}}\}$. If $\frac{\overline{m}}{\underline{m}} \geq 1 - \frac{\underline{\phi}}{\mu}$, $p_0^* = \frac{\underline{\phi}}{\underline{m}}$. If this holds for type $\underline{\phi}$, it also holds for type $\overline{\phi}$. It is also straightforward to show that with this contract, no client wishes to misreport downwards.

If $\frac{\overline{m}}{\underline{m}} < 1 - \frac{\underline{\phi}}{\mu}$, no types can screen. Let $\hat{t}(\tau, \phi, m)$ be the ranking of a client with types τ, ϕ and m. The Fairness Constraint with Assigned Consumption implies

that $\hat{t}(\tau, \underline{\phi}, m) = (1 \succ 0)$, $\hat{t}(1, \overline{\phi}, m) = (1 \succ 0)$, and $\hat{t}(0, \overline{\phi}, m) = (0 \succ 1)$ for all m. Then $\hat{t}(\tau, \underline{\phi}, m) = (1 \succ 0)$ requires $\frac{\mu}{\underline{m}} - p_1 \geq \frac{\underline{\phi}}{\underline{m}} - p_0$. This holds with equality so $p_1 - p_0 = \frac{\mu - \underline{\phi}}{\underline{m}}$. We show below that for any assigned consumption mechanism, $prob(t = 0|\tau) > 0$ for all τ as $q < 1$. Because $T(\tau, 0) = 0$ for $\tau = 1 \implies r_0 \leq 0$. Letting this hold with equality implies $P^{da} = \{0, \frac{\mu - \underline{\phi}}{\underline{m}}\}$ and $p^{da} = \frac{0 + \frac{\mu - \underline{\phi}}{\underline{m}}}{2} = \frac{\mu - \underline{\phi}}{2\underline{m}}$. $\square$

Proof of Proposition 1: Expected surplus is $ES^* = \int_{\underline{\epsilon}}^{\epsilon} \frac{\mathbb{D} + \epsilon}{w^A - p^A} \left(v + T^A - \frac{\delta(\mathbb{D} + \epsilon) f(x) dx}{2(w^A - p^A)} \right)$, or

$$ES^* = \frac{D}{w^A - r^A} \left(v + T^A - \frac{\delta D}{2(w^A - p^A)} \right) - \frac{\delta \sigma_\epsilon^2}{2(w^A - p^A)^2}, \tag{40}$$

where $\sigma_\epsilon^2 = \int_{\underline{\epsilon}}^{\epsilon} x^2 f(x) dx$. The firm has exogenous randomness in client coverage parameterized by σ_ϵ^2. This plays no role in the analysis for the non-profit. As a result, for expositional ease we considered the case where $\sigma_\epsilon^2 = 0$ above, but all the proofs here allow $\sigma_\epsilon^2 > 0$ and do not change qualitative outcomes.

From Lemma 1, $p = p^{da}$. Let $\hat{\tau}^a$ be the agent's posterior. Then as $t \perp (\hat{\tau}^a | \tau^a) \implies \hat{U}^a = \overline{U}^a \implies w = \overline{w}$ and (5) holds. Then $I = \frac{D + \epsilon}{\overline{w} - p^{da}}$.

The mechanism begins with clients $i = 1, 2, ...$reporting $\hat{t}$. The firm offers $I_i = 1 \forall i = 1, 2, ... \frac{D + \epsilon}{\overline{w} - p^{da}}$. $P^{da} = \{0, \frac{\mu - \underline{\phi}}{\underline{m}}\} \implies \hat{t}(\tau, \underline{\phi}, m) = (2 \succ 1 \succ 0)$, $\hat{t}(2, \overline{\phi}, m) = (2 \succ 1 \succ 0)$, $\hat{t}(1, \overline{\phi}, m) = (1 \succ 2 \succ 0)$, and $\hat{t}(0, \overline{\phi}, m) = (0 \succ 2 \succ 1)$ for all m. As a result

$$\hat{t} = \begin{cases} 1 \succ 0 & \text{with probability } 1 - \frac{h}{2} \\ 0 \succ 1 & \text{with probability } \frac{h}{2}. \end{cases} \tag{41}$$

1. *The firm allocates $t = 1$ to all clients for which $\tau^a = 1$ and $\hat{t} = (1 \succ 0)$.* Consider all who report $\hat{t} = (1 \succ 0)$. Of these, $\frac{1}{2}$ are $\tau = 1$ and $\frac{(1-h)}{2}$ are $\tau = 0$. $Prob(\tau^a = 1 | \hat{t} = (1 \succ 1 \succ 0)) = 1 - h(1 - q)$. Let $O(t|\tau^a, \hat{t})$ be the probability that a client with $\hat{t}$ and signal τ^a receives an offer of treatment t. Step 1 of the algorithm $\implies O(1|1, (1 \succ 1 \succ 0)) = 1$. All offers are accepted.

2. *The firm randomly allocates $t = 1$ to clients for which $\tau^a = 0$ and $\hat{t} = (1 \succ 0)$ until there are no more $t = 1$ slots.* There are then $\frac{h(1-q)}{2}$ remaining $t = 1$ slots available per capita, and $\frac{q(1-h) + (1-q)}{2}$) clients for whom $\hat{t} = (1 \succ 0)$ and $\tau^a = 0$. Step 2 $\implies O(1|0, (1 \succ 0)) = r = \frac{h(1-q)}{q(1-h) + 1 - q} = \frac{h(1-q)}{1 - hq}$. These offers are all accepted.

3. *The firm allocates $t = 0$ to all other clients.*

Expected surplus is then given by

$$ES^{da} = \frac{D}{\overline{w} - p^{da}} \left(v + T^{da} - \frac{\delta D}{2(\overline{w} - p^{da})} \right) - \frac{\delta \sigma_\epsilon^2}{2(\overline{w} - p^{da})^2}. \tag{42}$$

where

$$T^{da} = \frac{\mu}{2} + q\hat{\phi} + \frac{h(1-q)\overline{\phi}}{2} + \frac{r((1-q)\hat{\phi} - q(1-h)\underline{\phi})}{2}. \tag{43}$$

Proof or Proposition 3: Consider an agent who receives a signal τ^a on a client. With probability λ, $t_i = \tau_i^a$. With probability $1-\lambda$, $t_i = \tilde{\tau}_i$. The expected value of the treatment from the perspective of the agent is then $\hat{U}^{ca} \approx \lambda q\hat{\phi} + (1-\lambda)\hat{\phi}$. The firm's belief of surplus is $\lambda q\hat{\phi} + (1-\lambda)\frac{\hat{\phi}}{2}$. The agent's belief of the surplus from a random alternative agent is

$$\overline{U}_i^{ca} = \begin{cases} q\hat{\phi} & \text{with probability } \lambda^2 \\ \rho\hat{\phi} & \text{with probability } (1-\lambda)^2 \\ q\hat{\phi} & \text{with probability } \lambda(1-\lambda). \\ \frac{\hat{\phi}}{2} & \text{with probability } (1-\lambda)\lambda, \end{cases} \tag{44}$$

where $\hat{U}^{ca} - \overline{U}^{ca} = (1-\lambda)(1-\frac{\lambda}{2} - (1-\lambda)\rho)\hat{\phi}$. Then from (5), $w^* = \overline{w} - \alpha(1-\lambda)(1-\frac{\lambda}{2} - (1-\lambda)\rho)\hat{\phi}$ (we ignore the approximation). Then as prices are $P^{ac} = \{0, \frac{\mu}{\underline{m}}\}, I^a = \frac{D+\epsilon}{w^* - p^b}$. The surplus from treatment is then

$$T^a = \mu + \lambda q\hat{\phi} + (1-\lambda)\frac{\hat{\phi}}{3} < T^b. \tag{45}$$

and

$$S^a = \frac{D}{w^* - p^{ac}}\left(v + T^a - \frac{\delta D}{2(w^* - p^{ac})}\right) - \frac{\delta\sigma_\epsilon^2}{2(w^* - p^{ac})^2}.\square \tag{46}$$

Pareto Efficient Agent Assignment or Agent Deferred Acceptance: We now allow the client to report $\hat{t}$ at stage 1 of the mechanism and set prices at the level at which $\hat{t}(\tau, \underline{\phi}, m) = (1 \succ 0), \hat{t}(1, \overline{\phi}, m) = (1 \succ 0)$ and $\hat{t}(0, \overline{\phi}, m) = (0 \succ 1)$, namely by choosing P^{da}. Pareto efficiency implies that $\hat{t} = 0 \succ 1 \implies t = 0$, as the agent's posterior is 1. After allocating these clients, the firm follows the Deferred Acceptance algorithm. There is a fraction $\frac{1}{2} - q\frac{h}{2}$ of clients with $\tau_i^a = 0$ of whom $\tau_i^a \neq \tilde{\tau}_i^a$ with probability $(1-\lambda)$. There is a fraction $\frac{1}{2} - (1-q)\frac{h}{2}$ of clients with $\tau_i^a = 1$ of whom $\tau_i^a \neq \tilde{\tau}_i^a$ with probability $(1-\lambda)$.

the agent has a fraction $1 - \frac{h}{2}$ clients for which

$$\tau_i^a = \begin{cases} 1 & \text{with probability } \frac{1}{2} - (1-q)\frac{h}{2} \\ 0 & \text{with probability } \frac{1}{2} - q\frac{h}{2}. \end{cases} \tag{47}$$

The agent has beliefs that

$$\tilde{\tau}_i^a = \begin{cases} 1 & \text{with probability } \lambda(\frac{1}{2} - (1-q)\frac{h}{2}) + (1-\lambda)(\frac{1}{2} - \frac{h}{4}) \\ 0 & \text{with probability } \lambda(\frac{1}{2} - q\frac{h}{2}) + (1-\lambda)(\frac{1}{2} - \frac{h}{4}). \end{cases} \tag{48}$$

Per client, there are $\frac{1}{2}$ slots for $t=1$, and $1-\frac{h}{2}$ slots for $t=0$. The agent assigns $t=1$ to all clients with $\tilde{\tau}_i^a=1$. There are then $\frac{1}{2}-\lambda(\frac{1}{2}-(1-q)\frac{h}{2})+(1-\lambda)(\frac{1}{2}-\frac{h}{4})$ remaining $t=1$ slots. He randomly offers those to the $\lambda(\frac{1}{2}-q\frac{h}{2})$ clients for whom he agrees with the $\tau_i^a=0$ signal. This implies randomizing with probability r above. The remaining clients are assigned $t=0$.

If he accepts the position, the agent's ex ante belief of the client's welfare is

$$\mu+\frac{h}{2}\overline{\phi}+(\frac{1}{2}-(1-q)\frac{h}{2})\lambda q\hat{\phi}+(1-\frac{h}{2})(1-\lambda)\frac{\hat{\phi}+\underline{\phi}}{2}+ \\ (\frac{1}{2}-\frac{qh}{2})\lambda q\underline{\phi}+\lambda r\left((1-q)\hat{\phi}-q\underline{\phi}\right). \quad (49)$$

If another agent accepts the position, the agent's ex ante belief of the client's welfare is

$$\mu+\frac{h}{2}\overline{\phi}+(\frac{1}{2}-(1-q)\frac{h}{2})\lambda^2 q\hat{\phi}+(\frac{1}{2}-(1-q)\frac{h}{2})\lambda(1-\lambda)\frac{\hat{\phi}}{2}+ \\ (1-\lambda)^2(\frac{1}{2}-\frac{h}{4})\rho\hat{\phi}+(1-\lambda)\lambda(\frac{1}{2}-\frac{h}{4})q\hat{\phi}+ \\ (1-\lambda)^2(\frac{1}{2}-\frac{h}{4})\rho\underline{\phi}+(1-\lambda)\lambda(\frac{1}{2}-\frac{h}{4})q\underline{\phi}+(\frac{1}{2}+q\frac{h}{2})\lambda^2 q\underline{\phi}+ \\ (\frac{1}{2}+q\frac{h}{2})\lambda(1-\lambda)\frac{\underline{\phi}}{2}+\lambda r\left((1-q)\hat{\phi}-q\underline{\phi}\right), \quad (50)$$

which simplifies to

$$\mu+\frac{h}{2}\overline{\phi}+(\frac{1}{2}-(1-q)\frac{h}{2})\lambda^2 q\hat{\phi}++(\frac{1}{2}-(1-q)\frac{h}{2})\lambda(1-\lambda)\frac{\hat{\phi}}{2}+ \\ (1-\lambda)^2(\frac{1}{2}-\frac{h}{4})(\rho+\frac{1-\rho}{2})(\hat{\phi}+\underline{\phi})+(1-\lambda)\lambda(\frac{1}{2}-\frac{h}{4})q(\hat{\phi}+\underline{\phi})+ \\ +(\frac{1}{2}+q\frac{h}{2})\lambda^2 q\underline{\phi}+(\frac{1}{2}+q\frac{h}{2})\lambda(1-\lambda)\frac{\underline{\phi}}{2}+\lambda r\left((1-q)\hat{\phi}-q\underline{\phi}\right). \quad (51)$$

The wage that must be paid to the agent is then given by

$$w^{**}=\overline{w}- \\ (1-\lambda)\left(\frac{\lambda}{2}\left(\frac{1}{2}-(1-q)\frac{h}{2})\right)\hat{\phi}+(\frac{1}{2}+q\frac{h}{2})\underline{\phi}\right)+\frac{\hat{\phi}+\underline{\phi}}{2}(1-\frac{h}{2})(1-(1-\lambda)\rho)\right)>w^*. \quad (52)$$

□

Proof of Proposition 4: Let $\hat{t}$ now refer to the good chosen by the client. Unrestricted Choice arises when the client chooses $t_i = \tau_i$ at a price p_i for all i. As $p_0 \geq 0$, this implies that the Fairness Constraint can only be satisfied if $\frac{V'(\hat{y}+\overline{\eta})}{V'(\hat{y}+\underline{\eta})} \equiv \frac{\overline{m}}{\underline{m}} \geq \frac{\mu-\underline{\phi}}{\mu+\underline{\phi}}$. Note that $\frac{d\left(\frac{V'(\hat{y}+\overline{\eta})}{V'(\hat{y}+\underline{\eta})}\right)}{d\hat{y}} > 0$ if $\frac{V''(\hat{y}+\overline{\eta})}{V'(\hat{y}+\overline{\eta})} > \frac{V''(\hat{y}+\underline{\eta})}{V'(\hat{y}+\underline{\eta})}$ as occurs under Decreasing Absolute Risk Aversion. This implies that there exists a critical average income $\hat{y}^{uc}$ such that $\frac{V'(\hat{y}+\overline{\eta})}{V'(\hat{y}+\underline{\eta})} \geq \frac{\mu-\underline{\phi}}{\mu+\underline{\phi}}, \forall \hat{y} \geq \hat{y}^{uc}$. Lemma 1 generates the optimal contract, given by $P^{uc} = \{p_0^{uc}, \frac{\mu+\underline{\phi}}{\underline{m}}\}$, for all $\hat{y} \geq \hat{y}^{uc}$. Straightforward calculations show that $p_0^* = \frac{\underline{\phi}}{\underline{m}}$ for $\hat{y} \geq \tilde{y}^{uc}$ where $\frac{V'(\tilde{y}^{uc}+\overline{\eta})}{V'(\tilde{y}^{uc}+\underline{\eta})} = 1 - \frac{\underline{\phi}}{\mu}$. Then revenue per client from the mechanism is $p^{uc} = \frac{\mu+p_0^{uc}+\underline{\phi}}{2V'(y^{uc}-\underline{\eta})}$ for $y^{uc} \leq \hat{y} \leq \tilde{y}^{uc}$ and $p^{uc} = \frac{\frac{\mu}{2}+\underline{\phi}}{V'(y^{uc}-\underline{\eta})}$ for $y^{uc} > \tilde{y}^{uc}$. As $T^{uc} = T^* = \mu + \underline{\phi}$ and $I^{uc} = \frac{D}{\overline{w}-r^{uc}}$,

$$ES^{uc} = \frac{D}{\overline{w}-p^{uc}}\left(v + T^* - \frac{\delta D}{2(\overline{w}-p^{uc})}\right) - \frac{\delta\sigma_\epsilon^2}{2(\overline{w}-p^{uc})^2}.\square \tag{53}$$

The Non-Profit Outcome when $\sigma_\epsilon^2 > 0$: Let $\Delta p = p^{ac} - p^{da} = \frac{\underline{\phi}}{2\underline{m}}$ be the additional per client revenue from Bureaucratic Assignment. The firm chooses Bureaucratic Assignment for all $D \leq D_b^*$, where D_b^* is uniquely defined by

$$(T^{da} - T^b)D_b^* = \frac{D_b^*\Delta p}{(\overline{w}-p^{da})(\overline{w}-p^{ac})}\left(v + T^b - \frac{\delta(D_b^* + \frac{\Delta p}{2})}{2(\overline{w}-p^{ac})}\right) - \frac{\delta[(\overline{w}-p^a c)^2 - (\overline{w}-p^{da})^2]\sigma_\epsilon^2}{2(\overline{w}-p^{ac})^2(\overline{w}-p^{da})^2}. \tag{54}$$

Similarly, the firm chooses Agent Assignment over Bureaucratic Assignment for all $D \leq D_a^*$, where D_a^* is uniquely defined by

$$(T^b - T^a)D_a^* = \frac{D_a^*\Delta w}{(w^*-p^{ac})(\overline{w}-p^{ac})}\left(v + T^a - \frac{\delta(D_a^* + \frac{\Delta w}{2})}{2(w^*-p^{ac})}\right) - \frac{\delta[(\overline{w}-p^{ac})^2 - (w^*-p^{ac})^2]\sigma_\epsilon^2}{2(w^*-p^{ac})^2(\overline{w}-p^{ac})^2}. \tag{55}$$

The Restricted Choice Mechanism: The firm knows I^{rc} when allocating services to all clients. The Restricted Choice mechanism is:

1. The firm chooses quantities κ_1^* and unit prices $P^{rc} = (p_0^{rc}, p_1^{rc})$.
2. All clients $i \leq I^{rc}$ report $\hat{t}$ and receive their first choice.
3. The firm assigns $I_I = 1$ the remaining $t = 2$ at price r_2^{rc*} until there is no more remaining supply.

The outcome is given in Proposition 6.

Proof of Proposition 6: *The firm chooses quantities of each* $t_i = (1, \kappa_1^*)$ *at unit prices* (p_0^{rc}, p_1^{rc}). Under Restricted Client Choice, let $\hat{t}(\tau, \underline{\phi}, m)$ be the good chosen by a client of type $\tau, \underline{\phi}$ with income m. Then

- $\hat{t}(0, \underline{\phi}, m) = 0 \succ \hat{t}(0, \underline{\phi}, m) = 1$ if $\underline{\phi} - m_i p_0 \geq \kappa_1(\mu - m_i p_1), \forall m$, for $\frac{\underline{\phi}}{\underline{m}} \geq p_0 \geq 0$ and $p_1 \leq \frac{\mu + \underline{\phi}}{\underline{m}}$.

If the Fairness Constraint can be overcome then for the same reasons as in Lemma 1, $p_1^* = \frac{\mu + \underline{\phi}}{\underline{m}}$ as this maximizes revenue and the first two constraints. For $\hat{y} < \tilde{y}^{uc}$, the constraints on p_0 bind and $\underline{\phi} + \overline{m} p_0^* = \kappa_1^*(\mu + \overline{m}\frac{\mu + \underline{\phi}}{\underline{m}})$ for $p_0^* \geq 0$. The benefits and costs of reducing p_0^* from $\underline{\phi}$ to 0 are linear, as the cost of diminishing returns is also linear, so there are two relevant corner solutions as optima, where $p_0^* = \frac{\underline{\phi}}{\underline{m}}$ and where $p_0^* = 0$. We consider each in turn.

First, consider κ_1 at prices P^{uc}. Then (37) with $k_0 = 1$ implies that $(\hat{t}|\tau = 0) = 0 \succ (\hat{t}|\tau = 0) = 1$ iff $(1 - \frac{m}{\underline{m}})\underline{\phi} \geq \kappa_1(\mu - \frac{m}{\underline{m}}(\mu + \underline{\phi}))$ for all m. This requires that

$$\kappa_1^* \leq \frac{1}{\frac{\mu}{\underline{\phi}} - \frac{\overline{m}}{\underline{m} - \overline{m}}}. \tag{56}$$

Now consider the assignment component. There is remaining supply of $I^{rc}(\frac{1 - \kappa_1^*}{2})$. This is assigned to all clients $i = I^{rc} + 1, ..., I^{rc}(1 + \frac{1 - \kappa_1^*}{2})$ at a price of $p_1^{rc*} = \frac{\mu}{\underline{m}}$. Then revenue per client is $(1 - \frac{1 - \kappa_1^*}{2})p^{uc} + \frac{1 - \kappa_1^*}{2} p^{ac} = p^{rc}$. Restricted choice then has surplus of

$$\begin{aligned} S^{rc} = I^{rc} \Bigg(& v + \mu + \hat{\phi}(1 - \frac{1 - \kappa_1^*}{2}) + \frac{1 - \kappa_1^*}{2}\frac{\hat{\phi}}{2} \Bigg) \\ & - \delta \frac{I^{rc2}}{2} \left((1 - \frac{1 - \kappa_1^*}{2}) + (1 - \kappa_1^*)(1 + \frac{1 - \kappa_1^*}{4}) \right) \\ & - \frac{\delta \sigma_\epsilon^2 \left((1 - \frac{1 - \kappa_1^*}{2}) + (1 - \kappa_1^*)(1 + \frac{1 - \kappa_1^*}{4}) \right)^2}{2(\overline{w} - p^{rc})^2}. \end{aligned} \tag{57}$$

The Artificial Currency Game: The firm offers supply of $t_0 = \frac{I^{rc}}{3}$ and $t_1 = \frac{\kappa_1^* I^{rc}}{3}$ at unit fiat money prices of $P = P^*$. In addition, each client $i = 1, 2, ...I^{rc}$ is given 1 unit of artificial currency to spend only on services. This highest bidders for each service have the option to purchase t_i and unit prices in P^*. Each client reports minimum quantities λ_i of each t_i that they are willing to offer for their unit of currency at a unit fiat money price of p_i. Any currency not spent has no value. Any over-subscribed treatments are assigned equally to all paying the equilibrium artificial currency price.

Let the equilibrium quantities per client be $(\lambda_0^*, \lambda_1^*)$. Then $(\lambda_0^*, \lambda_1^*) = (1, \frac{1}{\kappa_1^*})$ is a competitive equilibrium outcome. Consider a client i of type τ_{ik}, $k = 0, 1$. Then consider the following equilibrium strategy: $\lambda_k = 0$ and $\lambda_{-k} = 1$. With these strategies, there is demand for each good of $\frac{I^{rc}}{2}$ at a minimum allocation of 0, so with equal rationing the equilibrium quantities of each good are $(t_0^*, t_1^*) = (1, \frac{1}{\kappa_1^*})$. Finally consider the equilibrium strategies of each agent. By the definition of κ_i^*, client i of type $\tau_i = k$ prefers the quantity κ_i^* at unit price p_i, where $\kappa_0^* = 1$ to any κ_{-i}^*. Hence this is an equilibrium.

Proof of Lemma 1a: Unrestricted Choice requires that for a client i of type $\tau_i = a$

$$a^2 + \phi - mp_a \geq ab - mp_b, \tag{58}$$

for all a, b, ϕ and m. Under Client Choice, let $\hat{t}(\tau, \underline{\phi}, m)$ be the treatment chosen (meaning their first choice) by a client of type $\tau, \underline{\phi}$ with income m. Consider all cases where a client misreports $\hat{t}^1 > \tau$. This implies that

- $\hat{t}(\tau, \underline{\phi}, m) = \tau \succ \hat{t}(\tau, \underline{\phi}, m) = \tau'$ if $\tau\mu + \underline{\phi} - mp_0 \geq \tau'\mu - mp_1, \forall m$, or $p_{\tau'} - p_\tau \geq \frac{\tau'\mu - \underline{\phi}}{\tau\mu\overline{m}}$, for $\frac{\underline{\phi}}{\underline{m}} \geq p_\tau \geq 0$ and $p_{\tau'} \leq \frac{\tau\ \mu + \underline{\phi}}{\underline{m}}$. The most binding is $\tau = 0$ and $\tau' = n-1$ and can be satisfied if

$$\overline{m}p_0^* \leq (n-1)\mu - \underline{\phi} + \overline{m}\frac{(n-1)\mu + \underline{\phi}}{\underline{m}}, \tag{59}$$

  for $p_0 \geq 0$. As $p_0 \geq 0$, this implies that the Fairness Constraint can only be satisfied if

$$\frac{\overline{m}}{\underline{m}} \geq \frac{(n-1)\mu - \underline{\phi}}{(n-1)\mu + \underline{\phi}}. \tag{60}$$

It also implies that if (60) holds,

$$p_0^* = \min\{\frac{\underline{\phi} - (n-1)\mu}{\overline{m}} + \frac{(n-1)\mu + \underline{\phi}}{\underline{m}}, \frac{\underline{\phi}}{\underline{m}}\}. \tag{61}$$

$p_0^* = \frac{\underline{\phi}}{\underline{m}}$ if $\frac{\overline{m}}{\underline{m}} \geq 1 - \frac{\underline{\phi}}{(n-1)\mu}$. If this holds for type $\underline{\phi}$, it also holds for type $\overline{\phi}$. It is also straightforward to show that with this contract, no client wishes to misreport downwards.

Now consider Deferred Acceptance. Let $\hat{t}(\tau, \phi, m)$ be the ranking of a client with types τ, ϕ and m. The Fairness Constraint with Assigned Consumption implies that $\hat{t}(\tau, \underline{\phi}, m) = (n \succ (n-1) \succ .. \succ 0)$ and $\hat{t}(\tau, \overline{\phi}, m) = (\tau \succ n \succ (n-1)... \succ 0)$ for all m. But this implies that for a client of type τ

$$\frac{\mu}{\underline{m}} - p_{\tau+1} \geq \frac{\underline{\phi}}{\underline{m}} - p_\tau, \tag{62}$$

These hold with equality and $p_{\tau+1} - p_\tau = \frac{\mu-\underline{\phi}}{\underline{m}}$ and $p_0 = 0$. This implies $P^{da} = \{0, \frac{\mu-\underline{\phi}}{\underline{m}}, \frac{2(\mu-\underline{\phi}))}{\underline{m}}, ...\}$ and $p^{da} = \frac{(n-1)(\mu-\underline{\phi})}{2\underline{m}}$. For Bureaucratic Assignment, the firm does not have the client incentive compatibility constraint and revenue is $p^b = \frac{(n-1)\mu}{2\underline{m}}$.

We now show the Deferred Acceptance outcome when $p_j^{da} = \frac{(j-1)(\mu-\underline{\phi})}{\underline{m}}$ so it can be implemented by non-random Serial Dictatorship. We show this for h low, but the conceptual results generalize to all values of h. The mechanism is as follows. From the timing assumptions, the firm knows ϵ and hence I^{da} when allocating t_i. It offers $I_i = 1$ to all $i \leq I^{da}$ and 0 to all other clients.

1. The firm chooses $P^{da} = (p_0, p_1, ..., p_{n-1})$ and all clients $i < I^{da}$ report $\hat{t}_i = (\hat{t}_1, \hat{t}_2, ..\hat{t}_{n-1})$.

2. The firm *offers* $t_i \in \{0, 1, ...n-1\}$ to all clients $i < I^{da}$ for which $\tau_i^a = n-1$. If there are remaining $t = n-1$ slots, it randomly assigns them to clients for which $\tau_i^a = n-2$, then $\tau_i^a = n-3$..., until there are no more $t = n-1$ slots.

3. The firm *offers* $t_i \in \{0, 1, ...n-1\}$ to all clients $i < I^{da}$ for which $\tau_i^a = n-1$. If there are remaining $t = n-1$ slots, it randomly assigns them to clients for which $\tau_i^a = n-1$, then $\tau_i^a = n-1$..., until there are no more $t = n-1$ slots.

4. The firm *offers* $t_i \in \{0, 1, ...n-2\}$ to all clients $i < I^{da}$ for which $\tau_i^a = n-2$. If there are remaining $t = n-1$ slots, it randomly assigns them to clients for which $\tau_i^a = n-3$, then $\tau_i^a = n-4$..., until there are no more $t = n-2$ slots.

5. .....

6. The firm *offers* $t_i \in \{0, 1\}$ to all clients $i < I^{da}$ for which $\tau_i^a = 1$. If there are remaining $t = 1$ slots, it randomly assigns them to clients for which $\tau_i^a = 0$ there are no more $t = 1$ slots.

7. The firm allocates $t = 0$ to all other clients.

The outcome is as follows for $h \leq h^*$ defined below:

1. The firm offers $t \in \{0, 1, .., n-1\}$ to all the $\frac{1}{n}$ clients for which $\tau^a = n-1$. A fraction $1 - \frac{h(1-q)}{n}$ choose $t = n-1$ and a fraction $\frac{h(1-q)}{n(n-2)}$ choose each $t_i, i = 1, ..., n-2$.

2. The firm offers $t = n-1$ randomly to clients for which $\tau^a = n-2$ until there is no more supply of $t = n-1$ slots. This arises only if there are enough clients with $\tau^a = n-2$ who accept the $\frac{h(1-q)}{n}, j = n-1$ slots. This arises if

$$\frac{h(1-q)}{n} \leq \frac{1-q}{n^2} + \frac{(1-h)(1-q)(n-2)}{n^2} + \frac{(1-h)q}{n^2}. \tag{63}$$

Let $r_{n-1} = \frac{h(1-q)}{n}\frac{1-q}{n^2} + \frac{(1-h)(1-q)(n-2)}{n^2} + \frac{(1-h)q}{n^2}$ be the probability of a client with $\tau^a = n-2$ being offered $t = n-1$. This must be less than 1. This arises if or $h \leq h^* = \frac{2(n-1)+\frac{q}{1-q}}{n-1+\frac{q}{1-q}} \geq \frac{1}{2}$.

3. The firm offers $t \in \{0, 1, .., n-2\}$ to all the $\frac{1}{n}$ clients for which $\tau^a = n-2$. A fraction $1 - \frac{h(1-q))(n-1)}{n^2}$ choose $t = n-2$ and a fraction $\frac{h(1-q)}{n(n-1)}$ choose each $t_i, i = 1, ..., n-3$.

4. The firm offers $t = n-2$ randomly to clients for which $\tau^a = n-3$ until there is no more supply of $t = n-2$ slots. This arises only if there are enough clients with $\tau^a = n-3$ who accept or $r_{n-2} \leq 1$

5. This algorithm continues for the remaining treatments. r_{n-j} is simply computed to be
$$r_{n-j} = \frac{n-j}{\frac{q(1-h)}{(1-q)h} - j + 2 + \frac{n-1}{n}}. \tag{64}$$
Note that r_{n-j} is decreasing in j so that if $r_{n-1} \leq 1$ this also holds for all r_{n-j}. Hence if $h \leq h^*$ this is the outcome of the Deferred Acceptance mechanism.

Now let $h > h^* \implies \frac{h(1-q)}{n} > \frac{1-q}{n^2} + \frac{(1-h)(1-q)(n-2)}{n^2} + \frac{(1-h)q}{n^2}$. A client with $t_i^a = n-2$ is $t_i = n-1$ with probability 1. Let $n-k_1$ be defined by the highest value of τ^a such that not all clients with $\tau_i^a = \tau_{n-k_1}$ receive an offer of $t_i = n-1$. The game then continues by allocating $t = n-2$ using the residual supply of clients until no supply remains.

Unbalanced Demand and Supply: Here we identify the key outcomes for the case of $\zeta > 0$. First, Bureaucratic and Agent Assignment have the same prices as above, as they are set at the minimum valuations of those goods. Surplus from Bureaucratic Assignment is given by

$$\tilde{T}^b = \mu + \frac{1-\zeta}{2}q\hat{\phi} + \frac{1}{2}\frac{q(1+\zeta)}{q(1+\zeta)+(1-q)(1-\zeta)}\hat{\phi} + (q(1+\zeta)+(1-q)(1-\zeta)-\frac{1}{2})\frac{(1-q)(1-\zeta)}{q(1+\zeta)+(1-q)(1-\zeta)}\hat{\phi}. \tag{65}$$

Surplus from Agent Assignment $\mu + \frac{1-\zeta}{2}(\lambda q + \frac{1-\lambda}{2})\hat{\phi} + \frac{\zeta}{2}(\lambda(1-q)\hat{\phi} + \frac{1-\lambda}{2})\hat{\phi} + (\frac{1-(1-\lambda)(1+\zeta)}{2})q\hat{\phi}$, as the agent always assigns those clients he is certain about to $\tau = t$. As above, $\tilde{T}^b > \tilde{T}^a$. The wage reduction from Agent Assignment is $\tilde{\Delta}w = \alpha\hat{\phi}(\lambda(1-\lambda)(q-\frac{1}{2}) + (1-\zeta)(1-\lambda)^2\rho + \frac{(1-\lambda)\lambda}{2} + (1+\zeta)(1-\lambda)^2\rho + (1+\zeta)(1-\lambda)\lambda(1-q))$. Unrestricted Choice now requires that $p_0 = 0$ as some $\tau = 1$ clients are assigned $t = 0$. Then the condition for Unrestricted Choice is (9) but the maximum revenue from Unrestricted Choice is now $\frac{\mu+\phi}{2}$ rather than $\mu + \frac{\phi}{2}$ in the balanced case. This still exceeds the

revenue from any other mechanism. The κ_i^* for the case of Restricted Choice remain as above.

Finally, the Deferred Acceptance mechanism yields higher allocative efficiency than any other Assigned Consumption mechanisms. There is one potential change. In the balanced case, r represented the fraction of clients receiving $\tau^a = 0$ who are assigned $t = 1$. This qualitatively remains the case if $q\frac{1+\zeta}{2} + (1-h)(1-q)\frac{1-\zeta}{2} < \frac{1}{2}$ so that there are not enough $\tau^a = 1$ signals to fill the $\frac{1}{2}$ slots. In that case the probability that a client who reports $\hat{t} = 1 \succ 0$ but has $\tau^a = 0$ receives $t = 1$ is $\tilde{r} = \frac{1-q(1+\zeta)-(1-h)(1-q)(1-\zeta)}{(1-q)(1+\zeta)+(1-h)q(1-\zeta)}$. Treatment surplus per client is then given by

$$\tilde{T}^{da} = \mu + q\frac{1+\zeta}{2}\hat{\phi} + \tilde{r}\frac{(1-q)(1+\zeta)}{(1-q)(1+\zeta)+(1-h)q(1-\zeta)} + \frac{h}{2}\underline{\phi} + (1-\tilde{r})\frac{(1-h)q(1-\zeta)}{(1-q)(1+\zeta)+(1-h)q(1-\zeta)} > \tilde{T}^b. \quad (66)$$

If however, $q\frac{1+\zeta}{2} + (1-h)(1-q)\frac{1-\zeta}{2} > \frac{1}{2}$, then the firm randomly assigns $t = 1$ to clients with $\tau^a = 1$ and assigns the residual to $t = 0$. This occurs with probability $\overline{r} = 1 - q(1+\zeta) - (1-h)(1-q)(1-\zeta)$. $\tilde{T}^{da}$ is defined analogously and exceeds $\tilde{T}^b$. Hence all the features of the balanced model are retained.

Proof of Proposition 7: The setting now has $t \in \{0, 1, 2\}$. The program now becomes identical to that above except:

The Program: The firm begins the game by choosing A. The firm's objective given A is to choose I_i^{*A}, P^A and w^{*A} maximize

1. Prices $p_j^A, j = 0, 1, 2$ are given by $T(i, j|\underline{\phi}) \geq \underline{m}p_j^A$, in any equilibrium where type τ_i receives treatment $t = j$ with positive probability. Revenue per client is $p^A = \frac{p_0^A + p_1^A + p_2^A}{3}$.

2. Client Reporting in A:

   - If A involves Assigned Consumption and the allocation mechanism is strategy proof, this simplifies to

   $$T(i, 2|\underline{\phi}) - m_i p_2^A \geq T(i, j|\underline{\phi}) - m_i p_j^A, \forall m \in [\underline{m}, \overline{m}], \quad (67)$$

   $$T(i, 1|\underline{\phi}) - m_i p_1^A \geq T(i, 0|\underline{\phi}) - m_i p_0^A, \forall m \in [\underline{m}, \overline{m}], \quad (68)$$

   and

   $$T(i, i|\overline{\phi}) - m_i p_i^A \geq T(i, j|\overline{\phi}) - m_i p_j^A, \forall m \in [\underline{m}, \overline{m}]. \quad (69)$$

The firm then chooses the A that maximizes ES given these constraints.

We first show existence of a Partial Choice equilibrium where (i) all $\tau = 1$ clients choose $t = 1$, (ii) all $\underline{\phi}$ clients of type $\tau \neq 2$ do not, and (ii) revenue raised exceeds r^b. This equilibrium has the following form. All $\tau = 1$ receive $t = 1$. Then the firm allocates all $t_i = \tau_i^a$ with probability 1 and allocates $t_i = 2$ with probability $\frac{1}{2}$ and $t_i = 0$ with probability $\frac{1}{2}$ when $\tau_i^a = 1$. To show existence of this equilbrium we consider incentive compatibility in this proposed equilibrium at the contract R^b. First, type $\tau = 2$ does not choose $t = 1$ if

$$\mu(1 - \frac{\overline{m}}{\underline{m}}) \leq (q + \frac{1-q}{4})(2\mu(1 - \frac{\overline{m}}{\underline{m}}) + \phi). \tag{70}$$

which always holds. Second, type $\tau = 1$ chooses $t = 1$ if

$$\mu(1 - \frac{\overline{m}}{\underline{m}}) + \phi \geq (\frac{q}{2} + \frac{1-q}{2})2\mu(1 - \frac{\overline{m}}{\underline{m}}), \tag{71}$$

which always holds. Finally, a client with preferences $\underline{\phi}$ and $\tau = 0$ does not choose $t = 1$ only if

$$\mu(1 - \frac{\overline{m}}{\underline{m}}) \leq \frac{3(1-q)}{4}2\mu(1 - \frac{\overline{m}}{\underline{m}}) + (q + \frac{1-q}{2})\underline{\phi}, \tag{72}$$

which holds only if

$$\frac{\overline{m}}{\underline{m}} \geq 1 - \frac{(1+3q)\underline{\phi}}{2\mu(3q-1)}. \tag{73}$$

Note that $1 - \frac{(1+3q)\underline{\phi}}{2\mu(3q-1)} < 1 - \frac{\underline{\phi}}{\mu}$, which defined $\tilde{y}^{uc}$. The condition for Unrestricted Choice remains $\tilde{y}^{uc}$ with three treatments. If these conditions hold, then an equilibrium exists at prices P^b. But if the final constraint is slack, the firm can increase p_1 above $p_1 = \frac{\mu}{\underline{m}}$, thereby showing $p^{1st} \geq p^b$. Note also that $T^{1st} = \frac{2q\hat{\phi}}{3} + \frac{\hat{\phi}}{3} > q\hat{\phi} = T^b$.

Partial Choice entails that one $\underline{\phi}$ type chooses their allocation $t = \tau_i$ and the other two are assigned. We conclude by showing that the only Partial Choice equilibrium is where type $\tau = 1$ choose $t = 1$.

1. Type $\tau = 0$. In order for type $\tau = 0$ to choose $t = 0$ it must be that

$$\underline{\phi} - m_i p_0 \geq min\{\mu - m_i p_1, 2\mu - m_i p_2\} \tag{74}$$

for all m. As $t = 1$ and $t = 2$ are assigned by DA, $p_1 \leq \frac{\mu}{\underline{m}}$ and $p_2 \leq \frac{2\mu}{\underline{m}}$ as services are misallocated with positive probability. Similarly, $p_1 \geq 0$. This implies that

$$\frac{\overline{m}}{\underline{m}} \geq min\{\frac{2\mu - \underline{\phi}}{2\mu}, \frac{\mu - \underline{\phi}}{\mu}\}. \tag{75}$$

Note however that the firm can use Unrestricted Choice, which has higher payoff, if $\frac{\overline{m}}{\underline{m}} \geq \frac{2\mu - \underline{\phi}}{2\mu + \underline{\phi}}$. But $\frac{2\mu - \underline{\phi}}{2\mu + \underline{\phi}} < min\{\frac{2\mu - \underline{\phi}}{2\mu}, \frac{\mu - \underline{\phi}}{\mu}\}$ and hence the firm never uses Partial Choice where $\tau = 0$ chooses.

2. Type $\tau = 2$. In order for only type $\tau = 2$ to choose $t = 0$ it must be all $\tau = 0$ types do not also do so. But this is the condition for Unrestricted Choice, and so the firm never implements Partial Choice with $\tau = 2$ choosing. □

Proof of Proposition 8: For any assignment mechanism A, $S^* = \int_{\underline{\epsilon}}^{\overline{\epsilon}} \frac{(\mathbb{D}(b)-\pi^A(x)+x)}{w^A-p^A}(v+T^A - \frac{\delta(\mathbb{D}(b)-\pi^A(x)+x)}{2(w^A-p^A)})f(x)dx$, where $\pi(x) = \pi(\tilde{I} - E\tilde{I})$, as $E\pi = 0$. The firm chooses $\pi^*(\tilde{I})$ holding donations fixed, though in equilibrium $\hat{\pi}^A = \int_{\underline{\epsilon}}^{\overline{\epsilon}} \pi^{*'A}(\tilde{I}(x))f(x)dx$. As $E\pi = 0$ this simplifies to $ES^* = \frac{\mathbb{D}(b)}{w^A-p^A}(v + T^A - \frac{\delta\mathbb{D}(b)}{2(w^A-p^A)}) - \int_{\underline{\epsilon}}^{\overline{\epsilon}} \frac{\delta(-\pi(x)+x)^2}{w^A-p^A}f(x)dx$. This is maximized at $\pi^*(\epsilon) = \epsilon$ or $\pi^{*'} = 1$ with equilibrium surplus $S^* = \frac{\mathbb{D}(b)}{w^A-p^A}(v + T^A - \frac{\delta\mathbb{D}(b)}{2(w^A-p^A)})$. As $\hat{\pi}^A = \int_{\underline{\epsilon}}^{\overline{\epsilon}} \pi^{*'A}(\tilde{I}(x))f(x)dx = 1$, this implies that $\mathbb{D}(b) = 0$ if $b = b^*$. □

Proposition 10 *The link between profit status and assignment mechanisms is as follows:*

- *If $\tilde{m}^{da} < V'(y^{uc} + \underline{\eta})$, no non-profit uses Unrestricted Choice. Then if*
  - *If $\tilde{m}^{uc} < V'(y^{uc} + \underline{\eta})$, some non-profit firms and all for-profit firms use Unrestricted Choice,*
  - *If $\tilde{m}^{uc} \geq V'(y^{uc} + \underline{\eta})$, all firms using Unrestricted Choice are for-profit.*
- *If $\tilde{m}^{da} \geq V'(y^{uc} + \underline{\eta})$, for profit firms can use any assignment mechanism.*

Proof of Proposition 10:

- *If $\tilde{m}^{da} < V'(y^{uc} + \underline{\eta})$, no non-profit uses Unrestricted Choice.* The firm always uses Unrestricted Choice if $\underline{m} \geq V'(y^{uc} + \underline{\eta})$. As a result, if $\tilde{m}^{da} < V'(y^{uc} + \underline{\eta}), \nexists \underline{m} \in [V'(y^{uc} + \underline{\eta}), \tilde{m}^{da}]$ where a for-profit firm uses DA.
  Then if
  - *If $\tilde{m}^{uc} < V'(y^{uc} + \underline{\eta})$, some non-profit firms and all for-profit firms use Unrestricted Choice,* The firm always uses Unrestricted Choice if $\underline{m} \geq V'(y^{uc} + \underline{\eta})$. If $\tilde{m}^{uc} < V'(y^{uc} + \underline{\eta}), \exists \underline{m} \in [V'(y^{uc} + \underline{\eta}), \tilde{m}^{uc}]$ where a non-profit firm uses Unrestricted Choice.
  - *If $\tilde{m}^{uc} \geq V'(y^{uc} + \underline{\eta})$, all firms using Unrestricted Choice are for-profit.* The firm always uses Unrestricted Choice if $\underline{m} \geq V'(y^{uc} + \underline{\eta})$. If $\tilde{m}^{uc} \geq V'(y^{uc} + \underline{\eta}), \nexists \underline{m} \in [V'(y^{uc} + \underline{\eta}), \tilde{m}^{uc}]$ where a non-profit firm uses Unrestricted Choice.
- If $\tilde{m}^{da} \geq V'(y^{uc} + \underline{\eta})$, for profit firms can use any assignment mechanism. The firm always uses Unrestricted Choice if $\underline{m} \geq V'(y^{uc} + \underline{\eta})$. If $\tilde{m}^{da} \geq V'(y^{uc} + \underline{\eta}), \exists \underline{m} \in [\tilde{m}^{da}, V'(y^{uc} + \underline{\eta})]$ where a for-profit firm uses DA or First Choice.

The Agency Interpretation: The firm now chooses non-contractible quality e at cost $C(e)$ per unit of I, where $C(0) = 0, C' > 0, C'' > 0$. We assume that $\delta = 0$ to isolate agency concerns. Clients value quality at $U_i^c = eT(\tau_i, t)I_i - V(y_i - r_j I_i)$. Quality is not observed by clients when they purchase, but is inferred in equilibrium and for perceived quality $\hat{e}$ will pay expected revenue of $r^A\hat{e}$. Firm utility is $\pi + S$ where $S = \sum_{i=1}^{N}(v_i + eT(\tau_i, t_i))I_i$. As above, we assume that $E\pi = 0$. Uncertainty is no longer given by ϵ but by $I(\epsilon; e)$ where $I(\epsilon; e) \sim F(\epsilon), f(\epsilon)$ over a support $(0, \overline{I})$, with mean

$$\tilde{I} = \frac{eD}{w^A + C(e^*) - p^A e}. \tag{76}$$

We assume that f_e is increasing in ϵ, and $\frac{f_e(x)}{f(x)}$ is increasing in x. The timing of the model is adapted to allow the choice of quality in *italics*:

1. The firm chooses A P^A, w^A, π^A and Π, which is observed by clients and the agent.

2. *The firm chooses* e and observes τ^a, the agent observes τ^a, and clients report $\hat{t}$.

3. $\tilde{I}$ *is revealed.*

4. I_i is implemented, p^A paid by clients, and w^A and π^A are paid.

To offer the best case for incentive contracting, we restrict attention to the surplus maximizing Bayes Nash equilibrium.

The analog to the donor-funded organization is $p^A\hat{e} - C(e) < w^A$. For a given allocation mechanism A, consider the first best choice of e, which would arise if e was contractible. The cost of each client served is $w^A - p^A e + C(e)$. Expected surplus is $ES = \frac{eD}{w^A + C(e) - p^A e}(v + T^A)$. This yields a first order condition for optimal quality

$$\frac{e^*(C'(e^*) - p^A)}{w^A + C(e^*) - p^A e^*} = 1. \tag{77}$$

Now consider the non-profit firm's incentives, where $\tilde{e}$ maximizes $S = \frac{\tilde{e}D}{w^A + C(\tilde{e}) - p^A\hat{e}}\left(v + T^A\right)$, where $\hat{e}$ is expected quality. Here the first order condition is $\frac{e^*C'(\tilde{e})}{w^A + C(\tilde{e}) - p^A\hat{e}} = 1$, or as beliefs are correct in equilibrium,

$$\frac{\tilde{e}C'(\tilde{e})}{w^A + C(\tilde{e}) - p^A\tilde{e}} = 1. \tag{78}$$

Comparing (78) to (77) shows that the non-profit chooses quality too low, as it does not internalize the impact of quality on revenue.

Now consider the outcome for a for-profit firm in this setting. The contract offered is $\pi(\tilde{I}) \leq \mathbb{D}(b)$. Implemented treatments are now given by $I \leq \frac{e(\mathbb{D}(b) - \pi(\tilde{I}))}{w^A + C(e) - p^A}$, where

$E\pi = 0$. The firm chooses e to maximize $\int_0^{\overline{I}} \pi(\tilde{I}(e))f(x)dx + e(v+T^A)\frac{\mathbb{D}(b)}{w^A+C(e)-p^A\hat{e}}$. The firm's choice of e is given by the first order condition

$$\int_0^{\overline{I}} \pi(x)f_e(x)dx + \frac{\mathbb{D}(b)(v+T^A)(w^A+C(e)-p^A\tilde{e}-eC'(e))}{(w^A+C(e)-p^A\hat{e})^2} = 0. \tag{79}$$

The firm wishes to implement e^* in (77). We consider Residual Claimant contracts, where the firm keeps all revenue above a certain level of $\tilde{I}$, which we call I^*. Given risk neutrality and MLRP, the optimal contract (not observed by donors) has a simple form:

$$\pi(\tilde{I}(x)) = max\{\tilde{I}(x)) - I^*, 0\} - \overline{\pi} \tag{80}$$

where $\overline{\pi} = \int_{I^*}^{\overline{I}} \pi(\tilde{I}(x))f(x)dx$ to ensure that $E\pi = 0$. Then I^* is set such that the analog to (79) holds:

$$\int_{I^*}^{\overline{I}} (\tilde{I}(x) - I^*)f_e(x)dx =$$
$$\mathbb{D}(b)(v+T^A)\left(\frac{w^A+C(e^*)-r^Ae^*-e^*C'(e^*)}{(w^A+C(e^*)-p^Ae^*)^2} - \frac{w^A+C(\tilde{e})-r^A\tilde{e}-\tilde{e}C'(\tilde{e})}{(w^A+C(\tilde{e})-p^A\tilde{e})^2}\right). \tag{81}$$

If

$$\int_0^{\overline{I}} \tilde{I}(x)f_e(x)dx <$$
$$\mathbb{D}(b)(v+T^A)\left(\frac{w^A+C(e^*)-p^Ae^*-e^*C'(e^*)}{(w^A+C(e^*)-p^Ae^*)^2} - \frac{w^A+C(\tilde{e})-p^A\tilde{e}-\tilde{e}C'(\tilde{e})}{(w^A+C(\tilde{e})-p^A\tilde{e})^2}\right), \tag{82}$$

then full residual claimant status is not enough to induce efficient effort and efficiency is not feasible. I^* is set such that the expected marginal income balances the need for additional quality (the term in large brackets), applied to the surplus for all clients. Holding donations fixed, this optimally corrects the quality provision problem, improving allocative efficiency.

While this increases allocative efficiency, it deters donors. The firm chooses π^* holding donations constant However, in equilibrium the donor's contribution funds mission-related activities with probability $F(I^*)$. Donations are $\mathbb{D}(b) = \mathbb{D}(F(I^*))$. When $I^* = 0$, all donations are crowded out. The firm then chooses between non-profit status with return S^n, or for-profit with return $ES^* = \frac{e^*F(I^*)D(\underline{m})}{\overline{U}^a+C(e^*)-p^Ae^*}\left(v+T^A\right)$.When $I^* = 0$, the firm can only operate as a for-profit if it can self-fund: $w^a + C(e^*) \leq p^Ae^*$. In that case, optimal quality is no longer given by (77), as the cost there is fewer clients served.