The Structure of the Social Organization

Canice Prendergast*

University of Chicago

February 4, 2025

Abstract

We offer a theory of the social organization. We model the social organization as allocating services to clients subject to the need to fund their mission. Without funding constraints, we replicate the outcomes of the existing market design literature. When funding is linked to service provision, in line with empirical evidence, we offer a much broader taxonomy of outcomes and optimal allocation mechanisms. These range from different forms of Assigned Consumption to (subsidized) Unrestricted Choice. We also identify a novel constraint facing firms on whether to incorporate as a non-profit rather than as a for-profit Social Enterprise.

*I am grateful to Jacob Leshno and seminar participants at the 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, p.93.

The private social sector is enormous, with non-profits alone accounting for over 10% of US employment, generating revenues of $3 trillion per annum.[1] There are also estimated to be ten million for-profit Social Enterprises worldwide. Social organizations play a fundamental role in society - they provide medical care to the sick, feed the hungry, offer mentoring and educational support, house those in need of shelter, and on and on. The vast majority of social organizations distribute goods and services to clients in areas such as social assistance, housing, education and health care.[2] Our focus is on how social organizations should be structured to most effectively distribute these services.

The relationship between allocation mechanisms, prices and profit has been perhaps *the* dominant theme of modern microeconomics.[3] There is no analog for the social organization.[4] The field of market design has of course addressed how best to allocate students to school, kidneys to patients, and so on. However, that literature does not address the fact that social organizations need funding. Social organizations consistently report financial issues as their greatest concern. Our starting point is that most social organizations fund themselves through how they distribute services, with non-profit firms receiving 49% of their revenues from fees from clients, with a further 22% from donors.[5]

The firm we study needs to effectively distribute services to clients and to raise funds to do so. When funding does not depend on service provision, we replicate the outcomes of the market design literature. When it does, we show a far wider range of outcomes and preferred allocation mechanisms. This arises because optimal allocation mechanisms both determine funding and are themselves generated by the firm's capacity to raise revenue. First, we show how different forms of *Assigned Consumption* arise, where the firm chooses services obtained by the client. The gold standard of the market design literature, Deferred Acceptance, is one example. The other end of the spectrum is where clients have *Unrestricted Choice* at subsidized

[1]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]This is through vast literatures on practices such as optimal auction design, non-linear pricing, quality distortions, bundling, platform competition, tie-ins, and so on.

[4]Closest is the classic work of Rose Ackerman, 1982, who addresses how non-profits allocate excessive resources to fund-raising, as a way of business stealing from other non-profit organizations.

[5]The remainder largely derives from grants from foundations, endowment income, and public funding.

prices. There are also intermediate outcomes. *Partial Choice* arises when a client can choose service A if it is under-subscribed, but service B is assigned by the firm. Finally, the firm could allow *Restricted Choice*, where the client is not allowed to choose between A and B, but rather between A and a less desirable service B'. For each of these ways of distributing services, the firm could do so as a non-profit or as a Social Enterprise. To illustrate the optimality of these outcomes, we use insights from market design, agency theory, and efficient screening.

We address the social organization through two lenses. First, that its mission is limited by its capacity to raise revenue. Second, that existing models of pricing are inappropriate for understanding how social organizations operate. Profit-maximizing firms typically design allocation mechanisms to sell to high "willingness to pay" clients, while excluding clients willing to pay less. For social organizations, who typically exist to serve those who cannot pay, this is anathema to their mission. We propose an alternative pricing model: that they are instead governed by a *Fairness Constraint*, where no client can be denied for lack of income. For many service providers, this is a lexicographic constraint that defines their mission of serving those in need.

It is useful to see the objectives of the firm in terms of extensive and intensive margins. The extensive margin of the number of clients served is limited by funding. To fund itself, the firm charges clients for services but relies on donors for residual funding. The intensive margin is allocative efficiency, the accuracy with which services match client needs. Clients are privately informed of their need, which varies in both ordinal and cardinal dimensions. Clients may be able to report their preferences. To allocate services, the firm hires an agent who obtains a noisy signal on the client's needs. The firm chooses an organizational form and an allocation mechanism - for example, a non-profit using Deferred Acceptance - where prices, donations and agent compensation depend on both choices. The firm's problem then becomes the quote that begins the paper: how best to allocate a scarce resource to clients who know more than the firm does.

Prices usually solve this problem. However, social organizations typically exist as (at least some) clients cannot pay enough to cover service costs. Prices play three roles in the canonical economic transaction. First, efficient pricing allows the firm to delegate choice to clients. 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 below, leading to the outcomes described above.

The objective of the existing market design literature is to maximize some measure of allocative efficiency. The key addition here is the interaction with funding. For a standard for-profit firm, more efficient firms capture greater market size and revenue. A natural inference could be that the same is true for the social organization. We show that this is not the case. We offer two distinct conceptual reasons why these are likely substitutes for many social organizations. In other settings, firm revenue is

independent of how efficiently it distributes services, with revenue no higher than if the firm tossed a coin to allocate services. Finally, in some settings, the traditional positive relationship applies. This leads to very different outcomes than in the existing literature.

Central to outcomes are the three roles of prices. First, we characterize Assigned Consumption as arising when optimal pricing of services solves none of the issues above, by revealing no more information on client preferences than with no prices. We first show that all Assigned Consumption mechanisms raise the same revenue from clients, independent of how allocatively efficient they are. A potentially appealing allocation mechanism in that setting is *Deferred Acceptance*, as it can at least incorporate ordinal client preferences in allocations. As in the existing literature, its outcome is Pareto efficient and strategy proof. However, more directly related to funding issues, Deferred Acceptance has the lowest level of client coverage of all mechanisms that we consider. On this dimension, it inverts the usual pecking order of mechanisms in the market design literature.

Disagreement among stakeholders is a characteristic of many mission-based organizations. The absence of credible information on client preferences (often because of the absence of prices) exacerbates this. The second Assigned Consumption outcome that we consider is *Agent Assignment*, where a social service provider assigns services at his discretion. Bureaucracies are often accused of ignoring the concerns of its clients: even ordinal client information is not used here. The agency literature on mission-based settings (described below) typically has the outcome that firms offer otherwise excessive autonomy to agents. Agents, who are motivated by altruism towards clients, are assumed to disagree with the firm over client needs in the absence of credible information. Because of such disagreement, Agent Assignment reduces the intensive margin of allocative efficiency. However, allowing autonomy reduces compensation, which allows for higher client coverage. The intensive and extensive margins are substitutes here. We describe an extensive empirical literature on the importance of worker autonomy among social service providers. For firms sufficiently short on resources, the firm chooses Agent Assignment over Deferred Acceptance. This arises even though implemented allocations are not Pareto efficient, inverting what is usually a minimal requirement in the market design literature. This arises because of a characteristic outcome in agency settings, where ex post efficiency is distorted to improve ex ante outcomes.

Rather than assigning services, many social organizations instead allow choice to clients at subsidized prices. *Unrestricted Choice* arises when all relevant client information is revealed through optimal prices. This has an obvious appeal in overcoming the "how shall I choose better than you?" component of the opening quote. We show how pricing is necessary but not sufficient for Unrestricted Choice. The problem is that the Fairness Constraint may be violated, as price menus that induce efficient choices for poorer clients can fail for richer ones. How measures of risk aversion change with wealth has been frequently used to predict savings and investment be-

havior. We show that it is the key factor the determines the capacity to use Choice. Specifically, if clients exhibit Decreasing Absolute Risk Aversion, (subsidized) Unrestricted Choice can screen efficiently if and only if average client income is sufficiently high. If so, revenue and allocative efficiency are positively related. This shows how better resourced social organizations can both use Choice and additionally generate even more revenue from clients.

Many services assigned by social organizations are vertically differentiated. Families prefer larger homes to smaller ones, patients prefer more tests, and almost all foods banks prefer to receive chicken than produce. Without prices, it is particularly hard to allocate these efficiently. Progress can be made when services are divisible. *Restricted Choice* offers 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 generates the benefits of client choice. However, it fails to appropriately target the highest priority clients. We use two recent settings, Feeding America and Covax, to illustrate this tradeoff. We also show that this outcome can be implemented by a common market design mechanism, namely by purchasing (or bidding for) services with artificial currency.

Partial Choice arises when a client can choose among any undersubscribed services but oversubscribed ones are assigned by the firm. This is not an outcome of Deferred Acceptance but is instead analogous to what is known as a First Choice mechanism. There is a small market design literature arguing for the merits of prioritizing matching first choices. We show two outcomes. The first is well known: that despite not being strategy proof, First Choice mechanisms can offer higher allocative efficiency than Deferred Acceptance. The second is not: Partial Choice raises more client revenue than either Deferred Acceptance or Agent Assignment. For firms short on revenue, this makes Partial Choice appealing. Despite the rarity of formal First Choice mechanisms, we argue that many social organizations routinely offer Partial Choice options to clients, between a less-than-optimal option for sure and an endogenous random alternative.

Viewing the social organization through the lens of efficient allocation mechanisms in the shadow of funding leads to a variety of novel implications. First, the market design literature has largely coalesced around the merits of Deferred Acceptance. All impacts of funding render Deferred Acceptance less likely. Going further, we show a simple and intuitive condition showing when no firm uses Deferred Acceptance *for any level of client ability to pay.* This arises because of two funding issues. Given this outcome, it is notable that funding is not central in any of the arenas where Deferred Acceptance has been successfully implemented: school choice, resident matching and kidney exchange.

Beyond the relationship to the previous literature, we offer a taxonomy of the non-profit. First, poorly resourced non-profits allocate services in a subjective manner where clients have little input. This is not done through incompetence or malfeasance but because the primary priority of small non-profits is to better recruit poorly paid

employees. Furthermore, firms targeting the poorest clients must choose Assigned Consumption as they cannot be trusted to choose without violating fairness. Second, as firms become better funded, they transition to more formal mechanisms. These take better account for the preferences of clients but still make bureaucratic error, which limits revenue from clients. Finally, wealthier non-profits can allow clients choose at appropriate prices without harming those who more financially needy. This has a complementary benefit that the firm earns enough revenue, potentially by enough to become self-funded.

So far, we have described strategies for the non-profit. Yet there are 10 million for-profit Social Enterprises worldwide, which account for 3% of global employment (World Economic Forum, 2024). The key characteristic of the non-profit is the Non-distribution Constraint, where residual income cannot be taken by the firm (Hansmann, 1980). The existing literature has focused on harmful firm incentives that arises from the for-profit social organization.[6] Yet the ability of the firm to take residual income as profit is a cornerstone of the capitalist system. Reflecting recent empirical evidence, we consider a setting where there is a value to the firm acting as residual claimant, such that the optimal organizational form is for-profit.

Despite this, we show that due to the behavior of donors (the other major source of income for the social organization), the sector is distorted towards too many non-profits. Our starting point here is that to a first approximation, small private donors do not give to for-profit enterprises.[7] Yet many have a clear social mission, and a large fraction of expenditures are on that mission. Why then do Social Enterprises not attract undirected donor funding? Donors overwhelmingly say that the primary consideration in giving is the impact of *their* gift on the mission. A donor gives holding the expected donations of others constant. As a result, she evaluates her potential gift at her belief of the use of the marginal dollar received by the firm. If the marginal dollar increases profit, she will not give as it has no impact on the mission. This arises even if the vast majority of funds are used on the mission.

As a result, an organizational form that increases allocative efficiency deters donors, such that the Social Enterprise can only exist if it is fully client-funded. This becomes the second conceptual reason why allocative efficiency and funding are substitutes, and offers a non-behavioral reason why a firm that predominantly uses resources for social objectives cannot attract donors. We show this in two settings where residual claimant status increase allocative efficiency: where the firm offers financial security and where incentive contracting is needed. Rather than deter donors, the sector remains distorted towards too many non-profits. This also means that the non-profit firm has to live "hand-to-mouth" to attract donors, never sure of how well it can cover its mission.

[6]The concern of the existing literature is that a for-profit firm will cut corners on unobserved quality to increase its profit.

[7]While Social Enterprises are typically funded by clients, it is not unusual for foundations or states to offer seed money to allow these firms to get off the ground.

We begin in Section 1 by describing the model. Section 2 considers the optimal form of the non-profit, while we address the for-profit outcome in Section 3. Section 4 addresses robustness and extensions while Section 5 provides an agency interpretation of the results. Section 6 concludes.

1 Model

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

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

The parameter $\mu > 0$ scales vertical differentiation and $\phi > 0$ reflects horizontal differentiation. We assume that the supply of types $\tau \in \{0, 1, 2\}$ and the distribution of treatments $t = \{0, 1, 2\}$ is balanced and uniform, where there are $\frac{1}{3}$ of all types and treatments. We also assume that the number of clients is large to avoid sampling error. Surplus to the firm is

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

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, $i = 1, 2, ...$[8] To allocate services to clients, the firm chooses an assignment mechanism A described below. The firm charges prices $R = \{r_0, r_1, r_2\}$, where $r_j \geq 0$ is the unit charge for treatment t_j.

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

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

There is uncertainty over the client's capacity to pay for services, which we model as variation in m_i, where m_i is privately known.[9]

We model this as variation in client income.[10] Let a client's indirect utility function be $V(y)$, where y is income and $V'(0) = \infty, V' > 0, V'' < 0$. We assume that

[8]There is a cost below to serving a client given by the wage cost of the agent. We interpret this cost as a diagnosis cost incurred to determine the appropriate t to offer the client.

[9]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.

[10]This could alternatively be interpreted as savings, liquidity, credit card debt, or access to alternative support.

$y \sim (\hat{y} + \underline{\eta}, \hat{y} + \overline{\eta})$, where $\hat{y} + \underline{\eta} \geq 0$ and $E\eta = 0$. Average income is $\hat{y}$. Let $V'(y) \equiv m(y)$ where $V'(\hat{y} + \underline{\eta}) \equiv \underline{m}(\hat{y})$ and $V'(\hat{y} + \overline{\eta}) \equiv \overline{m}(\hat{y})$. The poorest clients have marginal utility $\underline{m}(\hat{y})$ and the richest $\overline{m}(\hat{y})$. We assume Decreasing Absolute Risk Aversion, where $-\frac{V''(x)}{V'(x)}$ is decreasing in x. Income and type τ are independent.

The problem is only of interest if client preferences do not lead to trivial outcomes. The technology $T(\tau_i, t)$ and supply of types implies that τ_i should receive treatment $t = \tau_i$. If clients always preferred that allocation, the firm would simply ask them what they want. We choose a a combination of horizontal and vertical preferences such that client preferences are partially informative. We assume that $\phi > 0$ 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. Let $\hat{\phi} = h\overline{\phi} + (1 - h)\underline{\phi}$. To render the problem interesting, we make Assumption 1.[11]

Assumption 1: $\overline{\phi} > 2\mu$ and $\underline{\phi} < \mu$.

Assumption 1 implies that absent pricing those with $\phi = \underline{\phi}$ prefer $t = 2$, while those with $\overline{\phi}$ prefer τ_i.[12]

Note that the model set-up differs from the canonical matching setting in two ways. In that literature, both sides of the match have preferences for each others, and the objective is to find a stable match. Here the environment is closer to the traditional public finance structure where the planner's optimum depends on a privately known characteristic, here client need. In addition, the social organization is designed to provide goods to those who cannot afford them. It is not a setting where the principal has preferences beyond that. For instance, in a school choice setting, the school may have preferences over the optimal mix of students that even well enough resourced families would resolve.

[11]Note that client preferences have been chosen to reflect the two approaches to market design in the social sector. First, there is variation in the ordinal preferences of clients, where a fraction $1 - \frac{2h}{3}$ clients have $t = 2$ as there utility maximizing treatment, with $\frac{h}{3}$ for each of the other two. This reflects the market design literature on matching. Second, there is variation in the intensity of preferences. For instance, a client of type $\tau = 2$ values $t = 2$ over $t = 1$ at $\mu + \phi$ whereas a client of type $\tau = 1$ values $t = 2$ over $t = 1$ at $\mu - \phi$. Differences in marginal rates of substitution across clients are the focus of the market design literature using artificial currencies.

[12]This assumption implies that for clients with type $\underline{\phi}$, $t = 2 \succ t = 1 \succ t = 0$. Their vertical preferences are for a"higher" treatment. We do not address the intermediate case where $\mu < \underline{\phi} < 2\mu$. In this case, there is only one instance where client preferences are distorted relative to firm preferences, which is where type $\tau = 0$ and $\underline{\phi}$ prefers $t = 2$ to $t = 0$. Otherwise the firm can delegate choice to the client. This case leads to a greater region where client choice is used - as they can be trusted more - but is otherwise conceptually identical to the outcomes below.

Agent: The firm hires an agent to implement t_i. Agent a receives a (contractible) signal τ_i^a on each served client's type:

$$\tau_i^a = \begin{cases} \tau_i & \text{with probability } q \\ \tau_j \neq \tau_i & \text{with probability } \frac{1-q}{2}, \end{cases} \tag{4}$$

where $\frac{1}{3} \leq q < 1$. Signals are uncorrelated across clients conditional on τ_i. The agent has preferences over wages and client welfare:

$$U^a = \sum_{i=1}^{N} (w + \alpha \hat{U}_i^c) I_i. \tag{5}$$

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 $\overline{U}$ per unit of I.[13]

A central component of the analysis is that stakeholders can disagree on appropriate actions. We model this as differences in priors. We assume that τ_i^a is a sufficient statistic for τ_i with respect to the firm's prior. As a result, the firm's maximum likelihood estimate of τ_i conditional on τ_i^a is τ_i^a. 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. λ is independently distributed across agents. Conditional on a given agent holding a belief $\tilde{\tau}$, if a randomly chosen other agent also does so, there is a probability $\rho \geq \frac{1}{3}$ that they agree.

An agent is randomly offered the position from a pool of potential agents at a wage of w per unit of I. If he turns it down, another random agent is chosen. Let $\overline{U}^{ca}$ be an agent's belief of expected client utility if he does not work at the firm, and $\hat{U}^{ca}$ be beliefs if he does. The agent accepts any offer where

$$w + \alpha \hat{U}^{ca} \geq \overline{U} + \alpha \overline{U}^{ca}. \tag{6}$$

Note that this characterization does not simply map the agent's belief of client welfare into his wage. Instead, it depends on the marginal impact he has on client welfare relative to another potential hire. This becomes clear below.

Donations: Donations are given by $\mathbb{D}(b) + \epsilon$, 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 $f > 0$. b is donor belief of the probability that their donation is spent on I_i. We assume that $\mathbb{D}(0) = 0$, $\mathbb{D}'(b) \geq 0$, $\mathbb{D}(1) \equiv D$. We describe donor information below which is used to derive b.

[13] 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.

Client Coverage: Client coverage $I = \sum I_i$ depends on funding. Let r be average revenue per client. The budget constraint depends on two cases:

1. When $w > r$, donor-funding limits client coverage. Let π be the net profit of the firm (described below). Then

$$I = \frac{\mathbb{D} - \pi + \epsilon}{w - r}. \tag{7}$$

2. When $r \geq w$, the firm is self-funded without donors. This will only be relevant for the for-profit firm below.

Organizational Form: The firm can commit to not take any profit by incorporating as a non-profit. Let $\Pi \in \{0, 1\}$, where $\Pi = 0$ implies non-profit status. Let $\tilde{I} = \frac{D+\epsilon}{w-r}$. We assume that

$$\pi(\tilde{I}) \begin{cases} = 0 & \text{if } \Pi = 0 \\ \leq \mathbb{D} + \epsilon & \text{if } \Pi = 1. \end{cases} \tag{8}$$

We assume that $E\pi = 0$ (this is simply to distinguish the firm from being a donor) and that π is contingent on $\tilde{I}$.[14] The firm cannot take any revenue if is a non-profit.[15] If $\Pi = 1$, the only constraint is budget balance.

Each client is defined by two attributes that affect her willingness to pay in (3), her value $T(\tau_i, t)$ and her capacity to pay m_i. Pricing by the firm then has the danger of excluding high m_i clients. However, 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 income y:

Fairness Constraint: $t_i(\tau_i|m_i) = t_i(\tau_i), I_i(\tau_i|m_i) = I_i(\tau_i)$, for all m_i.

Note that this is an *ex post* measure of fairness. We discuss this below. As clients cannot be excluded for low income, in any equilibrium where type τ_i receives treatment $t = j$ with positive probability, the Fairness Constraint requires that

$$r_j \leq \frac{T(\tau_i, j)}{\underline{m}}. \tag{9}$$

Below we consider a less stark version of the Fairness Constraint, where the firm excludes clients below some ability to pay (for example, where a housing association would require a minimum rent above the capacity of some families), but treats all clients above that level equally.

[14] This is simply a shorthand for π to be contingent on any source of revenue or costs.

[15] Assuming $\pi = 0$ is an exaggeration of the legal standard. This is discussed below.

This describes payoffs and constraints, but not the mechanism by which treatments are assigned to clients. Let A reflect an assignment mechanism that maps treatments to clients. We describe the set of A considered below. All firm choices are indexed by A below.

Strategies and timing:

1. The firm chooses A, R^A, w^A, π^A and Π, which is observed by clients and agents.
2. The firm offers the position to one agent subject to (6).
3. The firm and agent observes τ_i^a for all i.
4. ϵ is revealed.
5. A is implemented, r_i^A paid by clients, and w^A and π^A are paid.

Given this timing, the objective of the firm is to choose $A, w^A, \Pi, \pi^A(\tilde{I})$ and R^A to maximize ES subject to (3), (5), (6), (7), (9), $E\pi = 0$, and Assumption 1. The equilibrium concept used is Bayesian Nash. Throughout, we restrict attention to pure strategies.[16] The program is formally defined in the Appendix.[17]

A Benchmark: Without describing A, the modeling assumptions allow tractable solutions for surplus. Consider the first best outcome for a given level of funding and assignment mechanism A. Let T^A be the per capita treatment surplus from assignment mechanism A, let r^A be average revenue per client, and let w^A be the wage paid per unit of I. Then $I^A = \frac{D-\pi(\tilde{I})+\epsilon}{w^A-r^A}$ is the number of services distributed by the donor constrained firm. To avoid trivial outcomes, we assume that donor funding does not exhaust all efficient treatments, or $\mathbb{D} + \bar{\epsilon} < \frac{v+T^A}{\delta}$. As returns are linear in $I_i \in (0,1)$, optimal service is $I_i \in \{0,1\}$, where $I_i = 1$ for all $i \leq I^A$ and 0 otherwise. As $v_i = v - \delta i$, the average value of a client with realization ϵ and firm profit π is $v + T^A - \frac{\delta(\mathbb{D}-\pi(\tilde{I})+\epsilon)}{2(w^A-r^A)}$. As a result, expected surplus is $ES^* = \int_{\underline{\epsilon}}^{\epsilon} \frac{\mathbb{D}-\pi(\tilde{I})+\epsilon}{w^A-r^A}\left(v + T^A - \frac{\delta(\mathbb{D}-\pi(\tilde{I})+\epsilon)f(x)dx}{2(w^A-r^A)}\right)$, or

$$ES^* = \frac{\mathbb{D}}{w^A - r^A}\left(v + T^A - \frac{\delta\mathbb{D}}{2(w^A - r^A)}\right) - \frac{\delta\int_{\underline{\epsilon}}^{\bar{\epsilon}}(-\pi(\tilde{I}) + x)^2}{2(w^A - r^A)^2}f(x)dx, \qquad (10)$$

[16] We do not consider mixed strategies. 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. We exclude mixing as social organizations are unlikely to impose these costs on their clients.

[17] The client has three pieces of private information, m, τ and ϕ. In general, this could render the problem complex. However, the modeling assumptions imply that it is much simplified. Although ability to pay varies between $\underline{m}$ and $\overline{m}$, the Fairness Constraint merely means that we need to check outcomes at the boundaries. Second, Assumption 1 implies that type $\overline{\phi}$ always wants $\tau = t$ as its preferred outcome. As a result, for most of the analysis below we merely need to check for the preferences of type $\underline{\phi}$ for different τ.

This characterization of the problem focuses the analysis.[18]

Note two further characterization of the first best. First, first best assignment entails $\tau_i = t_i$, with surplus $T^* = \mu + \hat{\phi}$. Second, the firm can eliminate uncertainty in funding by choosing $\pi(\tilde{I}) = \epsilon$. This offers financial certainty to the mission. First best surplus conditional on $\mathbb{D}, w^A$ and r^A is

$$ES^{**} = \frac{\mathbb{D}}{w^A - r^A}\left(v + \mu + \hat{\phi} - \frac{\delta \mathbb{D}}{2(w^A - r^A)}\right). \tag{11}$$

1.1 Outcomes and Assignment Mechanisms

We now address different assignment mechanisms A. We consider three classes of outcomes, and mechanisms for each of these. Each of the mechanisms is described in full detail below. We consider the following outcomes:

1. Assigned Consumption arises when pricing reveals no more information on client preferences than no pricing. We consider two mechanisms that reflect this outcome:
   - *Deferred Acceptance*: This implements a standard DA mechanism.[19]
   - *Agent Assignment*: The agent allocates based solely on his signal.
2. Choice arises when prices reveal all client preferences. We consider two mechanisms that reflect this outcome:
   - *Unrestricted Choice*: The client receives $I_i = 1$ of her highest ranked preference.
   - *Restricted Choice*: The client receives I_i of her highest ranked preference, where $I_i < 1$ for at least one i.
3. Partial Choice arise when pricing partially reveals client preferences, so some goods are assigned and other are chosen. One mechanism reflects this outcome:
   - *First Choice:* Priority is given to reported first preferences. For over-subscribed choices, the firm uses the Deferred Acceptance mechanism.
4. For each of these, the firm decides whether to operate as a non-profit or as a Social Enterprise.

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

[19]We are bundling Top Trading Cycles with Deferred Acceptance here. In our setting, there is no difference between Top Trading Cycles and Deferred Acceptance.

2 The Non-Profit

We begin by considering the optimal structure of the non-profit. This implies that all funding is spent on I, so $b = 1$ and $\mathbb{D} = D$. Furthermore, $\pi(\tilde{I}) = 0$. As a result, expected surplus is

$$ES^* = \frac{D}{w^A - r^A}\left(v + T^A - \frac{\delta D}{2(w^A - r^A)}\right) - \frac{\delta\sigma_\epsilon^2}{2(w^A - r^A)^2}, \tag{12}$$

where $\sigma_\epsilon^2 = \int_{\underline{\epsilon}}^{\epsilon} x^2 f(x)dx$.[20] This characterization of the problem allows us to evaluate the non-profit in terms of three endogenous variables, T^A, r^A, and w^A. For any non-profit with donations $D + \epsilon$, the chosen allocation mechanism affects not only the intensive margin of allocation efficiency T^A but also the extensive margin of clients covered through revenue from clients r^A (determined by (9)) and the cost of service provision w^A (determined by (6)).

2.1 Assigned Consumption

Assigned Consumption arises when the capacity of the firm to price does not change the ranking of clients compared to a no-pricing regime. Let $\hat{t}_i$ be a report by client i of her preference ranking where $\hat{t}_i = (t_i^1 \succ t_i^2 \succ t_i^3)$, where t_i^j is her jth preferred treatment. Assigned Consumption arises when pricing does not affect $\hat{t}(\tau, \underline{\phi}, m) = (2 \succ 1 \succ 0), \hat{t}(2, \overline{\phi}, m) = (2 \succ 1 \succ 0), \hat{t}(1, \overline{\phi}, m) = (1 \succ 2 \succ 0)$ and $\hat{t}(0, \overline{\phi}, m) = (0 \succ 2 \succ 1)$. Lemma 1 simplifies.[21]

Lemma 1: For any Assigned Consumption mechanism A, $R^A = \{0, \frac{\mu - \underline{\phi}}{\underline{m}}, \frac{2(\mu - \underline{\phi})}{\underline{m}}\}$, and average revenue raised per client is $r^A = \underline{r} = \frac{\mu - \underline{\phi}}{\underline{m}}$.

The Fairness Condition requires that any client offered a treatment accepts that treatment, as in (9). The maximum that can be charged for any treatment is the minimum utility attained by any client offered that treatment in equilibrium. As $q < 1$, all clients receive all treatments. As a result, prices are bounded above by the minimum utility obtained by any client. This implies that revenue is invariant to the average efficiency of the Assigned Consumption allocation mechanism, and that the same revenue could be obtained by randomly assigning services.

We now turn to the two Assigned Consumption outcomes, Deferred Acceptance and Agent Assignment.

[20] Note the separability of the cost of uncertainty. This arises as we have assumed that $v_i = v - \delta i$. Uncertainty plays no substantive role until the for-profit decision is considered so this allows us to concentrate on the more relevant features of the problem.

[21] Lemma 1 arises through the combination of maximizing expected revenue subject to (3), (9), $\hat{t}(\tau, \underline{\phi}, m) = (2 \succ 1 \succ 0), \hat{t}(2, \overline{\phi}, m) = (2 \succ 1 \succ 0), \hat{t}(1, \overline{\phi}, m) = (1 \succ 2 \succ 0)$ and $\hat{t}(0, \overline{\phi}, m) = (0 \succ 2 \succ 1)$.

2.1.1 Deferred Acceptance

Deferred Acceptance is often considered the gold standard for how best to assign without prices (Roth, 2008, Pathak and Sonmez, 2008, Abdulkadiroglu, 2005, etc). We use it as the benchmark to compare the other mechanisms. 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. The Deferred Acceptance mechanism is as follows:

1. The firm chooses $R^{da} = \{r_0^{da}, r_1^{da}, r_2^{da}\}$ and all clients $i < I^{da}$ report $\hat{t}_i$.

2. The firm allocates $t = 2$ to all clients $i < I^{da}$ for which $\tau_i^a = 2$ and $\hat{t}_i = (2 \succ 1 \succ 0)$ and randomly allocates $t = 2$ to clients for which $\tau_i^a = 1$ and $\hat{t}_i = (2 \succ 1 \succ 0)$ until there are no more $t = 2$ slots.

3. The firm allocates $t = 1$ to all remaining clients for which $\tau^i = 2$ and $\hat{t}_i = (2 \succ 1 \succ 0)$ and to all clients with $\tau_i^a = 1$ and $\hat{t}_i \neq (0 \succ 2 \succ 1)$ and randomly allocates $t = 1$ for which $\tau_i^a = 0$ and $\hat{t}_i \neq (0 \succ 2 \succ 1)$ until there are no more $t = 1$ slots.

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

This is a standard DA mechanism for an imperfectly informed firm.[22] The outcome depends on the extent of horizontal differentiation h. Proposition 1 provides the outcome for $h \leq \frac{2}{3-q}$ for brevity. The cases with $h > \frac{2}{3-q}$ is given in the Appendix and is conceptually similar.

Proposition 1 *Let $h \leq \frac{2}{3-q}$. Expected surplus from Deferred Acceptance is*

$$ES^{da} = \frac{D}{\overline{U} - \underline{r}} \left(v + T^{da} - \frac{\delta D}{2(\overline{U} - \underline{r})} \right) - \frac{\delta \sigma_\epsilon^2}{2(\overline{U} - \underline{r})^2}. \tag{13}$$

where

$$T^{da} = \mu + \frac{1}{3} \left(\hat{\phi}(q + p_2 \frac{1-q}{2}) + \overline{\phi} h (q + \frac{1-q}{2}) + (1-h)(\underline{\phi} q (1-p_2) + \hat{\phi} \frac{1-q}{2} p_1) \right) + \\ + \frac{1}{3} \left(h\overline{\phi} + q(1-p_1)(1-h)\underline{\phi} \right), \tag{14}$$

where $p_2 = \frac{h(1-q)}{q(1-h) + (2-h)\frac{1-q}{2}}$ *and* $p_1 = \frac{\frac{h(1-q)}{2}}{(1-h)q + (1-q)}$.

[22] Deferred Acceptance here is analogous to Non-Random Serial Dictatorship. After the firm observes its signal, it offers a choice of $t = \{0, 1, 2\}$ to all clients who receive a $\tau_i^a = 2$ signal. Those who do not choose $t = 2$ split equally between $t = 1$ and $t = 0$. The firm then offers a choice of $t = \{0, 1, 2\}$ to those with $\tau_i^a = 1$ until the remaining $t = 2$ slots are filled. It does similarly for $t = 1$ slots among those who have not chosen.

Expected surplus in (12) depends on w^A, r^A and T^A. Consider each component. Revenue is given by Lemma 1. Allocations are independent of the agent's prior belief, so $\hat{U}^{ca} = \overline{U}^{ca}$ and $w^{da} = \overline{U}$ from (6). This implies that expected client coverage is $\frac{D}{\overline{U}-\underline{r}}$.

Now consider T^{da}. This depends on q and the p_i terms. Reported preferences are

$$\hat{t} = \begin{cases} 2 \succ 1 \succ 0 & \text{with probability } 1 - \frac{2h}{3} \\ 1 \succ 2 \succ 0 & \text{with probability } \frac{h}{3} \\ 0 \succ 2 \succ 1 & \text{with probability } \frac{h}{3}. \end{cases} \tag{15}$$

The p_i terms reflect the benefits of Pareto efficiency, accruing from the $\frac{2h}{3}$ clients who do not rank $\hat{t} = (2 \succ 1 \succ 0)$. Finally, the allocative optimum is $\tau = t$ yet $\frac{(1-h)(1-q)}{3}$ are misidentified by the firm as type $\tau = 2$ and are given $t = 2$.[23] Errors are also made for the other services. We refer to these outcomes as assignment error.

2.1.2 Agent Allocation

The second Assigned Consumption outcome that we address is the Agent Allocation mechanism. This delegates control over t to the agent alone. The firm knows I^a when allocating t. It offers $I_i = 1$ to all $i \leq I^a$.

1. The firm chooses $R^a = \{r_0^a, r_1^a, r_2^a\}$.
2. The firm and agent observe τ_i^a for each client $i < I^a$.
3. The agent allocates t to all clients $i < I^a$.

Note that the mechanism does not ask clients to report preferences. This is discussed below.

Proposition 2 *Expected surplus from Agent Allocation is*

$$ES^{da} = \frac{D}{w^* - \underline{r}} \left(v + T^a - \frac{\delta D}{2(w^* - \underline{r})} \right) - \frac{\delta \sigma_\epsilon^2}{2(w^* - \underline{r})^2}. \tag{16}$$

where

$$T^a = \mu + (\lambda q + \frac{1-\lambda}{3})\hat{\phi}, \tag{17}$$

and $w^* = \overline{U} - \alpha(1-\lambda)(1 - \frac{\lambda}{3} - (1-\lambda)\rho)\hat{\phi} < \overline{U}$.

[23] Note that when $q = \frac{1}{3}$, τ^a is irrelevant and Deferred Acceptance is identical to Random Serial Dictatorship, a common outcome in one-sided market design matching algorithms.

Consider the comparison to Deferred Acceptance. Revenue $r^a = \underline{r}$ is once again determined by Lemma 1. Three features distinguish the outcome from Deferred Acceptance. First, it uses a worse signal to allocate services. From the firm's perspective, the expected value of the agent's signal is $\lambda q + (1-\lambda)\frac{1}{3} < q$. Second, the outcome is not Pareto efficient. For example, consider those who are assigned $t = 2$ by the agent. A fraction $h(1 - (\lambda q + \frac{1-\lambda}{3}))$ would prefer $t = 1$ or $t = 0$ and $2(1-h)(1 - (\lambda q + \frac{1-\lambda}{3}))$ are willing to swap. This is reflected in the absence of analogs to the p_i terms in (14). Third, $w^a = w^* < \overline{U}$ as $\hat{U}^{ca} > \overline{U}^{ca}$. The reduction in wages below $\overline{U}$ is given by $\Delta w^* = \alpha(1-\lambda)(1 - \frac{\lambda}{3} - (1-\lambda)\rho)\hat{\phi}$. This is because his belief of his expected impact on the client's welfare relative to an alternative agent is $(1-\lambda)(1 - \frac{\lambda}{3} - (1-\lambda)\rho)\hat{\phi}$. 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 client believes that an alternative agent is less likely to share his views (ρ lower).

As a result, Agent Assignment covers more clients $EI^a = \frac{D}{w^* - \underline{r}} > EI^{da}$, where $\Delta I^a = EI^a - EI^{da} = \frac{D\Delta w^*}{(w^* - \underline{r})(\overline{U} - \underline{r})} > 0$. Agent Allocation is preferred to Deferred Acceptance if

$$I^{da} \leq \frac{\Delta I^a (v + T^a - \frac{\delta \Delta w^*}{2})}{T^{da} - T^a + \delta \Delta I^a}. \tag{18}$$

If this holds, the firm values the extensive margin of more clients covered more than the intensive margin of higher utility treatments. This condition is more likely to hold when clients are poorer ($\overline{m}$ higher) because the marginal value of additional clients is high. Let y^* be defined such that

$$I^{da}(y^*) = \frac{\Delta I^a (v + T^a - \frac{\delta \Delta w^*}{2})}{T^{da} - T^a + \delta \Delta I^a}. \tag{19}$$

Agent Assignment is preferred for all average income below y^*. Agent Assignment is also more likely to hold when h is low. This is because the value of Deferred Acceptance in terms of returns to Pareto efficiency (the p_i terms above) is low.[24]

Note the importance of Lemma 1 for (18). Average client treatment surplus is lower under Agent Assignment. If revenue from clients depended on that surplus, Agent Assignment would reduce funding, perhaps by enough to overturn the outcome in favor of Deferred Acceptance. Yet Lemma 1 implies that revenue under Assigned Consumption is independent of allocative efficiency. This is where the Fairness Constraint bites for Assigned Consumption.

Note that we have assumed that the agent holds a prior $\tilde{\tau}$ where he is almost certain of the client's type with probability $1-\lambda$. This is for simplicity. Consider an alternative where with probability $1-\lambda$, the agent holds a prior where with probability

[24] A simple illustrative case arises when preferences are vertical, $h = 0$ and $\delta = 0$. In that instance, Deferred Acceptance is equivalent to Random Serial Dictatorship, and Agent Allocation is preferred if $\Delta w^*(\mu + v + q\hat{\phi}) - (q - q')(\overline{U} - \frac{\mu}{\underline{m}})\hat{\phi} + \frac{\delta D}{2}\frac{(\overline{U} - \frac{\mu}{\underline{m}} - \Delta w^*)^2 - (\overline{U} - \frac{\mu}{\underline{m}})^2}{(\overline{U} - \frac{\mu}{\underline{m}} - \Delta w^*)(\overline{U} - \frac{\mu}{\underline{m}})} > 0$. This is more likely to hold for D low, so client income reduces the need for Agent Assignment.

q^a the client is type $\tilde{\tau}$ where $q^a \sim U[\frac{1}{3}, 1]$. Then the agent's prior is pivotal after observing τ_i^a with probability $\frac{3(1-q)}{2}$ and $\Delta w^* = \frac{3(1-q)}{2}\alpha(1-\lambda)(1-\frac{\lambda}{3}-(1-\lambda)\rho)\hat{\phi}$. The qualitative results remain.

As we described above, this outcome is not Pareto efficient. Pareto efficiency is typically seen as a minimal condition for a desirable assignment mechanism so this result may seem surprising. Yet it reflects a conceptually typical outcome in agency theory. From its inception, agency theory has emphasized distorting *ex post* efficiency in order to improve *ex ante* incentives (in its earliest incarnation by distorting risk sharing). The results here apply that logic.

Related Evidence: The paper is meant to be anchored in the empirical concerns of the social organization. First, surveys consistently point 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). Concern about funding is even higher for small non-profits: 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). Second, 29% of non-profits report "difficulty recruiting and retaining quality staff" as their biggest challenge, likely as they are paid below market wages.[25]

The key ingredient above is that what matters to employees is not simply working in a social organization, but what they do. To put it more simply, it is hard to be attracted to a job where outcomes are determined by algorithm, even if the algorithm is efficient. There is a large empirical literature illustrating the importance of autonomy in the social sector. This is often identified by outcomes when firms choose to bureaucratize, using more rule-based procedures. Pascoe et al, 2023, survey almost forty academic publications on the impact of bureaucracy on the experiences of 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). Langer et al, 2019, study a large set of public sector workers and conclude that "when autonomy is reduced because of high levels of routineness and centralization in the workplace negative outcomes such as decreased job satisfaction result". For more studies corroborating this outcome, see Adler and Borys, 1996, Cartarelli et al, 2016, and Wright and Davis, 2003. As a result of all of these concerns, it should not be surprising that the primary focus of many social organizations is keeping the job attractive to their employees rather than some of the concerns of the market design literature.[26]

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

[26] We assume that agents know the assignment mechanism and their impact on client welfare

Pareto Efficient Agent Assignment: The Agent Assignment mechanism above has two bundled components: the agent allocating the good and outcomes not being Pareto efficient. We now show that the firm can implement Pareto Efficient Agent Assignments but will not do so for low funding levels. To see this, we now change the mechanism above to allow the client to report $\hat{t}$ by an additional stage:

0. Clients report $\hat{t}_i$.

Further consider the case where reports are given by (15).[27] In the introduction, we discussed how social organizations are characterized by disagreement among stakeholders. This change to the mechanism reduces disagreement: any client who reports $t = 0(1)$ as her first choice is type $\tau = 0(1)$ with probability 1. This changes the posterior of the agent.[28] The agent now implements Pareto efficient outcomes, where $\hat{t} = 1 \succ 2 \succ 0 \implies t = 1$, and $\hat{t} = 0 \succ 2 \succ 1 \implies t = 0$. This increases treatment surplus per client by $\left(\frac{2(1-h)}{9}\frac{1-q}{2}\right)\hat{\phi}+\frac{2h}{3}\overline{\phi}(1-(\lambda q+\frac{1-\lambda}{3}))+2(\frac{1-\frac{2(1-h)}{3}}{3})(q-(\lambda q+\frac{1-\lambda}{3}))\underline{\phi} > 0$, illustrating the utilitarian benefit to Pareto efficient outcomes.

Despite client welfare being higher, and the agent's altruism, he must be paid a higher wage than in Proposition 2. This is because Pareto efficient assignment reduces the marginal value of the agent. In effect, a fraction $\frac{2h}{3}$ clients determine their own service and the agent is less pivotal. The wage must increase by $(1-\frac{1-\frac{2(1-h)}{3}}{3})\alpha(1-\lambda)(1-\frac{\lambda}{3}-(1-\lambda)\rho)\hat{\phi}$, and the number of clients covered falls accordingly. The firm prefers non-Pareto efficient Agent Assignment for low enough funding. As a result, the absence of Pareto efficiency is not generated by the assumptions used in Proposition 2.[29]

This last extension identifies an implication that may not be apparent at first

when accepting a position at the firm. Realistically, the margin on which this matters is more likely retention of skilled labor, where existing employees understand the technology and their impact on outcomes. 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 that he is offered on the job (for examples, see Liu at al., 2011, and Zhang et al., 2024).

[27]We address the possibility of non-strategy proof mechanisms under the First Choice assignment mechanism.

[28]Note above that the agent was almost certain about the client's type. This was made to ensure that when the client expressed their preferences with certainty here, it supersedes the prior.

[29]We have not considered an intermediate mechanism between DA and Agent Allocation, where the firm implements a DA mechanism using the agent's signal instead of its own. This is because there is little conceptually novel about it (and we know of no empirical examples). Instead, it offers an intermediate outcome where it allows Pareto efficient outcomes, but at the cost of increasing wages compared to Agent Assignment. This is shown in the Appendix. Agent allocation remains preferred when funding is low.

glance. The agent is motivated by altruism towards the client. The agent knows that the welfare of the client is higher under Pareto Efficient Agent Assignment than in Proposition 2. Yet the agent demands a higher wage with the Pareto efficient outcome. The reason is that agent compensation does not depend on the agent's belief of the client's welfare. Instead, it depends on the *marginal* impact of the agent on client welfare relative to an alternative agent. For this reason, Agent Assignment both implies commonly known lower welfare for the client and lower wages for an agent who cares about client welfare.

Agency theory has studied incentive provision in mission-oriented organizations. A common outcome is the need to offer more autonomy to agents than in a standard profit-seeking firm. 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, 2016). A particularly relevant antecedent for this section is Aghion and Tirole, where control is delegated to an agent to induce more effort. There are two important differences here. First, the objective of the firm in delegating control is to reduce compensation, by increasing the agent's perceived impact on the client. Second, and more importantly, the agent here is not motivated by private benefits but by altruism, in a setting where these benefits are endogenous to the information revealed by the mechanism. This is shown in the extension showing the Pareto efficient outcome above.

Above, we mentioned the need for assigned consumption when the price system fails. Prices fail here in all the ways outlined in the introduction. First, prices charged here do not pay the bills, and delegating control is the response to harness worker preferences. Second, prices are not set to allow client choice so Assignment is necessary. Finally, the Hayekian view that prices orient the actions of suppliers fails, where the absence of credible signals from clients about what they want allows differing views.[30] The interaction between shortness of funds and disagreement gives rise to the appeal of simply delegating control to the agent.

Allocative efficiency and funding are substitutes here. A similar outcome might arise if, for example, the firm chose to hire lower quality labor when it is short of money. Our objective is not to make that simple point. Instead it is that more efficient allocation mechanisms themselves become compromised by shortage of resources. For example, there would be no reason to implement non-Pareto efficient outcomes simply because of a cheap labor force. That is not true in Proposition 2, as wage reductions depend on the otherwise free choice to ignore client information. This is why we interpret the implication of financially constrained firms through the

[30]In 2005, Feeding America transitioned to an assignment mechanism where food banks bid on truckloads of food. This lead to a large increase in donations. Part of this was through its ability to overcome disagreement about the value of different kinds of foods. Specifically, the existence of prices helps those involved in encouraging donation to focus on the most valuable foods, which could be verified with donors by showing them relative prices (Prendergast, 2022).

lens of delegation.

2.2 Choice

Many social organizations offer subsidized choice to their clients.[31] We characterize Choice arising when each client's highest ranked outcome is $\tau = t$ for any m, and all receive that choice. Note that this involves multi-dimensional screening, as client i varies in both τ_i and y_i. The requirement for screening is that t_i does not depend on y_i. There are two forms of Choice: Unrestricted and Restricted. We consider each in turn.

2.2.1 Unrestricted Choice

Unrestricted choice arises when the equilibrium entails all clients of type τ receiving $I_i = 1$ of $t = \tau$. The mechanism for Unrestricted Choice is as follows. The firm knows I^{uc} when allocating services to all clients and offers $I_i = 1$ to all $i \leq I^{uc}$. The mechanism is:

1. The firm chooses $R^{uc} = \{r_0^{uc}, r_1^{uc}, r_2^{uc}\}$.
2. Each client $i \leq I^{uc}$ reports $\hat{t}_i$.
3. The firm allocates all clients their first choice if i is undersubscribed.
4. For all $t = i'$ that are oversubscribed, the firm implements a Deferred Acceptance Algorithm among those where $\hat{t}^k = i' \succ ...$

The final stage of the mechanism is only used out-of-equilibrium as all agents receive $t_i = \tau_i$ in equilibrium.

Lemma 2: Unrestricted Choice satisfies the Fairness Constraint if and only if $\hat{y} \geq \hat{y}^{uc}$, where

$$\frac{V'(\hat{y}^{uc} + \overline{\eta})}{V'(\hat{y}^{uc} + \underline{\eta})} \equiv \frac{\overline{m}^{uc}}{\underline{m}^{uc}} = \frac{2\mu - \underline{\phi}}{2\mu + \underline{\phi}}. \tag{20}$$

For $\hat{y} \geq \hat{y}^{uc}$, optimal prices are $R^* = \{r_0^*, \frac{\underline{\phi}+\mu}{\underline{m}}, \frac{\underline{\phi}+2\mu}{\underline{m}}\}$, where $r_0^* = \min\{\frac{\underline{\phi}-2\mu}{\overline{m}} + \frac{2\mu+\underline{\phi}}{\underline{m}}, \frac{\underline{\phi}}{\underline{m}}\}$. Average revenue per client is $r^* = \frac{\mu + \frac{r_0^* + 2\underline{\phi}}{3}}{\underline{m}}$, where $\underline{m} \leq \underline{m}^{uc}$.

The Fairness condition requires that wealthier clients do not benefit in their allocations. This arises for average income above $\hat{y}^{uc}$. As an example, with logarithmic

[31] For example, housing associations often offer vouchers to qualified clients, where clients pay subsidized rents, and are offered choice over available housing units. As another example, 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.

utility $V(y) = ln(y)$, $\hat{y}^{uc} = \overline{\eta}(\frac{\mu}{\underline{\phi}} - \frac{1}{2}) - \underline{\eta}(\frac{\mu}{\underline{\phi}} + \frac{1}{2})$. Lemma 2 also determines revenue from clients. The central issue with Fairness is that all clients of type $\tau = 0$ do not choose $t = 2$. The firm can induce efficient choice at $\hat{y}^{uc}$ with $r_0 = 0$. As average income rises above $\hat{y}^{uc}$, the firm can increase the price of the $t = 0$ service, until it reaches its maximum of $\frac{\underline{\phi}}{\underline{m}}$ above average income of $\tilde{y}^{uc} > \hat{y}^{uc}$, where $\frac{V'(\tilde{y}^{uc}+\overline{\eta})}{V'(\tilde{y}^{uc}+\underline{\eta})} = 1 - \frac{\underline{\phi}}{2\mu}$.

How risk preferences vary with wealth has been extensively used in finance to predict investment behavior and portfolio choice.[32] Diminishing Absolute Risk Aversion is also key to allowing clients Choice. It reflects the following simple intuition: when clients have low average income, small differences in income translate into large differences in utility. This declines as average income rises, which allows efficient screening here. Note also the importance of horizontal differentiation, as $\hat{y}^{uc}$ is decreasing in $\underline{\phi}$. Proposition 3 immediately follows.

Proposition 3 *Unrestricted Choice is used only if $\hat{y} \geq \hat{y}^{uc}$. Surplus from Unrestricted Choice is*

$$ES^{uc} = \frac{D}{\overline{U} - r^*}\left(v + T^* - \frac{\delta D}{2(\overline{U} - r^*)}\right) - \frac{\delta\sigma_\epsilon^2}{2(\overline{U} - r^*)^2}, \tag{21}$$

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

Wage are given by $\overline{U}$ as $\hat{U}^{ca} = \overline{U}^{ca}$. Lemma 1 does not apply as $\hat{t}(\tau, \underline{\phi}, m) = (\tau \succ ...)$. Instead, $R^* = \{r_0^*, \frac{\mu+\underline{\phi}}{\underline{m}}, \frac{2\mu+\underline{\phi}}{\underline{m}}\}$, as all clients receive the efficient level of t, with expected value T^*. At $\hat{y} = \hat{y}^{uc}$, revenue discretely jumps by $\frac{5\underline{\phi}}{3\underline{m}}$ and then increases by a further $\frac{\underline{\phi}}{3\underline{m}}$ as $\hat{y} \to \tilde{y}^{uc}$. If Unrestricted Choice overcomes the Fairness Constraint, it is always preferred to Deferred Acceptance.[33] Note how (21) gets closer to the first best benchmark given in (11) as treatment surplus is T^*.

Uncertainty: The support of ability to pay above does not change with average income. 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).[34] If this translates to the support of outcomes, Decreasing Absolute Risk Aversion is sufficient but not necessary. Second, an alternative could be where randomness

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

[33] We cannot rank Unrestricted Choice relative to Agent Assignment, as agents may value discretion so much that the wage reductions they accept pay for enough additional clients to be served that it overcomes its clear allocative cost.

[34] This appears to be because at low income levels, loss of employment is common.

scales in a linear way with $\hat{y}$, $\eta(\hat{y}) = \eta\hat{y}$. This would be analogous to some finance settings which 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.[35]

2.2.2 Restricted Choice

In the rubric proposed above, the second intermediate outcome is *Restricted Choice*, where a distorted treatment can be chosen. For illustrative purposes only, initially consider a case where the *quality* of all goods in the analysis so far has been normalized to 1, but where the firm can choose any quality $\kappa_i \in (0, 1)$ at no cost and client utility is $U_i^c = (T(\tau_i, t_j) - m_i r_j)\kappa_i$. We showed above how the outcome of the two-dimensional screening problem requires that $\hat{y} \geq \hat{y}^{uc}$. Rather than show required income for $I_i = 1$ in (20), we now consider what quality of each service is consistent with screening for Choice satisfying the Fairness Condition for any given income level. Let the firm allows clients to choose between $I_0 = 1$, $I_1 = \kappa_1 < 1$, and $I_2 = \kappa_2 < \kappa_1$. We show in the Appendix that for $\hat{y} < \hat{y}^{uc}$ with per unit quality prices R^*, Choice satisfies the Fairness Constraint for any $\kappa_i \leq \kappa_i^*$ where

$$\kappa_i^* = \frac{\underline{\phi}}{i\mu - \frac{\overline{m}\underline{\phi}}{\underline{m}-\overline{m}}}, i = 1, 2. \tag{22}$$

This has the implication that higher t treatments receive lower quality but the reduction declines in client income. The outcome is given in Figure 1, where we plot required restrictions on quality as a function of client income. The intercept when prices cannot be used ($\underline{m} = \infty$) is $\kappa_i^* = \frac{\underline{\phi}}{i\mu}, i = 1, 2$. For the component of Figure 1 to the left of $\hat{y}^{uc}$, it shows how client income relaxes the need to distort the κ_i^*. This is given by the second term in the denominator of (22). This is because prices are a substitute to distorted quality. The κ_i^* are increasing in $\hat{y}$, where $k_2^* \to 1$ as $\hat{y} \to \hat{y}^{uc}$. To the right of $\hat{y}^{uc}$, the firm can increase prices charged for $t = 0$ until it reaches income of $\tilde{y}^{uc}$ as above. This shows how the screening problem involves both the use of two instruments, quality distortions and prices. Expected treatment surplus in this setting is given by $\tilde{T}^{rc} = \frac{\hat{\phi}+\kappa_1^*(\hat{y})(\hat{\phi}+\mu)+\kappa_2^*(\hat{y})(\hat{\phi}+2\mu)}{3}$ and revenue is $\frac{\tilde{T}^{rc}}{\underline{m}}$.

Figure 1 shows that the distinction between Choice and Assignment is not as stark as outlined with Unrestricted Choice when the good can be changed in a continuous manner. As $\hat{y}$ increases to $\hat{y}^{uc}$, the firm transitions from cases where both κ_1 and κ_2 need to be distorted to one where κ_1 is at its efficient level, and then $\kappa_2 \to 1$ as client income increases further. This shows how Choice, even Restricted Choice, becomes

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

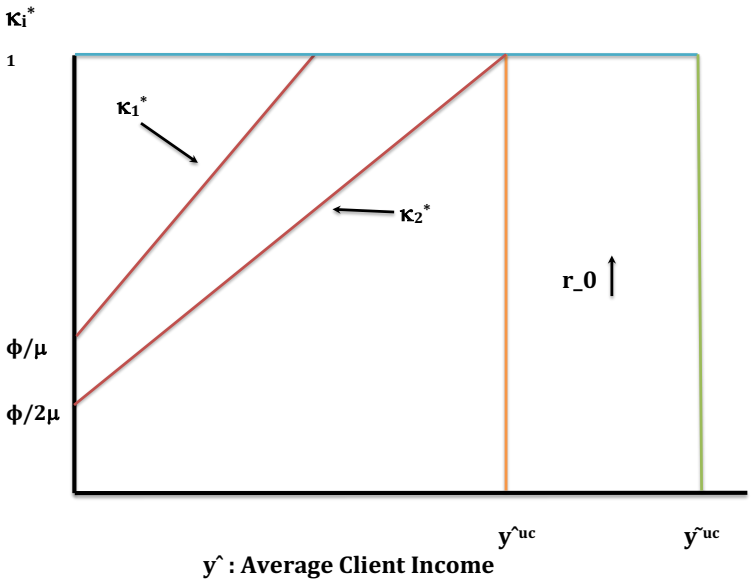


Figure 1: **Restricted and Unrestricted Choice**

preferable to Deferred Acceptance whenever client income is sufficiently high. This is because the distortions to quality are decreasing in $\hat{y}$ in (22) and Figure 1.

Typically Restricted Choice market designs (described below) do not entail destroying quality but rather quantities of divisible goods are redirected though the mechanism. We have chosen quality distortions above as it shows most of the trade-offs with Restricted Choice in a simple manner. Now consider a case where the client chooses between *quantities* $I_0 = 1$, $I_1 = \kappa_1 < 1$, and $I_2 = \kappa_2 < \kappa_1$. The κ_i^* are identical. Let I^{rc} be the number of clients given this choice. As the distribution of supply is uniform, after client $i = I^{rc}$ there is no more remaining supply of $t = 0$, but there is $I^{rc}(1 - \kappa_i)$ of treatment i. The firm uses a Deferred Acceptance mechanism to allocate the remaining $t = 1$ and $t = 2$.[36] The mechanism for distributing divisible quantities and Proposition 4 illustrate this.

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

1. The firm chooses quantities κ_1^*, κ_2^* and unit prices $R^{rc} = (r_0^{rc}, r_1^{rc}, r_2^{rc})$.
2. All clients $i \leq I^{rc}$ report $\hat{t}$ and receive their first choice.
3. The firm uses a Deferred Acceptance mechanism to allocate the remaining $t = 1$ and $t = 2$.

[36] In the Appendix we consider an alternative where the firm uses distorted quantities to allow choice for these goods. It does not change the qualitative outcomes.

The outcome is given in Proposition 4.

Proposition 4 *Let $\hat{y} < y^{uc}$ and $I^{rc} = \frac{D}{\overline{U} - r^{rc}}$, where r^{uc} is defined in the Appendix. Assume that $R = R^*$ for all clients $i \leq I^{rc}$. Surplus from Restricted Choice is*

$$S^{rc} = I^{rc}[(v + (\frac{1+\kappa_1^*+\kappa_2^*}{3})\left(T^* - \frac{\delta I^{rc}}{2}\right) + (\frac{1-\kappa_1^*}{3} + \frac{1-\kappa_2^*}{3})\left(\tilde{T}^{da} - \frac{\delta I^{rc}(1 + \frac{1-\kappa_1^*}{3} + \frac{1-\kappa_2^*}{3})}{2}\right)] \quad (23)$$

where

- *If* $(q + (2-h)\frac{1-q}{2})(1 + \frac{1-\kappa_1^*}{3} + \frac{1-\kappa_2^*}{3}) < (\frac{1-\kappa_2^*}{3})$

$$\tilde{T}^{da} = \mu + (\frac{1-\kappa_2^*}{2-\kappa_1^*-\kappa_2^*})(\frac{q}{q+(2-h)\frac{1-q}{2}})\hat{\phi} + (\frac{1-\kappa_1^*}{2-\kappa_1^*-\kappa_2^*})(\hat{p}_1(\frac{1-q}{2}(1-h)\underline{\phi}) + (1-\hat{p}_1)(h\overline{\phi} + (1-h)q\underline{\phi})) \quad (24)$$

where $\hat{p}_1 = \frac{\frac{1-\kappa_2^*}{3}}{(q+(2-h)\frac{1-q}{2})(1+\frac{1-\kappa_1^*}{3}+\frac{1-\kappa_2^*}{3})}$.

- *If* $(q + (2-h)\frac{1-q}{2})(1 + \frac{1-\kappa_1^*}{3} + \frac{1-\kappa_2^*}{3}) < (\frac{1-\kappa_2^*}{3})$, *then*

$$\tilde{T}^{da} = \mu + (\frac{1-\kappa_2^*}{2-\kappa_1^*-\kappa_2^*})[\hat{p}_2(\frac{q}{q+(2-h)\frac{1-q}{2}})\hat{\phi} + (1-\hat{p}_2)\frac{1-q}{2}\hat{\phi}] + (\frac{1-\kappa_1^*}{2-\kappa_1^*-\kappa_2^*})(h\overline{\phi} + (1-h)q\underline{\phi}). \quad (25)$$

where $\hat{p}_2 = \frac{\frac{1-\kappa_2^*}{3} - (q+(2-h)\frac{1-q}{2})(1+\frac{1-\kappa_1^*}{3}+\frac{1-\kappa_2^*}{3})}{(1-h)q+\frac{1-q}{2}(2-h)} < 1.$

The outcome of Restricted Choice has two parts. First, all clients $i \leq I^{rc}$ choose $t_i = \tau_i$ and pay unit prices from R^*. The remaining supply is allocated by Deferred Acceptance. Surplus and Revenue are a combination of these two groups.

The same implications as with the quality interpretation continue to hold. There is one new implication: client targeting. All other mechanisms entail $I_i^A = 1$ for all $i \leq I^A$. Here all clients $i = 1, ..., I^{rc}$ receive the efficient type of treatment $t = \tau$, but $\tau \in \{1, 2\}$ receive lower than optimal quantities. While the additional supply of treatments $1 - \kappa_i$ is consumed by another client, these are lower priority clients. This is what we mean below by poor targeting of clients. These is the final part of the bracketed term in (23). We discuss empirical targeting issues below.

Proposition 4 assumes that the firm charges R^*. The firm has two instruments to implement fair allocations; reducing quantities (κ_i) and reducing the price of $t = 0$ (r_0), as that makes lower ranked services more appealing. This implies that there is one other possible equilibrium outcome, where $\hat{R}^{rc} = \{0, \frac{\mu+\phi}{\underline{m}}, \frac{2\mu+\phi}{\underline{m}}\}$ to all clients $i = 1, ...I^{rc}$, so that treatment $t = 0$ is free. This is done to increase κ_i above the levels in (22). This does not change the tradeoffs, and is described in the Appendix.

2.2.3 Restricted Choice Mechanisms and Artificial Currency

Another interpretation of the outcome above is that it implements competitive equilibrium with equal incomes (Foley, 1967). Competitive equilibrium with artificial currency is the most common "market design" alternative to Deferred Acceptance (Budish, 2011, Budish and Cantillon, 2012, Prendergast, 2017, 2022, Goedde, 2024, and Wong, 2024). To see the analogy to the outcome above, consider the following game using artificial currency.

Artificial Currency Game: The firm offers supply of $t_0 = \frac{I^{rc}}{3}$, $t_1 = \frac{\kappa_1^* I^{rc}}{3}$ and $t_2 = \frac{\kappa_2^* I^{rc}}{3}$ at unit fiat money prices of $R = R^*$. Each client $i = 1, 2, ...I^{rc}$ is given 1 unit of artificial currency to spend only on services. Each client reports a 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 r_i^{1st}. 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 (t_0^*, t_1^*, t_2^*). Then $(t_0^*, t_1^*, t_2^*) = (1, \frac{1}{\kappa_1^*}, \frac{1}{\kappa_2^*})$ is a competitive equilibrium outcome. Consider a client i of type τ_{ik}, $k = 0, 1, 2$. Then consider the following equilibrium strategy: $\lambda_k = 0$ and $\lambda_{-k} = 1$. With these strategies, there is demand for each good of $\frac{I^{rc}}{3}$ at a minimum allocation of 0, so with equal rationing the equilibrium quantities of each good are $(t_0^*, t_1^*, t_2^*) = (1, \frac{1}{\kappa_1^*}, \frac{1}{\kappa_2^*})$. 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 r_i^{1st}, where $\kappa_0^* = 1$ to any κ_{-i}^*. Hence this is a competitive equilibrium.

Food banking reflects many of these issues. It has two components: larger food depositories and smaller food pantries. Food depositories distribute food of different desirability to food pantries. Restricted Choice implies quantity restrictions. Some foods - produce for instance - are free to food pantries in unlimited quantities. Other food can be purchased with fiat money. Yet for the most desirable food - typically meats - food pantries are rationed in the amount they can buy. Food pantries purchase these subject to a total food demand. As a result, the outcome is a combination of pricing and quantity restrictions.

Second, our interest is in the way in which fiat money payments affect outcomes, as the role of prices is central to the analysis. Consider the artificial currency market design by Feeding America, In addition to paying artificial currency for food, food

banks must also use fiat money to pay for transportation of food to the food bank. On average, food banks travel 800 miles to collect food. Satisfying the Fairness Constraint was central to its success. Consider an interpretation where service $t = 0$ is where a food bank does not acquire food in a given week while $t \neq 0$ involves acquiring food of different quality. Consider the role of paying for transportation. This has the standard benefit of screening out those who do not need food in a given week. Yet this must be designed to satisfy Fairness which generates the κ_i^* above. A central concern of that market design was to ensure that these monetary costs would not harm smaller (typically poorer) food banks. The danger that the redesign would violate this Fairness Constraint was a major concern in the redesign. This is where the Fairness Constraint was most relevant in this design. The design was successful on this dimension as smaller food banks ended up acquiring more truckloads of food than before.

2.3 Partial Choice and the First Choice Mechanism

Above we described the market design literature as prioritizing Deferred Acceptance over other assignment mechanisms. This is typically implemented using an ordinal representation of preferences. A smaller literature uses cardinal preferences, as we do, to address the merits of non-strategy proof mechanisms that prioritize matching as many first choices as possible. The potential advantage of first choice mechanisms is that they allow the intensity of cardinal preferences to be represented in outcomes. 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 analog in our setting is Partial Choice. This arises when clients are asked to rank preferred outcomes, and the firm assigns any client to her first choice if that treatment is not oversubscribed, where oversubscribed treatments are assigned. We assume here that assignment is by Deferred Acceptance.

When $\hat{y} \geq \hat{y}^{uc}$, the firm can choose prices such that no treatment is oversubscribed. This is not the case when $\hat{y} < \hat{y}^{uc}$. We show in the Appendix that the only case where Partial Choice can be preferred to Deferred Acceptance is when all type $\tau = 1$ clients rank $t = 1$ as their first choice but $\hat{t}(0, \underline{\phi}, m) = \hat{t}(2, \underline{\phi}, m) = 2$. The key difference to Assigned Consumption is that now all type $\tau = 1$ clients report $t = 1$ as their first choice, whereas previously $1 - h$ of these reported $t = 2$. This outcome can be implemented by a First Choice mechanism, which we now describe. The firm knows I^{1st} when allocating t. It offers $I_i = 1$ to all $i \leq I^{1st}$. The First Choice Allocation Mechanism is:

1. The firm chooses $R^{1st} = \{r_0^{1st}, r_1^{1st}, r_2^{1st}\}$.

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 a Deferred Acceptance Algorithm among those where $\hat{t}^k = i' \succ ...$

Proposition 5 shows the outcome. In the interests of brevity, we show the case where $h \leq \frac{1}{2}$. The case where $h > \frac{1}{2}$ is considered in the Appendix and is conceptually similar.

Proposition 5 *Let $h \leq \frac{1}{2}$. If $\hat{y}$ exceeds a critical value, y^{1st}, a First Choice equilibrium exists with expected surplus*

$$ES^{1st} = \frac{D}{\overline{U} - r^{1st}} \left(v + T^{1st} - \frac{\delta D}{2(\overline{U} - r^{1st})} \right) - \frac{\delta \sigma_\epsilon^2}{2(\overline{U}^* - r^{1st})^2} > ES^{da}, \tag{26}$$

where $r^{1st} = \underline{r} + \frac{2\underline{\phi}(1-(\frac{1-q}{2}+q(1-p_2^*)))}{3\underline{m}}$, $T^{1st} = \frac{\mu+h\overline{\phi}+\hat{\phi}+(q+\frac{1-q}{2}(1-p_2^*)(1-h))\underline{\phi}+\hat{\phi}(q+\frac{1-q}{2}p_2^*)}{3}$, and $p_2^* = \frac{1+h}{2-h}$. The critical income level y^{1st} is defined in the Appendix, and can be $y^{1st} = \underline{\eta}$.

Part of Proposition 5 is familiar, where expected allocation surplus from First Choice is higher than with Deferred Acceptance.[37] The other outcome is novel: the First Choice mechanism raises more funding than Deferred Acceptance. Revenue increases by $\frac{2\underline{\phi}(1-(\frac{1-q}{2}+q(1-p_2^*)))}{3\underline{m}}$ and $EI^{1st} = \frac{D}{\overline{U}-r^{1st}} > EI^{da}$.[38]

The equilibrium in Proposition 5 only exists for average income high enough.[39] In the Appendix, we show the outcome for lower levels of client income. For average income sufficiently low, there is no pure strategy First Choice equilibrium. However, for higher levels of average income, the firm can implement the allocative outcome above, but at the expense of lower revenue than r^{1st}. This is shown in the Appendix.[40]

[37] Intensity is revealed through comparing the certainty of $t = 1$ with the equilibrium distribution of outcomes from reporting $\hat{t} = 2$. Allocative efficiency is higher above under First Choice because the technology in (1) has the objective of matching $\tau = i$ to $t = i$. For other matching technologies, allocative efficiency under DA can be higher. To see this, let $T(t_i, \tau_j) = t_{ij}$, where $t_{22} \approx t_{12} > 0$ and all other $t_{ij} \approx 0$. In this case, what matters is ensuring type $\tau = 2$ receives either $t = 1$ or $t = 2$.

[38] Revenue increases by just enough to deter type $\tau = 1$ from choosing the Deferred Acceptance lottery between $t = 2$ and $t = 0$.

[39] The reason is that any client can receive $t = 1$ by ranking it first, but the equilibrium requires that only type $\tau = 1$ does so. In order for this to arise, the firm may need to price good $t = 1$ high enough to deter type $\tau = 0$ clients from claiming $t = 1$.

[40] The market design literature has emphasized the importance of strategy proof mechanisms. We have ignored that cost here. It is straightforward to allow the firm to incurs a fixed cost if its mechanism is not strategy proof. A concern with non-strategy proof mechanisms is that some clients may not be able to strategize. In the Appendix, we address the case where a fraction χ always reporting the treatment that give them highest utility without considering the actions of others. This reduces the value of the First Choice Allocation, but does not change the conceptual results.

Formal First Choice mechanisms are rare. Yet we argue below that their central feature: offering a client a choice between the sure thing of an under-subscribed service and an alternative uncertain outcome are ubiquitous. This is discussed below.

2.4 The Non-Profit Outcome

We now aggregate our results in two ways to address the optimal structure of the non-profit. First, Figure 2 offers a taxonomy based on the capacity of clients to pay $\hat{y}$ and the information held by the firm on client need, q.

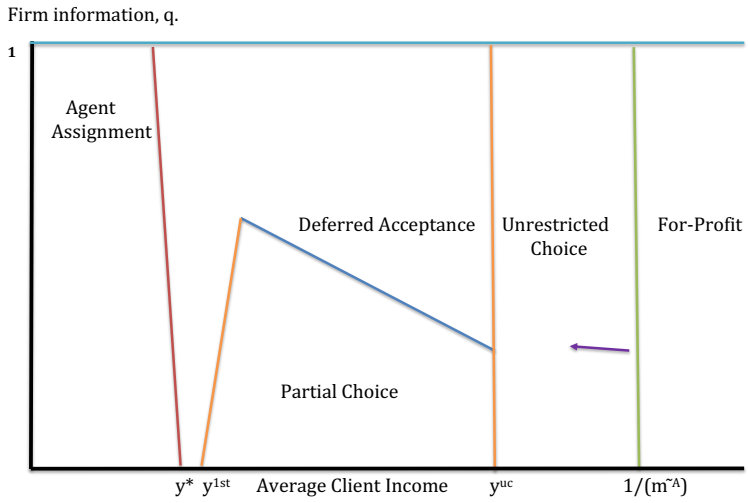


Figure 2: **The Outcome for Indivisible Goods.**

We have focused on r^A and D as the primary sources of funds for the social organization in (12). Non-profits receive 49% of their revenues from fees to clients, with another 22% coming from donors (National Council of Nonprofits, 2024). Consider first how the preferred allocation mechanism changes with $\hat{y}$ which generates r^A. When the non-profit is poorly funded, it cannot use Unrestricted Choice. This implies that the outcome has assignment error. The Fairness Constraint also means its revenue is no higher than if it tossed a coin to assign services. As a result, its only choice to expand the mission becomes to further distort outcomes by delegating control to its service providers. Firms that are sufficiently low on resources use Agent Assignment. For income below y^*, the firm uses Agent Assignment. However, y^* is decreasing in q as higher q makes DA more appealing as it entails less allocation error. In this sense, optimal assignment mechanisms are generated by the inability to raise funds.

As client income levels increase, the firm is less reliant on cutting costs for employees and the importance of allocative efficiency rises. This implies transitioning to

rule-based mechanisms with indivisible services: Deferred Acceptance or First Choice. The firm is more likely to choose Deferred Acceptance is it is well informed, q high, as there are fewer allocation errors. Once y^{uc} is reached, the firm transitions to Unrestricted Choice over either First Choice or Deferred Acceptance. (For wage reductions large enough, it may continue to prefer Agent Assignment.) The outcome for indivisible goods is given in Figure 4 in the Appendix. The only change is that Restricted Choice becomes a viable outcome, and is preferred when income is higher, as the κ_i^* are increasing in income, and when the firm is poorly informed, as the alternative to client choice is worse.

Our second approach is to offer a comparison to the existing literature, where client coverage is exogenous to the allocation mechanism used. This is given in Figure 3. This replicates the outcomes of the existing market design literature where the choice is between Deferred Acceptance and the First Choice mechanism. As we view

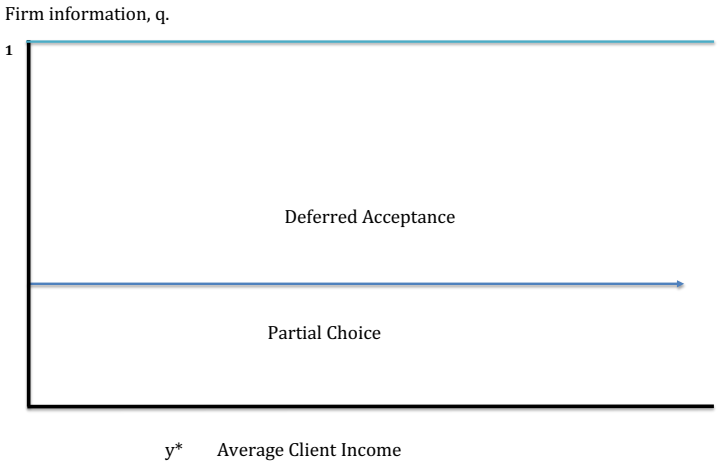


Figure 3: **The Outcome for Indivisible Goods When Clients Coverage is Exogenous**

outcomes in utilitarian terms, Deferred Acceptance is preferred above some critical level of q as allocative errors fall. The distinction between Figures 1 and 2 shows clearly how the addition of funding renders Choice and Agent Assignment more likely, and Deferred Acceptance less likely. Yet the outcome is potentially more clearcut than this, as shown in Proposition 6.

Proposition 6 *If $y^* > \hat{y}^{uc}$, no firm chooses Deferred Acceptance.*

This arises for two contrasting effects of prices. Agent Assignment is preferred to Deferred Acceptance if $\hat{y} < y^*$, and Unrestricted Choice is preferred to Deferred Acceptance if $\hat{y} > \hat{y}^{uc}$. The Proposition immediately follows. In this sense, funding nullifies

the priority given to Deferred Acceptance by the market design literature.[41] Deferred Acceptance mechanisms have made few inroads among the two million non-profits in the United States.[42] It is also notable that successfully implemented Deferred Acceptance mechanisms have either been in settings where funding occurs through taxation (school choice) or where the potential funds received are trivial compared to overall funding for the institutions (medical resident matching and kidney exchange).

3 The Social Enterprise

Many social organizations operate as Social Enterprises that can make profit but whose actions are oriented by social objectives. 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). More generally, the World Economic Forum has estimated that there are 10 million such enterprises worldwide, accounting for 3% of world employment (World Economic Forum, 2024). Here we address a stark feature of the Social Enterprise, namely that despite its social mission, it cannot exist unless it is entirely client-funded. To address this, we need to consider the motivation of donors.

Donors: Donors overwhelmingly report that their motivation is the impact that their gift has on the mission of the institution.[43] Yet any behavioral model must take account of the limited information of donors.[44] We make three assumptions. First, donors do not observe the assignment mechanism or contracts, nor the actions of other donors. Instead, all they observe is the profit status of the firm. Second, donations are given by $\mathbb{D}(b)$, where b is donor belief about the probability that their donation is spent on I, the mission of the firm. Finally, we assume that donor beliefs about b are correct in equilibrium.

[41] Another potential benchmarks is when clients cannot pay, which would arise if $\underline{m} = \infty$ (when some clients have no access to money). This can be incorporated by assuming that the support of income for any level of average income includes $y = 0$. The outcome is in Figure 5. Two differences arises from Figure 1. Both Deferred Acceptance and Agent Assignment are more likely. The hashed line represents the outcome in Figure 2, which shifts to the left as the absence of funds from clients makes cost cutting more important.

[42] The number of non-profits has increased from 1.25 million in 2000 to two million today.

[43] 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.

[44] This is manifested in the literature on whether donations are responsive to firm efficiency, where the outcome is ambiguous. See Buchheit and Parsons, 2006, Parson, 2007, and Yoruk, 2013.

Earlier, we assumed that

$$\pi(\tilde{I}) \begin{cases} = 0 & \text{if } \Pi = 0 \\ \leq \mathbb{D}(b) + \epsilon & \text{if } \Pi = 1. \end{cases} \tag{27}$$

Let $\pi(\epsilon) \equiv \pi(\tilde{I}|\mathbb{D}(b), w^A, r^A)$ be the net profit of the firm given $\epsilon, \mathbb{D}(b), w^A$ and r^A, and let $\pi^*(\epsilon)$ be the equilibrium contract, where $\int_{\underline{\epsilon}}^{\overline{\epsilon}} \pi^*(x) f(x) dx = 0$. Note that all donors give holding the donations of others constant. As a result, they evaluate their donation at the use of the marginal dollar received by the firm, as their donation increases donations from d to $d+1$. Then for a realization ϵ, a fraction $1 - \pi^{*'}(\epsilon)$ of the marginal dollar is spend on I rather than on π. As b is assumed to be correct in equilibrium, Lemma 4 follows.

Lemma 4: $b = \int_{\underline{\epsilon}}^{\overline{\epsilon}} (1 - \hat{\pi}^{*'}(x)) f(x) dx$

Now consider the outcome for an allocation mechanism A. If $\Pi = 0$, by definition $\pi = 0$ and the outcome is given by (10). If $\Pi = 1$, the firm chooses π^*, but does so holding donations fixed as donors cannot observe contracts. Surplus is

$$ES = \int_{\underline{\epsilon}}^{\overline{\epsilon}} \frac{(\mathbb{D}(b|\Pi) - \pi^A(x) + x)}{w^A - r^A} (v + T^A - \frac{\delta(\mathbb{D}(b|\Pi) - \pi^A(x) + x)}{2(w^A - r^A)}) f(x) dx. \tag{28}$$

This simplifies to

$$ES = \frac{\mathbb{D}(b|\Pi)}{w^A - r^A} (v + T^A - \frac{\delta \mathbb{D}(b|\Pi)}{2(w^A - r^A)}) - \int_{\underline{\epsilon}}^{\overline{\epsilon}} \frac{\delta(-\pi(x) + x)^2}{w^A - r^A} f(x) dx. \tag{29}$$

Proposition 7 immediately follows.

Proposition 7 *The Social Enterprise chooses* $\pi^{*'}(\epsilon) = 1$. *As a result,* $b = 0$ *and* $\mathbb{D}(0) = 0$ *and the Social Enterprise must be fully funded by clients.*

Proposition 7 shows starkly how a for-profit firm cannot attract donations despite its social mission. The firm chooses π^* holding b fixed (as donors do not observe the contract). It responds by providing full insurance to the mission. In this way, the cost that arose consistently above from randomness in funding ($\frac{\delta \sigma_\epsilon^2}{2(w^A - r^A)^2}$) disappears. However, as b is correctly inferred in equilibrium, $(b|\Pi = 1) = \int_{\underline{\epsilon}}^{\overline{\epsilon}} (1 - \hat{\pi}^{*'}(x)) f(x) dx = 0$. The central conclusion of Proposition 7 is then the impossibility of a for-profit charity.[45] For empirical context, we described above how non-profits rely on donors for 22% of funding. This is more pronounced for small non-profits, where 30% of

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

income comes from donors (Tomasko et al., 2021). The result above shows that the Social Enterprise must make up that difference.[46]

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. 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 what is necessary to encourage donations, as a donor must know that her donation affects the mission.[47]

There is a simpler behavioral interpretation of this outcome. Imagine that donors simply do not give to for-profit firms as they fear that their donation will be simply used to enrich the owners of the firm. That would offer a similar outcome. The purpose of this section is that (i) rational donors should value their gift at the marginal rather than average dollar, (ii) they correctly choose not to give if $\pi^{*'}(\epsilon) = 1$, and (ii) there is a simple reason why the firm will indeed choose $\pi^{*'}(\epsilon) = 1$.

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 in the charitable sector. 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 allowed.[48] 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".[49] As a result, it is widely believed

[46]In that case, coverage is at its efficient level, $I^* = \frac{v+T^A}{\delta}$. In settings where $\frac{r^A}{\underline{m}} \geq w^a$, the model is silent on equilibrium pricing. The firm above only cares about surplus, and so is indifferent about any prices between w^a and $\frac{r^A}{\underline{m}}$.

[47]A useful analogy here is to budget balance and the team production problem. In that setting, efficient effort only arises if all parties can be residual claimant on output, which is impossible with budget balance. A similar issue arises here. Donors give more when the mission-related activities are the use of residual funds, while financial security only arises when the firm is residual claimant. The assumption that donors do not see the contract offered implies that the outcome is "1/0".

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

[49]WeConserve also argue that "some foundations, corporations, and private donors will not make

in the sector that using donations as marginal compensation deters donors.[50]

3.1 The Non-Profit Choice

Firms choose whether to incorporate as non-profit or for-profit. Here that choice is between financial security without donors, or financial uncertainty with donors.[51] Yet the capacity to cover costs is not sufficient for becoming a Social Enterprise: additionally, it must not compromise its mission. To see this, note that the self-funded firm is not limited in client coverage, choosing $I^* = \frac{v+T^A}{\delta}$ for assignment mechanism A. If the firm can self-fund with assignment mechanism A, it never operates as a non-profit with that assignment mechanism. 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 $w^* < \underline{r}$. The firm chooses a non-profit with mechanism A over a self-funded one with A' if $ES^A \geq \frac{(v+T^{A'})^2}{2\delta}$.

People often complain about bureaucracies as unresponsive to their needs and concerns. The contrast is to the "consumer is always right" ethos of the typical for-profit setting. We now show a condition under which this dichotomy arises, namely where all non-profit firms assign client consumption while all for-profits allow choice. This arises if

$$\frac{\mu - \underline{\phi}}{V'(\hat{y}^{uc} + \underline{\eta})} < \overline{U} < \frac{\mu + \frac{r_0^* + 2\underline{\phi}}{3}}{V'(\hat{y}^{uc} + \underline{\eta})}. \tag{30}$$

Consider the outcome with indivisible goods at $\hat{y}^{uc}$, when the firm transitions to Unrestricted Choice. Client payments increase discretely from $\frac{\mu - \underline{\phi}}{V'(\hat{y}^{uc} + \underline{\eta})}$ to $\frac{\mu + \frac{r_0^* + 2\underline{\phi}}{3}}{V'(\hat{y}^{uc} + \underline{\eta})}$. Then if the extra revenue allows the firm to cover its costs, all non-profits make allocation errors but the for-profit never does, as it allows Choice.

More generally, the ability to self-fund is a comparison between the cost of the agent w and the ability to raise funds r. Let $\tilde{m}^A$ be the marginal utility of the poorest client such that $w^A = r^A$, defining the average income at which the firm becomes self-funded. This is given by the arrowed line in Figure 2, where we trace outcomes based on where that line lies. In general, the firm can be self-funded for any assignment mechanism if $\overline{U}$ is low enough. However, (30) offers a stark case where Choice is only observed among Social Enterprises. In Proposition 10 in the Appendix we show

a contribution if a portion of the donation will be paid as a commission. In some cases, they will not give to a campaign that pays any commission at all".

[50] 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.

[51] This echoes the recent message of Brumme and Trelstad, 2023, whose study of the for-profit decision concludes that "the central question here is: Can market conditions support a for-profit approach, or does the cost to serve...exceed potential revenue?".

that while for-profit firms can exist with any allocation method, it more likely when client choice is involved. Choice and the for-profit sector are correlated within in the social sector, solely because of the implications of funding constraints for assignment mechanisms. In this sense, it is a generalization of (30) and the view that non-profit firms are inherently more likely to make allocative errors.

4 Observations and Extensions

Here we discuss some modeling choices and their robustness.

Relaxed Fairness Constraint: The outcomes above are premised on the assumption that no client can be denied coverage for lack of income. Yet some social organizations may feel that they have no choice but to exclude some low income clients in order to serve more clients.[52] As an example, private housing associations will not rent units to those who cannot pay at all, but instead have a minimum rent that while it includes the low income, does not include all.

We now consider a less stark Fairness Constraint to offer an alternative interpretation. Assume that the firm continues to use a Fairness Constraint among all clients that it serves, but that it only covers clients whose income exceeds $\hat{y} + \eta^* \geq \hat{y} + \underline{\eta}$. Let $V'(\hat{y} + \eta^*) = m^*$, where the firm chooses m^*, the lowest income client that it covers. Assume that a fraction $G(m^*)$ clients have income between $\underline{m}$ and m^*, where $G(\underline{m}) = 0$ and $G' \geq 0$. As there are an infinite number of potential clients, this includes the revenue maximizing choice by the firm, where it only serves clients of type $\overline{m}$.

The tradeoff for the firm is between raising more funds and failing to serve high priority clients. For illustrative purposes, consider the case where the firm uses Unrestricted Choice to allocate services. If the firm chooses $m^* = \underline{m}$, surplus is given by (21). If instead it chooses $m^* > \underline{m}$, it charges prices of $\tilde{R}^* = \{\tilde{r}_0^*, \frac{\phi+\mu}{m^*}, \frac{\phi+2\mu}{m^*}\}$, where $\tilde{r}_0^* = \min\{\frac{\phi-2\mu}{m^*} + \frac{2\mu+\phi}{m^*}, \frac{\phi}{m^*}\}$ with revenue of $\tilde{r}^{uc} = \frac{\tilde{r}_0^* + \frac{\phi+\mu}{m^*} + \frac{\phi+2\mu}{m^*}}{3}$. Surplus is

$$ES^{uc} = \frac{D}{\overline{U} - \tilde{r}^{uc}} \left(v + T^* - \frac{\delta \frac{D}{1-G(m^*)}}{2(\overline{U} - \tilde{r}^{uc})} \right) - \frac{\delta \sigma_\epsilon^2}{2(\overline{U} - \tilde{r}^{uc})^2}, \tag{31}$$

This illustrates the firm's tradeoff. Excluding some clients increases total coverage, by $\frac{D(\tilde{r}^{uc} - r^{uc})}{(\overline{U} - \tilde{r}^{uc})(\overline{U} - r^{uc})}$. The problem is that it covers the wrong clients. With probability $G(m^*)$ a client is skipped in the priority ranking, so that while $\frac{D}{\overline{U} - \tilde{r}^{uc}}$ clients are covered, their average value is $v + T^{da} - \frac{\delta \frac{D}{1-G(m^*)}}{2(\overline{U} - \tilde{r}^{uc})}$, lower than when clients are not excluded. For example, the revenue maximizing choice, $m^* = \overline{m}$, yields $ES^{uc} =$

[52] As an example, Antecol and Smith, 2012, and Murphy, 2023, show that the use of Early Decision in university admissions excludes lower income students.

$\frac{D}{\overline{U}-\tilde{r}^{uc}}\left(v+T^*-\frac{\delta\frac{D}{g(\overline{m})}}{2(\overline{U}-\tilde{r}^{uc})}\right)-\frac{\delta\sigma_\epsilon^2}{2(\overline{U}-\tilde{r}^{uc})^2}$, where $\tilde{r}^{uc}=\frac{\tilde{r}_0^*+\frac{\phi+\mu}{\overline{m}}+\frac{\phi+2\mu}{\overline{m}}}{3}$. The firm chooses where to locate on the choice between more clients and higher priority clients. For δ high enough, the firm will cover its highest priority clients over expanding coverage, which is the premise that underlies much of the analysis here.

Donor Information: We have assumed that donations are exogenously given by D. This has allowed us to map firm resources in terms of the revenue they receive from clients. In reality, donors likely know the kind of clients who will be benefit, so D may be indexed by $\underline{m}$. All results continue to hold if $D(\underline{m})+\underline{r}$ is decreasing in $\underline{m}$. 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).

A Managerial Implication: Social organizations have many stakeholders: clients, donor, and employees. A central issue for governance of the social organization is how to weight the interests of each of these in the absence of the usual objective of profit maximization. Consider a welfare weight approach to this problem where the firm's objective is to maximize $\beta\sum_{i=1}^{\infty}U_i^c+(1-\beta)U^a$, subject to the worker being paid according to (6). How should the firm prioritize the interests of its employees relative to its clients? The implication of the analysis above is that for $\hat{y}<y^*$, the firm should choose $\beta=0$ and for $\hat{y}\geq y^*$, the firm should choose $\beta=1$. This is manifested in Agent Assignment. In words, small social organizations should prioritize the interests of its employees, yet when the firm becomes large enough, it should reorient towards the interests of its clients. Of course, ultimately the actions taken towards employees for $\hat{y}<y^*$ will benefit clients, but this offers a managerial implication for the social organization.

Diminishing Returns: We have assumed that diminishing returns are linear: $v_i=v-\delta i$. This is simply to allow the impact of uncertainty in funding, which plays no role until the for-profit decision, to be separable from funding levels. If instead we do not make this assumption, the analog to (10) is

$$ES^*=\frac{D}{w^A-r^A}T^A+\sum_{i=1}^{\frac{D}{w^A-r^A}}v_i-\int_{\underline{\epsilon}}^{\overline{\epsilon}}\frac{(x-\pi(x))(v_{D+x-\pi(x)}-v_D)}{w^A-r^A}f(x)dx. \tag{32}$$

This only changes the outcome to one where now the cost of uncertainty depends on where on the distribution of v funding is located, but does not change any conceptual results above.

Matching Returns: The technology in (1) has been chosen to simply distinguish between vertical differentiation μ and horizontal differentiation ϕ. This is largely for ease of interpretation of results. Consider a case where instead we assume

$$T(\tau_i, t) = \mu_{\tau_i t} + \phi_{\tau_i} \mathbb{1}(t = \tau_i), \tag{33}$$

with $\phi_{\tau_i} \in \{\underline{\phi}_{\tau_i}, \overline{\phi}_{\tau_i}\}$ and $\mu_{\tau_i t}$ is increasing in t for all τ_i. The analog to Assumption 1 becomes $\overline{\phi}_{\tau_i} > \mu_{\tau_i 2}$ and $\underline{\phi}_{\tau_i} < \mu_{\tau_i 2}$ for all τ_i. The assumption of the optimal match entailing $t = \tau$ requires that the maximum match utility across all permutations of matched is $\sum_{j=0}^{2}(\mu_{jj} + \phi_j)$. Then the same qualitative results continue to hold. For example, the analog to the critical average income level allowing Unconstrained Choice becomes

$$\frac{V'(\hat{y}^{uc} + \overline{\eta})}{V'(\hat{y}^{uc} + \underline{\eta})} = \frac{max\{\mu_{20} - \underline{\phi}_0, \mu_{21} - \underline{\phi}_1\}}{\mu_{22} + \underline{\phi}_2}. \tag{34}$$

It is to avoid extensive notation that the simpler structure above has been chosen.

Compensation and Altruism: The agent has been assumed to value the extent to which he believes he affect client outcomes relative to an alternative agent. The primary reason for this is to reflect the fact that when firms transition to rule-based allocation mechanisms, even though rule based allocations are likely more efficient, the evidence is clear that it is overwhelmingly disliked by social service providers. It has one more subtle implication described above, namely that the agent must be paid a higher wage when Pareto efficient Agent Allocations are implemented, despite client welfare being higher in the eyes of the agent. To address robustness, assume instead that the condition to hire the agent under Agent Assignment is instead $w + \alpha\hat{U}^{ca} \geq \overline{U} + \alpha\tilde{U}^{c}$, where $\tilde{U}^{c}$ is an exogenous belief that the agent holds about the client's utility without the agent. This changes the calculus in one substantive way: the agent now accepts a *lower* wage under Pareto Efficient Agent Assignment than in Proposition 2. This is because the agent does not internalize the fact that if he turns down the job, the benefits of Pareto efficiency will accrue when the next agent implements outcomes. Under this alternative preference for the agent, the firm always chooses the Pareto efficient benchmark over Agent Assignment.

Ex Ante and Ex Post Fairness: Note one feature of the Restricted Choice outcome. We have described this as relevant for settings where services are divisible. This is because we have characterized the Fairness Constraint as *ex post*, where observed allocations must satisfy it. An alternative is to characterize Fairness as ex ante. In this case, with indivisible goods the κ_i^* would be interpreted as client stating probabilities of receiving the service. An outcome that takes this approach would mimic the insights in Hylland and Zeckhauser, 1979. We have ignored it here as we believe that most social organizations simply will not fail to offer service its most needy clients, even if it has appealing *ex ante* properties.

Partial Choice outcomes: Formal First Choice mechanisms are not common. However, social organizations routinely offer informal non-strategy proof choices akin to the mechanism above. This arises through presenting a client with a non-strategy proof choice between a less-than-optimal option and a random alternative generated by the allocation process. We give two examples.

First, social housing is typically allocated through waiting lists, where a client is matched to a house as it becomes available.[53] The house may not be the client's most desired type or in her preferred location. The offer made to a client typically comes with a cost: if the client turns down a "reasonable" offer, she is placed at the bottom of the waiting list going forward. With repeated turn-downs, she will be taken off the list.[54] As a result, the client needs to strategize about whether to accept. This it is conceptually analogous to the First Choice mechanism above, in that there is a choice between a sure, if less preferred, thing or a random alternative.

Second, the process for applying to universities in the US has two distinct options. First, students can apply through Regular Decision, where the student is not committed to a school if accepted. Second, a student has an earlier option to identify a single school to apply to via Early Decision. If the student is accepted by that school, she is obliged to accept the offer.[55] Among elite colleges, probabilities of acceptance are often four times higher than with Regular Decision and the use of ED has increased significantly over the last decade (Murphy, 2022). Despite not being designed as a First Choice mechanism above, students choose between a more certain option and a random alternative. More relevantly for our setting, it is well known that Early Decision students pay higher average tuition (Antecol and Smith, 2012, Murphy, 2023). As a result, the funding component of the assignment mechanism seems key.

Targeting and Restricted Choice: Above we described the tradeoff for Restricted Choice between reducing assignment error and worse targeting of clients. We use two examples.

Feeding America allocates food using an artificial currency that food banks use to bid. They choose between higher quantities of a less desired good and lower quantities of a more desired one. Before the design, each food bank was offered a random assignment of food, where each food bank received the same number of pounds of

[53]See Leshno, 2022, for a price-theoretic approach to addressing housing allocations through waiting lists.

[54]For example, the Citizens Advice Bureau for housing in Britain advises "refusing could also make your situation worse, for example if your local council reduces your points or removes you from their priority list" (https://www.citizensadvice.org.uk/). In Warwick in Britain, "If you refuse two suitable offers of accommodation within a 12-month period, you will be disqualified from the housing register for a period of 12 months" (https://www.warwickdc.gov.uk.). McDermott, 2019, reports that over 350 clients had this penalty enforced in the Irish social housing system.

[55]This is not a legally binding commitment, but is treated as binding by most students. Avery and Levin, 2011, have shown that Early Decision can be used to screen students on their enthusiasm for a school in the same way as described above.

food, or in our language $I_i = 1$. The outcome with bidding with artificial currency is a Restricted Choice outcome, with prices (the analog to $\frac{1}{\kappa_i^*}$) given in Figure 5 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.

The targeting issue is that under the old allocation mechanism, some 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? During the redesign at Feeding America, the designers cared greatly about who would benefit in terms of more pounds of food and who would get less. There was clearly a desire that the amount of food would be reoriented towards smaller, poorer food banks.[56] 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 over 125 countries after early 2021. Here $S = \sum_{i=1}^{N}(v_i + T(t_i, \tau_i))I_i$ 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 arise was that many vaccines were close to their expiry date by the time they were to be distributed. In the language of the model, $t = 0$ vaccines are close to expiry while $t = 2$ 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 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.[57]

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-

[56]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.

[57]This led to well-publicized complaints from the Organization of African States and the Africa CDC Chief that "the majority of the donations have been ad hoc, provided with little notice and short shelf lives" (Al Jazeera, 2022, Organization of African States, 2021). Anecdotes abound: South Sudan destroying 60,000 vaccines, or the Democratic Republic of Congo returning 1m vaccines close to its expiration. It is also reported that 681m vaccines remained unused in low income countries (Al Jazeera, 2022).

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.

There is one 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. By definition, targeting is then harmed by Restricted Choice, the final two terms in (23). In reality, political considerations implied that both Feeding America and Covax had to offer equal access to all food banks or to all participating countries. Yet Feeding America's priorities were to reallocate food to poorer food banks, and Covax wished to vaccinate the poorest countries (and to be perceived to do so). This is why targeting improved for Feeding America. The purpose of this section is not to match the outcome tightly to the model but instead to illustrate how impacts on targeting have a critical influence on whether the mechanism is implemented.

Pricing: We have ignored one possibility for the firm to raise funds. The firm could potentially charge a fixed fee to clients before treatments are allocated: $R = \{\hat{r}, r_0, r_1, r_2\}$, where $\hat{r}$ is a charge before allocations are made. This allows the firm to raise more funds, which we describe in the Appendix. The results do not change in any conceptually substantive way. We have ignored it here because any additional implications rely on clients making inferences about the equilibrium, as the fee that the firm can charge depends on their beliefs of likely outcomes. Many clients of social services are not well informed about likely outcomes when they encounter the system, so we have ignored it here.

We have also assumed $r_i \geq 0$. This is for the usual reason that when clients are paid for receiving services, the supply elasticity of undeserving candidates only interested in the money may be large. Without that assumption of non-negativity, the cost to the firm is that by paying those who receive $t = 0$ (to generate efficient screening), there are less resources to increase the scope of the mission I. It does not change the essential tradeoff.

Donor Preferences: It is straightforward to provide a micro-foundation for donations. Assume that there are N potential donors, where donor l has marginal utility of income m_l^d. Assume $m^d \sim (0, \infty)$ according to a CDF H. Each donor is considering donating \$1 to the firm. He has preferences $U^d = \alpha^d E S^* - m_l^d \mathbb{I}_l^d$, where α^d represents altruism towards a client and $\mathbb{I}_l^d \in \{0, 1\}$ and equals 1 if donor l gives. Donors observe Π. All donor beliefs are that donations of others are $\hat{D}$, allocative

efficiency is $\tilde{T}$, wages are $\tilde{w}$ and revenue is $\tilde{r}$. With correct equilibrium beliefs, donor l gives if $\frac{1}{\tilde{w}-\tilde{r}}\int_{\underline{\epsilon}}^{\overline{\epsilon}}(1-\hat{\pi}'(x))(v+\tilde{T}-\frac{\delta(\hat{D}+x)}{2(\tilde{w}-\tilde{r})})f(x)dx \geq \frac{1}{m_l^d}$. Total donations are $\mathbb{D}(b) \equiv NH((\frac{1}{\tilde{w}-\tilde{r}}\int_{\underline{\epsilon}}^{\overline{\epsilon}}(1-\hat{\pi}'(x))(v+\tilde{T}-\frac{\delta(\hat{D}+x)}{2(\tilde{w}-\tilde{r})})f(x)dx)^{-1})$. Then as Π is observed, donations for $\Pi = 1$ are 0 as $\pi^{*'}(\epsilon) = 1$, while $D \equiv NH((\frac{1}{\tilde{w}-\tilde{r}}\int_{\underline{\epsilon}}^{\overline{\epsilon}}(v+\tilde{T}-\frac{\delta(\hat{D}+x)}{2(\tilde{w}-\tilde{r})})f(x)dx)^{-1})$ when $\Pi = 0$ as $\pi = 0$.

5 Agency Interpretation

The value of allowing the firm to be residual claimant above is to offer financial security to the mission. The literature on non-profit status has instead focused on agency problems (Hansmann, 1980, Glaeser and Schleifer, 2001). We now interpret the choice of profit status through an agency lens. We do so for two reasons. First, to show a novel agency interpretation of the for-profit choice. Second, to show that outcomes need not be so stark as to eliminate all donations when the firm operates as a Social Enterprise.

The literature on the for-profit choice has studied a setting where a firm controls unobserved quality, and cuts corners on unobservable dimensions of quality in order to increase its profits at the expense of clients. The firm then potentially commits to non-profit status to eliminate that incentive. 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.[58]

We also assume that (i) the firm invests in unobservable quality, and (ii) when clients pay for treatment they do not see quality but instead make an equilibrium inference. 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.

[58]This is not just to reflect the empirical evidence above, but because there remain many areas - fund raising, cost cutting by firing workers - where it remains difficult to motivate the management of non-profits. For example, Angrist et al, 2013, and Campos and Kearns, 2024, identify that the strongest predictors of charter school quality are increased student suspension rates and attendance rates, which school administrators are often reluctant to implement.

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

$$\tilde{I} = \frac{eD}{w^A + C(e^*) - r^A e}. \tag{35}$$

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 R, w, π 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, r_i paid by clients, and w and π 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 $r^A\hat{e} - C(e) < w^A$. 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 now $w^A - r^A e + C(e)$. Expected surplus is now given by $ES = \frac{eD}{w^A + C(e) - r^A e}(v + T^A)$.[59] Note the comparison to (12). As $\delta = 0$ there is no longer a diminishing returns cost to uncertainty. However, quality scales both the value of surplus and the potential revenue from clients. This yields a first order condition for optimal quality

$$\frac{e^*(C'(e^*) - r^A)}{w^A + C(e^*) - r^A e^*} = 1. \tag{36}$$

Now consider the non-profit firm's incentives where $\pi = 0$, taking account of the fact that e is not observable to clients. Then $\tilde{e}$ is set to maximize $ES = \frac{\tilde{e}D}{w^A + C(\tilde{e}) - r^A\hat{e}}\left(v + T^A\right)$, where $\hat{e}$ is expected quality. Here the first order condition is $\frac{e^*C'(\tilde{e})}{w^A + C(\tilde{e}) - r^A\hat{e}} = 1$, or as beliefs are correct in equilibrium,

$$\frac{\tilde{e}C'(\tilde{e})}{w^A + C(\tilde{e}) - r^A\tilde{e}} = 1. \tag{37}$$

Comparing (37) to (36) shows that the non-profit chooses quality too low, as it does not internalize the impact of quality on revenue.

[59] We have not indexed e by the allocation mechanism A for notational simplicity.

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 $I \leq \frac{e(\mathbb{D}(b)-\pi(\tilde{I}))}{w^a+C(e)-r}$, where $E\pi = 0$. The firm chooses e to maximize

$$\int_0^{\overline{I}} \pi(\tilde{I}(e))f(x)dx + e(v+T^A)\frac{\mathbb{D}(b)}{w^A + C(e) - r^A\hat{e}}, \tag{38}$$

The firm's choice of e is given by the first order condition

$$\int_0^{\overline{I}} \pi(x)f_e(x)dx + \frac{\mathbb{D}(b)(v+T^A)(w^A + C(e) - r^A\tilde{e} - eC'(e))}{(w^A + C(e) - r^A\hat{e})^2} = 0. \tag{39}$$

The firm wishes to implement e^* in (36). We consider Residual Claimant contracts, where the firm keeps all revenue above a certain level of $\tilde{I}$, which we call I^*.[60] Given risk neutrality and MLRP, the optimal contract (not observed by donors) has a simple form:

$$\pi(\tilde{I}(x)) = max\{\tilde{I}(x)) - I^*, 0\} - \overline{\pi} \tag{40}$$

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 (39) holds:

$$\int_{I^*}^{\overline{I}} (\tilde{I}(x) - I^*)f_e(x)dx =$$
$$\mathbb{D}(b)(v+T^A)\left(\frac{w^A + C(e^*) - r^Ae^* - e^*C'(e^*)}{(w^A + C(e^*) - r^Ae^*)^2} - \frac{w^A + C(\tilde{e}) - r^A\tilde{e} - \tilde{e}C'(\tilde{e})}{(w^A + C(\tilde{e}) - r^A\tilde{e})^2}\right). \tag{41}$$

If

$$\int_0^{\overline{I}} \tilde{I}(x)f_e(x)dx <$$
$$\mathbb{D}(b)(v+T^A)\left(\frac{w^A + C(e^*) - r^Ae^* - e^*C'(e^*)}{(w^A + C(e^*) - r^Ae^*)^2} - \frac{w^A + C(\tilde{e}) - r^A\tilde{e} - \tilde{e}C'(\tilde{e})}{(w^A + C(\tilde{e}) - r^A\tilde{e})^2}\right), \tag{42}$$

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

[60] The relevant issue is what set of contracts are feasible within that class. One possibility would be to allow the firm to pay the firm D if $I \geq I^*$. These are likely to be infeasible. The theoretically most desirable contract is to offer the firm all revenue from the firm if I exceeds some level. We describe this contract in the Appendix. These contracts are empirically infeasible for two reasons. First, no clients will be served in these instances. Second, realistically the way in which this information becomes available is that at the end of the year, excess revenue is seen, and the firm can claim some of that. It cannot claim all revenue for the entire year.

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 d constant, as donors do not observe π.[61] However, in equilibrium the donor's contribution funds mission-related activities only with probability $F(I^*)$. Total donations are $\mathbb{D}(b) = \mathbb{D}(F(I^*))$. When $I^* = 0$, all donations are crowded out.[62]

This section has been provided both to illustrate the possibility of a for-profit firm continuing to attract donors ($I^* > 0$), but also to offer an agency based interpretation. Since Hansmann, 1980, the literature has argued that agency issues cause the firm to be non-profit, as harmful incentives causes the firm to cut corners and harm clients. This is reversed here: only if a firm can self-fund through clients can it be for-profit. From this perspective, the limit to for-profit social entrepreneurship is not quality concerns, but rather the simpler problem of whether the firm can pay the bills after scaring donors away.

6 Conclusion

Social organizations do not use market clearing prices to allocate their services. The academic literature has championed Deferred Acceptance as a preferred alternative to choice with prices. Its appeal is couched in terms of strategy proofness and stability rather than utilitarian gains.[63] 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. We also demonstrated how revenue is bounded by that which would arises by randomly assigning services. Finally, our approach offers an explanation for the co-existence of for-profit and non-profit firms that does not rely on the for-profit cutting corners on quality.

[61] This analysis assumes that the optimal assignment mechanism A does not depend on the for-profit status of the firm. It can do so, of course, but this does not change the logic of the choice being between efficient quality and low donations, or less efficient quality with more donors.

[62] The firm then chooses between non-profit status with return S^n, or for-profit with return $ES^* = \frac{e^* F(I^*) D(\underline{m})}{\overline{U}^a + C(e^*) - r^A e^*} \left(v + T^A\right)$.When $I^* = 0$, the firm can only operate as a for-profit if it can self-fund: $w^a + C(e^*) \leq r^A e^*$. In that case, optimal quality is no longer given by (36), as the cost there is fewer clients served. With the self-funded firm, quality is instead constrained by the budget constraint, $w^a + C(e^*) \leq r^A e^*$. As the objective of the firm is to maximize surplus, the firm will choose the maximum possible e where $w^a + C(e^*) = r^A e^*$, and $C'(e^*) = r^A$.

[63] See Li, 2017, on this.

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 on objectives. 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 agency literature that distorting ex post allocations can improve ex ante incentives may also apply to attributes such as *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 rather than through donors and clients. In addition, many social organizations do not used the price system not because its clients cannot afford it, but because there are externalities to client choice. A good example here is that schools care about the composition of their student body for reasons beyond anything that could be revealed by pricing. This paper is not meant to be a criticism of those allocation mechanisms. Finally, we also do not consider instances where the objectives of the firm are a hybrid of both surplus and private returns. The main reason for this is to understand the complex tradeoff that social sector firms face even when their objectives are "pure".

We began the paper by describing the importance of prices for allocative efficiency. 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, Volume86, Issue2, 391-444.

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

Al Jazeera, "Poorer Nations Forced to Dump Close-to-expiry Covid Vaccines", 2022. https://www. aljazeera.com/news/2022/1/13/poorer-nations-dump-close-to-expiry-covid-vaccines-unicef.

Alatas Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken,

Ririn Purnamasari and Matthew Wai-Poi "Ordeal Mechanisms In Targeting: Theory And Evidence From A Field Experiment In Indonesia", 2013, NBER Working Paper 19127.

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.

Angrist, Josh, , P. A. Pathak, and Chris Walters, "Explaining Charter School Effectiveness", *American Economic Journal: Applied Economics*, vol. 5, no. 4, October 2013.

Angrist, Josh, , Guthrie Gray-Lobe, Clemence M. Idoux and Parag A. Pathak, "Still Worth the Trip? School Busing Effects in Boston and New York", 2024, NBER Working Paper, 30308.

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.

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.

Bruhn, Jesse, Christopher Campos, Eric Chyn, and Anh Tran, "Who Benefits from Remote Schooling? Self-Selection and Match Effects", mimeo, July, 2024.

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. https://www.bls.gov/opub/mlr/2024/article/nonprof a-look-at-national-trends-in-establishment-size-and-employment.htm.

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.

Campos, Christopher and Caitlin Kearns, "The Impact of Public School Choice: Evidence from Los Angeles Zones of Choice", *Quarterly Journal of Economics*, 2024, 139, 1–43.

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. https://www.ccsfundraising

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

Cheng, M., and Hinnant, L., "Rich Nations Dip into COVAX Supply While Poor Wait for Shots", Associated Press, 2021, https://apnews.com/article/joe-biden-middle-east-africa-europe-corona- virus-pandemic-5e57879c6cb22d96b942cbc973b9296c.

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.

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

Davis, Jonathan M.V. , Kyle Greenberg and Damon Jones, "An Experimental Evaluation of Deferred Acceptance: Evidence from Over 100 Army Officer Labor Markets", 2023, NBER Working Paper, 31612.

Dewatripont, Matthias, Ian Jewitt, and Jean Tirole, "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies", *Review of Economic Studies*, 1999, vol. 66, issue 1, pages 199–217.

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.

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.

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.*

Foley, Duncan, "Resource Allocation and the Public Sector", 1967, *Yale Economic Essays*, 7, 45-98.

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.

Goldhill, O., '"Naively Ambitious": How COVAX Failed on its Promise to Vaccinate the World", STAT News, 2021, https://www.statnews.com/2021/10/08/how-covax-failed-on-its-promise-to-vaccinate-the- world/.

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, 6(6), 1109–1150.

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, Adam J., Christopher A. Neilson and Seth D. Zimmerman, "Heterogeneous Beliefs and School Choice Mechanisms", *American Economic Review*, 2020, vol. 110, no. 5, 1274–1315.

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.

Leshno, Jacob, "Dynamic Matching in Overloaded Waiting Lists",*American Economic Review*, 2022, vol. 112, no. 12, 3876–3910.

Li, Shengwu, "Ethics and Market Design", mimeo, 2017.

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.

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

McDermott, S, "Hundreds suspended from housing lists for refusing multiple offers", *The Journal*, 14 Apr 2019. https://www.thejournal.ie/housing-application-refusals-4587311-Apr2019.

Murphy, James, "The Future of Fair Admissions", *Education Reform Now*, 2022. https://edreformnow.org/The-Future-of-Fair-Admissions-Report-1-Early-Decision.pdf

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

Neal, Derek and Joseph Root, "The Provision of Information and Incentives in School Assignment Mechanisms", 2024, BFI Working Paper 2024-55.

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.

Mark Ottoni-Wilhelm, 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 .

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, "Bureaucratic Responses", *Journal of Labor Economics*, 2016, 34 (S2), S183-215.

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.

Radford, R.A., "The Economic Organization of a P.O.W. Camp", *Economica*, 1945, 12, 189-201.

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.

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

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

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.

Teamstage, "Volunteering Statistics for 2024", 2024, https://teamstage.io/volunteering-statistics/.

Terrier, Camille, Parag Pathak, and Kevin Ren, "From Immediate Acceptance to Deferred Acceptance: Effects on School Admissions and Achievement", 2021, NBER Working Paper 29600.

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

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.

WeConservePA, "Commission-Based Compensation for Fundraising",

library.weconservepa.org/guides/99-commission-based-compensation-for-fundraising.

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.

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.

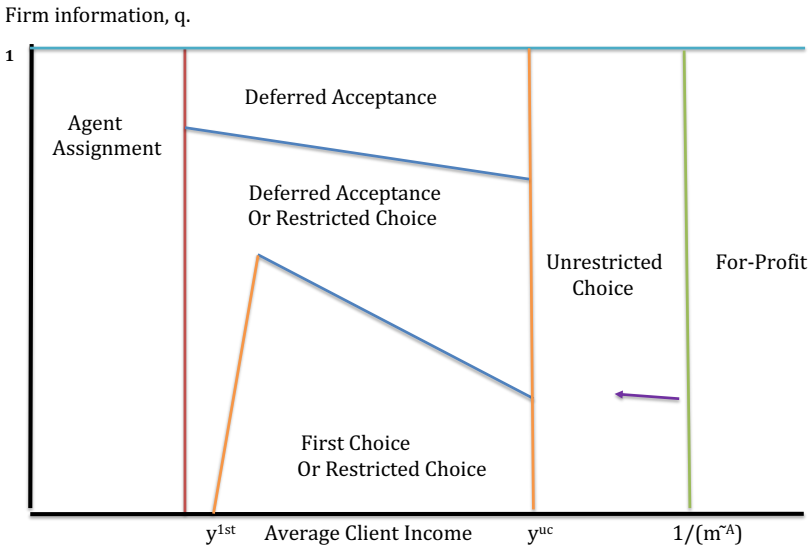


Figure 4: **The Choice of Assignment Mechanisms for Divisible Goods.**

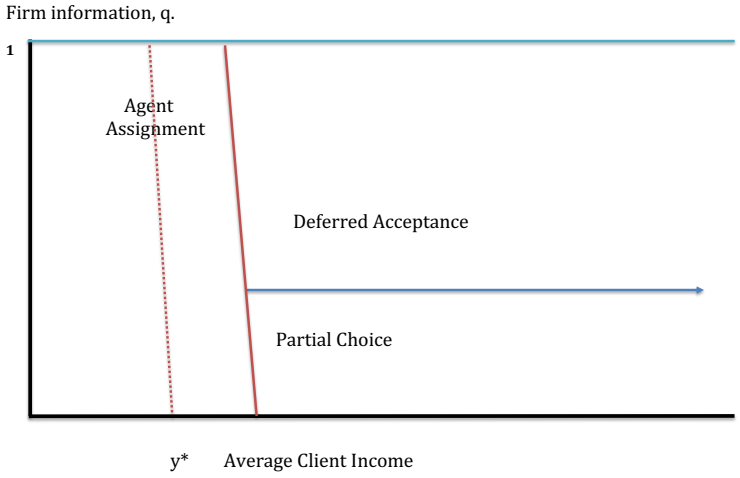


Figure 5: **The Outcome for Indivisible Goods When Clients Cannot Pay.**

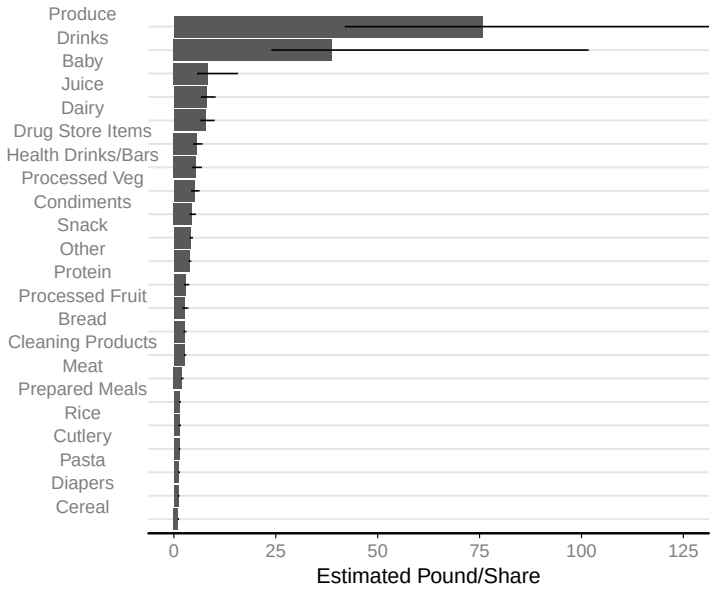


Figure 6: **Feeding America Price of Different Goods.**

7 Appendix

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

$$ES = \int_{\underline{\epsilon}}^{\overline{\epsilon}} \sum_{i=1}^{\infty} (v_i + T^A(t_i(\tau_i)) I_i(x) f(x) dx - \Upsilon I_\chi, \tag{43}$$

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

$$I^A(\epsilon) \leq \frac{\mathbb{D}(b)^A - \pi^A(\tilde{I}) + \epsilon}{w^A - r^A}, \tag{44}$$

where $\tilde{I}^A(\epsilon) = \frac{\mathbb{D}^A + \epsilon}{w^A - r^A}$. This is subject to the following constraints:

1. $E\pi^A = 0$. $\pi^A(\tilde{I}^A) = 0$ if $\Pi = 0$ and $\pi^A(\tilde{I}^A) \leq \mathbb{D}$ if $\Pi = 1$.
2. Wages are given by (6).
3. Prices r_j^A are given by $T(i, j|\underline{\phi}) \geq m r_j^A, \forall m \in [\underline{m}, \overline{m}]$, in any equilibrium where type τ_i receives treatment $t = j$ with positive probability. Revenue per client is $r^A = \frac{r_0^A + r_1^A + r_2^A}{3}$.
4. Client Reporting in A:

- If A involves Assigned Consumption, let $P_{ij}^{1st}(\hat{t}_l|\hat{t}_{-l})$ 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

$$\sum_j P_{ij}^{1st}(\hat{t}_l|\hat{t}_{-l})(T(i,j|\phi)-mr_j^{1st}) \geq \sum_j P_{ij}^{1st}(\hat{t'}_l|\hat{t}_{-l})(T(i,j|\phi)-mr_j^{1st}), \forall m \in [\underline{m}, \overline{m}], \tag{45}$$

  where P_{ij}^{1st} are generated by Bayes rule.

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

$$T(i,2|\underline{\phi}) - mr_2^A \geq T(i,j|\underline{\phi}) - mr_j^A, \forall m \in [\underline{m}, \overline{m}], \tag{46}$$

$$T(i,1|\underline{\phi}) - mr_1^A \geq T(i,0|\underline{\phi}) - mr_0^A, \forall m \in [\underline{m}, \overline{m}], \tag{47}$$

  and

$$T(i,i|\overline{\phi}) - mr_i^A \geq T(i,j|\overline{\phi}) - mr_j^A, \forall m \in [\underline{m}, \overline{m}]. \tag{48}$$

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

$$k_i^A(T(i,i|\underline{\phi}) - mr_i^A) \geq k_j^A(T(i,j|\underline{\phi}) - mr_j^A), \forall m \in [\underline{m}, \overline{m}]. \tag{49}$$

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

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

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

$$a^2 + \phi - mr_a \geq ab - mr_b, \tag{50}$$

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 the case where a client misreports "up". There are three cases:

- $\hat{t}(0,\underline{\phi},m) = 0 \succ \hat{t}(0,\underline{\phi},m) = 1$ if $\underline{\phi} - mr_0 \geq \mu - mr_1, \forall m$, or $r_1 - r_0 \geq \frac{\mu-\underline{\phi}}{\overline{m}}$, for $\frac{\underline{\phi}}{\underline{m}} \geq r_0 \geq 0$ and $r_1 \leq \frac{\mu+\underline{\phi}}{\underline{m}}$.

- $\hat{t}(0,\underline{\phi},m) = 1 \succ \hat{t}(0,\underline{\phi},m) = 2$ if $\underline{\phi} + \mu - mr_1 \geq 2\mu - mr_2, \forall m$, or $r_2 - r_1 \geq \frac{\mu-\underline{\phi}}{\overline{m}}$, for $r_1 \leq \frac{\mu+\underline{\phi}}{\underline{m}}$ and $r_2 \leq \frac{2\mu+\underline{\phi}}{\underline{m}}$.

- $\hat{t}(0,\underline{\phi},m) = 0 \succ \hat{t}(0,\underline{\phi},m) = 2$ if $\underline{\phi} - mr_0 \geq 2\mu - mr_2$ for $r_0 \geq 0$ and $r_2 \leq \frac{2\mu+\underline{\phi}}{\underline{m}}$. If the Fairness Constraint can be overcome, where all three conditions hold,

then $r_2^* = \frac{2\mu+\underline{\phi}}{\underline{m}}$, and $r_1^* = \frac{\mu+\underline{\phi}}{\underline{m}}$ as this maximizes revenue and the first two constraints. The final constraint is the most binding and can be satisfied if

$$\overline{m} r_0^* \leq 2\mu - \underline{\phi} + \overline{m}\frac{2\mu+\underline{\phi}}{\underline{m}}, \tag{51}$$

for $r_0 \geq 0$. As $r_0 \geq 0$, this implies that the Fairness Constraint can only be satisfied if

$$\frac{V'(\hat{y}+\overline{\eta})}{V'(\hat{y}+\underline{\eta})} \equiv \frac{\overline{m}}{\underline{m}} \geq \frac{2\mu-\underline{\phi}}{2\mu+\underline{\phi}}. \tag{52}$$

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 $\frac{V'(\hat{y}+\overline{\eta})}{V'(\hat{y}+\underline{\eta})} \geq \frac{2\mu-\underline{\phi}}{2\mu+\underline{\phi}}, \forall \hat{y} \geq \hat{y}^{uc}$. It also implies that if (52) holds,

$$r_0^* = \min\{\frac{\underline{\phi}-2\mu}{\overline{m}} + \frac{2\mu+\underline{\phi}}{\underline{m}}, \frac{\underline{\phi}}{\underline{m}}\}. \tag{53}$$

Straightforward calculations show that $r_0^* = \frac{\underline{\phi}}{\underline{m}}$ above a critical average income $\tilde{y}^{uc}$ where $\frac{V'(\tilde{y}^{uc}+\overline{\eta})}{V'(\tilde{y}^{uc}+\underline{\eta})} = 1 - \frac{\underline{\phi}}{2\mu}$.

- 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. □

Proof of Lemma 2: There are two cases that need to be considered. First, 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 $\hat{t}(\tau, \underline{\phi}, m) = (2 \succ 1 \succ 0)$, $\hat{t}(1, \overline{\phi}, m) = (1 \succ 2 \succ 0)$, and $\hat{t}(0, \overline{\phi}, m) = (0 \succ 2 \succ 1)$ for all m. From (46) and (47), $\hat{t}(\tau, \underline{\phi}, m) = (2 \succ 1 \succ 0)$ requires

$$\frac{\mu}{\underline{m}} - r_1 \geq \frac{\underline{\phi}}{\underline{m}} - r_0, \tag{54}$$

and

$$\frac{2\mu}{\underline{m}} - r_2 \geq \frac{\underline{\phi}+\mu}{\underline{m}} - r_1. \tag{55}$$

These hold with equality and $r_1 - r_0 = \frac{\mu-\underline{\phi}}{\underline{m}}$ and $r_2 - r_1 = \frac{\mu-\underline{\phi}}{\underline{m}}$. 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 $\tau \in 1, 2 \implies r_0 \leq 0$. Letting this hold with equality implies $R^* = \{0, \frac{\mu-\underline{\phi}}{\underline{m}}, \frac{2(\mu-\underline{\phi}))}{\underline{m}}\}$ and $\underline{r} = \frac{0+\frac{\mu-\underline{\phi}}{\underline{m}}+\frac{2(\mu-\underline{\phi}))}{\underline{m}}}{3} = \frac{\mu-\underline{\phi}}{\underline{m}}$.

Now let $1-\frac{\underline{\phi}}{\mu} < \frac{\overline{m}}{\underline{m}} < 1-\frac{\underline{\phi}}{2\mu}$. This offers the possibility of partial choice. As $\hat{t}(0,\underline{\phi},m)=2$ as shown above, the client cannot be allowed to choose $\hat{t}=2$. However, it raises the possibility that the firm could assign the client to either $t=2$ or $t\in\{0,1\}$ where the client chooses between $t=0$ and $t=1$. We show that this is impossible when $\frac{\overline{m}}{\underline{m}} < 1-\frac{\underline{\phi}}{2\mu}$. $\hat{t}(0,\underline{\phi},m)=0\forall m$ if $m(r_1-r_0)\geq \mu-\underline{\phi}$. $\hat{t}(1,\underline{\phi},m)=1\forall m$ if $m(r_1-r_0)\leq \mu+\underline{\phi}$. But as $q=1$, clients of type τ are assigned to a choice between $t=0$ and $t=1$. As a result, $r_0=0 \implies mr_1\geq \mu-\underline{\phi}$ and $mr_1\leq \mu+\underline{\phi}$. The first condition is most binding for $m=\overline{m}$ and the second for $m=\overline{m}$. Then it must be the case that $\frac{\overline{m}}{\underline{m}} \geq \frac{\mu+\underline{\phi}}{\mu-\underline{\phi}} = \frac{1}{1-\frac{2\mu}{\mu+\underline{\phi}}}$. However as we are considering the case where $1-\frac{\underline{\phi}}{\mu} < \frac{\overline{m}}{\underline{m}} < 1-\frac{\underline{\phi}}{2\mu}$, we have a contradiction, which implies that for $\frac{\overline{m}}{\underline{m}} < 1-\frac{\underline{\phi}}{2\mu}$ there is Assigned Choice and $r=\underline{r}$. $\square$

Proposition 8 *Expected surplus from Deferred Acceptance is*

$$ES^{da} = \frac{D}{\overline{U}-\underline{r}}\left(v+T^{da}-\frac{\delta D}{2(\overline{U}-\underline{r})}\right)-\frac{\delta\sigma_\epsilon^2}{2(\overline{U}-\underline{r})^2}. \tag{56}$$

where (i) if $h\leq\frac{2}{3-q}$,

$$T^{da}=\mu+\frac{1}{3}\left(\hat{\phi}(q+p_2\frac{1-q}{2})+\overline{\phi}h(q+\frac{1-q}{2})+(1-h)(\underline{\phi}q(1-p_2)+\hat{\phi}\frac{1-q}{2}p_1)\right)+$$
$$+\frac{1}{3}\left(h\overline{\phi}+q(1-p_1)(1-h)\underline{\phi}\right), \tag{57}$$

where $p_2=\frac{h(1-q)}{q(1-h)+(2-h)\frac{1-q}{2}}$ *and* $p_1=\frac{\frac{h(1-q)}{2}}{(1-h)q+(1-q)}$, *and*
(ii) if $h>\frac{2}{3-q}$,

$$T^{da}=\mu+\frac{1}{3}\left(\hat{\phi}(q+(1+\tilde{p}_2\frac{1-q}{2})+\overline{\phi}h(q+\frac{1-q}{2})+(1-h)\underline{\phi}\frac{1-q}{2}\tilde{p}_1\right)+$$
$$+\frac{1}{3}\left(h\overline{\phi}+q(1-\tilde{p}_1)(1-h)\underline{\phi}\right), \tag{58}$$

where $\tilde{p}_2=\frac{h(1-q)-(q(1-h)+(2-h)\frac{1-q}{2})}{q(1-h)+(2-h)\frac{1-q}{2}}$ *and* $\tilde{p}_1=\frac{h(1-q)}{(1-\tilde{p}_2)(q(1-h)+(2-h)\frac{1-q}{2})}$

Proof of Proposition 8 (and 1): From Lemma 1, $r=\underline{r}$. Let $\hat{\tau}^a$ be the agent's posterior. Then as $t\perp(\hat{\tau}a|\tau^a) \implies w=\overline{U}$ so $\hat{U}^a=\overline{U}^a \implies w=\overline{U}$ and (6) holds. Then $I=\frac{D+\epsilon}{\overline{U}-\underline{r}}$.

The mechanism begins with clients $i=1,2,$ ...reporting $\hat{t}$. The firm offers $t\forall i=1,2,...\frac{D+\epsilon}{\overline{U}-\underline{r}}$. At prices $R^*=\{0,\frac{\mu-\underline{\phi}}{\underline{m}},\frac{2(\mu-\underline{\phi}))}{\underline{m}}\} \implies \hat{t}(\tau,\underline{\phi},m)=(2\succ 1\succ 0)$,

$\hat{t}(2,\overline{\phi},m) = (2 \succ 1 \succ 0)$, $\hat{t}(1,\overline{\phi},m) = (1 \succ 2 \succ 0)$, and $\hat{t}(0,\overline{\phi},m) = (0 \succ 2 \succ 1)$ for all m. As a result, (46) and (47) hold. Note that

$$\hat{t} = \begin{cases} 2 \succ 1 \succ 0 & \text{with probability } 1 - \frac{2h}{3} \\ 1 \succ 2 \succ 0 & \text{with probability } \frac{h}{3} \\ 0 \succ 2 \succ 1 & \text{with probability } \frac{h}{3}. \end{cases} \tag{59}$$

1. *The firm allocates $t = 2$ to all clients for which $\tau^a = 2$ and $\hat{t} = (2 \succ 1 \succ 0)$.* Consider all who report $\hat{t} = (2 \succ 1 \succ 0)$. Of these, $\frac{1}{3}$ are $\tau = 2$ and $1 - \frac{h}{3}$ are each $\tau = 1$ and $\tau = 0$. $Prob(\tau^a = 2|\hat{t} = (2 \succ 1 \succ 0)) = 1 - h(1-q)$. 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 $\implies O(2|2,(2 \succ 1 \succ 0)) = 1$. All offers are accepted.

2. *The firm randomly allocates $t = 2$ to clients for which $\tau^a = 1$ and $\hat{t} = (2 \succ 1 \succ 0)$ until there are no more $t = 2$ slots.* There are then $h(1-q)$ remaining $t = 2$ slots available per capita, and $(q(1-h) + (2-h)\frac{1-q}{2})$ clients for whom $\hat{t} = (2 \succ 1 \succ 0)$ and $\tau^a = 1$. Note that $h(1-q) \leq (q(1-h) + (2-h)\frac{1-q}{2})$ if $h \leq \frac{2}{3-q}$. If this is the case then Step 2 $\implies O(2|2,(2 \succ 1 \succ 0)) = p_2 = \frac{h(1-q)}{q(1-h)+(2-h)\frac{1-q}{2}}$. These offers are all accepted.

3. *The firm allocates $t = 1$ to all remaining clients for which $\tau^a = 1$ and $\hat{t} = (2 \succ 1 \succ 0)$ and to all clients with $\tau^a = 1$ and $\hat{t} \neq (0 \succ 2 \succ 1)$.* There are $(1 - p_2)(q(1-h) + (2-h)\frac{1-q}{2})$ remaining clients per capita with for $\tau^a = 1$ and $\hat{t} = (2 \succ 1 \succ 0)$. There are $hq + \frac{h(1-q)}{2}$ for whom $\tau^a = 1$ or $\tau^a = 2$ and $\hat{t} = (1 \succ 2 \succ 0)$. Step 3 $\implies O(1|2,(2 \succ 1 \succ 0)) = O(1|1,(1 \succ 2 \succ 0)) = 1$. All are accepted.

4. *The firm randomly allocates $t = 1$ to clients for which $\tau^a = 0$ and $\hat{t} \neq (0 \succ 2 \succ 1)$ until there are no more $t = 1$ slots.* There are then $\frac{h(1-q)}{2}$ unfilled $t = 1$ slots, and $(1-h)q + (1-q)$ clients where $t^f = 0$ and $\hat{t} \neq (0 \succ 2 \succ 1)$. Step 4 $\implies O(1|0,(2 \succ 1 \succ 0)) = O(1|0,(1 \succ 2 \succ 0)) = p_1 = \frac{\frac{h(1-q)}{2}}{(1-h)q+(1-q)}$. All offers are accepted.

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

Expected surplus per client from treatment when $h \geq \frac{2}{3-q}$ is

$$T^{da} = \mu + \frac{1}{3}\left(\hat{\phi}(q + p_2\frac{1-q}{2}) + \overline{\phi}h(q + \frac{1-q}{2}) + (1-h)(\underline{\phi}q(1-p_2) + \hat{\phi}\frac{1-q}{2}p_1)\right) + \\ + \frac{1}{3}\left(h\overline{\phi} + q(1-p_1)(1-h)\underline{\phi}\right) \tag{60}$$

so expected surplus is

$$ES^{da} = \frac{D+\epsilon}{\overline{U}-\underline{r}}\left(v + T^{da} - \frac{\delta(D+\epsilon)}{2(\overline{U}-\underline{r})}\right), \tag{61}$$

or

$$ES^{da} = \frac{D}{\overline{U}-\underline{r}}\left(v + T^{da} - \frac{\delta D}{2(\overline{U}-\underline{r})}\right) - \frac{\delta\sigma_\epsilon^2}{2(\overline{U}-r)^2}. \tag{62}$$

This addressed the case where $h < \frac{2}{3-q}$. Now assume $h \geq \frac{2}{3-q}$. Then Step 2 $\implies$ $O(2|2,(2 \succ 1 \succ 0)) = 1$. These offers are all accepted. There are then $h(1-q) - (q(1-h) + (2-h)\frac{1-q}{2})$ remaining $t=2$ slots, and $(1-p_2)(q(1-h)+(2-h)\frac{1-q}{2})$ remaining clients per capita with for $\tau^a = 0$ and $\hat{t} = (2 \succ 1 \succ 0)$. For $h < 1$, $h(1-q) \geq 2(q(1-h)+(2-h)\frac{1-q}{2})$, so Step 2 also $\implies$ $O(2|0,(2 \succ 1 \succ 0)) = \tilde{p}_2 = \frac{h(1-q)-(q(1-h)+(2-h)\frac{1-q}{2})}{q(1-h)+(2-h)\frac{1-q}{2}}$. There are then $(1-\tilde{p}_2)(q(1-h)+(2-h)\frac{1-q}{2})$ clients for which $\hat{t} = (2 \succ 1 \succ 0))$ and $\tau^a = 0$ but who were not offered $t=2$ in the previous step. Step 3 $\implies$ $O(2|1,(1 \succ 2 \succ 0)) = 1$, so there are $h(1-q)$ unfilled $t=1$ slots. Step 4 $\implies$ $O(1|0,(2 \succ 1 \succ 0)) = \tilde{p}_1 = \frac{h(1-q)}{(1-\tilde{p}_2)(q(1-h)+(2-h)\frac{1-q}{2})}$. All offers are accepted. All other clients receive $t=0$. Expected surplus per client from treatment when $h \geq \frac{2}{3-q}$ is given by

$$\tilde{T}^{da} = \mu + \frac{1}{3}\left(\hat{\phi}(q + (1+\tilde{p}_2\frac{1-q}{2}) + \overline{\phi}h(q+\frac{1-q}{2}) + (1-h)\underline{\phi}\frac{1-q}{2}\tilde{p}_1\right) + \\ + \frac{1}{3}\left(h\overline{\phi} + q(1-\tilde{p}_1)(1-h)\underline{\phi}\right), \tag{63}$$

so

$$ES^{da} = \frac{D}{\overline{U}-\underline{r}}\left(v + \tilde{T}^{da} - \frac{\delta D(\underline{m})}{2(\overline{U}-\underline{r})}\right) - \frac{\delta\sigma_\epsilon^2}{2(\overline{U}-r)^2}.\square \tag{64}$$

Proof or Proposition 2: Consider an agent who receives a signal τ^a on a client. If his prior on that client is disparate, his posterior maximum likelihood estimate is that $\tau = \tau^a$ and implements that. If the agent holds a tight prior on the client's need, given by $\tilde{\tau}$, and implements that. These are offered at prices $R^* = \{0, \frac{\mu-\underline{\phi}}{\underline{m}}, \frac{2(\mu-\underline{\phi}))}{\underline{m}}\}$. The expected value of the treatment from the perspective of the agent is then $\hat{U}^{ca} \approx \lambda q\hat{\phi} + (1-\lambda)\hat{\phi}$. This implies that the agent's belief of surplus created is $\hat{U}^{ca} \approx \lambda q\hat{\phi} + (1-\lambda)\hat{\phi}$. The firm's belief on surplus is $\hat{U}^{ca} = \lambda q\hat{\phi} + (1-\lambda)\frac{\hat{\phi}}{3}$ as $\tilde{\tau}$ is uniformly distributed. The agent's belief of the surplus from a random alternative agent is

$$T = \begin{cases} q\hat{\phi} & \text{with probability } \lambda^2 \\ \rho\hat{\phi} & \text{with probability } (1-\lambda)^2 \\ q\hat{\phi} & \text{with probability } \lambda(1-\lambda). \\ \frac{\hat{\phi}}{3} & \text{with probability } (1-\lambda)\lambda, \end{cases} \tag{65}$$

with expected value $(1-\lambda)(1-\frac{\lambda}{3}-(1-\lambda)\rho)\hat{\phi}$ of the agent's value over his alternative (from the perspective of the agent). Then from (6), $w^* \geq \overline{U} - \alpha(1-\lambda)(1-\frac{\lambda}{3} - (1-\lambda)\rho)\hat{\phi}$. The firm maximizes I by this holding with equality (we ignore the approximation). Then $I^a = \frac{D+\epsilon}{w^*-\underline{r}}$, as $r^a = \underline{r}$ from Lemma 1. The surplus from treatment is then

$$T^a = \mu + \lambda q\hat{\phi} + (1-\lambda)\frac{\hat{\phi}}{3} < T^{da}. \tag{66}$$

and

$$ES^a = \frac{D}{w^*-\underline{r}}\left(v + T^a - \frac{\delta D}{2(w^*-\underline{r})}\right) - \frac{\delta\sigma_\epsilon^2}{2(w^*-\underline{r})^2}. \tag{67}$$

Then

$$E(I^a - I^{da}) = \Delta I^a = \frac{D\alpha(1-\lambda)(1-\frac{\lambda}{3}-(1-\lambda)\rho)\hat{\phi}}{(w^*-\underline{r})(\overline{U}-r)}, \tag{68}$$

and $ES^a - ES^{da} > 0$ iif

$$I^{da} \leq \frac{\Delta I^a(v + T^a - \frac{\delta\Delta w^*}{2})}{T^{da} - T^a + \delta\Delta I^a}, \tag{69}$$

where $\Delta w^* = \alpha(1-\lambda)(1-\frac{\lambda}{3}-(1-\lambda)\rho)\hat{\phi}$.

Note that $h = 0$ and $\delta = 0 \implies ES^{da} = \frac{D}{\overline{U}-\underline{r}}(\mu + q\hat{\phi} + v - \frac{\delta D}{2(\overline{U}-\underline{r})})$ and $ES^a = \frac{D}{\overline{U}-\underline{r}-\Delta w^*}(\mu + (\lambda q + \frac{1-\lambda}{3})\hat{\phi} + v - \frac{\delta D}{2(\overline{U}-\underline{r}-\Delta w^*)})$. Let $q' = \lambda q + \frac{1-\lambda}{3}$. Then $ES^a - ES^{da} > 0$ if

$$\Delta w^*(\mu + v + q\hat{\phi}) - (q-q')(\overline{U}-\underline{r})\hat{\phi} + \frac{\delta D}{2}\frac{(\overline{U}-\underline{r}-\Delta w^*)^2 - (\overline{U}-\underline{r})^2}{(\overline{U}-\underline{r}-\Delta w^*)(\overline{U}-\underline{r})} > 0. \tag{70}$$

□

Pareto Efficient Agent Allocations: Pareto efficiency implies that $\hat{t}(1,\overline{\phi},m) = 1 \succ 2 \succ 0$ and $\hat{t}(0,\overline{\phi},m) = 0 \succ 2 \succ 1$ with probability 1. As a result, $\hat{t} = 1 \succ 2 \succ 0 \implies t = 1$, and $\hat{t} = 0 \succ 2 \succ 1 \implies t = 0$, as the agent's posterior is 1 in each case. After allocating these clients, the agent has a fraction $1 - \frac{2h}{3}$ clients for which

$$\tau^a = \begin{cases} 2 & \text{with probability } \frac{1-\frac{2h}{3}}{3} \\ 1 & \text{with probability } \frac{1-\frac{2h}{3}}{3} \\ 0 & \text{with probability } \frac{1-\frac{2h}{3}}{3}. \end{cases} \tag{71}$$

It allocates these to $\frac{1}{3}$ slots for $t = 2$, $\frac{1-h}{3}$ slots for $t = 1$ and $\frac{1-h}{3}$ slots for $t = 0$. To maximize client welfare from the perspective of the agent, all clients with $\tau^a = i$ are allocated $t = 1$. There are $\frac{2h}{9}(1-h)$ remaining $t = 2$ slots, and $\frac{h}{9}$ clients with $\tau^a = 1$ and $\tau^a = 0$. These are matched. The expected treatment surplus per client is then

$$\mu + \left(\frac{1-\frac{2(1-h)}{3}}{3}(\lambda q + \frac{1-\lambda}{3}) + \frac{2(1-h)}{9}\frac{1-q}{2}\right)\hat{\phi} + \frac{2h}{3}\underline{\phi} + 2(\frac{1-\frac{2(1-h)}{3}}{3})q\underline{\phi}. \tag{72}$$

The increase in expected treatment surplus over Agent Assignment is $\left(\frac{2(1-h)}{9}\frac{1-q}{2}\right)\hat{\phi}+\frac{2h}{3}\overline{\phi}(1-(\lambda q+\frac{1-\lambda}{3}))+2(\frac{1-\frac{2(1-h)}{3}}{3})(q-(\lambda q+\frac{1-\lambda}{3}))\underline{\phi}>0$. However, now the reduction in wages is $\frac{1-\frac{2(1-h)}{3}}{3}\alpha(1-\lambda)(1-\frac{\lambda}{3}-(1-\lambda)\rho)\hat{\phi}$, which is lower than in Proposition 2 by $(1-\frac{1-\frac{2(1-h)}{3}}{3})\alpha(1-\lambda)(1-\frac{\lambda}{3}-(1-\lambda)\rho)\hat{\phi}$.□

Partial Choice: From Lemma 1, $\hat{t}(0,\underline{\phi},m)=0\succ\hat{t}(0,\underline{\phi},m)=2$ if $\underline{\phi}-mr_0\geq 2\mu-mr_2$ for $r_0\geq 0$ and $r_2\leq\frac{2\mu+\underline{\phi}}{\underline{m}}$. This results in condition (52). But this is also the necessary condition for Partial Choice where either type $\tau=0$ or $\tau=2$ are the only ones to choose $t=0$ or $t=2$ respectively. This arises because if (52) fails and type $\tau=0$ chooses $\hat{t}(0,\underline{\phi},m)=2$, so also does type $\tau=2$ in a DA mechanism. As a result, the necessary and sufficient condition for Unrestricted Choice is also the condition for Partial Choice for types $t=2$ and $t=0$. As allocative efficiency and revenue are higher under Unrestricted Choice, there is never Partial Choice unless it involves choice by type $\tau=1$. □

Proof of Proposition 3: Let $\hat{t}$ now refer to the good chosen by the client. *Unrestricted Choice arises when the client chooses $I_i=1$ at a price r_i for all i.* Lemma 1 generates the optimal contract, given by $R^*=\{r_0^*,\frac{\mu+\underline{\phi}}{\underline{m}},\frac{2\mu+\underline{\phi}}{\underline{m}}\}$, for all $\hat{y}\geq\hat{y}^{uc}$ and $R^{uc}=\{\frac{\underline{\phi}}{\underline{m}},\frac{\mu+\underline{\phi}}{\underline{m}},\frac{2\mu+\underline{\phi}}{\underline{m}}\}$ for all $\hat{y}\geq\tilde{y}^{uc}$. Then revenue per client from the mechanism is $r^{uc}=\frac{3\mu+r_0^*+5\underline{\phi}}{3V'(y^{uc}-\underline{\eta})}$ for $y^{uc}\leq\hat{y}\leq\tilde{y}^{uc}$ and $r^{uc}=\frac{\mu+\underline{\phi}}{V'(y^{uc}-\underline{\eta})}$ for $y^{uc}>\tilde{y}^{uc}$.

Then $T^{uc}=T^*=\mu+\underline{\phi}$, $I^{uc}=\frac{D}{\overline{U}-r^{uc}}$, and

$$ES^{uc}=\frac{D}{\overline{U}-r^{uc}}\left(v+T^*-\frac{\delta D}{2(\overline{U}-r^{uc})}\right)-\frac{\delta\sigma_\epsilon^2}{2(\overline{U}-r^{uc})^2}.\square \tag{73}$$

Proof of Proposition 4: *The firm chooses quantities of each $t_i=(1,\kappa_1^*,\kappa_2^*)$ at unit prices $(r_0^{rc},r_1^{rc},r_2^{rc})$.* 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. There are three cases:

- $\hat{t}(0,\underline{\phi},m)=0\succ\hat{t}(0,\underline{\phi},m)=1$ if $\underline{\phi}-mr_0\geq\kappa_1(\mu-mr_1),\forall m$, for $\frac{\underline{\phi}}{\underline{m}}\geq r_0\geq 0$ and $r_1\leq\frac{\mu+\underline{\phi}}{\underline{m}}$.
- $\hat{t}(0,\underline{\phi},m)=0\succ\hat{t}(0,\underline{\phi},m)=2$ if $\underline{\phi}-mr_0\geq\kappa_2(2\mu-mr_2)$ for $r_0\geq 0$ and $r_2\leq\frac{2\mu+\underline{\phi}}{\underline{m}}$.
- $\hat{t}(0,\underline{\phi},m)=1\succ\hat{t}(0,\underline{\phi},m)=2$ if $\kappa_1(\underline{\phi}+\mu-mr_1)\geq\kappa_2(2\mu-mr_2),\forall m$, for $r_1\leq\frac{\mu+\underline{\phi}}{\underline{m}}$ and $r_2\leq\frac{2\mu+\underline{\phi}}{\underline{m}}$.

If the Fairness Constraint can be overcome, where all three conditions hold, then for the same reasons as in Lemma 1, $r_2^* = \frac{2\mu+\underline{\phi}}{\underline{m}}$, and $r_1^* = \frac{\mu+\underline{\phi}}{\underline{m}}$ as this maximizes revenue and the first two constraints. For $\hat{y} < \tilde{y}^{uc}$, the constraints on r_0 bind and

$$\underline{\phi} + \overline{m} r_0^* = \kappa_1^*(\mu + \overline{m}\frac{\mu+\underline{\phi}}{\underline{m}}), \tag{74}$$

and

$$\underline{\phi} + \overline{m} r_0^* = \kappa_2^*(2\mu + \overline{m}\frac{2\mu+\underline{\phi}}{\underline{m}}), \tag{75}$$

for $r_0^* \geq 0$.

The firm then trades off higher revenue from type $\tau = 0$ with greater coverage for types $\tau = 1$ and $\tau = 2$ where $\frac{d\kappa_i^*}{dr_0^*} = \frac{\overline{m}}{i\mu+\overline{m}\frac{i\mu+\underline{\phi}}{\underline{m}}}$. The benefits and costs of reducing r_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 $r_0^* = \frac{\underline{\phi}}{\underline{m}}$ and where $r_0^* = 0$. We consider each in turn.

First, consider κ_i at prices R^{uc}. Then (49) with $k_0 = 1$ implies that $(\hat{t}|\tau = 0) = 0 \succ (\hat{t}|\tau = 0) = 2$ iff

$$(1 - \frac{m}{\underline{m}})\underline{\phi} \geq \kappa_2(2\mu - \frac{m}{\underline{m}}(2\mu + \underline{\phi})). \tag{76}$$

for all m. Then (49) implies that maximum κ_2^* is

$$\kappa_2^* = \frac{1}{\frac{2\mu}{\underline{\phi}} - \frac{\overline{m}}{\underline{m}-\overline{m}}} = \frac{\underline{\phi}}{2\mu - \frac{\overline{m}\underline{\phi}}{\underline{m}-\overline{m}}}. \tag{77}$$

There are two necessary conditions for κ_1. (49) with $k_0 = 1$ implies that $(\hat{t}|\tau = 0) = 0 \succ (\hat{t}|\tau = 0) = 1$ iff

$$(1 - \frac{m}{\underline{m}})\underline{\phi} \geq \kappa_1(\mu - \frac{m}{\underline{m}}(\mu + \underline{\phi})) \tag{78}$$

for all m. This requires that

$$\kappa_1^* \leq \frac{1}{\frac{\mu}{\underline{\phi}} - \frac{\overline{m}}{\underline{m}-\overline{m}}}. \tag{79}$$

Finally, (49) with $k_0 = 1$ implies that $(\hat{t}|\tau = 1) = 1 \succ (\hat{t}|\tau = 1) = 2$ iff

$$\kappa_1(\mu + \underline{\phi} - \frac{m}{\underline{m}}(\mu + \underline{\phi})) \geq \kappa_2(2\mu - \frac{m}{\underline{m}}(2\mu + \underline{\phi})). \tag{80}$$

for all m. This implies that

$$\kappa_1^* \geq \frac{\kappa_2(2\mu - \frac{m}{\underline{m}}(2\mu + \underline{\phi}))}{(\mu + \underline{\phi} - \frac{m}{\underline{m}}(\mu + \underline{\phi}))}. \tag{81}$$

The two requirements for κ_1 hold for $\kappa_1 \in [\frac{\kappa_2(2\mu-\frac{m}{\underline{m}}(2\mu+\underline{\phi}))}{(\mu+\underline{\phi}-\frac{m}{\underline{m}}(\mu+\underline{\phi}))}, \frac{1}{\frac{\mu}{\underline{\phi}}-\frac{\overline{m}}{\underline{m}-\overline{m}}}]$ if $\frac{1}{\frac{\mu}{\underline{\phi}}-\frac{\overline{m}}{\underline{m}-\overline{m}}} > \frac{\kappa_2(2\mu-\frac{m}{\underline{m}}(2\mu+\underline{\phi}))}{(\mu+\underline{\phi}-\frac{m}{\underline{m}}(\mu+\underline{\phi}))}$. This always holds as $\kappa_1(\mu+\underline{\phi}-\frac{m}{\underline{m}}(\mu+\underline{\phi})) \geq \kappa_2(2\mu-\frac{m}{\underline{m}}(2\mu+\underline{\phi})$ when evaluated at κ_2^* and κ_1^* in (79). Hence κ_i^* are given by (77) and (79). As v_i is decreasing in i, the firm has (77) and (79) hold with equality.

Now consider the Deferred Acceptance component. The mechanism begins with clients $i = I^{rc}+1, I^{rc}+2, ...I^{rc}(1+\frac{1-\kappa_1^*}{3}+\frac{1-\kappa_2^*}{3})$ reporting $\hat{t}$, where they only rank $t=1$ and $t=2$. The firm chooses prices $\tilde{R}^{da} = \{\infty, \frac{\mu}{\underline{m}}, \frac{2\mu}{\underline{m}}\}$. At prices $\tilde{R}^{da}$ (46) and (47) hold so $\hat{t}(\tau, \underline{\phi}, m) = (2 \succ 1)$, $\hat{t}(2, \overline{\phi}, m) = (2 \succ 1)$, $\hat{t}(1, \overline{\phi}, m) = (1 \succ 2)$, and $\hat{t}(0, \overline{\phi}, m) = (2 \succ 1)$ for all m. Note that

$$\hat{t} = \begin{cases} 2 \succ 1 & \text{with probability } 1-\frac{h}{3} \\ 1 \succ 2 & \text{with probability } \frac{h}{3} \end{cases} \tag{82}$$

There are $I^{rc}(\frac{1-\kappa_2^*}{3})$ available $t=2$ slots and $I^{rc}(1-\frac{h}{3})(1+\frac{1-\kappa_1^*}{3}+\frac{1-\kappa_2^*}{3})$ clients for whom $\hat{t} = 2 \succ 1$. Of those who have $\hat{t} = 2 \succ 1$, there are $I^{rc}(1+\frac{1-\kappa_1^*}{3}+\frac{1-\kappa_2^*}{3})(q+(2-h)\frac{1-q}{2})$ for whom $\tau^a = 2$. There are two cases. First, if $(q+(2-h)\frac{1-q}{2})(1+\frac{1-\kappa_1^*}{3}+\frac{1-\kappa_2^*}{3}) \geq (\frac{1-\kappa_2^*}{3})$, then the firm allocates $t=2$ to $I^{rc}\frac{1-\kappa_2^*}{3}$ clients for which $\tau^a = 2$ and $\hat{t} = (2 \succ 1)$ and $t=1$ randomly to all other clients. Let $\hat{p}_1 = \frac{\frac{1-\kappa_2^*}{3}}{(q+(2-h)\frac{1-q}{2})(1+\frac{1-\kappa_1^*}{3}+\frac{1-\kappa_2^*}{3})}$. This has expected per capita treatment surplus of

$$\begin{aligned}\tilde{T}^{da} = \mu &+ (\frac{1-\kappa_2^*}{2-\kappa_1^*-\kappa_2^*})(\frac{q}{q+(2-h)\frac{1-q}{2}})\hat{\phi} \\ &+ (\frac{1-\kappa_1^*}{2-\kappa_1^*-\kappa_2^*})(\hat{p}_1(\frac{1-q}{2}(1-h)\underline{\phi}) + (1-\hat{p}_1)(h\overline{\phi}+(1-h)q\underline{\phi}))\end{aligned} \tag{83}$$

Second, if $(q+(2-h)\frac{1-q}{2})(1+\frac{1-\kappa_1^*}{3}+\frac{1-\kappa_2^*}{3}) < (\frac{1-\kappa_2^*}{3})$, then the firm allocates $t=2$ to all clients for which $\tau^a = 2$. It offers $t=2$ randomly to $I^{rc}(\frac{1-\kappa_2^*}{3} - (q+(2-h)\frac{1-q}{2})(1+\frac{1-\kappa_1^*}{3}+\frac{1-\kappa_2^*}{3}))$ clients for whom $\hat{t} = 2 \succ 1$ and $\tau^a = 1$. The probability that a such a client receives an offer of $t=2$ is $\hat{p}_2 = \frac{\frac{1-\kappa_2^*}{3} - (q+(2-h)\frac{1-q}{2})(1+\frac{1-\kappa_1^*}{3}+\frac{1-\kappa_2^*}{3})}{(1-h)q+\frac{1-q}{2}(2-h)} < 1$. This has expected per capita treatment surplus of

$$\begin{aligned}\tilde{T}^{da} = \mu &+ (\frac{1-\kappa_2^*}{2-\kappa_1^*-\kappa_2^*})[\hat{p}_2(\frac{q}{q+(2-h)\frac{1-q}{2}})\hat{\phi} + (1-\hat{p}_2)\frac{1-q}{2}\hat{\phi}] \\ &+ (\frac{1-\kappa_1^*}{2-\kappa_1^*-\kappa_2^*})(h\overline{\phi}+(1-h)q\underline{\phi}).\end{aligned} \tag{84}$$

This has expected per capita treatment surplus of

$$\tilde{T}^{da} = \mu + (\frac{1-\kappa_2^*}{2-\kappa_1^*-\kappa_2^*})(\frac{q}{q+(2-h)\frac{1-q}{2}})\hat{\phi}$$
$$+(\frac{1-\kappa_1^*}{2-\kappa_1^*-\kappa_2^*})(h\overline{\phi}+(1-h)q\underline{\phi})+(1-h)\frac{1-q}{2}\underline{\phi}\left(1-\frac{\frac{1-\kappa_2^*}{3}}{(q+(2-h)\frac{1-q}{2})(1+\frac{1-\kappa_1^*}{3}+\frac{1-\kappa_2^*}{3})}\right). \tag{85}$$

Proof of Proposition 5: Let $r^{1st} = \frac{r_1+2(\mu-\underline{\phi})}{3\underline{m}}$ be revenue per client. Finally, note that $t \perp (\tau^a|\tau^a) \implies w = \overline{U}$ so $\hat{U}^a = \overline{U}^a \implies w = \overline{U}$. As a result, $I = \frac{D+\epsilon}{\overline{U}-r^{1st}}$. The firm receives $\hat{t}$ for all $i = 1, 2, ...\frac{D+\epsilon}{\overline{U}-r^{1st}}$.

We address the existence of an equilibrium where the firm raises the maximum possible revenue where $\hat{t}(1,\overline{\phi},m) = \hat{t}(1,\underline{\phi},m) = 1 \succ 2 \succ 0$, $\hat{t}(2,\overline{\phi},m) = \hat{t}(2,\underline{\phi},m) = \hat{t}(0,\underline{\phi},m) = 2 \succ 1 \succ 0$ and $\hat{t}(0,\overline{\phi},m) = 0 \succ 2 \succ 1$ for all m. We do this in two steps, by first assuming such an equilibrium and then showing necessary conditions for its existence.

1. *Each client reports $\hat{t}$.* First assume that there exists a Bayes Nash equilibrium exists where $\hat{t}(1,\overline{\phi},m) = \hat{t}(1,\underline{\phi},m) = 1 \succ 2 \succ 0$, $\hat{t}(2,\overline{\phi},m) = \hat{t}(2,\underline{\phi},m) = \hat{t}(0,\underline{\phi},m) = 2 \succ 1 \succ 0$ and $\hat{t}(0,\overline{\phi},m) = 0 \succ 2 \succ 1$ for all m. Then

$$\hat{t} = \begin{cases} 2 \succ 1 \succ 0 & \text{with probability } \frac{2-h}{3} \\ 1 \succ 2 \succ 0 & \text{with probability } \frac{1}{3} \\ 0 \succ 2 \succ 1 & \text{with probability } \frac{h}{3}. \end{cases} \tag{86}$$

2. *The firm allocates $t = i$ to all clients k for which $\hat{t}^k = i \succ ...$ if i is undersubscribed.* A fraction $\frac{2-h}{3}$ have $\hat{t} = 2 \succ 1 \succ 0$, $\frac{1}{3}$ have $\hat{t} = 1 \succ 2 \succ 0$ and $\frac{h}{3}$ have $\hat{t} = 0 \succ 2 \succ 1$. Then only $\hat{t} = 2 \succ 1 \succ 0$ is oversubscribed. Step 1 $\implies O(\tau^a, 1 \succ 2 \succ 0) = O(\tau^a, 0 \succ 2 \succ 1) = 1$.

3. *For all $t = i'$ that are oversubscribed, the firm implements a Deferred Acceptance Algorithm among those where $\hat{t}^k = i' \succ ...$* By step 2, a Deferred Acceptance mechanism is used for all clients where $\hat{t} = 2 \succ 1 \succ 0$. Among those, per client there are $\frac{q+\frac{1-q}{2}(1-h)}{3}$ where $\tau^a = 2$. As $q + \frac{1-q}{2}(1-h) < 1$, step 1 of the DA algorithm $\implies O(2|2, 2 \succ 1 \succ 0) = 1$. There are $\frac{\frac{1-q}{2}(1+h)}{3}$ remaining $t = 2$ slots per client, and $\frac{(\frac{1-q}{2})(2-h)}{3}$ clients where $\tau^a = 1$. $\frac{\frac{1-q}{2}(1+h)}{3} < \frac{(\frac{1-q}{2})(2-h)}{3}$ if $h < \frac{1}{2}$. If $h \leq \frac{1}{2}$, step 2 of the DA algorithm $\implies O(2|1, 2 \succ 1 \succ 0) = p_2^*$, where $p_2^* = \frac{1+h}{2-h}$. Under step 4 of the DA algorithm all other clients receive $t = 0$. The

surplus from treatments under First Choice with $h \leq \frac{1}{2}$ is given by

$$T^{1st} = \frac{\mu + h\overline{\phi} + \hat{\phi} + (q + \frac{1-q}{2}(1-p_2^*)(1-h))\underline{\phi} + \hat{\phi}(q + \frac{1-q}{2}p_2^*)}{3}. \tag{87}$$

The surplus from treatments under First Choice $h \leq \frac{1}{2}$ is given by

$$T^{1st} = \frac{\mu + h\overline{\phi} + \hat{\phi} + (q(1-h)(1-p_1^*) + \frac{1-q}{2})\underline{\phi} + \hat{\phi}(q + \frac{1-q}{2})}{3}. \tag{88}$$

As a result,

$$ES^{1st} = \frac{D}{\overline{U} - r^{1st}}\left(v + T^{1st} - \frac{\delta D}{2(\overline{U} - r^{1st})}\right) - \frac{\delta\sigma_\epsilon^2}{2(\overline{U}^* - r^{1st})^2} - \Upsilon \tag{89}$$

When $h < \frac{1}{2}$, $T^{1st} - T^{da} = h\overline{\phi}(1-p_1)\frac{1-q}{2} + \hat{\phi}(p_2^* - p_2) + (1-h)\underline{\phi}(1 - \frac{1-q}{2}p_1) > 0$ as $q \geq \frac{1}{3}$.

Existence: We check incentive compatibility for types $\tau = \{1, 2, 3\}$ with $\underline{\phi}$ for all m. To do so, begin by considering the case where $r_0 = 0, r_1 = \frac{\mu - \underline{\phi}}{\overline{m}}$ and $r_2 = \frac{2(\mu - \underline{\phi})}{\overline{m}}$.

1. Type $\tau = 1$: (45) implies that $\hat{t}(1, \underline{\phi}, m) = 1 \succ 2 \succ 0$ if

$$(\frac{1-q}{2} + qp_2^*)(2\mu - mr_2) \leq \mu + \underline{\phi} - mr_1, \tag{90}$$

as $P_{12}^{1st}(2 \succ 1 \succ 0) = (\frac{1-q}{2} + qp_2^*)$ and $P_{11}^{1st}(1 \succ 2 \succ 0) = 1$. There are then two cases:

- When $(\frac{1-q}{2} + qp_2^*)(2\mu - \overline{m}\frac{2(\mu-\underline{\phi})}{\underline{m}}r_2) \leq \mu + \underline{\phi} - \overline{m}\frac{\mu-\underline{\phi}}{\underline{m}}$. In this instance, clients of all incomes with $\tau = 1$ and $\underline{m}$ prefer to report $\hat{t}(1, \underline{\phi}, m) = 1 \succ 2 \succ 0$. Then as the cost of increasing r_1 is increasing in m, the firm sets $r_1 = \frac{\mu + \underline{\phi} - 2\underline{\phi}(\frac{1-q}{2} + qp_2^*)}{\underline{m}}$, as this maximizes revenue and relaxes the other IC constraints. Rearranging, this implies that the firm can charge $r_1 = \frac{\mu + \underline{\phi} - 2\underline{\phi}(\frac{1-q}{2} + qp_2^*)}{\underline{m}}$ when

$$\frac{\overline{m}}{\underline{m}} \geq \frac{(\frac{1-q}{2} + qp_2^*)(2\mu + \underline{\phi}) - (\underline{\phi} - \mu)}{2(\mu - \underline{\phi})(\frac{1-q}{2} + qp_2^*) - (\underline{\phi} + \mu)}. \tag{91}$$

  Let $\tilde{y}^{1st}$ be the average income level at which (91) holds with equality.

- When $(\frac{1-q}{2}+qp_2^*)(2\mu-\overline{m}\frac{2(\mu-\underline{\phi})}{\overline{m}}r_2) \geq \mu+\underline{\phi}-\overline{m}\frac{\mu-\underline{\phi}}{\overline{\phi}}$. In this instance, clients of all incomes with $\tau=1$ and $\underline{m}$ prefer to report $\hat{t}(1,\underline{\phi},m)=2\succ 1\succ 0$. In this instance, r_1 must be reduced below $\frac{\mu-\underline{\phi}}{\overline{\phi}}$. The highest charge for $t=1$ is here given by $r_1=\frac{m}{\overline{m}}\left(\underline{\phi}+\mu-(2\mu+\underline{\phi}-\frac{\overline{m}}{\underline{m}}(2(\mu-\underline{\phi})))(\frac{1-q}{2}+qp_2^*)\right)$. If (91) fails, the firm must reduce prices for the first price equilibrium to arise. We describe the equilibrium in this case below. If

$$\frac{\overline{m}}{\underline{m}} \geq \frac{(\frac{1-q}{2}+qp_2^*)(2\mu+\underline{\phi})-(\underline{\phi}-\mu)}{2(\mu-\underline{\phi})(\frac{1-q}{2}+qp_2^*)} \tag{92}$$

  the required price is negative which cannot arise and an equilibrium cannot exist.

2. Type $\tau=0$: A price of r_1 above deters type 0 from reporting $\hat{t}(0,\underline{\phi},m)=1\succ 2\succ 0$ instead of $\hat{t}(0,\underline{\phi},m)=2\succ 1\succ 0$ if

$$\frac{1-q}{2}(1+p_2^*)(2\mu-mr_2)+(1-\frac{1-q}{2}(1+p_2^*))\underline{\phi} \geq \mu-m(\frac{\mu+\underline{\phi}-2\underline{\phi}(\frac{1-q}{2}+qp_2^*)}{\underline{m}}), \tag{93}$$

as $P_{02}^{1st}(2\succ 1\succ 0)=\frac{1-q}{2}(1+p_2^*)$, $P_{00}^{1st}(2\succ 1\succ 0)=(1-\frac{1-q}{2}(1+p_2^*))$ and $P_{01}^{1st}(1\succ 2\succ 0)=1$. As $r_2=\frac{2(\mu-\underline{\phi})}{\underline{m}}$, this simplifies to

$$\frac{1-q}{2}(1+p_2^*)(2\mu-m\frac{2(\mu-\underline{\phi})}{\underline{m}})+(1-\frac{1-q}{2}(1+p_2^*))\underline{\phi} \geq \mu-m(\frac{\mu+\underline{\phi}-2\underline{\phi}(\frac{1-q}{2}+qp_2^*)}{\underline{m}}). \tag{94}$$

This condition holds for $m=\underline{m}$. It must hold for all income levels, and $\hat{t}(0,\underline{\phi},m)=1\succ 2\succ 0$ for all m if

$$\underline{\phi}+\frac{1-q}{2}(1+p_2^*)(2\mu-\underline{\phi})-\mu \geq \frac{m}{\overline{m}}\left(\mu-3\underline{\phi}+2\underline{\phi}\frac{1-q}{2}(1+p_2^*)\right). \tag{95}$$

This implies that $\hat{t}(0,\underline{\phi},m)=1\succ 2\succ 0$ if $\hat{y}\geq \hat{y}^{1st}$, where

$$\frac{V'(\hat{y}^{1st}+\overline{\eta})}{V'(\hat{y}^{1st}+\underline{\eta})}=\frac{\underline{\phi}+\frac{1-q}{2}(1+p_2^*)(2\mu-\underline{\phi})-\mu}{\frac{1-q}{2}(1+p_2^*)2(\mu-\underline{\phi})-\mu-\underline{\phi}+2\underline{\phi}(\frac{1-q}{2}+qp_2^*)}. \tag{96}$$

Note that $V'\geq 0$ so $\hat{t}(0,\underline{\phi},m)=1\succ 2\succ 0$ for any m if $\frac{1-q}{2}(1+p_2^*)2(\mu-\underline{\phi})-\mu-\underline{\phi}+2\underline{\phi}(\frac{1-q}{2}+qp_2^*)<0$.

3. Type $\tau=2$:
   (45) implies that $\hat{t}(2,\underline{\phi},m)=2\succ 1\succ 0$ iff

$$(q+\frac{1-q}{2}p_2^*)(2\mu+\underline{\phi}-mr_2)\geq \mu-mr_1, \tag{97}$$

as $P_{22}^{1st}(2 \succ 1 \succ 0) = (q + \frac{1-q}{2}p_2^*)$ and $P_{21}^{1st}(1 \succ 2 \succ 0) = 1$. As $r_2 > r_1$, this conditions binds most for $m = \underline{m}$. At $m = \underline{m}$, this is $(q + \frac{1-q}{2}p_2^*)3\underline{\phi} \geq \underline{\phi}$ which always holds.

Combining these three outcomes implies that an equilibrium exists and raises revenue of $r^{1st} = \frac{\mu+\underline{\phi}-2\underline{\phi}(\frac{1-q}{2}+qp_2^*)+2(\mu-\underline{\phi})}{3\underline{m}} > \underline{r}$ if $\hat{y} \geq y^{1st} = max\{\tilde{y}^{1st}, \hat{y}^{1st}\}$.

Consider the case above where $(\frac{1-q}{2} + qp_2^*)(2\mu - \overline{m}\frac{2(\mu-\underline{\phi})}{\overline{m}}r_2) \geq \mu + \underline{\phi} - \overline{m}\frac{\mu-\underline{\phi}}{\overline{\phi}}$. The highest charge for $t = 1$ is here given by $\tilde{r}_1 = \frac{\underline{m}}{\overline{m}}\left(\underline{\phi} + \mu - (2\mu + \underline{\phi} - \frac{\overline{m}}{\underline{m}}(2(\mu - \underline{\phi})))(\frac{1-q}{2} + qp_2^*)\right) < \frac{\mu-\underline{\phi}}{\overline{m}}$. Then revenue per client is $\tilde{r}^{1st} = \frac{\tilde{r}_1+2(\mu-\underline{\phi})}{3\underline{m}} < \underline{r}$. Then the expected surplus is given by

$$ES^{1st} = \frac{D}{\overline{U} - \tilde{r}^{1st}}\left(v + T^{1st} - \frac{\delta D}{2(\overline{U} - \tilde{r}^{1st})}\right) - \frac{\delta\sigma_\epsilon^2}{2(\overline{U}^* - \tilde{r}^{1st})^2}, \tag{98}$$

which cannot be signed relative to ES^{da} as the firm must trade lower coverage against higher allocative efficiency.

Proposition 9 *Let $h > \frac{1}{2}$. Then a First Choice equilibrium exists if $\hat{y} \geq Y^{1st}$ where (26) holds but*

$$T^{1st} = \frac{\mu + h\overline{\phi} + \hat{\phi} + (q(1-h)(1-p_1^*) + \frac{1-q}{2})\underline{\phi} + \hat{\phi}(q + \frac{1-q}{2})}{3}, \tag{99}$$

where $p_1^* = \frac{2h-1}{(1-h)q+\frac{1-q}{2}}$.

Proof: Steps 1 to 3 in Proposition 11 apply in exactly the same way. Then as $h > \frac{1}{2}$, step 2 of the DA algorithm $\implies O(2|1, 2 \succ 1 \succ 0) = 1$. There are then $\frac{\frac{1-q}{2}(2h-1)}{3}$ remaining $t = 2$ slots per client. The firm has already matched the h clients who choose $\hat{t} = (0 \succ 2 \succ 1)$. There are $\frac{1-h}{3}$ remaining $t = 0$ slots and $\frac{(1-h)q+\frac{1-q}{2}}{3}$ signals of $\tau^a = 0$ per worker. As $\frac{\frac{1-q}{2}(2h-1)}{3}$ receive $t = 1$, step 3 of the DA algorithm $\implies O(2|0, 2 \succ 1 \succ 0) = p_1^*$ where $p_1^* = \frac{2h-1}{(1-h)q+\frac{1-q}{2}}$.

In this instance, the IC constraint for type $\tau = 1$ is given by

$$(\frac{1-q}{2}(1 + p_1^*) + q)(2\mu - mr_2) \leq \mu + \underline{\phi} - mr_1, \tag{100}$$

which with $r_2 = 2(\mu - \underline{\phi})$ implies that the maximum possible revenue is now $r_1 = \frac{\mu+\underline{\phi}-2\underline{\phi}(\frac{1-q}{2}(1+p_1^*)+q}{\underline{m}}$. Funding per client is now given by $r^{1st} = \frac{\mu+\underline{\phi}-2\underline{\phi}(\frac{1-q}{2}(1+p_1^*)+q)+2(\mu-\underline{\phi})}{3\underline{m}} > \underline{r}$. Analogous conditions to the Proof of Proposition 5 show required Y^{1st} for existence. □

Proof of Proposition 7: For any assignment mechanism A,

$$S^* = \int_{\underline{\epsilon}}^{\overline{\epsilon}} \frac{(\mathbb{D}(b) - \pi^A(x) + x)}{w^A - r^A}(v + T^A - \frac{\delta(\mathbb{D}(b) - \pi^A(x) + x)}{2(w^A - r^A)})f(x)dx, \tag{101}$$

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 $\hat{\pi}^A = \int_{\underline{\epsilon}}^{\overline{\epsilon}} \pi^{*\prime A}(\tilde{I}(x))f(x)dx$. As $E\pi = 0$ this simplifies to

$$S^* = \frac{\mathbb{D}(b)}{w^A - r^A}(v + T^A - \frac{\delta\mathbb{D}(b)}{2(w^A - r^A)}) - \int_{\underline{\epsilon}}^{\overline{\epsilon}} \frac{\delta(-\pi(x) + x)^2}{w^A - r^A} f(x)dx. \tag{102}$$

This is maximized at $\pi^*(\epsilon) = \epsilon$ or $\pi^{*\prime} = 1$ with equilibrium surplus

$$S^* = \frac{\mathbb{D}(b)}{w^A - r^A}(v + T^A - \frac{\delta\mathbb{D}(b)}{2(w^A - r^A)}). \tag{103}$$

As $\hat{\pi}^A = \int_{\underline{\epsilon}}^{\overline{\epsilon}} \pi^{*\prime A}(\tilde{I}(x))f(x)dx = 1$, this implies that $d = 0$. □

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.

Agency Contracting when all revenue can be paid out: Here the firm pays itself all revenue as a Bonus if I high enough: $\pi(\overline{I}) = d$ for enough high I.

$$\pi = \begin{cases} \mathbb{D}(b) - \overline{\pi} & \text{if } I \geq \hat{I} \\ -\overline{\pi} & \text{if } I < \hat{I} \end{cases} \tag{104}$$

where $\overline{\pi} = \int_{I^*}^{\overline{I}} d(\underline{m}) f(x) dx$ to ensure that $E\pi = 0$. Then $\hat{I}$ is set such that

$$f_e(\hat{I})\mathbb{D}(b) = \\ \mathbb{D}(b)(v + T^A)\left(\frac{w^A + C(e^*) - r^A e^* - e^* C'(e^*)}{(w^A + C(e^*) - r^A e^*)^2} - \frac{w^A + C(\tilde{e}) - r^A \tilde{e} - \tilde{e} C'(\tilde{e})}{(w^A + C(\tilde{e}) - r^A \tilde{e})^2}\right). \tag{105}$$

Here $\mathbb{D}(b) = \mathbb{D}(b)F(\hat{I})$. $\square$

Extensions: We now consider a series of extensions described in the text. For brevity reasons, we only consider the case where $h < \frac{2}{3-q}$ in each case.

The First Choice Mechanism when some clients cannot strategize: Assume that a proportion χ of all clients report their preferred allocation in the sense of maximizing $T(t_j, \tau_i) - r_j$, independent of the actions of other clients, while the remaining $1 - \chi$ can strategize in the usual Bayesian way.

Proposition 11 *Assume a fraction χ of clients cannot strategize. A separating equilibrium exists for q low enough. If it exists, the expected surplus from First Choice Allocation is*

$$ES^{1st} = \frac{D}{\overline{U} - \underline{r}}\left(v + T^{1st} - \frac{\delta D}{2(\overline{U} - \underline{r})}\right) - \frac{\delta \sigma_\epsilon^2}{2(\overline{U}^* - \underline{r})^2}. \tag{106}$$

where

$$T^{1st} = \mu + \frac{1}{3}\left(2h\overline{\phi} + (q + \frac{1-q}{2}p_2^*)\underline{\phi}\right) + \\ \frac{1}{3}\left(((1-h)(1-\chi) + \chi(1-h)qp_1^*)\underline{\phi} + \underline{\phi}(q + \frac{1-q}{2}(1-p_1^*))\right), \tag{107}$$

$p_2^* = \frac{(\frac{1-q}{2})(2-(1+\chi)(1-h))}{(\frac{1-q}{2})(2-h)+\chi(1-h)q)}$ *and* $p_1^* = \frac{\chi(1-h))}{\frac{1-q}{2}(1-2h+\chi(1-h))+\chi(1-h)q)}$.

When a Fixed Fee can be charged to clients: Assume that the firm chooses prices $\hat{R} = \{\hat{r}, \hat{r}_0, \hat{r}_1, \hat{r}_2\}$. Assume further that clients understand the equilibrium being played. The choice of the fixed fee $\hat{r}$ is given in Lemma 3, where $U^A(0, \underline{m})$ is the expected utility of a client with type $\tau = 0$ and income $\underline{m}$ under assignment mechanism A.

Lemma 3: Assume that a fixed fee can be charged. For any Assigned Consumption allocation with uniform pricing and $q < 1$,

$$\hat{R} = \{\hat{r}, 0, \frac{\underline{\phi} + \mu}{\underline{m}}, \frac{\underline{\phi} + 2\mu}{\underline{m}}\}, \tag{108}$$

where $\hat{r}^a = \frac{U^a(0,\underline{m}) - E(r_i|\tau=0)}{\underline{m}}$.

Now consider the outcome of Deferred Acceptance:

Proposition 12 *Assume that a fixed fee can be charged. The expected surplus from Deferred Acceptance is*

$$ES^{da} = \frac{D}{\overline{U} - r^{da}} \left(v + T^{da} - \frac{\delta D}{2(\overline{U} - r^{da})} \right) - \frac{\delta \sigma_\epsilon^2}{2(\overline{U} - r^{da})^2}. \tag{109}$$

where

$$T^{da} = \mu + \frac{1}{3} \left(\hat{\phi}(q + p_2 \frac{1-q}{2}) + \overline{\phi} h (q + \frac{1-q}{2}) + (1-h)(\underline{\phi} q (1-p_2) + \hat{\phi} \frac{1-q}{2} p_1) \right) + \frac{1}{3} \left(h\overline{\phi} + q(1-p_1)(1-h)\underline{\phi} \right), \tag{110}$$

$p_2 = \frac{h(1-q)}{q(1-h) + (2-h)\frac{1-q}{2}}$, $p_1 = \frac{\frac{h(1-q)}{2}}{(1-h)q + (1-q)}$ *and* $r^{da} = \frac{\mu + \underline{\phi}(1-q)(1+p_2)}{\overline{m}}$.

Then consider the outcome with Agent Allocation:

Proposition 13 *Assume a fixed fee can be charged. The expected surplus from Agent Allocation is*

$$ES^{da} = \frac{D}{w^* - r^a} \left(v + T^a - \frac{\delta D}{2(w^* - r^a)} \right) - \frac{\delta \sigma_\epsilon^2}{2(w^* - r^a)^2}. \tag{111}$$

where

$$T^a = \mu + \hat{\phi}(q\rho + \frac{(1-q)(1-\rho)}{2}), \tag{112}$$

$w^* = \overline{U} - \alpha(1-\rho)(3q-1)\hat{\phi}$ *and* $r^a = \frac{\mu + \phi(1 - (q\rho + \frac{(1-q)(1-\rho)}{2}))}{\underline{m}}$.

Deferred Acceptance with the Agent's Signal: Let $\tilde{q} = (q\rho + \frac{(1-q)(1-\rho)}{2})$.

Proposition 14 *The expected surplus from Deferred Acceptance using the agent's signal is* $ES^{da} = \frac{D}{\tilde{w} - \underline{r}} \left(v + T^{da} - \frac{\delta D}{2(\tilde{w} - \underline{r})} \right) - \frac{\delta \sigma_\epsilon^2}{2(\tilde{w} - \underline{r})^2}$, *where*

$$T^{da} = \mu + \frac{1}{3} \left(\hat{\phi}(\tilde{q} + \tilde{p}_2 \frac{1-\tilde{q}}{2}) + \overline{\phi} h (\tilde{q} + \frac{1-\tilde{q}}{2}) + (1-h)(\underline{\phi} \tilde{q} (1-\tilde{p}_2) + \hat{\phi} \frac{1-\tilde{q}}{2} \tilde{p}_1) \right) + + \frac{1}{3} \left(h\overline{\phi} + \tilde{q}(1-\tilde{p}_1)(1-h)\underline{\phi} \right), \tag{113}$$

$\tilde{p}_2 = \frac{h(1-\tilde{q})}{\tilde{q}(1-h)+(2-h)\frac{1-\tilde{q}}{2}}$, $\tilde{p}_1 = \frac{\frac{h(1-\tilde{q})}{2}}{(1-h)\tilde{q}+(1-\tilde{q})}$, *and*

$$\tilde{w} = \overline{U} - \frac{\alpha(3q-1)(1-\rho)}{2}\left(\frac{\hat{\phi}(1-\tilde{p}_2)+\frac{h\overline{\phi}}{2}-(1-h)\underline{\phi}\tilde{p}_2-\frac{\hat{\phi}\tilde{p}_1}{2}+(1-\tilde{p}_1)(1-h)\underline{\phi}}{3}\right) > w^*. \tag{114}$$

The treatment surplus arises by running the same algorithm as in Proposition 1 but with the agent's signal quality $\tilde{q}$. From the agent's perspective, the quality of her signal over that of his replacement is $\frac{\alpha(3q-1)(1-\rho)}{2}$. Differentiating T^{da} above with respect to $\tilde{q}$ and multiplying by $\frac{\alpha(3q-1)(1-\rho)}{2}$ yields the wage discount and hence $\tilde{w}$. The marginal effect of q on w^* is $\hat{\phi}$. But

$$\hat{\phi} - \frac{1}{3}\left(\hat{\phi}(\tilde{q}+\tilde{p}_2\frac{1-\tilde{q}}{2})+\overline{\phi}h(\tilde{q}+\frac{1-\tilde{q}}{2})+(1-h)(\underline{\phi}\tilde{q}(1-\tilde{p}_2)+\hat{\phi}\frac{1-\tilde{q}}{2}\tilde{p}_1)\right) = \frac{\hat{\phi}\tilde{p}_2+h\overline{\phi}+(1-h)\underline{\phi}\tilde{p}_1}{6}+\frac{(1-h)\overline{\phi}(1-\frac{\tilde{p}_2}{2})+(1-h)\underline{\phi}\tilde{p}_2}{3} > 0. \tag{115}$$

As a result, the discount on wages is greater for Agent Allocation and $\tilde{w} > w^*$.

An Alternative Restricted Choice Mechanism: In the text, Deferred Acceptance is used to allocate services when $t = 0$ is depleted. Here we consider an alternative, where Choice is used throughout, which offers similar conclusions:

1. The firm chooses quantities $\kappa_1^*, \kappa_2^*, \hat{\kappa}_2$.
2. The firm chooses unit prices $R^{uc} = (r_0^{rc}, r_1^{rc}, r_2^{rc}, \hat{r}_1, \hat{r}_2, \tilde{r}_2)$
3. Each client $i = 1, 2, ..N_0$ chooses t_i, $i \in \{0, 1, 2\}$. At N_0, there is no more remaining supply of $t = 0$.
4. The firm chooses quantities of $I_1 = 1, I_2 = \hat{\kappa}_2$ and unit prices $(\hat{r}_1, \hat{r}_2)$.
5. Each client $i = N_0 + 1, ..., N_1$ chooses t_i, $i \in \{1, 2\}$ until there is no more remaining supply of $t = 1$.
6. The firm allocates $I_2 = 1$ to all clients at price $\tilde{r}_2$ until all $t = 2$ is exhausted.

Proposition 15 *Let $\hat{y} < y^{uc}$ and $I^{rc} = \frac{D}{\overline{U}-r^{rc}}$. Assume that $R = R^*$. Surplus from Restricted Choice is*

$$S^{rc} = I^{rc}\left(v + T^{rc} - \frac{\delta I^{rc}}{2} - \frac{5\delta(1-\kappa_1^*)}{8} - \delta[\frac{5}{8}+\frac{3}{8}\kappa_1^* - \kappa_2^*]\right) - \frac{\delta\sigma_\epsilon^2}{2(\overline{U}-r^{rc})^2}. \tag{116}$$

where

$$T^{rc} = T^* - \frac{2\hat{\phi}(1-\kappa_1^*)}{5} - \frac{2\hat{\phi}[\frac{5}{8}+\frac{3}{8}\kappa_1^*-\kappa_2^*]}{3}, \tag{117}$$

and

$$r^{rc} = \frac{r^* - \frac{2-(\kappa_1+\kappa_2)}{3}}{\underline{m}}. \tag{118}$$

Proof of Proposition 15:

1. *The firm chooses quantities of each* $t_i = (1, \kappa_1^*, \kappa_2^*)$ *at unit prices* $(r_0^{rc}, r_1^{rc}, r_2^{rc})$.. This is exactly as in Proposition 4.

2. *Each client* $i = N^*+1, ...\hat{N}$ *chooses* t_i, $i = 1, 2$ *until there is no more remaining supply of* $t = 1$. This continues until $t = 1$ is exhausted, which occurs at $I^{rc}(1-\kappa_1^*)\frac{3}{2}$. For $i = I^{rc}+1, ...I^{rc}(1-\kappa_1^*)\frac{3}{2}$, the expected value per client is

$$T_1 = \mu + \frac{3\hat{\phi}}{5}. \tag{119}$$

3. *The firm allocates* $t = 2$ *to all clients at price* $\tilde{r}_2$ *until all* $t = 2$ *is exhausted.* At client $i = I^{rc}(1-\kappa_1^*)\frac{3}{2}$, the firm's supply of $t = 0$ and $t = 1$ is exhausted. At that point, there remains $1 - \kappa_2^* - \frac{3}{8}(1-\kappa_1^*)$ supply of $t = 2$. For $i = I^{rc}(1-\kappa_1^*)\frac{3}{2}+1, ..., I^{rc}[(1-\kappa_1^*)\frac{3}{2}+1-\kappa_2^*-\frac{3}{8}(1-\kappa_1^*)$, the firm offers only $t = 2$ at a price of $\frac{2\mu}{\underline{m}}$, the minimum utility received from that treatment. The expected value per client in this region is $T_3 = 2\mu + \frac{\hat{\phi}}{3}$.

Weighting these prices and surpluses yields that the outcome is $r^{rc} = \frac{T^* - \frac{2-(\kappa_1+\kappa_2)}{3}}{\underline{m}}$, so that $I^{rc} = \frac{D(\underline{m})}{\overline{U}-r^{rc}}$, and

$$S^{rc} = I^{rc}\left(v + T^{rc} - \frac{\delta I^{rc}}{2} - \frac{5\delta(1-\kappa_1^*)}{8} - \delta[\frac{5}{8}+\frac{3}{8}\kappa_1^*-\kappa_2^*]\right). \tag{120}$$

where

$$T^{rc} = T^* - \frac{2\hat{\phi}(1-\kappa_1^*)}{5} - \frac{2\hat{\phi}[\frac{5}{8}+\frac{3}{8}\kappa_1^*-\kappa_2^*]}{3}. \tag{121}$$

Now consider the case where $R^{rc} = \{0, \frac{\mu+\underline{\phi}}{\underline{m}}, \frac{2\mu+\underline{\phi}}{\underline{m}}\}$. Then

$$\kappa_2^* = \frac{\underline{\phi}}{2\mu - \frac{\overline{m}}{\underline{m}}(\mu+\underline{\phi})}. \tag{122}$$

and

$$\kappa_1^* = \frac{\underline{\phi}}{\mu - \frac{\overline{m}}{\underline{m}}(\mu+\underline{\phi})}. \tag{123}$$

Funding is lower than above, given by

$$r^{rc} = \frac{T^* - \frac{\phi}{3} - \frac{2-(\kappa_1+\kappa_2)}{3}}{\underline{m}}. \tag{124}$$

□