Maximizing the Effect of Altruism*

Nicholas Wu † Nicole Immorlica‡ Brendan Lucier §

July 31, 2023

Abstract

We study the problem faced by an altruistic donor who wants to allocate a budget to influence an agent's action. The agent has a private type that determines their (convex and increasing) costs, and their action is publicly observed. We propose a mechanism design-based model that incorporates two new features. Firstly, the altruist's utility primarily comes from the agent's action. Secondly, the altruist is endowed with a limited budget. We show that the optimal mechanism takes the form of a subsidy schedule, where the altruist offers payment depending on the action of the agent. The optimal schedule induces the same action from sufficiently efficient types, and withholds funds from less efficient agents. We also show that with a separability condition on the agent's cost functions, the threshold for sufficiently efficient agents is invariant to the size of the budget.

*We would like to thank Dirk Bergemann, Laura Doval, Johannes Hörner, Kai Hao Yang, Jidong Zhou, Tan Gan, Yingkai Li, and participants at the Yale microeconomic theory lunch seminar and breakfasts for helpful comments and discussions.

†Department of Economics, Yale University, New Haven, CT 06511, nick.wu@yale.edu

‡Microsoft Research

§Microsoft Research

1 Introduction

With the rise of government, corporate and individual social responsibility, nontrivial sums of money are donated to address environmental and social issues. Individuals and corporations seeking to address environmental issues by offsetting carbon emissions will offer money to private agents such as biochar projects, conservation efforts, and nascent carbon removal technologies. Others donate to their alma maters, seeking to advance research and education, or to poverty alleviation organizations. Governments subsidized production of socially beneficial technologies such as the COVID-19 vaccine, provide school or housing vouchers to the poor, or give developers incentives to provide low-income housing.The amount of money given away is nontrivial; the top 25 individual givers, according to Forbes, have given over $169 billion to charitable causes over their lifetimes.[1]

In particular cases however, market limitations can cloud the impact of donations. Consider the example of a philanthropist seeking to combat climate change by removing carbon dioxide. The current state of technology for carbon dioxide removal is young and premature, and carbon credit markets do not operate at the impact scale the philanthropist would like. More broadly, subsidies issued to profit-maximizing firms and are often deemed necessary to spur growth in these nascent industries.[2] Given a budget, how does one make an impact when a market does not yet exist?

We model this problem as a strategic interaction between an altruist and a private firm with private information about its own operations. This induces a principal-agent problem. This scenario differs from the standard economic delegation literature in two important ways. Firstly, the altruist is budget-constrained; that is, the altruist cannot spend more than an upper bound regardless of the private information of the firm. Secondly, the altruist has a primary concern in the change in quantity induced, and the altruist has no value herself for the budget. The altruist herself only cares about her impact through increasing the production of the firm. Although we formulate the model with the altruism motivation in mind, the model is sufficiently general to also characterize other related design problems; we provide additional examples in Section 2.

In this paper, we extend the mechanism design and delegation literature by proposing a model for budget-constrained investment into an industry under a primary concern for output, and characterize the solution. We start with the single-firm case, where the altruist

[1] See https://www.forbes.com/sites/forbeswealthteam/2022/01/19/americas-top-givers-2022-the-25-most-philanthropic-billionaires/

[2] See https://www.politico.com/news/agenda/2021/12/22/carbon-removal-advance-market-commitments-525988

only cares about the output of one firm in particular. We show that the optimal solution is implementable as a subsidy schedule, where the altruist compensates the firm based on the observed change in production. We show that the schedule partitions the agent types into more and less efficient types, offering the entire budget for types that meet a threshold and discriminating on less efficient types. We also find that the principal never excludes a positive measure of types.

Intuitively, the primary concern for output induces a fully *endogenous* value of money for the principal. The principal's schedule splitting the agent types can be understood as selecting some threshold type and splitting the agent types. All the more efficient agents are pooled; they are required to satisfy a single output threshold and given the entire budget. The more efficient agent types determine the shadow value of money; the less efficient agents face a schedule that optimizes with respect to this shadow value. Intuitively, the threshold trades off a welfare loss from withholding the budget from the less efficient agents with information rents it must pay the more efficient types. That is, an incorrect intuition for the "pooling" region for the most efficient types is that the budget is insufficient to separate the types. However, regardless of the size of the budget, the threshold never degenerates to the lower bound, and so a positive measure of the most efficient types are always pooled. To illustrate this point starkly, we consider agent cost functions that satisfy a separability condition. We show that under this separability condition, the shape of the schedule is invariant to the size of the budget. That is, under the separability assumption, no matter how large the budget is, the threshold for the pooling region never changes.

We also consider some extensions and illustrate some of the limitations of our model. We first extend to allow a competitive "outside option" for the altruist; that is, the unused residual budget has a linear value. We show that provided the value of the budget is not high, the same pooling region of the optimal schedule persists. We also show that this region is sensitive to the nature of the budget constraint; such a pooling feature does not appear if the budget is only required ex-ante. We finally extend to a broader class allowing nonlinear, convex budget cost, and provide a mathematical characterization.

The paper is structured as follows. In the next subsection, we preview the related literature in mechanism design and delegation and distinguish this particular situation from similar work. Section 2 details the model and assumption. Section 3 formalizes the general mechanism design problem and highlights the important features of the model. Section 4 restricts attention to a specific subclass of cost functions, and provides stronger characterizations. Section 5 discusses variants of the model, and Section 6 concludes.

1.1 Literature

Our model is related to the literature on contract design in settings with moral hazard and private actions going back to Grossman and Hart (1983) and Holmström (1979). In this line of work, a principal contracts with an agent to take a costly action that is not directly observable, but that results in an observable outcome that is payoff-relevant to the principal and upon which payments can be contingent. In contrast to this line of work, in our model the principal can observe the action (i.e., production) and the uncertainty is over the cost function of the agent. Our work also introduces a constraint on the maximum transfer.

Additionally, our work is related to literature on monopoly regulation. Closely related work to ours is that of Baron and Myerson (1982), which studies a social planner problem for regulating a monopolist, and Laffont and Tirole (1986), which introduces noisy cost observations. However, our work is distinct in multiple ways. Firstly, our environment considers a budget-constrained principal, whereas there are no constraints on transfers in the other models in the literature. Additionally, the only objective for the principal in our scenario is to maximize production, where as the principal in the monopoly regulation environment typically maximizes a weighted sum of consumer surplus and producer profits. There are other minor distinctions in our model; we allow for a broad class of cost functions for the agent, whereas the literature typically assumes particular functional forms.

Given that our work involves nonstandard value for transfers, we complement the literature on delegation. The original delegation framework was formulated to Holmstrom (1980), which introduced the general class of delegation problems and provides conditions for existence of a solution. Delegation models typically assume that contingent transfers are unavailable to the principal, and instead the principal retains the right to take a state-dependent action that impacts utilities of both the principal and agent. Holmstrom (1980) characterizes optimal delegation sets for a series of examples assuming that the delegation sets take an interval form; later delegation literature generalized the interval assumption Alonso and Matouschek (2008), added money burning Amador and Bagwell (2013); Amador et al. (2018), and added costly verification Halac and Yared (2020). In contrast to much of the delegation literature, our model does allow for contingent transfers between the principal and agent, but in a special way; the transfers do not factor into the principal objective and face a budget constraint.

Given that the altruist has no intrinsic value for the budget herself, our problem relates to the literature on budget-feasible mechanism design (see Singer (2014) for a survey). The seminal work in that space studied a procurement problem where the designer wishes to

purchase a utility-maximizing set items from sellers subject to a budget constraint Singer (2010). The main result shows that the designer can achieve a bounded approximation to the first-best so long as the utility function is submodular. Recent work in this area has investigated approximation ratios under alternate assumptions on the utility function Amanatidis et al. (2019); Gravin et al. (2020); Leonardi et al. (2017). Our work contributes to this literature in two significant ways. First, the altruist is maximizing production from a single firm rather than maximizing a utility function over a combinatorial set of items. Second, the firm does not induce a fixed cost for providing a good, but rather selects revenue-optimal production levels in a downstream market. Thirdly, other papers typically impose budget constraints on the agent side, whereas the constraints in our environment are on the principal's side.

Our paper is topically related to the large literature on mechanism design for social good Abebe and Goldner (2018). This literature seeks to use mechanism design to solve social problems. In particular, our primary motivation of an altruist funding carbon removal technologies relates to the environmental economics literature (see Greaker and Popp (2022) for a survey). Papers on carbon reduction specifically address a range of related issues, including optimal cap and trade policies for firms with uncertain production costs Goldner et al. (2019), the design of carbon credits in the presence of an uncontrolled downstream market Babaioff et al. (2023), and optimal contract design for a budget-constrained altruist attempting to maximize afforestation Dai Li et al. (2022).

One interpretation of our model focusing on the impact of markets on an altruist objective is echoed in Kang (2022). In that paper, the social welfare optimizer – a policymaker – determines the quality and allocation of a good that she produces with an eye towards redistribution. She must contend with the fact that her production of this good impacts the market for inputs to the production process and hence the price (and demand) for the privately-provided goods in the downstream market. This can create distortions that run against the policymaker's objective. In contrast, our work focuses on a funder, rather than an additional producer, of goods, and assumes these goods are produced subject to a type-dependent production cost function and then traded in a downstream market.

2 Model

Consider a principal-agent design problem where the principal (she) would like to provide incentives for an agent (he) to take a costly action. The action space is $X = \mathbb{R}_+$; an action

$x \in X$ can be interpreted as an amount of a good produced by the agent. Without loss, we normalize the space X so that the principal derives utility x from the agent taking action x. The agent has a private cost function $\Psi : X \times \Theta \to \mathbb{R}_+$, which determines the cost the agent incurs from producing $x \in X$ given the agent's type $\theta \in \Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}_+$. Without loss, we normalize the cost function $\Psi(0, \theta) = 0$ for all θ.

The agent's type is drawn prior to the start of the game by Nature from a full-support distribution $\mu \in \Delta(\Theta)$. The type is privately observed only by the agent, but the principal has belief μ over the agent type. We assume the belief μ admits a cumulative density function F with a differentiable mass function f.

The principal has a budget T of a transferable resource that the agent values. The agent's utility is quasilinear in the cost and the transfer: that is, if the agent of type θ produces x and receives a transfer t, his utility is $u_A(x, \theta) = t - \Psi(x, \theta)$. To be clear, we suppose the principal can observe the action x, as opposed to the moral hazard problem in Holmström (1979) and Grossman and Hart (1983); instead, the principal must decide how to award funds given the observed action, without precisely knowing the agent's cost *function*.

To highlight the key economic forces, we will present the baseline model where the principal has no value for the transferable resource. In Section 5 we consider extensions where the principal does value the resource.

Assumption 1 *We assume the cost function Ψ is C^2, supermodular, and strictly increasing in both arguments. Additionally, we assume that Ψ and its partial derivative in θ, Ψ_θ, are both convex in x.*

A couple of remarks about the assumptions and game structure are in order. The supermodularity assumption is indespensible, and intuitively requires that higher types suffer steeper costs. The C^2 differentiability assumption is technical and allows us to use cross-derivatives. The convexity assumptions ensure validity and sufficiency of first-order optimization conditions. The interpretation of the convexity condition in Ψ_θ is that the *convexity* of the agent's cost function is increasing in the type. For instance, in an example where an investor offers an investment budget to a firm, the assumption requires that the higher θ firms are less scalable (i.e. their production costs grow faster). In this setting, the type θ can be interpreted as an inefficiency parameter.

By standard arguments, we invoke the revelation principle, so we consider mechanisms where the agent reports his type to the principal subject to incentive compatibility and individual rationality constraints.

More precisely, the mechanism design problem requires the principal to select functions $x : \Theta \to X$ and $t : \Theta \to \mathbb{R}_+$, in the following optimization problem:

$$\begin{aligned}
\max_{x,t} \quad & \int_{\theta \in \Theta} x(\theta) f(\theta)\, d\theta && && (1)\\
\text{subject to} \quad & t(\theta) - \Psi(x(\theta), \theta) \geq t(\theta') - \Psi(x(\theta'), \theta) && \forall \theta, \theta' \in \Theta && \text{(IC)}\\
& t(\theta) - \Psi(x(\theta), \theta) \geq 0 && \forall \theta \in \Theta && \text{(IR)}\\
& t(\theta) \leq T && \forall \theta \in \Theta && \text{(B)}
\end{aligned}$$

To understand the problem, we also provide some examples of the model outside of our motivating example.

Example (Non-monetary incentives) Consider an academic institution interested in increasing its research output by hiring a new researcher. The institution can grant the researcher a non-monetary award (title, prize, tenure) which has no intrinsic value to the institution itself. The researcher values the award at T, and has a private cost of producing output $\Psi(\cdot, \theta)$. The institution can assess the output of the researcher and decide to give the award or not (or some randomization). The mechanism in this example corresponds to the institution committing to a schedule that maps research output into a probability of receiving the award.

Example (Collective action) Consider a policymaker trying to induce a population to take costly action, like reducing electricity consumption. Individuals in the population have private types $\theta \sim F$ which determine how severely the action reduces their utility; that is, $\Psi(x, \theta)$ is how costly a household of type θ finds it to reduce its electricity consumption by x. The policymaker cannot pay any individual agent more than an upper bound T, but would like to induce the largest possible collective action from the population. We show later that the optimal mechanism can be implemented by a subsidy schedule, which determines how much compensation an individual receives as a function of their electricity consumption reduction.

3 General Optimal Mechanism

In this section, we first present the mathematical characterization of the optimal mechanism, including the existence and uniqueness of the optimal solution. We then consider the case

when Ψ is separable in x and θ, and present a stronger characterization.

Solving our mechanism relies on a lemma which characterizes the feasible schedules, or the set of allocations $x : \Theta \to \mathbb{R}_+$ for which there exists a transfer schedule $t : \Theta \to [0, T]$ that satisfies (IC), (IR), and (B). The lemma follows the standard mechanism design techniques to derive the feasible set.

Lemma 1 *$x : \Theta \to X$ is a feasible schedule if and only if the following two conditions hold:*

1. *x is nonincreasing.*
2. *$\Psi(x(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi_\theta(x(s), s)\ ds \leq T$*

Further, given a feasible schedule x, a transfer function that supports the schedule is given by

$$t(\theta) = \Psi(x(\theta), \theta) + \int_{\theta}^{\bar{\theta}} \Psi_\theta(x(s), s)\ ds \tag{2}$$

Condition (1) is a monotonicity constraint, and condition (2) constrains the magnitude of x. Note that the normalization constraint (2) comes from the transfer bound, the IC constraints, and the envelope theorem, and so the feasible set characterization does not depend on the prior. Note the first part of the Lemma implies that these two conditions are necessary and sufficient for a feasible schedule, and the second part explicitly determines the transfer schedule that implements any feasible schedule.

The proof of Lemma 1 is standard; we use supermodularity to show that x must be nonincreasing, and then apply the envelope theorem to the IC constraints to derive the transfer schedule. The transfer schedule in Lemma 1 is analogous to the Myersonian transfer schedule, since the transfer schedule is also derived from using the envelope theorem to integrate out interim utility of the firm from the incentive compatibility conditions.

The integral expression in the transfer expression is the information rent that type θ receives; since $\Psi(x^*(\theta), \theta)$ is exactly the loss the firm experiences from producing $x^*(\theta)$, the transfer compensates the firm for the loss and provides additional rent, which is larger for more efficient types (lower θ).

As a corollary, since Ψ and Ψ_θ are convex, conditions (1) and (2) jointly determine a strictly convex subset of the space of functions from $\Theta \to \mathbb{R}_+$. Thus, as a corollary of our lemma, since the mechanism design problem maximizes a linear functional over a strictly convex set.

Corollary 1 *If a solution exists, the solution to the generalized mechanism design problem in (1) is unique.*

With some more work, we can show the existence of a solution and characterize it with a differential equation system.

Theorem 1 *An optimal mechanism (x, t) for (1) exists and is unique. The optimal mechanism, together with a nonnegative Lagrange multiplier λ and a nonnegative, absolutely continuous costate function ρ, jointly solve the following system of differential equations with boundary constraints:*

$$\rho(\theta) > 0 \implies \dot{x}(\theta) = 0 \tag{3}$$

$$\dot{\rho}(\theta) = \lambda \Psi_{x\theta}(x(\theta), \theta) - f(\theta) \tag{4}$$

$$\rho(\bar{\theta}) = 0 \tag{5}$$

$$\rho(\underline{\theta}) = \lambda \Psi_x(x(\underline{\theta}), \underline{\theta}) \tag{6}$$

$$t(\underline{\theta}) = \Psi(x(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi_\theta(x(s), s)\ ds = T \tag{7}$$

where (4) holds wherever ρ is differentiable.

We first interpret the conditions. The costate ρ denotes the shadow cost of the monotonicity constraint. The first condition (3) is the complementary slackness condition; wherever the constraint's shadow cost ρ^* is positive, the constraint is binding (i.e. x is constant). The second condition (4) dictates the costate evolution. The constraints (5) and (6) establish boundary conditions. The final constraint (7) establishes that the budget binds at the most efficient / least costly type.

The proof intuition is as follows; we write out the Lagrangian of the problem using the feasible set characterization from Lemma 1. We then formulate the optimization as an optimal control problem, where the control variable governs the derivative of x and is constrained to be nonpositive wherever x is differentiable. Corollary 1 implies uniqueness, and so we have to show existence. We first fix a Lagrange multiplier λ, and then invoke Pontryagin on the Lagrangian dual problem. We show that for any λ there exists a unique solution to the differential equations system given by (3)-(6). Finally, we show that there exists a unique Lagrange multiplier value such that the unique solution to (3)-(6) satisfies the budget binding condition (7).

To see how the characterization in Theorem 1 pins down mechanism, suppose λ were exogenously fixed. Consider the illustration in Figure 1. Complementary slackness implies that when ρ is positive, x must be constant, and whenever x is decreasing, ρ must be constant at zero, and hence x must follow the trajectory implied by the implicit equation derived from

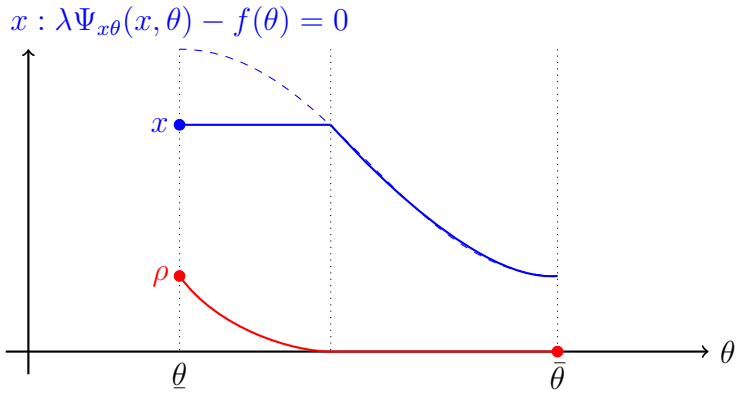


Figure 1: Illustration of the conditions in Theorem 1. The dotted blue line satisfies the implicit equation. When ρ is zero, x must coincide with the dotted blue line.

$\dot{\rho} = 0$. The boundary conditions dictate the start and end values of the costate ρ. Note that altering λ either uniformly increases or decreases x; intuitively, at the solution value of the Lagrange multiplier, the budget constraint exactly binds.

A remark on the convexity assumption on Ψ_θ is in order. The result in Theorem 1 relies on the Pontryagin maximum principle providing a unique solution; this occurs given that the maximized Hamiltonian is concave. In the model, the Arrow sufficiency condition is that for all θ,

$$xf(\theta) - \lambda\Psi_\theta(x,\theta)$$

is concave in x. Assuming Ψ_θ is convex in x is equivalent to this sufficiency condition. [3]

As a corollary, we can note that Theorem 1 implies that the mechanism is implementable by a subsidy schedule or action-contingent transfers.

Corollary 2 *Suppose $\theta < \theta'$, and let x^*, t^* be the optimal mechanism. If $x^*(\theta) = x^*(\theta')$, then $t^*(\theta) = t^*(\theta')$.*

Proof. By Lemma 1, since x^* must be nonincreasing and $x^*(\theta') = x^*(\theta)$, x^* must be constant

[3]There are other sufficiency conditions for Pontryagin's maximum principle with weaker assumptions on the Hamiltonian. Relaxing assumptions on Ψ would require another sufficiency argument for the uniqueness of a solution characterized by the maximum principle.

on (θ, θ'). As a minor abuse of notation, let x^* be that value on (θ, θ'). Then

$$\begin{aligned} t^*(\theta) &= \Psi(x^*, \theta) + \int_{\theta}^{\bar{\theta}} \Psi_\theta(x^*(t), t)\ dt \\ &= \Psi(x^*, \theta) + \int_{\theta}^{\theta'} \Psi_\theta(x^*, t)\ dt + \int_{\theta'}^{\bar{\theta}} \Psi_\theta(x^*(t), t)\ dt \\ &= \Psi_\theta(x^*, \theta') + \int_{\theta'}^{\bar{\theta}} \Psi_\theta(x^*(t), t)\ dt = t^*(\theta') \end{aligned}$$

■

3.1 Full Award

We first discuss when the principal awards the maximum T to the agent (i.e., for what types $t(\theta) = T$). From the budget binding constraint, we know that the most efficient type $\underline{\theta}$ always receives the full transfer. The following result shows that a positive measure of types always receive the entire transfer.

Proposition 1 *There exists a $\hat{\theta} > \underline{\theta}$ such that the optimal mechanism x^* is constant on $[\underline{\theta}, \hat{\theta}]$. The transfer t^* is equal to T for these types.*

Proof. This is a direct consequence of Theorem 1. Since $\rho(\underline{\theta}) > 0$, complementary slackness (3) implies that x^* must be constant until ρ reaches zero. Corollary 2 implies that the optimal t^* is also constant on this set of types, and the budget binding condition (7) implies that the transfer is T. ■

To intuitively understand why this pooling region always occurs, consider the loss in principal utility that occurs in the optimal mechanism relative to perfect information. With perfect information, the principal always gives away the entirety of the budget T and obtains the maximum action x from the agent such that $\Psi(x, \theta) = T$. The optimal mechanism, relative to perfect information, differs in two ways. Either the principal cannot obtain the highest action because the principal deliberately withholds funds, or the principal loses utility to information rent. Hence, the trade-off between the relative loss from these two effects dictates the size of the pooling region of the most efficient types.

Proposition 2 *Suppose the optimal x^* is not constant on the entire interval of types. Let $\underline{x} \equiv x^*(\underline{\theta})$. Then the largest type $\hat{\theta}$ that receives the full transfer satisfies the following:*

$$\frac{\Psi_{x\theta}(\underline{x}, \hat{\theta})}{\Psi_x(\underline{x}, \hat{\theta})} = \frac{f(\hat{\theta})}{F(\hat{\theta})} \tag{8}$$

Proof. By optimality, there exists some Lagrange multiplier and costate function λ, ρ that satisfy the conditions of Theorem 1. At the cutoff, since x^* is not constant on the entire interval, it must be that x^* stops being constant at $\hat{\theta}$. Complementary slackness (3) implies that $\rho(\hat{\theta}) = 0$. Integrating out the costate evolution (4) gives

$$\lambda \Psi_x(\underline{x}, \hat{\theta}) = F(\hat{\theta})$$

Since we assumed f was differentiable, x^* is continuous at $\hat{\theta}$ but the right derivative must be negative; hence, $\dot{\rho}(\hat{\theta}) = 0$ since complementary slackness implies ρ must be constant at zero for some interval above $\hat{\theta}$. Using the costate evolution (4) and taking the upper limit $\theta \to \hat{\theta}$, we get

$$\lambda \Psi_{x\theta}(\underline{x}, \hat{\theta}) = f(\hat{\theta})$$

Substituting out λ gives (10). ■

To interpret this condition, rewrite (10) as

$$\Psi_x(\underline{x}, \hat{\theta}) = \Psi_{x\theta}(\underline{x}, \hat{\theta}) \frac{F(\hat{\theta})}{f(\hat{\theta})}$$

Under this rewriting, the condition is equivalent to the marginal benefit from paying the threshold type to be equal to the marginal information rent cost demanded.

It is possible that the principal always awards the entire T to the agent, regardless of the type. The following result characterizes when this happens. Let $\bar{x}$ be such that $T = \Psi(\bar{x}, \bar{\theta})$ (that is, the highest action that least efficient type $\bar{\theta}$ would be willing to take if given the entire budget T). Then

Proposition 3 *The optimal mechanism pools all types (that is, the optimal x is constant) if and only if for all θ,*

$$\frac{\Psi_x(\bar{x}, \theta)}{\Psi_x(\bar{x}, \bar{\theta})} \geq F(\theta).$$

Further, if the optimal mechanism pools all types, then $x(\theta) = \bar{x}$ and $t(\theta) = T$ for all θ.

Proof. First, suppose that $x^*(\theta) = x^*$ is constant. Let λ, ρ be the Lagrange multiplier and

costate variable satisfying Theorem 1. By the budget binding condition (7),

$$\begin{aligned} T &= \Psi(x^*, \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi_\theta(x^*, s)\ ds \\ &= \Psi(x^*, \underline{\theta}) + \Psi(x^*, \bar{\theta}) - \Psi(x^*, \underline{\theta}) \\ &= \Psi(x^*, \bar{\theta}) \end{aligned}$$

Hence, $\bar{x} = x^*$. By the upper bound condition (5), integrating the costate evolution equation (4), we must have

$$\rho(\bar{\theta}) = \lambda \Psi_x(\bar{x}, \bar{\theta}) - 1 = 0$$

So $\lambda = 1/\Psi_x(\bar{x}, \bar{\theta})$. Then since ρ satisfies the costate evolution condition (4),

$$\rho(\theta) = \lambda \Psi_x(\bar{x}, \theta) - F(\theta) \geq 0$$

and so the result follows.

For the converse, suppose that $\bar{\lambda}\Psi_x(\bar{x}, \theta) \geq F(\theta)$ for all θ. Then construct $\rho(\theta) = \lambda\Psi_x(\bar{x}, \theta) - F(\theta)$, and define $x^*(\theta) = \bar{x}$ and $\lambda = 1/\Psi_x(\bar{x}, \bar{\theta})$. It is easy to check that these satisfy the condition of Theorem 1. Hence the optimal mechanism is $x(\theta) = \bar{x}$ and $t(\theta) = T$ for all θ. ■

3.2 No Exclusion

Since the resource that can be used to incentivize the agent is constrained between upper and lower bounds, and the previous result showed that the upper bound binds on a positive measure of types (i.e. the ex-ante probability of the optimal mechanism granting the full award T is positive), it is natural to ask whether the mechanism excludes types; i.e., whether a positive measure of types is asked to take the lowest action 0. The following result shows that this can never be the case:

Proposition 4 (Non-exclusion) *The optimal mechanism does not exclude a positive measure of types: $x(\theta) > 0\ \forall \theta < \bar{\theta}$.*

Proof. First, take a $\theta < \bar{\theta}$. Suppose, for sake of contradiction, that $x(\theta) = 0$. By monotonicity, this implies $x(s) = 0$ for all s in $(\theta, \bar{\theta})$. This implies that $\dot{x} = 0$ on the interval $(\theta, \bar{\theta})$, and so by Theorem 1, the costate evolution condition together with the boundary condition

$\rho(\bar{\theta}) = 0$ implies that:

$$\begin{aligned}\rho(\theta) &= -\int_{\theta}^{\bar{\theta}} \left(f(s) - \lambda\Psi(x(s), s)\right)\, ds \\ &= -\int_{\theta}^{\bar{\theta}} \left(f(s)\right)\, ds \\ &= -(1 - F(\theta)) < 0\end{aligned}$$

which is a contradiction, since ρ must be nonnegative. ■

This shows that no positive measure of types takes the lowest action, and so IR implies that no positive measure of types receives zero transfer, or is excluded. This is in contrast to Proposition 1; the upper bound on t always binds at a positive measure of types, and the lower bound never binds for a positive measure of types.

3.3 Split Population and Shadow Cost

Having established that the mechanism always admits a pooling region up to some $\hat{\theta}$, we provide some economic intuition for the optimal x given $\hat{\theta}$. That is, throughout this section, suppose the optimal mechanism has a cutoff type $\hat{\theta} < \bar{\theta}$ determined by Proposition 2. We establish two results

First, define $\bar{x} \equiv x(\theta)$ for $\theta \in [\underline{\theta}, \hat{\theta}]$. The following result holds:

Proposition 5 *The highest action $\bar{x}$ solves the following maximization:*

$$\bar{x} = \arg\max_{x} \left\{ xF(\hat{\theta}) - \lambda\Psi(x, \hat{\theta}) \right\}.$$

Proof. Integrating out the costate evolution from Theorem 1, we find that

$$\lambda\Psi_x(\bar{x}, \hat{\theta}) - F(\hat{\theta}) = 0$$

which is exactly the FOC for the maximization. The FOC exactly characterizes the solution since Ψ is convex in x. ■

Proposition 5 characterizes the tradeoff for the highest action $\bar{x}$; intuitively, by increasing the highest action, the principal gets more production from $F(\hat{\theta})$ types of agents, but pays the shadow cost $\lambda\Psi(\cdot, \hat{\theta})$. As mentioned before, the value of the Lagrange multiplier λ can be interpreted as the shadow value of the resource to the principal.

Now, we show that the types from $(\hat{\theta}, \bar{\theta}]$ intuitively face a *different* shadow cost to the

principal. This insight helps provide us an intuition for what the optimal mechanism looks like for types $\theta > \hat{\theta}$.

Proposition 6 *Suppose $\Psi_{x\theta}(x,\theta)/f(\theta)$ is increasing in θ and $\hat{\theta} < \bar{\theta}$. Then on $[\hat{\theta}, \bar{\theta}]$, the optimal x satisfies the point-wise maximization:*

$$x(\theta) = \max_{x} \{x - \lambda \Phi(x,\theta)\} \tag{9}$$

for the shadow cost function $\Phi(x,\theta) = \Psi_{\theta}(x,\theta)/f(\theta)$.

Proof. Recall that by definition, since pooling ends at $\hat{\theta}$, $\rho(\hat{\theta}) = 0$. By the Bellman principle of optimality, the restriction of x to the domain $[\hat{\theta}, \bar{\theta}]$, must be optimal for the following control subproblem:

$$\begin{aligned} \max_{x} \quad & \int_{\hat{\theta}}^{\bar{\theta}} \left(x(\theta) f(\theta) - \lambda \Psi_{\theta}(x(\theta),\theta)\right) \, d\theta \\ \text{subject to} \quad & \dot{x} = u \leq 0 \qquad\qquad \text{(Monotonicity)} \end{aligned}$$

Note that the solution to (9) point-wise maximizes the objective of the subproblem. To confirm that this is optimal, it suffices to show that $x(\theta)$ satisfies the monotonicity constraint. Note that since Ψ_{θ} is convex, Φ is convex in x and hence, the first-order condition for x gives

$$1 = \frac{\Psi_{x\theta}(x,\theta)}{f(\theta)}$$

Note that the RHS is an increasing function of x. Since the RHS is assumed to be increasing in θ, it follows that $x(\theta)$ must be decreasing in θ. ■

As a remark, note that the main obstacle to point-wise maximization is the monotonicity constraint; hence, one can obtain a more general variant of Proposition 6 by ironing the shadow cost function Φ. To interpret the shadow cost function Φ, note that by increasing $x(\theta)$, the principal gains a marginal benefit of $f(\theta)$ but must pay $\Psi_{\theta}(x(\theta),\theta)$ more information rent to the best types (or equivalently, must procure $\Psi_{\theta}(x(\theta),\theta)$ of the resource). Note that procuring more resource comes at the shadow cost λ; the shadow cost function Φ thus captures the marginal benefit of x vs. information rent trade off.

In sum, the optimal mechanism can be interpreted as the principal splitting the agent types into two populations, and maximizing x against different shadow cost functions for the two populations.

4 Separable Cost Function

Theorem 1 characterizes the solution form for a general specification of the function Ψ. In this section, we show that for specific forms of Ψ (namely, for multiplicatively separable Ψ) the mechanism admits a closed form solution, where the shape of the solution (including the pooling region from Proposition 1) is determined by the belief and independent of T.

To motivate the separability assumption, consider the following example:

Altruism Microfoundation Consider an large corporation (principal) seeking to reduce its carbon footprint by funding a carbon removal startup (agent). The startup functions as a Cournot monopolist with a linear demand curve $p(q) = K - q$. The startup has a quadratic cost function for production: $C(q;\theta) = v_c + v_b q + v_a q^2$, where v_a, v_b, v_c are all unknown to the principal. The startup initially operates alone, and maximizes profits:

$$\max_q \{qp(q) - C(q;\theta)\}$$

Some computation reveals that the laissez faire quantity and profits are:

$$q_{LF} = \frac{K - v_b}{2(v_a + 1)}$$

$$\Pi_{LF}(\theta) = (v_a + 1)q_{LF}^2$$

We can compute the profits of the monopolist producing q in terms of the laissez-faire profits and quantities:

$$\Pi(q) = qp(q) - C(q) = \Pi_{LF} - (v_a + 1)\left(q - q_{LF}\right)^2$$

The principal observes q_{LF} and wishes to contract with the startup to maximize q (equivalently, $x = q - q_{LF}$); however, observation of q_{LF} is not sufficient to identify the "scalability" of the startup, v_a. Renormalizing the agent's utility such that $\Psi(0) = 0$, the cost function becomes $\Psi(x, v_a) = (v_a + 1)x^2$. Notably, Ψ is multiplicatively separable.

4.1 Characterization

The optimal mechanism exhibits special features when Ψ is separable in x and θ (i.e, $\Psi(x,\theta) = B(\theta)\Gamma(x)$ for convex Γ). Without loss, since Ψ is assumed increasing in θ, we can renormalize the type space and redefine Θ so that $\Psi(x,\theta) = \theta\Gamma(x)$ where Γ is a strictly

convex function. In this case, we can actually provide a closed-form characterization of the shape of the schedule. Let $\tilde{f}$ denote the density associated with cav F, the concavification of the cumulative distribution function F on $[0, \infty)$.

Theorem 2 *Suppose that $\Psi(x, \theta) = \theta\Gamma(x)$, where Γ is convex. Then the optimal mechanism (x^*, t^*) induces a production schedule*

$$x^*(\theta) = (\Gamma')^{-1}\left(\frac{\tilde{f}(\theta)}{\lambda}\right)$$

where λ is the nonnegative Lagrange multiplier chosen so that

$$\underline{\theta}\Gamma(x^*(\underline{\theta})) + \int_{\underline{\theta}}^{\bar{\theta}} \Gamma(x^*(\theta))\ d\theta = T$$

The transfer schedule is still given by Lemma 1.

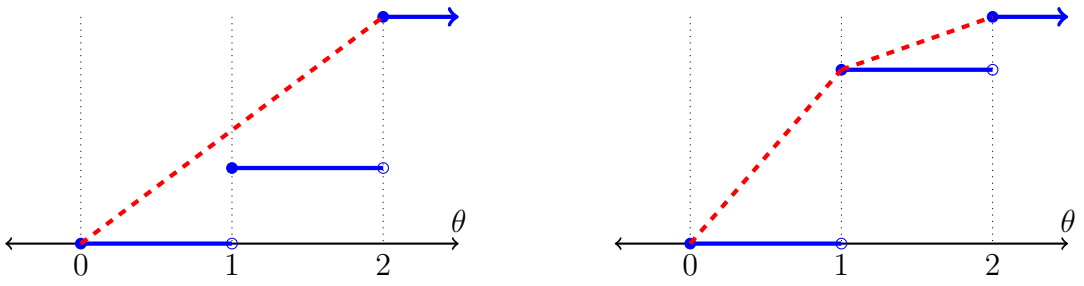
Intuitively, this holds because the linearity of Ψ in θ allows the separation of the monotonicity constraint into a standard ironing problem. Thus, for this specific case, the optimal schedule depends on the density of the concavified cumulative distribution function of the belief, $\tilde{f}$. In general, if a mechanism $x(\theta) = x(\theta')$ for types θ, θ', we say the mechanism **pools** the types. Otherwise, the mechanism **separates** θ, θ'. Thus, since Theorem 2 shows that x^* is an increasing transformation of $\tilde{f}$, the content of Theorem 2 is that for separable cases, the shape of the optimal mechanism is entirely determined by the *belief* of the principal and not by the agent's cost function.

We illustrate how this insight applies to a discrete, two-type example. Suppose θ is either 1 or 2, and the penalty function takes the form $\Psi(x, \theta) = \theta x^2$. The principal believes that $\theta = 1$ with probability μ. Figure 2 illustrates $\tilde{f}$; for the two-type case, whether μ is larger than or less than $1/2$ determines whether $\tilde{f}(1) = \tilde{f}(2)$ or not. By Theorem 2, x^* is an increasing transformation of $\tilde{f}$, so the shape of $\tilde{f}$ also determines the shape of $\tilde{f}$. The optimal mechanism separates types 1 and 2 if and only if $\mu > 1/2$. [4]

Note that for the separable case, Propositions 2 and 3 simplify considerably:

Corollary 3 *Suppose $\Psi(x, \theta) = \theta\Gamma(x)$, and suppose the optimal x^* is not constant on the entire interval of types. Let $\underline{x} \equiv x^*(\underline{\theta})$. Then the largest type $\hat{\theta}$ that receives the full transfer*

[4] In our modeling assumptions, we took F as continuous and differentiable. We can extend this to scenarios where F is discontinuous (equivalently, f admits point masses) by taking the limits of a sequence of continuous F_n converging to F and observing that that for the separable case, the optimizer function is continuous in $\tilde{f}$, which is continuous in f in the sup norm.



(a) $\mu \leq 1/2$ (b) $\mu > 1/2$

Figure 2: The distribution function corresponding to the belief of the principal is plotted in solid lines, and the dashed line denotes the concavification. In the first case, the concavified distribution function is constant, and so the optimal mechanism pools the types. In the second case, the left-derivative of the concavified distribution function takes different values at 1 and 2, so the optimal mechanism separates types.

satisfies the following:

$$\frac{1}{\hat{\theta}} = \frac{f(\hat{\theta})}{F(\hat{\theta})} \tag{10}$$

Corollary 4 *Suppose $\Psi(x, \theta) = \theta\Gamma(x)$. Then the optimal mechanism pools all types if and only if for all θ,*

$$\theta/\bar{\theta} \geq F(\theta)$$

For the separable case, the optimal mechanism exhibits a few additional features that do not occur for generic Ψ. In the next subsections, we discuss these properties.

4.2 Reward Size

In some sense, the upper bound on the schedule implied by Proposition 1 feels natural, given that there is an upper bound on transfers. An initial intuition might be that the pooling region occurs because the constraint is binding, and that the binding region shrinks if T is large. However, this is *not* the case. The pooling arises because of the *relative* informational cost of separating low and high θ to *relative* production ability. In fact, for the separable case, we have the following result:

Proposition 7 *Suppose $\Psi(x, \theta) = \theta\Gamma(x)$. Let $\hat{\theta}$ be the largest type that receives the full transfer. Then $\hat{\theta}$ is invariant to the changing the value of T.*

Proof. By Theorem 2, consider x_T and $x_{T'}$, the optimal schedules for budgets T and T'.

Since $\tilde{f}$ is unchanged, and x^* is an increasing transformation of $\tilde{f}$, the pooling region for x_T and $x_{T'}$ are the same, and hence $\hat{\theta}$ is invariant to altering the size of the budget. ■

That is, the size of the pooling region does not change at all, even if the T were made more or less valuable. This implies that the *shape* of the contract is determined primarily by the belief of the principal, and the *scale* of the contract is determined by the size of the budget. Hence, the existence of the pooling region at the top speaks to the principal's tradeoff between giving up information rents and withholding a non-valuable resource.

5 Extensions and Applications

In this section, we discuss some extensions of our model. We first characterize the problem under an ex-ante budget constraint, and contrast the optimal mechanism to the solution characterized by Theorem 1. In the second subsection, we consider a general arbitrary convex $c(t)$.

5.1 Linear Value

In the baseline model, we supposed that the principal's objective as $\mathbb{E}_F[x]$, which was independent of t, which we interpret as the principal having no value for the motivational resource. In certain applications, it may be reasonable to suppose the principal does have some value for the resource (or alternatively, obtaining the resource incurs a cost to the principal per unit of resource).

Consider a variation of the design problem (1):

$$\begin{aligned}
\max_{x,t} \quad & \int_{\theta\in\Theta} \left[x(\theta) - kt(\theta)\right] f(\theta)\, d\theta && && (11)\\
\text{subject to} \quad & t(\theta) - \Psi(x(\theta),\theta) \geq t(\theta') - \Psi(x(\theta'),\theta) && \forall \theta,\theta' \in \Theta && \text{(IC)}\\
& t(\theta) - \Psi(x(\theta),\theta) \geq 0 && \forall \theta \in \Theta && \text{(IR)}\\
& t(\theta) \leq T && \forall \theta \in \Theta && \text{(B)}
\end{aligned}$$

where $k \geq 0$ is some constant.

Proposition 8 *An optimal mechanism (x,t) for design problem* (11) *exists, and it satisfies Lemma 1. There exists a nonnegative Lagrange multiplier λ, a nonnegative, absolutely continuous costate function ρ, and a nonnegative, absolutely continuous costate w, such that (x,t,λ,ρ,w) uniquely solve complementary slackness* (3)*, border conditions* (5) *and* (6)*, a*

modified costate evolution given by

$$\dot{\rho}(\theta) = \lambda \Psi_{x\theta}(x^*(\theta), \theta) - f(\theta) \left(1 - k \left(\Psi_x(x^*(\theta), \theta) + \Psi_{x\theta}(x^*(\theta), \theta) \frac{F(\theta)}{f(\theta)}\right)\right) - w(\theta) \quad (12)$$

that holds wherever ρ is differentiable, and an additional complementary slackness constraint

$$w(\theta) > 0 \implies x(\theta) = 0. \quad (13)$$

Further, the budget binding condition (7) *must hold if $\lambda > 0$.*

A couple of comments on the modifications are in order. Note that the altered costate evolution features two new terms. First,

$$\psi(x, \theta) = \Psi_x(x, \theta) + \Psi_{x\theta}(x, \theta) \frac{F(\theta)}{f(\theta)} \quad (14)$$

is the *virtual* marginal cost of the agent, similarly to the Myersonian virtual value. The second, w, is a complementary slackness variable that arises because the lower bound on x can bind (and types of agents can be excluded in this case).

5.2 Ex Ante Budget Constraints

An alternative formulation might require that the budget only needs to bind ex-ante (i.e., the expected transfer cannot exceed T). That is, consider the following maximization

$$\begin{aligned}
\max_{x,t} \quad & \int_{\theta \in \Theta} x(\theta) f(\theta)\, d\theta & & & (15)\\
\text{subject to} \quad & t(\theta) - \Psi(x(\theta), \theta) \geq t(\theta') - \Psi(x(\theta'), \theta) & \forall \theta, \theta' \in \Theta & \quad \text{(IC)}\\
& t(\theta) - \Psi(x(\theta), \theta) \geq 0 & \forall \theta \in \Theta & \quad \text{(IR)}\\
& \int_{\underline{\theta}}^{\bar{\theta}} t(\theta) f(\theta)\, d\theta \leq T & \forall \theta \in \Theta & \quad \text{(B')}
\end{aligned}$$

The optimal solution is characterized by the following proposition.

Proposition 9 *An optimal mechanism (x, t) for* (15) *exists, and is unique. Further, there exists a nonnegative Lagrange multiplier λ and a nonnegative, absolutely continuous costate functions ρ such that (x, λ, ρ) is unique solution that satisfies Lemma 1 and complementary*

slackness (3) *holds, with the modified costate evolution equation*

$$\dot{\rho}(\theta) = f(\theta)\left(1 - \lambda\left(\Psi_x(x(\theta),\theta) + \Psi_{x\theta}(x(\theta),\theta)\frac{F(\theta)}{f(\theta)}\right)\right) \tag{16}$$

upper boundary condition (5)*, modified lower bound condition* $\rho(\underline{\theta}) = 0$*, and modified budget binding condition*

$$\int f(\theta)\left(\Psi(x(\theta),\theta) + \int_{\theta}^{\bar{\theta}} \Psi_\theta(x(s),s)\ ds\right)\ d\theta = T \tag{17}$$

Notably, the lower boundary condition is modified to $\rho(\underline{\theta}) = 0$, and hence the pooling region that appears for the ex-post budget constraint does *not* appear here.

5.3 Generalized Transfer Value

Suppose the principal has an arbitrary continuous, weakly convex resource value $c(t)$. That is, suppose the designer's problem is now

$$\begin{aligned}
\max_{x,t} \quad & \int_{\theta\in\Theta} [x(\theta) - c(t(\theta))]\, f(\theta)\ d\theta && && (18)\\
\text{subject to} \quad & t(\theta) - \Psi(x(\theta),\theta) \geq t(\theta') - \Psi(x(\theta'),\theta) && \forall \theta,\theta' \in \Theta && \text{(IC)}\\
& t(\theta) - \Psi(x(\theta),\theta) \geq 0 && \forall \theta \in \Theta && \text{(IR)}\\
& t(\theta) \leq T && \forall \theta \in \Theta && \text{(B)}
\end{aligned}$$

and that there is no longer a hard budget constraint T. The theorem variant for this extension is:

Proposition 10 *An optimal mechanism* (x,t) *for design problem* (18) *exists and is unique. Further, there exists nonnegative, absolutely continuous costate functions* ρ_x, ρ_t, λ *such that*

$x, t, \rho_x, \rho_t, \lambda$ satisfy Lemma 1 and solve

$$\rho_x(\theta) + \rho_t(\theta)\Psi_x(x(\theta), \theta) > 0 \implies \dot{x}(\theta) = 0 \tag{19}$$

$$\dot{\rho}_x(\theta) = -\rho_t(\theta)\dot{x}(\theta)\Psi_{xx}(x(\theta), \theta) - f(\theta) \tag{20}$$

$$\lambda(\theta) > 0 \implies t(\theta) = T \tag{21}$$

$$\dot{\rho}_t(\theta) = c'(t(\theta))f(\theta) - \lambda(\theta) \tag{22}$$

$$\rho_x(\bar{\theta}) + \rho_t(\bar{\theta})\Psi_x(x(\bar{\theta}), \bar{\theta}) = 0 \tag{23}$$

$$\rho_x(\underline{\theta}), \rho_t(\underline{\theta}) = 0 \tag{24}$$

The key simplification is that even though the state space (x, t) is two-dimensional, the variable $\dot{x} = u$ jointly controls both state variables; the boundary condition (23) then comes from the orthogonality constraint for the boundary surface at $\bar{\theta}$ implied by the binding IR constraint $t(\bar{\theta}) = \Psi(x(\bar{\theta}), \bar{\theta})$.

6 Conclusion

We develop a mechanism design model for altruistic giving, where a principal altruist interacts with an agent. The design problem features an ex-post budget constraint and a first-order objective maximizing the agent's action. We derive the optimal contract, and show that the contract pools sufficiently efficient types, and withholds the budget from less efficient types. We show that with an additional separability restriction, the efficiency threshhold is invariant to the size of the budget.

Future theoretical work extending the model in technical directions could examine the problem with multiple agents or endogenize the outside option of the agent. Additionally, we took a Bayesian approach to the design problem, where the principal maximizes given a belief over agent types; however, there is room for interesting robust mechanism design work where the principal minimizes a regret measure. More broadly, we hope the model provides a foundation for more work focusing on early market investment.

References

Abebe, R. and Goldner, K. (2018), 'Mechanism design for social good', *AI Matters* **4**(3), 27–34.

Alonso, R. and Matouschek, N. (2008), 'Optimal delegation', *The Review of Economic Studies* **75**(1), 259–293.

Amador, M. and Bagwell, K. (2013), 'The theory of optimal delegation with an application to tariff caps', *Econometrica* **81**(4), 1541–1599.
URL: *http://www.jstor.org/stable/23524185*

Amador, M., Bagwell, K. and Frankel, A. (2018), 'A note on interval delegation', *Economic Theory Bulletin* **6**(2), 239–249.

Amanatidis, G., Kleer, P. and Schäfer, G. (2019), Budget-feasible mechanism design for non-monotone submodular objectives: Offline and online, *in* 'Proceedings of the 2019 ACM Conference on Economics and Computation', pp. 901–919.

Babaioff, M., Immorlica, N., Li, Y. and Lucier, B. (2023), 'Making auctions robust to aftermarkets', *ITCS* .

Baron, D. P. and Myerson, R. B. (1982), 'Regulating a monopolist with unknown costs', *Econometrica* **50**(4), 911–930.
URL: *http://www.jstor.org/stable/1912769*

Dai Li, W., Immorlica, N. and Lucier, B. (2022), Contract design for afforestation programs, *in* 'Web and Internet Economics: 17th International Conference, WINE 2021, Potsdam, Germany, December 14–17, 2021, Proceedings', Springer, pp. 113–130.

Goldner, K., Immorlica, N. and Lucier, B. (2019), 'Reducing inefficiency in carbon auctions with imperfect competition', *ITCS* .

Gravin, N., Jin, Y., Lu, P. and Zhang, C. (2020), 'Optimal budget-feasible mechanisms for additive valuations', *ACM Transactions on Economics and Computation (TEAC)* **8**(4), 1–15.

Greaker, M. and Popp, D. (2022), Environmental economics, regulation, and innovation, Working Paper 30415, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w30415*

Grossman, S. J. and Hart, O. D. (1983), 'An analysis of the principal-agent problem', *Econometrica* **51**(1), 7–45.
URL: *http://www.jstor.org/stable/1912246*

Halac, M. and Yared, P. (2020), 'Commitment versus flexibility with costly verification', *Journal of Political Economy* **128**(12), 4523–4573.

Holmström, B. (1979), 'Moral hazard and observability', *The Bell journal of economics* pp. 74–91.

Holmstrom, B. (1980), On the theory of delegation, Technical report, Discussion Paper.

Kamien, M. I. and Schwartz, N. L. (1971), 'Sufficient conditions in optimal control theory', *Journal of Economic Theory* **3**(2), 207–214.
URL: *https://www.sciencedirect.com/science/article/pii/0022053171900184*

Kang, Z. Y. (2022), The public option and optimal redistribution, Technical report, Working paper.

Laffont, J.-J. and Tirole, J. (1986), 'Using cost observation to regulate firms', *Journal of Political Economy* **94**(3), 614–641.
URL: *http://www.jstor.org/stable/1833051*

Leonardi, S., Monaco, G., Sankowski, P. and Zhang, Q. (2017), Budget feasible mechanisms on matroids, *in* 'International conference on integer programming and combinatorial optimization', Springer, pp. 368–379.

Singer, Y. (2010), Budget feasible mechanisms, *in* '2010 IEEE 51st Annual Symposium on foundations of computer science', IEEE, pp. 765–774.

Singer, Y. (2014), 'Budget feasible mechanism design', *ACM SIGecom Exchanges* **12**(2), 24–31.

Toikka, J. (2011), 'Ironing without control', *Journal of Economic Theory* **146**(6), 2510–2526.
URL: *https://www.sciencedirect.com/science/article/pii/S0022053111000871*

A Appendix: Supplemental Material

A.1 Omitted Proofs

Proof of Lemma 1. Consider any x that satisfies the two conditions. Take t given by (2). Since Ψ_θ is nonnegative, $t(\theta) \geq \Psi(x(\theta), \theta)$, so IR is satisfied. Now, we show that t must be nonincreasing. Suppose $\theta \leq \theta'$. Then

$$\begin{aligned}
t(\theta) &= \Psi(x(\theta), \theta) + \int_\theta^{\overline{\theta}} \Psi_\theta(x(s), s)\ ds \\
&= \Psi(x(\theta), \theta) + \int_\theta^{\theta'} \Psi_\theta(x(s), s)\ ds + \int_{\theta'}^{\overline{\theta}} \Psi_\theta(x(s), s)\ ds \\
&\geq \Psi(x(\theta), \theta) + \int_\theta^{\theta'} \Psi_\theta(x(\theta'), s)\ ds + \int_{\theta'}^{\overline{\theta}} \Psi_\theta(x(s), s)\ ds \\
&= \Psi(x(\theta), \theta) + \Psi(x(\theta'), \theta') - \Psi(x(\theta'), \theta) + \int_{\theta'}^{\overline{\theta}} \Psi_\theta(x(s), s)\ ds \\
&= [\Psi(x(\theta), \theta) - \Psi(x(\theta'), \theta)] + t(\theta')
\end{aligned}$$

where we used the fact that x is nonincreasing and supermodularity of Ψ in the 3rd step. Also, since x is nonincreasing, and Ψ is increasing in its first argument, the bracketed part is nonnegative, and so we have that t is nonincreasing. Hence, to check the budget constraint, we just have to check that the budget constraint holds at $t(\underline{\theta})$:

$$t(\underline{\theta}) = \Psi(x(\underline{\theta}), \underline{\theta}) + \int_\Theta \Psi_\theta(x(s), s)\ ds = \int_\Theta \Psi_\theta(x(s), s) \leq T$$

Finally, to check that IC is satisfied, consider IC for type θ. We consider the deviation of θ to θ'. We first show that θ does not want to deviate up, or $\theta \leq \theta'$:

$$\begin{aligned}
t(\theta) - \Psi(x(\theta), \theta) &= \int_{\theta}^{\overline{\theta}} \Psi_{\theta}(x(s), s)\ ds \\
&= \int_{\theta}^{\theta'} \Psi_{\theta}(x(s), s)\ ds + \int_{\theta'}^{\overline{\theta}} \Psi_{\theta}(x(s), s)\ ds \\
&= \int_{\theta}^{\theta'} \Psi_{\theta}(x(s), s)\ ds + t(\theta') - \Psi(x(\theta'), \theta') \\
&\geq \int_{\theta}^{\theta'} \Psi_{\theta}(x(\theta'), x) + t(\theta') - \Psi(x(\theta'), \theta') \\
&= \Psi(x(\theta'), \theta') - \Psi(x(\theta'), \theta) + t(\theta') - \Psi(x(\theta'), \theta') \\
&= t(\theta') - \Psi(x(\theta'), \theta)
\end{aligned}$$

Now we check that θ does not want to deviate down to type $\theta' \leq \theta$:

$$\begin{aligned}
t(\theta) - \Psi(x(\theta), \theta) &= \int_{\theta}^{\overline{\theta}} \Psi_{\theta}(x(s), s)\ ds \\
&= -\int_{\theta'}^{\theta} \Psi_{\theta}(x(s), s)\ ds + \int_{\theta'}^{\overline{\theta}} \Psi_{\theta}(x(s), s)\ ds \\
&= -\int_{\theta'}^{\theta} \Psi_{\theta}(x(s), s)\ ds + t(\theta') - \Psi(x(\theta'), \theta') \\
&\geq -\int_{\theta'}^{\theta} \Psi_{\theta}(x(\theta'), s)\ ds + t(\theta') - \Psi(x(\theta'), \theta') \\
&= -\Psi(x(\theta'), \theta) + \Psi(x(\theta'), \theta') + t(\theta') - \Psi(x(\theta'), \theta') \\
&= t(\theta') - \Psi(x(\theta'), \theta)
\end{aligned}$$

So IC is satisfied. Thus, if x satisfies the conditions, it is feasible.

Now, we show the two conditions are necessary. Consider any feasible $x : \Theta \to \mathbb{R}_+$. Suppose $\theta \leq \theta'$, and $\theta, \theta' \in \Theta$. From rewriting the IC constraints, we get that

$$\Pi(x(\theta); \theta) - \Pi(x(\theta'); \theta) \geq t(\theta') - t(\theta)$$

$$\Pi(x(\theta'); \theta') - \Pi(x(\theta); \theta') \geq t(\theta) - t(\theta')$$

Note that the $\Pi_{LF}(\theta)$ and $\Pi_{LF}(\theta')$ terms will cancel on the left hand sides. Adding these

together, we get

$$-\Psi(x(\theta),\theta)+\Psi(x(\theta'),\theta)-\Psi(x(\theta'),\theta')+\Psi(x(\theta),\theta')\geq 0$$

Rearranging

$$\Psi(x(\theta'),\theta)+\Psi(x(\theta),\theta')\geq \Psi(x(\theta),\theta)+\Psi(x(\theta'),\theta')$$

And so supermodularity of Ψ implies that $x(\theta')\leq x(\theta)$, and so the first condition must hold. To show that the transfers must take the form given, define the interim utility

$$U(\theta)=t(\theta)-\Psi(x(\theta),\theta)$$

IC implies that

$$U(\theta)=\max_{\theta'} t(\theta')-\Psi(x(\theta'),\theta)$$

By the envelope theorem,

$$U'(\theta)=-\Psi_\theta(x(\theta),\theta)$$

$$U(\theta)=U(\bar{\theta})+\int_\theta^{\bar{\theta}}\Psi_\theta(x(s),s)\ ds$$

So transfers must be of the form

$$\begin{aligned} t(\theta)&=U(\theta)\\ &=U(\bar{\theta})+\Psi(x(\theta),\theta)+\int_\theta^{\bar{\theta}}\Psi_\theta(x(s),s)\ ds \end{aligned}$$

In particular, the transfer for type $\underline{\theta}$ satisfies

$$\begin{aligned} T&\geq t(\underline{\theta})\\ &=U(\bar{\theta})+\Psi(x(\underline{\theta}),\underline{\theta})+\int_{\underline{\theta}}^{\bar{\theta}}\Psi_\theta(x(s),s)\ ds\\ &\geq \Psi(x(\underline{\theta}),\underline{\theta})+\int_{\underline{\theta}}^{\bar{\theta}}\Psi_\theta(x(s),s)\ ds \end{aligned}$$

since $U(\bar{\theta})\geq 0$ by IR. Hence, the two conditions are necessary and sufficient for a feasible schedule. ■

Proof of Theorem 1. By Lemma 1, we can rewrite the design problem as

$$\begin{aligned} \max_{x} \quad & \int_{\Theta} x(\theta) f(\theta)\ d\theta && (25) \\ \text{subject to} \quad & \Psi(x(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi_{\theta}(x(s), s)\ ds \leq T && \text{(Normalization)} \\ & x \text{ nonincreasing} && \text{(Monotonicity)} \end{aligned}$$

The Lagrangian relaxation is

$$\begin{aligned} \max_{x} \quad & \int_{\Theta} x(\theta) f(\theta)\ d\theta + \lambda \left(T - \Psi(x(\underline{\theta}), \underline{\theta}) - \int_{\underline{\theta}}^{\bar{\theta}} \Psi_{\theta}(x(s), s)\ ds \right) \\ \text{subject to} \quad & x \text{ nonincreasing} && \text{(Monotonicity)} \end{aligned}$$

By monotonicity, x must be differentiable almost everywhere. The monotonicity constraint then requires that wherever x is differentiable, $\dot{x} \leq 0$. Letting $u = \dot{x}$ and rewriting the objective function, the problem becomes

$$\begin{aligned} \max_{x} \quad & \int_{\Theta} \left(x(\theta) f(\theta) - \lambda \Psi_{\theta}(x(\theta), \theta) \right)\ d\theta + \lambda T - \lambda \Psi(x(\underline{\theta}), \underline{\theta}) \\ \text{subject to} \quad & \dot{x} = u \leq 0 && \text{(Monotonicity)} \end{aligned}$$

The problem now is an optimal control problem with a scrap value constraint; u is the control variable, and x is the state variable.

We first show the only-if direction of the proof. For a fixed $\lambda > 0$, Pontryagin's maximum principle implies that any optimal solution admits a control variable such that (3) - (6) are true.

Further, note that if $\lambda = 0$, (6) implies that $\rho(\underline{\theta}) = 0$, but (4) implies that $\dot{\rho}(\underline{\theta}) = -f(\underline{\theta}) < 0$, which is impossible since ρ must be nonnegative. Hence, $\lambda > 0$, so by Lagrangian duality we must have that the budget binds, so (7) holds.

For the if direction, note that fixing λ, the maximized Hamiltonian corresponding to the control problem is concave in the state variable x, so the Arrow sufficiency conditions hold.[5] Hence the conditions provided are also sufficient for an optimal solution given λ. Further, since the solution maximizes the Lagrangian and and the budget constraint (7) binds, the solution must be optimal in the original constrained maximization. ■

[5]See Kamien and Schwartz (1971) for details.

Proof of Theorem 2. As in the proof of Theorem 1, we write out the Lagrangian:

$$\max_x \qquad \int_\Theta x(\theta) f(\theta)\ d\theta + \lambda \left(T - \underline{\theta}\Gamma(x(\underline{\theta})) - \int_{\underline{\theta}}^{\bar{\theta}} \Gamma(x(s))\ ds \right)$$

$$\text{subject to} \quad x \text{ nonincreasing} \qquad \text{(Monotonicity)}$$

Define $f(\theta) = 0$ for $\theta < \underline{\theta}$. Then note that the optimization can be rewritten equivalently as

$$\max_x \qquad \int_0^{\bar{\theta}} x(\theta) f(\theta)\ d\theta + \lambda \left(T - \int_0^{\bar{\theta}} \Gamma(x(s))\ ds \right)$$

$$\text{subject to} \quad x \text{ nonincreasing} \qquad \text{(Monotonicity)}$$

To handle the monotonicity constraint, we use a generalized ironing technique. By Theorem 4.4 in Toikka (2011), the optimal $x(\theta)$ then maximizes $\overline{J}(\cdot, \theta; \lambda)$ at the optimal λ, where

$$\overline{J}(x, \theta; \lambda) = \int_0^s \frac{\partial}{\partial \theta^-} \text{cav}_\theta \left[F(\theta) - \lambda\theta\Gamma'(x) \right] d\theta$$

and cav_θ denote the concavification operation in θ, where the differentiation is taken from the left (as the concavification need not be smooth in θ). Note that since $(\theta - \underline{\theta})$ is linear in θ, we can rewrite $\overline{J}$ as

$$\begin{aligned} \overline{J}(x, \theta; \lambda) &= \int_0^x \frac{\partial}{\partial \theta^-} \text{cav}_\theta \left[F(\theta) \right] d\theta - \lambda\Gamma'(x) \\ &= \tilde{f}(\theta) - \lambda\Gamma'(x) \end{aligned}$$

Hence, the problem is reduced to a point-wise optimization at each θ: inverting Γ' implies that

$$x^*(\theta) = (\Gamma')^{-1} \left(\frac{\tilde{f}(\theta)}{\lambda} \right)$$

It remains to argue that a λ exists such that the budget constraint binds. Note that as $\lambda \to 0$, $x^* \to \infty$, and as $\lambda \to \infty$, $x^* \to 0$. Hence, the intermediate value theorem implies some finite value of λ makes the budget constraint bind. ■

Proof of Proposition 8. The analysis proceeds exactly as in Theorem 1; in this case, we plug in the transfers from Lemma 1 into the objective and integrate by parts. The

Hamiltonian becomes

$$\mathcal{H} = xf(\theta) - \lambda\Psi_\theta(x,\theta) - k\Psi(x,\theta)f(\theta) - k\Psi_\theta(x,\theta)F(\theta) + \rho u$$

Applying Pontryagin, the costate evolution (12) follows. ■

Proof of Proposition 10. Applying Lemma 1, and noting that (IR) binds at the highest type $\underline{\theta}$, the problem becomes

$$\begin{aligned} \max_x \quad & \int_\Theta (x(\theta) - c(t(\theta)))f(\theta)\ d\theta \\ \text{subject to} \quad & t(\theta) = \Psi(x(\theta),\theta) + \int_\theta^{\bar\theta} \Psi_\theta(x(s),s)\ ds \\ & x \text{ nonincreasing} \end{aligned}$$

As before, we interpret this as a control problem; however, we now have two state variables x, t. Note that unlike the linear case, we cannot substitute t into the objective and integrate out by parts; rather, the control problem maintains t in the state. To formulate the law of motion for t, note that

$$\dot t(\theta) = \Psi_x(x(\theta),\theta)\dot x(\theta)$$

Hence, as before, let $u = \dot x$ be the control variable. Then u simultaneously controls both x and t according to the following laws of motion:

$$\begin{bmatrix} \dot x(\theta) \\ \dot t(\theta) \end{bmatrix} = \begin{bmatrix} u(\theta) \\ \Psi_x(x(\theta),\theta)u(\theta) \end{bmatrix}$$

Then since IR binds at the top, we obtain the following boundary surface constraint:

$$t(\bar\theta) - \Psi(x(\bar\theta),\bar\theta) = 0$$

So all together, the control problem is

$$\begin{aligned} \max_x & \int_\Theta (x(\theta) - c(t(\theta)))f(\theta)\ d\theta \\ & \begin{bmatrix} \dot x(\theta) \\ \dot t(\theta) \end{bmatrix} = \begin{bmatrix} u(\theta) \\ \Psi_x(x(\theta),\theta)u(\theta) \end{bmatrix} \\ & t(\bar\theta) - \Psi(x(\bar\theta),\bar\theta) = 0 \end{aligned}$$

We apply Pontryagin again, and note that because c is convex, the Hamiltonian is concave in both state variables, and invoke the Arrow sufficiency conditions. The Hamiltonian is

$$\mathcal{H} = (x - c(t))f(\theta) + \rho_x u + \rho_t \Psi_x(x, \theta)u$$

Since u maximizes the unconstrained Hamiltonian, we get the complementary slackness condition (19). The costate equations (20) and (22) follow. The boundary condition lies on a surface, and so the orthogonality condition implies that the costate vector at the upper terminal must lie in the direction of the gradient vector:

$$\begin{bmatrix} \rho_x \\ \rho_t \end{bmatrix} (\bar{\theta}) \propto \begin{bmatrix} -\Psi_x(x(\bar{\theta}), \bar{\theta}) \\ 1 \end{bmatrix}$$

Using ρ_t to substitute out the constant of proportionality, we get the boundary condition (23). Since the lower endpoint is free, we get the other boundary condition (24). ■

Proof of Proposition 9. By Lemma 1, we can reduce (IC) and (IR) into a monotonicity constraint and a transfer schedule. Adding a Lagrange multiplier λ to the budget constraint, the control problem becomes

$$\begin{aligned} \max_x \quad & \int_\Theta x(\theta) f(\theta)\ d\theta + \lambda \left(T - \int_{\underline{\theta}}^{\bar{\theta}} \left(\Psi(x(\theta), \theta) + \int_\theta^{\bar{\theta}} \Psi_\theta(x(s), s)\ ds \right) f(\theta)\ d\theta \right) \\ & \dot{x} = u \leq 0 \end{aligned}$$

Integrating by parts, and writing the integral in one term, we get

$$\begin{aligned} \max_x \quad & \int_\Theta \left(x(\theta) - \lambda \left(\Psi(x(\theta), \theta) + \Psi_\theta(x(\theta), \theta) \frac{F(\theta)}{f(\theta)} \right) \right) f(\theta)\ d\theta + \lambda T \\ & \dot{x} = u \leq 0 \end{aligned}$$

Applying Pontryagin and the Arrow sufficiency result, we get the conditions in the proposition. ■