Strategy-Proof Resource Allocation with Punishment

Lining Han *[1] and Ruben Juarez †[2]

[1]Economics and Management School, Wuhan University

[2]Department of Economics and UHERO, University of Hawaii

May 4, 2023

Abstract

A planner chooses an allocation of a divisible resource and charges agents based on their reported type. We discover and describe the set of dominant-strategy incentive-compatible (strategy-proof) mechanisms when the planner has the ability to observe the true type of agents ex-post and punish those agents who misreported their type. This class of mechanisms depends on the punishment function available for the planner to use and expands previous characterizations of incentive-compatible mechanisms when punishment was not available. For any punishment function, an optimal mechanism for the planner is characterized as the convex combination of two mechanisms resembling the first-price and second-price mechanisms. When the planner has the ability to select the punishment function, the minimal punishment necessary to achieve incentive compatibility and first-best efficiency is provided.

Keywords: Strategy-proofness, Punishment, Resource Allocation

JEL classification: C70, D82, D85

*Email: hanlining@whu.edu.cn

†Email: rubenj@hawaii.edu

1 Introduction

Investors (buyers) often rely on seller's information to make investment decisions. Although sellers typically know more than investors at the time of investment, this information disparity often evaporates after the investment has occurred, potentially leading to compensations and punishments if the sellers disclosed inaccurate information.

Examples of these post-transaction punishments abound. Wall Street firms often settle cases over fraudulent and misleading information with the SEC and other federal and state attorney offices (e.g., Purdue-Pharma with multiple attorneys general, Luckin Coffee and Tesla 420 incident with the SEC). Public and private companies have been sanctioned with severe punishment for sellers (e.g., over $10 billion in fines have been ordered for pharmaceutical companies over the last 10 years for fraudulent or misleading information). In extreme cases, companies get dissolved due to fraud, ranging from unicorn startups (e.g., Theranos) to the largest Ponzi scheme in history (Bernard L. Madoff Investment Securities LLC). In the United States, there are state and federal false advertising laws that prohibit various types of deceptive advertising, misleading labeling, and similar practices. These laws provide important rights for investors, arming them with the ability to seek monetary damages when they have been misled. In addition, these problems were so persistent in markets of cars and other consumer goods that lemon laws have been written in half a dozen countries to provide explicit remedies to buyers of goods that fail to meet standards of quality and performance disclosed at the time of purchase.

In this paper, we study the role of ex-post punishment to induce accurate/truthful reporting ex-ante. As an example of our model, imagine an investor (planner or buyer) endowed with capital that he could allocate to a variety of privately-owned enterprises that produce some profit. Information is asymmetric, i.e., the owners of enterprises (agents or sellers) know the production possibilities of the enterprises they own, whereas the planner does not. Although the investor is not familiar with the actual production possibilities during the investing decision, he may rely on information revealed by the owners of the enterprises to make his determination. We study a model with an empowered investor in which, after the investment has been made, he observes the actual production possibilities of the enterprises, and is able to recover some of the investment if the owners of the enterprises misreported information as punishment (e.g., by law or by the threat of going to court). Although the investor may be able to implement truthful revelation with a traditional incentive-compatible mechanism, these mechanisms may significantly sacrifice profit or fairness when implemented.

Would the investor be better off when he is empowered with ex-post information and the ability to punish the owners of enterprises? What type of punishment functions allow for the implementation of mechanisms that improve the welfare of the investor? What type of mechanisms would be optimal given a punishment function? What type of punishment function allows for a mechanism to be incentive-compatible? In this paper, we answer these questions in a general model of resource allocation with quasilinear utilities that has an empowered planner who is able to observe the true abilities of agents post-allocation and punish them for misreporting.

In order to introduce our notion of implementation, we note that *strategy-proof* (dominant-strategy) mechanisms are desirable mechanisms that make it easier for participants to decide what to do regardless of the information context of other agents. To date, the traditional mechanism design literature has widely focused on ***ex-ante strategy-proofness***, whereby agents report their type to the planner, who makes his allocation knowing that these are truthful reports (see related literature for more information). Unfortunately, restricting to ex-ante strategy-proof mechanisms might come at a high cost to the planner in terms of efficiency and fairness, where impossibilities abound. In contrast, in our model where an empowered planner is able to observe truthful reports and punish agents who misreported, we introduce the notion of ***ex-post strategy-proofness*** whereby agents maximize their utility as the sum of ex-ante payment plus ex-post punishment. Notably, this notion of ex-post strategy-proofness opens opportunities for more incentive-compatible mechanisms beyond the ones discovered under ex-ante strategy-proofness. Our model takes a very general monetary punishment that depends on the reported type and the true type of the agents. To date, the theoretical literature has been silent about studying the *punishments* necessary to achieve truthful reporting, and our paper will address this gap.

1.1 Illustrative Example and Overview of Results

Consider an investor with a unit of capital, and n agents, denoted by $i = 1, \dots, n$, who own companies that can produce profit with a production function. As an illustration, assume that the production function of each company is $f_i(\theta_i, x_i) = \theta_i x_i$, where x_i is the capital allocated by the planner, and $\theta_i > 0$ is the type of agent i that is private information only known to this agent.

On a direct mechanism, the agents report their types $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_n)$ to the planner, who uses this information to make an allocation of his capital among agents $(x_1(\hat{\theta}), \dots, x_n(\hat{\theta})) \in \mathbb{R}^n_+$, such that $x_1(\hat{\theta}) + \cdots + x_n(\hat{\theta}) = 1$. We write $\hat{x}_i = x_i(\hat{\theta})$ for simplicity.

When the true type of the agents is $\theta = (\theta_1, \dots, \theta_n)$, each agent produces $f_i(\theta_i, \hat{x}_i)$ and gets charged $t_i(\hat{\theta})$. Agent i's (ex-ante) payment equals to $f_i(\theta_i, \hat{x}_i) - t_i(\hat{\theta})$.

It is well known that for a report $\hat{\theta}$ such that $\hat{\theta}_1 > \hat{\theta}_2 \geq \cdots \geq \hat{\theta}_n$, the *second-price* *mechanism* is the unique ex-ante strategy-proof mechanism that allocates all resources to the agent with the largest report (allocative efficiency), $\hat{x}_1 = 1$ and $\hat{x}_i = 0$ for $i > 1$, does not subsidize agents (no positive transfers), and is individually rational. Such a mechanism charges the second highest production to the more productive agent $t_1(\hat{\theta}) = \hat{\theta}_2$ (Vickrey[38], Clarke[13], Groves[16]).

The second-price mechanism may not maximize the profit of an investor who is able to punish agents ex-post. To see this, consider the punishment function $h_i(\theta_i, \hat{\theta}_i) = \pi(|\theta_i - \hat{\theta}_i|)^\gamma$ for $\gamma > 0$ and $0 < \pi \leq 1$. This punishment may be generated, for instance, by a court that will award with probability π an amount equal to $(|\theta_i - \hat{\theta}_i|)^\gamma$. In this new setting with punishment, agent i would maximize his ex-post payment equal to $f_i(\theta_i, \hat{x}_i) - t_i(\hat{\theta}) - h_i(\theta_i, \hat{\theta}_i)$.

We say that a mechanism is ex-post strategy-proof under punishment vector $h = (h_1, \ldots, h_n)$, or for short, h-strategy-proof, if it is a dominant strategy for agents to report their true types when maximizing their ex-post payment. Clearly, an ex-ante strategy-proof mechanism is h-strategy-proof, however, depending on the punishment vector h, new mechanism may emerge or not. Figure 1 shows the sequence of events in our model.

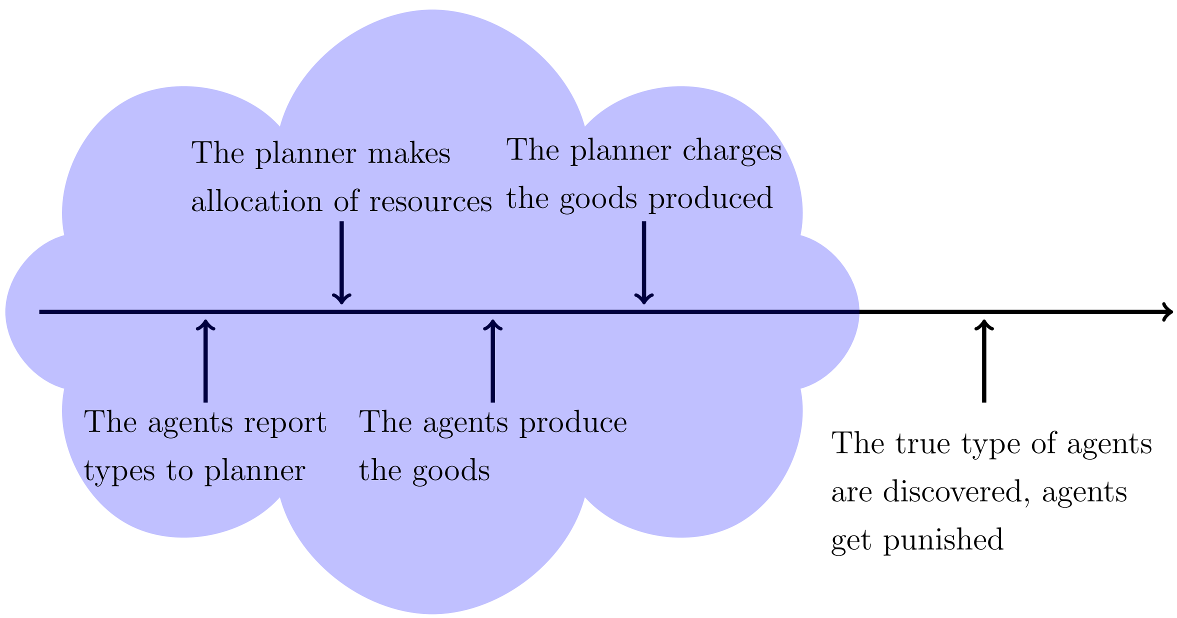

Figure 1: Timeline of our model: Events inside the cloud represent the original timeline studied for the implementation of ex-ante strategy-proof mechanisms. The last event outside the cloud, whereby the planner discovers agents true types and punish agents, is added for the study of ex-post strategy-proof mechanisms.

First, consider the case of $\gamma = 1$ and $0 < \pi \leq 1$. In this case, the combination mechanism that allocates the most productive agent all resource, $x_1(\hat{\theta}) = 1$, and charges a combination between the production of the most productive and second most productive agent, $t_1(\hat{\theta}) = \pi\hat{\theta}_1 + (1-\pi)\hat{\theta}_2$, is h-strategy-proof. To see this, note that the punishment for deviating equals $\pi(\hat{\theta}_1 - \theta_1)$, which is also the maximal gain that can be obtained from deviating in the combination mechanism. As such, a new mechanism, that was not ex-ante strategy-proof, is now h-strategy-proof. The extreme case when $\pi = 1$ represents the first-price mechanism, a paramount achievement for the investor, which extracts all surplus from agents.

Next, consider the case of $\gamma > 1$. In this case, regardless of the probability of award π, the class of h-strategy-proof mechanisms and ex-ante strategy-proof mechanisms coincide. That is, the punishment h is *ineffective* to generate new mechanisms. To get the intuition, consider the combination mechanism discussed above. For the top agent, the ex-ante gain by deviating is $\pi(\theta_1 - \hat{\theta}_1)$ for any $\hat{\theta}_1 > \hat{\theta}_2$. The ex-post punishment equals $\pi(\theta_1 - \hat{\theta}_1)^\gamma$, which is smaller than the ex-ante gain for a report $\hat{\theta}_1$ that is close enough to the true report θ_1.

The first result of the paper, Theorem 1, provides the necessary conditions for a punishment function to be effective. Roughly speaking, these conditions require that the marginal punishment at the true report be positive. Thus, the class of ineffective punishment functions resemble $h_i(\theta_i, \hat{\theta}_i) = \pi(|\theta_i - \hat{\theta}_i|)^\gamma$ for $\gamma > 1$, whereby the derivative with respect to $\hat{\theta}_i$ at $\hat{\theta}_i = \theta_i$ equals zero.

In addition, Lemma 1 provides necessary and sufficient conditions for the punishment function to be effective. Notably, this Lemma may be used to characterize the conditions that γ and π need to satisfy for the punishment function h_i to be effective.

In Section 4 we consider the case where the punishment h is fixed, for instance, when the government has already instituted a lemon law that awards damages upon discovery of the misreporting. For a profit-maximizing investor who cares only about the maximum amount collected from the agents, Theorem 2 discovers the unique h-strategy-proof mechanism that maximizes the investor's profit under a marginally non-decreasing condition on the punishment function. Notably, and similar to the example above, this mechanism would be a combination mechanism that charges the top agent an amount between $\hat{\theta}_1$ and $\hat{\theta}_2$. This provides a novel characterization of the combination mechanism.

Finally, in Section 5 we consider a situation in which a planner is interested in implementing a given mechanism, for instance, where there is a *customary* pre-determined share of the profits between investors and agents. These customary shares are typical in partnerships and often involve an equal share of the profit or some combination based on the power of the agents. The paper studies the minimal punishment h that needs to be contractually implemented to make the mechanism h-strategy-proof.

To illustrate this, suppose that the planner is allowed to charge $t_1(\hat{\theta}) = \alpha\hat{\theta}_1 + (1-\alpha)\hat{\theta}_2$.

Clearly, for $0 < \alpha \leq 1$, this mechanism is not ex-ante strategy-proof. However, for a large enough punishment function h, this mechanism is h-strategy-proof. Indeed, consider punishment function $h_i(\theta_i, \hat{\theta}_i) = \alpha|\theta_i - \hat{\theta}_i|$, if agent 1 with highest production reports $\hat{\theta}_1$, the profit $(1-\alpha)(\theta_1 - \hat{\theta}_2) + \alpha(\theta_1 - \hat{\theta}_1) - \alpha|\theta_1 - \hat{\theta}_1|$ is maximized with truthful report $\hat{\theta}_1 = \theta_1$.

In general, the last result of the paper shows the existence of a minimal punishment that the planner needed to implement to make an arbitrary mechanism incentive-compatible (Lemma 2). In addition, when the planner cares about first-best efficiency, the first-best efficiency allocation for the planner is $x_1(\hat{\theta}) = 1$ and $t_1(\hat{\theta}) = \hat{\theta}_1$ in the example above, the minimal punishment to achieve the first-best efficiency is $h_i(\theta_i, \hat{\theta}_i) = |\theta_i - \hat{\theta}_i|$. Proposition 1 characterizes the minimal punishment function to achieve first-best efficiency for general cases.

1.2 Related Literature

Most of the social choice literature has been concerned with non-manipulable mechanisms dating back from Arrow[1] and Gibbard[15]. See, Barberà[2], for an introduction to strategy-proof social choice functions. Such studies include the case of strategy-proof social choice functions in classical exchange economies (Barberà and Jackson[4]), matching with contracts (Hatfield and Kojima[20]), house allocation with prices (Miyagawa[29]), cost and resource sharing (Moulin and Shenker[31], Moulin[30], Sprumont[37], Juarez[22, 23], You and Juarez[39]), preference aggregation (Bossert and Sprumont[9]), social choice (Barberà, Dutta and Sen[3]). In particular, incentive compatibility, herein interpreted as (ex-ante) strategy-proofness has been widely explored when money is available. The traditional VCG mechanisms in Vickrey[38], Clarke[13], Groves[16] are characterized by Holmstrom[21] by ex-ante strategy-proof and efficient.

Although this extensive literature has been prolific, it has lacked the incorporation of punishments to induce behavior. Indeed, monetary punishments and penalties have been used for centuries as a deterrent for unwanted behavior. For instance, the deterrence hypothesis which dates back to at least the XVIII century (Beccaria[5], Bentham[7]), claims that *the introduction of penalties that leave everything else unchanged reduces the occurrence of the behavior* has been the cornerstone of legal, criminal and psychological studies. Theoretical work on fines has been more prominently applied to game-theoretical studies of antitrust fines to prevent collusion in auctions and cartels (Buccirossi and Spagnolo[10], Harrington and Chang[18], Katsoulacos et al.[24]) or fishery markets (Nøstbakken[34]). Notably, our work provides a general framework to study the effects of punishments and penalties in resource allocation problems, including auctions and fisheries.

In addition to this theoretical work, there is recent literature on costly verification with punishment. Ben-Porath et al.[6] study the problem of a principal allocating an indivisible good among agents without transfer payments. The principal can learn the types of agents with cost. If agents are found to lie, the good will not be allocated as punishment. Mylovanov and Zapechelnyuk[32] study the optimal allocation with verification and limited penalties, i.e. the principal can punish the agent for lying by recovering part of the prize. Li[26] studies mechanism design with costly verification and limited punishment, in which the punishment is a linear function of the agents' benefit. In contrast with this literature, we assume costless verification but widely expand the range of punishment functions available. Indeed, we study the mechanism design problem when the planner is allocating a divisible good with exogenous probabilistic verification and various forms of punishments. We discover the class of strategy-proof mechanisms under different punishments, characterize optimal mechanisms for the planner depending on the punishment functions, and study the minimal punishment necessary to achieve first-best efficiency.

We also note that recent literature has been focused on various relaxations and strengthening of ex-ante strategy-proofness. There are approximately strategy-proof mechanisms in voting (Birrell and Pass[8]), matching (Pathak and Sönmez[35]), and more generally, Carroll[11] finds that local strategy-proofness with single-crossing ordinal preferences implies full strategy-proofness. Obviously strategy-proof mechanisms in Li[25] refine the strategy-proof mechanisms by requiring the strategy to be obviously dominant. Pathak and Sönmez[35] develop a rigorous methodology to compare mechanisms based on their vulnerability to manipulation. Unlike this literature on ex-ante strategy-proofness, our notion of manipulation depends on the punishment function h. This allows for a different notion of manipulation that expands the class of strategy-proof mechanisms that a planner can use, hence providing flexibility for the planner when selecting mechanisms.

2 The Model

Each agent $i \in \mathcal{N} = \{1, \ldots, N\}$ is endowed with a **production function** $f_i : \mathbb{R}_+^2 \to \mathbb{R}_+^M$ that generates an outcome $f_i(\theta_i, x_i) \in \mathbb{R}_+^M$ based on his ability $\theta_i \in \mathbb{R}_+$ and the amount of resource allocated by the planner $x_i \in \mathbb{R}_+$ to agent i.[1] We study the case where each production function f_i is continuous and twice differentiable. We also assume that θ_i and x_i are complements, that is, $\frac{\partial^2 f_{ij}(\theta_i, x_i)}{\partial \theta_i \partial x_i} \geq 0$ for all θ_i and x_i.

A planner is endowed with 1 unit of resource that can be allocated among the agents

[1] If agent i is allocated zero resource by the planner, his production is $f_i(\theta_i, 0) = \mathbf{0} = (0, \ldots, 0)_{M\times 1}$. Moreover, the agent with lower bound of ability produces $f_i(0, x_i) = \mathbf{0}$.

to produce the outcome. Let $\Delta_+^N = \{x \in \mathbb{R}_+^N | \sum_j x_j = 1\}$ be the set of allocations of this resource to the agents. We assume that there is asymmetric information, the agents know their own abilities but do not know others'. The planner does not know the agents' abilities. We study mechanisms where the agents reveal their abilities to the planner, who makes an allocation of the resource among agents and charges outcomes based on their reports. This is formalized in the following definition.

Definition 1 (Mechanism)
A mechanism $\phi = (x(\cdot), t(\cdot))$ is a pair of functions $(x(\cdot), t(\cdot))$ such that

i. $x : \mathbb{R}_+^N \mapsto \Delta_+^N$ allocates the share of a resource to every agent based on the vector of reported abilities θ, and $x(\theta) = (x_1(\theta), \ldots, x_N(\theta))$, the amount $x_i(\theta) \in \mathbb{R}_+$ represents the resource allocated to agent i.

ii. $t : \mathbb{R}_+^N \mapsto \mathbb{R}^{N \times M}$ is the tax of the outcome produced by the agents that the planner charges. Thus, for ability θ and $t(\theta) = (t_1(\theta), \ldots, t_N(\theta))$, the vector $t_i(\theta) \in \mathbb{R}_+^M$ is the resource charged by the planner to agent i.

For the rest of paper, we assume $x(\theta)$ and $t(\theta)$ are fully differentiable with respect to θ, except at some points with measure 0.

The planner is self-interested and cares only about the vector of taxes $\sum_{i \in N} t_i(\theta) \in \mathbb{R}_+^M$.[2] *However, the agents care about the total outcomes produced that are not charged by the planner. That is, for a vector of reports $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_N)$ and true ability of agent i is θ_i, his payment equals $\sum_{j=1}^M f_{ij}(\theta_i, x_i(\hat{\theta})) - t_{ij}(\hat{\theta}) = (f_i(\theta_i, x_i(\hat{\theta})) - t_i(\hat{\theta}))^T \mathbf{1}$, where $\mathbf{1} = (1, \ldots, 1)_{M \times 1}$ is the unitarian vector. We denote by $\bar{f}_i(\theta_i, x_i) = f_i(\theta_i, x_i)^T \mathbf{1}$ the aggregate production of agent i and by $\bar{t}_i(\theta) = t_i(\theta)^T \mathbf{1}$ the aggregate tax of agent i. The mechanism ϕ is individually rational if $(f_i(\theta_i, x_i(\theta)) - t_i(\theta))^T \mathbf{1} \geq 0$. The mechanism satisfies no positive transfers if any agent does not receive any resource $x_i = 0$, then the agent is not charged $t_i = 0$. Here we focus on the mechanisms of individual rationality and no positive transfers.*[3]

The generality of our model allows us to capture and integrate a variety of settings that have been studied separately. Some of these are discussed in the next example.

Example 1

1. ***Private equity investment:*** *A private investor is endowed with a fixed level of capital that he can invest in projects. Each project $i \in \mathcal{N}$ can produce M goods with the production function $f_i(\theta_i, x_i)$, which depends on the ability of the project θ_i and*

[2] We allow for a wide range of preferences over the vector of goods, for instance, the perfect complements or perfect substitutes.

[3] These are two standard assumptions of mechanism in Hartline and Karlin[19].

investment x_i. The investor could elicit the ability from reports of the projects $\hat{\theta}$, invest $x(\hat{\theta})$ to the projects, and charges $t(\hat{\theta})$ from the outputs.

2. ***Procurement auctions:*** *The agent i produces M goods $f_i(\theta_i, x_i) \in \mathbb{R}_+^M$ based on his ability θ_i and receives resource x_i from the planner. In a mechanism, the planner allocates resource $x(\hat{\theta})$ to agents, and taxes production $t(\hat{\theta})$ from the agents based on the their reports $\hat{\theta}$ about ability of production. When $M = 1$ and $f_i(\theta_i, x_i) = \theta_i x_i$, this captures the traditional auction setting.*

3. ***Government funding with intermediation:*** *A government is interested in delivering goods to M targets via a set of N intermediaries (e.g., charities). The charities report their ability $\hat{\theta}$ of resource transmission, and the maximal amount of goods transmitted to targets through charity i is $f_i(\theta_i, x_i)$. The government makes an allocation $x(\hat{\theta})$ to charities based on their reports $\hat{\theta}$, and requires the targets to receive $t(\hat{\theta})$ from the intermediaries.*

Although the class of incentive-compatible mechanisms without punishment for these examples has been studied. The class of incentive-compatible mechanisms when ex-post punishment is available is not well understood. We introduce this notion in the following section.

3 h-Strategy-Proof Mechanisms

A punishment function for agent i, $h_i : \mathbb{R}_+^2 \mapsto \mathbb{R}_+$, allocates a monetary punishment $h_i(\theta_i, \hat{\theta}_i)$ when his reported ability is $\hat{\theta}_i \in \mathbb{R}_+$ but his true ability is (discovered to be) $\theta_i \in \mathbb{R}_+$. We assume there is no punishment to an agent for truthfully reporting his type, that is $h_i(\theta_i, \hat{\theta}_i) = 0$ if $\hat{\theta}_i = \theta_i$. Moreover, $h_i(\theta_i, \hat{\theta}_i)$ is non-decreasing of $\hat{\theta}_i$ when $\hat{\theta}_i \geq \theta_i$, and $h_i(\theta_i, \hat{\theta}_i)$ is non-increasing of $\hat{\theta}_i$ when $\hat{\theta}_i \leq \theta_i$. We denote by $h : \mathbb{R}_+^{2N} \mapsto \mathbb{R}_+^N$ the punishment functions and by $h(\theta, \hat{\theta}) = (h_1(\theta_1, \hat{\theta}_1), \ldots, h_i(\theta_i, \hat{\theta}_i), \ldots, h_N(\theta_N, \hat{\theta}_N))$ the vector of punishment functions when the true abilities are θ and reported abilities are $\hat{\theta}$.

We consider the punishment function $h_i(\theta_i, \hat{\theta}_i)$, which is differentiable on $\hat{\theta}_i$, except at the true ability $\hat{\theta}_i = \theta_i$. This derivative is denoted by $h'_{i2}(\theta_i, \hat{\theta}_i) = \lim_{q \to 0} \frac{h_i(\theta_i, \hat{\theta}_i + q) - h_i(\theta_i, \hat{\theta}_i)}{q}$. Furthermore, we assume the upper and lower limits at $\hat{\theta}_i = \theta_i$ exist, and such limits are denoted by $h'_{i2+}(\theta_i, \theta_i) = \lim_{\hat{\theta}_i \to \theta_i^+} \frac{h_i(\theta_i, \theta_i) - h_i(\theta_i, \hat{\theta}_i)}{\theta_i - \hat{\theta}_i}$ and $h'_{i2-}(\theta_i, \theta_i) = \lim_{\hat{\theta}_i \to \theta_i^-} \frac{h_i(\theta_i, \theta_i) - h_i(\theta_i, \hat{\theta}_i)}{\theta_i - \hat{\theta}_i}$.

Definition 2

The punishment function h_i is marginally non-increasing on the true ability (MNI), if for

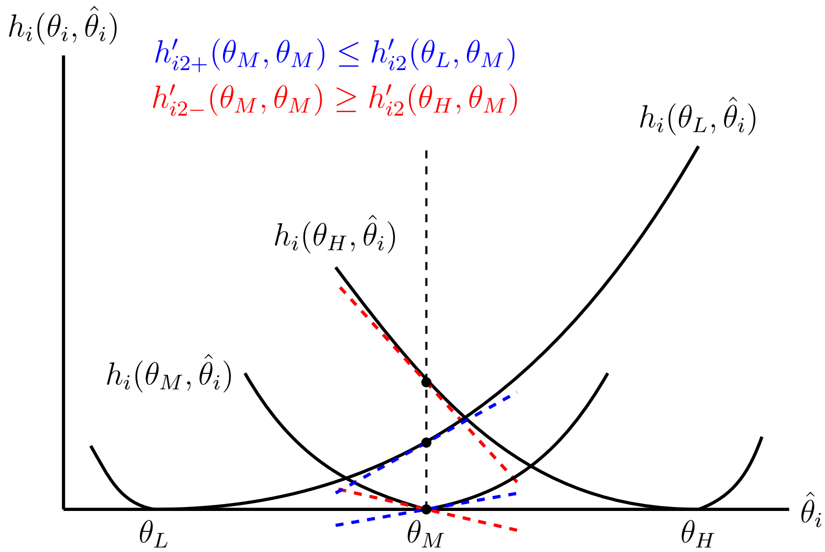

Figure 2: Marginally non-increasing

any three reports $\theta_L < \theta_M < \theta_H$*,*

$$h'_{i2}(\theta_L, \theta_M) \geq h'_{i2+}(\theta_M, \theta_M) \geq h'_{i2-}(\theta_M, \theta_M) \geq h'_{i2}(\theta_H, \theta_M)$$

Figure 2 illustrates MNI for three reports, $\theta_L < \theta_M < \theta_H$ in comparison to the report θ_M. This property is related to concavity on the cross-derivative, and it is satisfied in a variety of punishment functions. For instance, when the punishment function only depends on the deviation from the truthful report $h_i(\theta_i, \hat{\theta}_i) = f_i(\theta_i - \hat{\theta}_i)$ and f_i is weakly convex (e.g., $f_i(\theta_i - \hat{\theta}_i) = |\theta_i - \hat{\theta}_i|$ or $f_i(\theta_i - \hat{\theta}_i) = (\theta_i - \hat{\theta}_i)^2$). Alternatively, this is satisfied when the cross-derivative is not positive:

$$\frac{\partial^2 h_i(\theta_i, \hat{\theta}_i)}{\partial \theta_i \partial \hat{\theta}_i} \leq 0.$$

This definition is extended to take into account the potential discontinuity of the first order derivative at the true report $\hat{\theta}_i = \theta_i$, see Figure 2 for illustration. Although we don't assume MNI for a punishment function, this property will be important to guarantee the sufficiency of a couple important results.

Definition 3 (h-Strategy-Proof)
The mechanism $\phi = (x(\cdot), t(\cdot))$ *is h-strategy-proof (h-SP) if for any agent* i *of ability* θ_i *reports* $\hat{\theta}_i$*, the following condition holds:*

$$\bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta) \geq \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{t}_i(\hat{\theta}_i, \theta_{-i}) - h_i(\theta_i, \hat{\theta}_i), \forall \theta_i, \hat{\theta}_i, \theta_{-i}.$$

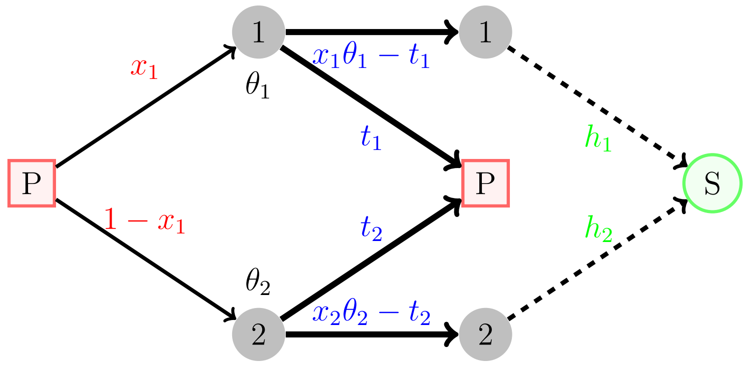

Figure 3: Two agents and one good.

h-strategy-proof mechanisms captures the planners' ability to punish agent i of true ability θ_i, who misreports his ability $\hat{\theta}_i$, with the amount of punishment $h_i(\theta_i, \hat{\theta}_i)$.

Example 2 (Linear Production)
Consider a planner who allocates 1 unit of resource to agents, who produce one kind of good with resource. The agents have linear production function $f_i(\theta_i, x_i) = \theta_i x_i$. Agent i reports his ability $\hat{\theta}_i$ to the planner, then the planner allocates resource $x_i(\hat{\theta})$ to agent i, and charges product t_i from agent i. The profit of agent i is $x_i(\hat{\theta})\theta_i - t_i(\hat{\theta})$. See Figure 3 for the case of two agents. Assume the allocation rule x allocates all the resource to the agent of highest reported ability. Without loss of generality, assume $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_N$, and agent $i \geq 2$ reports truthfully $\hat{\theta}_{-1} = \theta_{-1}$. Consider the report of agent 1, if agent 1 has highest report, $x_i(\hat{\theta}) = 0$ for any $i \geq 2$, no positive transfers requires $t_i(\hat{\theta}) = 0$. The condition of individual rationality requires $t_1(\hat{\theta}) \leq \hat{\theta}_1$.

If there is no punishment $h_i = 0$, the h-SP mechanism is ex-ante SP. To induce the truthful report of agent 1, the charge to agent 1 satisfies $\theta_1 - t_1(\hat{\theta}) \leq \theta_1 - t_1(\theta_1, \hat{\theta}_{-1})$, $\forall \hat{\theta}_1 \geq \theta_2$, then $\hat{\theta}_1 \geq t_1(\hat{\theta}_1, \hat{\theta}_{-1}) \geq t_1(\theta_1, \hat{\theta}_{-1})$. If $t_1(\hat{\theta}_1, \hat{\theta}_{-1}) > t_1(\theta_1, \hat{\theta}_{-1})$, then agent 1 has incentive to misreport θ_1, when his true ability is $\hat{\theta}_1$. Thus, $t_1(\hat{\theta}_1, \hat{\theta}_{-1}) = t_1(\theta_1, \hat{\theta}_{-1}) \leq \theta_2$. If $t_1(\theta_1, \hat{\theta}_{-1}) < \theta_2$, there exists $\tilde{\theta}_1$, such that $t_1(\theta_1, \hat{\theta}_{-1}) < \tilde{\theta}_1 < \theta_2$. Consider the case that true abilities of agents are $\theta_2 \geq \tilde{\theta}_1 \geq \theta_3 \cdots \geq \theta_N$, the abilities of agent 1 and 2 are $\tilde{\theta}_1$ and θ_2. Then agent 1 has incentive to misreport θ_1 and gets positive profit $\tilde{\theta}_1 - t_1(\theta_1, \hat{\theta}_{-1})$. Thus, the ex-ante SP mechanism charges $t_1(\hat{\theta}) = \theta_2$, which is the second price mechanism.

Lemma 1 (Conditions for h-SP)
If a mechanism $\phi = (x(\cdot), t(\cdot))$ is h-SP, then there exists a function $\Phi : \mathbb{R}_+^N \mapsto \mathbb{R}_+^N$ such that:

i. The aggregate charge to agent i equals $\bar{t}_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - \Phi_i(\theta)$.

ii. For each i and θ_i, $\bar{f}_{i\theta}(\theta_i, x_i(\theta)) + h'_{i2-}(\theta_i, \theta_i) \leq \frac{\partial \Phi_i(\theta_i, \theta_{-i})}{\partial \theta_i} \leq \bar{f}_{i\theta}(\theta_i, x_i(\theta)) + h'_{i2+}(\theta_i, \theta_i)$ with $\bar{f}_{i\theta}(\theta_i, x_i(\theta)) = \frac{\partial \bar{f}_i(\theta_i, x_i(\theta))}{\partial \theta_i}$.

Conversely, if the punishment functions h satisfies MNI and there exists a function Φ that satisfies conditions i and ii, then the mechanism is h-SP.

From part i, the function

$$\Phi_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta)$$

is the profit of agent i when he truthfully reports $\hat{\theta}_i = \theta_i$ at the profile θ. From part ii, the boundary of marginal profit is related to marginal production and marginal punishment.

These are local conditions of h-SP for deviation of θ to $\hat{\theta}$. The conditions are not sufficient for a mechanism to be h-SP, and global conditions of h-SP are needed. The profit of agent i is affected by aggregate production $\bar{f}_i$ regardless specific dimension of production f_{ij}.

Theorem 1 (Ineffective Punishment Functions)
Consider the mechanism $\phi = (x(\cdot), t(\cdot))$ such that the allocation rule $x_i : \mathbb{R}_+^N \mapsto \mathbb{R}_+$ is non-decreasing in the report $\hat{\theta}_i$ for all i. The following three conditions are equivalent:

i. The mechanism ϕ is ex-ante SP.

ii. The mechanism ϕ is h-SP for any punishment function $h_i(\theta_i, \hat{\theta}_i)$ such that the derivative at the truthful report is zero, that is $h'_{i2}(\theta_i, \theta_i) = 0$ for any θ_i.[4]

iii. The marginal profit of an agent equals to his marginal product. That is, $\frac{\partial \Phi_i(\theta)}{\partial \theta_i} = \frac{\partial \bar{f}_i(\theta_i, x_i(\theta))}{\partial \theta_i}$ for any report θ and any agent i.

The proof of this result is in the appendix. The intuition of the proof is as follows. In order to prove that part (i) implies part (iii), under the condition of individual rationality and no positive transfers, given any allocation rule, an ex-ante SP mechanism can be characterized by a local condition, that requires no incentive for marginal deviation. As we take the limit of this local condition, we arrive to part (iii). For the converse, as the allocation to an agent is non-decreasing in his reported ability, the complementarity of the resource and ability in production guarantees the mechanism satisfying a local condition of SP, which in turn satisfied a global condition of SP.

There are two important consequences of Theorem 1. On the one hand, it provides precise conditions for mechanisms to be ex-ante strategy-proof, which is discussed in the

[4]This happens, for instance, at the large class of polynomial punishment functions $h_i(\theta_i, \hat{\theta}_i) = \gamma|\theta_i - \hat{\theta}_i|^k$ with $k > 1$.

literature. First, the resource allocated to an agent depends on his aggregate ability of production $\bar{f}_i$. Second, the charged share of an agent depends on the average production over all the abilities $\bar{f}_i(\theta_i, x_i(\theta)) - \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq$. On the other hand, it shows that the punishment functions are effective only if there is a positive punishment at the margin (i.e., for an arbitrary small deviation). In other words, the punishment function h, whose derivative at the truthful report is zero, is ineffective, in the sense that all h-SP mechanisms are ex-ante SP.

Example 3 (Linear Production (Continued))
From Example 2, the agents have linear production function $f_i(\theta_i, x_i) = \theta_i x_i$, consider the allocation rule x_i allocates all the resource to the agent of highest reported ability. Assume the punishment function is $h_i(\theta_i, \hat{\theta}_i) = \pi(|\theta_i - \hat{\theta}_i|^\gamma)$, and $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_N$.

Consider $\gamma > 1$, $\pi > 0$, the h-SP requires $\theta_1 - t_1(\hat{\theta}) - \pi(|\theta_1 - \hat{\theta}_1|)^\gamma \leq \theta_1 - t_1(\theta_1, \hat{\theta}_{-1})$, for any $\hat{\theta}$, equivalent with $t_1(\theta_1, \hat{\theta}_{-1}) - t_1(\hat{\theta}) \leq \pi(|\theta_1 - \hat{\theta}_1|)^\gamma$. For $\theta_1 \geq \hat{\theta}_1$, $t_1(\theta_1, \hat{\theta}_{-1}) - t_1(\hat{\theta}) \leq \pi(\theta_1 - \hat{\theta}_1)^\gamma$, take the limit $\hat{\theta}_1 \to \theta_1$, $\lim_{\hat{\theta}_1 \to \theta_1^-} \frac{t_1(\theta_1, \hat{\theta}_{-1}) - t_1(\hat{\theta})}{\theta_1 - \hat{\theta}_1} \leq \lim_{\hat{\theta}_1 \to \theta_1^-} \frac{\pi(\theta_1 - \hat{\theta}_1)^\gamma}{\theta_1 - \hat{\theta}_1} = 0$. For $\theta_1 \leq \hat{\theta}_1$, there is $\lim_{\hat{\theta}_1 \to \theta_1^+} \frac{t_1(\theta_1, \hat{\theta}_{-1}) - t_1(\hat{\theta})}{\theta_1 - \hat{\theta}_1} \geq \lim_{\hat{\theta}_1 \to \theta_1^+} \frac{\pi(\hat{\theta}_1 - \theta_1)^\gamma}{\theta_1 - \hat{\theta}_1} = 0$. Then we have $\lim_{\hat{\theta}_1 \to \theta_1} \frac{t_1(\theta_1, \hat{\theta}_{-1}) - t_1(\hat{\theta})}{\theta_1 - \hat{\theta}_1} = 0$, which is the charge of second highest production, the ex-ante SP mechanism. In this case, the h-SP class of mechanism does not expand with h.

Consider $\gamma = 1$, $\pi = 1$. The punishment function is $h_i(\theta_i, \hat{\theta}_i) = |\theta_i - \hat{\theta}_i|$, the agents have no incentive to lie in the first-price mechanism, $t_1(\hat{\theta}) = \hat{\theta}_1$. Indeed, if agent 1 reports $\hat{\theta}_1$ lower than the true ability θ_1, his profit is $x_1\theta_1 - \hat{\theta}_1 - |\theta_1 - \hat{\theta}_1| = 0$. If agent 1 reports $\hat{\theta}_1$ higher than the true ability θ_1, he receives negative profit. Thus, there is no incentive for agent 1 to misreport, and the first-price mechanism is h-SP. The set of h-SP mechanisms with punishment h expands the set of ex-ante SP mechanisms.

Consider $0 < \gamma < 1$, $\pi = 1$. $h'_{i2-}(\theta_i, \theta_i) = \infty$, consider the charge $t_1(\hat{\theta}) = \hat{\theta}_2 + |\hat{\theta}_1 - \hat{\theta}_2|^\gamma$. Assume agents $2, \ldots, N$ report truthfully $\hat{\theta}_{-1} = \theta_{-1}$. The profit of agent 1 reporting $\hat{\theta}_1$ is $\theta_1 - \theta_2 - |\hat{\theta}_1 - \theta_2|^\gamma - |\theta_1 - \hat{\theta}_1|^\gamma$. If agent 1 reports truthfully, he receives $\theta_1 - \theta_2 - |\theta_1 - \theta_2|^\gamma$. For $\hat{\theta}_1 > \theta_1$, $|\hat{\theta}_1 - \theta_2|^\gamma > |\theta_1 - \theta_2|^\gamma$, agent 1 has no incentive to report higher. For $\theta_2 \leq \hat{\theta}_1 \leq \theta_1$, $|\theta_1 - \theta_2|^\gamma \leq |\hat{\theta}_1 - \theta_2|^\gamma + h_1(\theta_1, \hat{\theta}_1) = |\hat{\theta}_1 - \theta_2|^\gamma + |\theta_1 - \hat{\theta}_1|^\gamma$.[5] Thus, agent 1 reports truthfully $\hat{\theta}_1 = \theta_1$. The set of h-SP mechanisms with punishment h expands the set of ex-ante mechanisms.

Remark 1
For any punishment functions h, $\tilde{h}$, s.t. $h(\theta, \hat{\theta}) \leq \tilde{h}(\theta, \hat{\theta})$, $\forall \theta, \hat{\theta} \in \mathbb{R}_+^M$. Then, if the mechanism $\phi = (x(\cdot), t(\cdot))$ is h-SP, ϕ is $\tilde{h}$-SP.

[5]The inequality could be proved by dividing both sides by $\theta_1 - \theta_2$, to prove $x^\gamma + (1-x)^\gamma \geq 1$, for $x \in [0,1]$, $\gamma \in (0,1]$. $\frac{dx^\gamma}{d\gamma} = x^\gamma \ln x \leq 0$, monotonic decreasing for γ. $\gamma = 1$ is minimal, $x + (1-x) = 1$.

This remark presents the comparative static analysis of h-SP mechanisms. As the punishment function h increases, the set of h-SP mechanisms expands. The result is consistent with intuition that punishments decrease the incentives of agents to misreport. Furthermore, note that if the punishment function $h(\theta, \hat{\theta})$ approaches infinity, any mechanism is h-SP.

Until now, our study has focused on characterizing the class of h-SP mechanisms without considering the planner's preferences. The next section considers and characterizes optimal mechanisms for planner.

4 h-Optimal Mechanism

Recall that the planner cares about the *total tax* collected from the agents $\tau(\hat{\theta}) = \sum_{i=1}^{N} t_i(\hat{\theta}) \in \mathbb{R}_+^M$, where the agents report $\hat{\theta}$. In this section, we impose a minimal restriction on the preferences, namely, we assume that the planner's preferences $\succeq$ over τ are monotonic.

For an arbitrary set of planner's preferences, we find mechanisms that are optimal given a fixed punishment function h in this section. Our main result shows that the optimal mechanism for the planner, given an arbitrary punishment function, can be represented as a convex combination of two mechanisms: (i) full-charge mechanism, which charges the entire production of the agents, (ii) the truth-inducing mechanism, which is ex-ante SP. The weights of these two mechanisms depend on marginal punishments and marginal production of agents.

Two mechanisms are salient for the main results of this section. First, the full-charge mechanism ϕ^F given an allocation rule x charges the production $t_i^F(\theta) = f_i(\theta_i, x_i(\theta))$ to all the agents. When the allocation rule x allocates all resources to the agent of highest ability, this mechanism equals to the traditional first-price mechanism.

Second, the truth-inducing mechanism, ϕ^T given an allocation $x(\cdot)$ is an ex-ante SP mechanism for allocation rule x, the planner charges agent i with $\bar{t}_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq$. When the allocation rule x allocates all resources to the agent of highest ability, this mechanism equals to the traditional second-price mechanism.

Our following result characterizes the h-optimal mechanism ϕ^h as a convex combination of ϕ^F and ϕ^T for any punishment function h.

In particular, note that ϕ^T is ex-ante SP. Indeed, the agents are taxed only part of their production so that they have incentive to report truthfully. Without punishments, the tax is optimal for planner. The full-charge mechanism ϕ^F is not ex-ante SP, the agents are taxed the production according to their reports. Without punishments, the agents have incentive to under-report their abilities. However, if the punishments are high, the agents have no incentive to misreport. The conditions of individual rationality require that agents are not taxed more than their production under truthful reports.

Definition 4
*Consider an arbitrary monotonic preferences $\succeq$ of planner, and let h be an arbitrary punishment function. We say an h-SP mechanism ϕ is h-**optimal** if for any h-SP mechanism $\hat{\phi}$ with the same allocation rule x, we have that $\tau(\theta) \succeq \hat{\tau}(\theta)$ for any θ.*

The definition of h-optimal restricts the optimality of mechanism within a subset of mechanisms with same allocation rule x. The mechanisms with different x might not be comparable, see the following example.

Example 4 (Linear Production (Continued))
Consider the agents with linear production function $f_i(\theta_i, x_i(\theta)) = \theta_i x_i \in \mathbb{R}_+$. Assume there is no punishment, $h_i(\theta_i, \hat{\theta}_i) = 0$, for any θ_i, $\hat{\theta}_i$. The h-SP mechanism is ex-ante SP. Without loss of generality, assume θ satisfies $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_N$.

From Theorem 1, the ex-ante SP mechanism satisfies $\frac{\partial \Phi_i(\theta_i, \theta_{-i})}{\partial \theta_i} = \bar{f}_{i\theta}(\theta_i, x_i(\theta))$, and $\bar{t}_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq$.

We study two allocation rules. First, the planner allocates all the resource to the agent of highest reported ability, from Example 2, the truth-inducing mechanism ϕ^T is second-price mechanism. Under truthful reports θ, agent 1 receives the resource $x_1 = 1$, and planner charges θ_2 from agent 1.

Second, the allocation rule $\tilde{x}$ satisfies $\tilde{x}_i(\hat{\theta}) = \frac{1_{[\theta_2, \theta_1)}(\hat{\theta}_i)}{\sum_{i=1}^N 1_{[\theta_2, \theta_1)}(\hat{\theta}_i)}$ for reports $\hat{\theta}$ with $\theta_2 \leq \max_{i \in \mathcal{N}} \hat{\theta}_i < \theta_1$,[6] and $\tilde{x}(\hat{\theta}) = x^T(\hat{\theta})$ for other reports.

If the highest report is in $[\theta_2, \theta_1)$, $\tilde{x}$ allocates resource equally to all agents in $[\theta_2, \theta_1)$. If the highest report is not in $[\theta_2, \theta_1)$, all the resource is allocated to the agent with highest reported ability. From Theorem 1, the ex-ante SP mechanism $\tilde{\phi}$ with allocation rule $\tilde{x}$ satisfies $\tilde{t}_i(\theta) = \theta_i x_i(\theta) - \int_0^{\theta_i} x_i(q, \theta_{-i})dq$.

When true abilities of agents are θ, $\tilde{x}_1(\theta) = 1$, agent 1 is charged $\tilde{t}_1(\theta) = \theta_1 - \int_0^{\theta_1} x_1(q, \theta_{-i})dq = \theta_1 - \int_{\theta_2}^{\theta_1} \frac{1}{2}dq = \frac{\theta_1 + \theta_2}{2}$. Thus, $\tilde{t}_1(\theta) > \theta_2 = t_1^T(\theta)$, the planner prefers $\tilde{\phi}$.

When true abilities of agents are θ' with $\theta_2 < \theta_2' < \theta_1' < \theta_1$ and $\theta_j' = \theta_j$ for $j \geq 3$. $\tilde{\phi}$ charges $\tilde{t}_1(\theta') = \tilde{t}_2(\theta') = \frac{1}{2}\theta_2$, ϕ^T charges $t_1^T(\theta') = \theta_2'$. Thus, total tax $\tilde{\tau} = \theta_2 < \theta_2' = \tau^T$, the planner prefers ϕ^T. So neither of these two mechanisms is optimal for any θ.

In Example 4, we find that the h-SP mechanisms with different allocation rules might not be comparable for the planner. In the following example, we study the optimal mechanism

[6] $1_{[\theta_2, \theta_1)}(\cdot)$ is an indicator function.

for the planner, given a specific punishment function and the allocation rule that sends all the resource to the agent with highest reported ability.

Example 5 (Linear Production (Continued))
Consider the agents with linear production function $f_i(\theta_i, x_i) = \theta_i x_i$, the allocation rule x allocates all the resource to the agent of highest reported ability, and the punishment function is $h_i(\theta_i, \hat{\theta}_i) = \pi|\theta_i - \hat{\theta}_i|$. Without loss of generality, assume $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_N$. In this case, the full-charge mechanism ϕ^F is the first-price mechanism, which charges $t_1^F = \hat{\theta}_1$. The truth-inducing mechanism ϕ^T is the second-price mechanism, which charges $t_1^T = \hat{\theta}_2$.

If $\pi = 0$, ϕ^F is not h-SP from Example 3. The truth-inducing mechanism ϕ^T maximizes the total tax for the planner in the set of h-SP mechanisms.

If $0 < \pi < 1$, assume agents $2, \ldots, N$ report truthfully $\hat{\theta}_{-1} = \theta_{-1}$. The profit of agent 1 reporting $\hat{\theta}_1$ is $\theta_1 - t_1(\hat{\theta}) - \pi|\theta_i - \hat{\theta}_i|$. The truth-inducing mechanism ϕ^T charge $\hat{\theta}_2$ is h-SP. The condition of individual rationality requires that the charge is not more than the production $\hat{\theta}_1$. Thus, assume the planner charge $t_1 \in [\hat{\theta}_2, \hat{\theta}_1]$ in optimal mechanism, let $t_1(\hat{\theta}) = \delta\hat{\theta}_1 + (1-\delta)\hat{\theta}_2$, $\delta \in [0, 1]$. If agent 1 reports truthfully, he receives $(1-\delta)\theta_1 - (1-\delta)\hat{\theta}_2$. If agent 1 reports $\hat{\theta}_1 > \theta_1$, his profit is $(1+\pi)\theta_1 - (\delta+\pi)\hat{\theta}_1 - (1-\delta)\hat{\theta}_2$, which is less than the profit of truthful report. If agent 1 reports $\hat{\theta}_1 \leq \theta_1$, his profit is $(1-\pi)\theta_1 + (\pi-\delta)\hat{\theta}_1 - (1-\delta)\hat{\theta}_2$. To induce truthful report, there is $(1-\pi)\theta_1 + (\pi-\delta)\hat{\theta}_1 - (1-\delta)\hat{\theta}_2 \leq (1-\delta)\theta_1 - (1-\delta)\hat{\theta}_2$, equivalent with $(\pi-\delta)(\theta_1 - \hat{\theta}_1) \geq 0$, so $\delta \leq \pi$. Thus, the optimal mechanism of the planner charges $t_1(\hat{\theta}) = \pi\hat{\theta}_1 + (1-\pi)\hat{\theta}_2$.

In Example 5, we discover that the optimal mechanism is a convex combination of the first-price mechanism and second-price mechanism for a specific allocation rule and punishment function. In the following definition, we formally define the combination mechanisms based on the truth-inducing mechanism and full-charge mechanism, given any allocation rule and punishment function.

Definition 5
Consider an allocation rule x and an arbitrary punishment function h, let

$$\lambda_i^h(\theta) = \min\{\frac{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} -h'_{i2-}(q,q)dq}{\int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq}, 1\},$$

with $\underline{\theta}_i(\theta_{-i}) = \sup\{\theta_i | x_i(\theta_i, \theta_{-i}) = 0\}$. Define the combination mechanism ϕ^h for the punishment function h as $\phi^h = \lambda^h(\theta)\phi^F + (1-\lambda^h(\theta))\phi^T$, with $x^h(\theta) = x^T(\theta) = x^F(\theta)$, and $\bar{t}_i^h(\theta) = \lambda_i^h(\theta)\bar{t}_i^F(\theta) + (1-\lambda_i^h(\theta))\bar{t}_i^T(\theta)$.

The following theorem shows the optimality of the combination mechanisms.

Theorem 2 (h-Optimal Mechanism)
Consider an allocation rule x and a punishment function h that satisfies MNI, the combination mechanism ϕ^h is h-optimal.

Theorem 2 characterizes the h-optimal mechanism, which is optimal for the planner given any allocation rule x. In the following corollaries, we study the mechanisms with a special allocation rule, that allocates all resource to the agent with highest ability. Furthermore, assume the production function satisfies constant return to scale $f_i(\theta_i, x_i(\theta)) = x_i(\theta) f_i(\theta_i, 1)$.

$\underline{\theta}_i(\theta_{-i})$ satisfies $\bar{f}_i(\underline{\theta}_i(\theta_{-i}), 1) = \max_{j \neq i} \bar{f}_j(\theta_j, 1)$. For $\theta_i > \underline{\theta}_i(\theta_{-i})$, $x_i(\theta) = 1$. $\lambda_i^h(\theta)$ satisfies $\lambda_i^h(\theta) = \frac{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} -h'_{i2-}(q,q)dq}{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q,1)dq}$. The full-charge mechanism charges $\bar{t}_i^F(\theta) = \bar{f}_i(\theta_i, x_i(\theta))$, and the truth-inducing mechanism charges $\bar{t}_i^T(\theta) = \bar{f}_{-i}(\theta_{-i}, 1) x_i(\theta)$. For any punishment function h, the mechanism ϕ^h with charge $\bar{t}_i^h(\theta) = \lambda_i^h(\theta)\bar{t}_i^F(\theta) + (1 - \lambda_i^h(\theta))\bar{t}_i^T(\theta)$ is h-optimal.

For any ineffective punishment function h characterized in Theorem 1, the truth-inducing mechanism ϕ^T is h-optimal.

Corollary 1 (Optimality of Second-Price Mechanism)
Given the allocation rule x that allocates all resource to the agent with highest ability of production, the second-price mechanism ϕ^T is h-optimal for any punishment function h with derivative equal to zero, that is $h'_{i2}(\theta_i, \theta_i) = 0$, for any θ_i.

When all resource is allocated to the agent with highest reported ability, Corollary 1 is a natural extension of Theorem 1. The class of h-SP mechanisms does not expand if the punishment function h has derivative 0. There does not exist any h-SP mechanism charge more than the truth-inducing mechanism ϕ^T, which is the second-price mechanism.

Corollary 2 (Optimality of First-Price Mechanism)
Given the allocation rule x that allocates all resource to the agent with highest ability of production, the first-price mechanism ϕ^F is h-optimal for any punishment function h that satisfies $h_i(\theta_i, \hat{\theta}_i) \geq \bar{f}_i(\theta_i, 1) - \bar{f}_i(\hat{\theta}_i, 1)$, for any $\theta_i \geq \hat{\theta}_i$.

When all resource is allocated to the agent with highest reported ability, Corollary 2 characterizes the condition of punishment function h, such that the full-charge mechanism ϕ^F, which is the first-price mechanism, is h-optimal.

5 Minimal Punishment Function

It is often the case that a mechanism to allocate goods and services is given, whereas the planner/designer of mechanism has flexibility to select the punishment h. In this section, we ask the question: *What conditions do the punishment function h need to satisfy for a given mechanism to be h-SP?* Roughly speaking, in order to answer this question, we show that for a given mechanism there exists a minimal punishment function h that makes such a mechanism h-SP.

Definition 6 (Minimal Punishment Function)
Consider a mechanism ϕ. We say that the punishment function $h^{\min} = (h_1^{\min}, \ldots, h_N^{\min})$ is a minimal punishment function for ϕ if

- *ϕ is $h^{\min}$-SP, and*
- *for any other punishment function h such that ϕ is h-SP, we have that $h_i(\theta_i, \hat{\theta}_i) \geq h_i^{\min}(\theta_i, \hat{\theta}_i)$ for all $\theta_i, \hat{\theta}_i, i$.*

We now introduce a couple of definitions to formally characterize the minimal punishment function. For any mechanism $\phi = (x(\cdot), t(\cdot))$, the aggregate charge for agent i is defined as the sum of all tax payments made to the planner, or $\bar{t}_i(\theta) = \sum_{j=1}^{M} t_{ij}(\theta)$. The profit function for agent i is his gain from misreporting $\hat{\theta}_i$ when his true type is θ_i and other agents report $\hat{\theta}_{-i}$, formally, $v_i : \mathbb{R}_+^{N+1} \mapsto \mathbb{R}_+$ equals $v_i(\theta_i, \hat{\theta}_i, \hat{\theta}_{-i}) = \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \hat{\theta}_{-i})) - \bar{t}_i(\hat{\theta}_i, \hat{\theta}_{-i})$.

Lemma 2 (Existence and Properties of the Minimal Punishment Function)

i. *For any mechanism ϕ, there exists a unique minimal punishment function. The minimal punishment function equals*

$$h_i^{\min}(\theta_i, \hat{\theta}_i) = \max_{\hat{\theta}_{-i}} \{v_i(\theta_i, \hat{\theta}) - v_i(\theta_i, \theta_i, \hat{\theta}_{-i}), 0\}$$

ii. *If the mechanism is ex-ante SP, then $h_i^{\min}(\theta_i, \hat{\theta}_i) = 0$, for any θ_i, $\hat{\theta}_i$, i.*

iii. *If the mechanism is not ex-ante SP, then the minimal punishment function is nonzero. In other words, there exists θ_i, $\hat{\theta}_i$, i, such that $h_i^{\min}(\theta_i, \hat{\theta}_i) > 0$.*

Part (i) of this result shows that the minimal punishment function exists and it charges a given agent his maximal possible profit over all reports of the other agents. Parts (ii) and (iii) clarify that minimal punishment functions equal to 0 or non-zero, for ex-ante or non-ex-ante SP mechanisms, respectively.

We now apply this concept for the implementation of the first-best efficient mechanisms for the planner.

5.1 First-best Efficient Mechanisms

For this subsection, we consider a general setting where the planner has preferences $\succeq$ over the total tax τ collected from the agents. For simplicity, we assume that these preferences are represented by a utility function $u : \mathbb{R}_+^M \mapsto \mathbb{R}_+$. The first-best efficient outcome from the perspective of planner is formalized in the following definition.

Definition 7 (First-Best Efficient (FBE))
Given the preferences of the planner $\succeq$ and the utility function $u : \mathbb{R}_+^M \mapsto \mathbb{R}_+$. A mechanism ϕ is first-best efficient (FBE), if the total tax collected by the planner τ maximizes his utility for any profile of agents θ. That is, for any θ, the mechanism taxes all production $t_i(\theta) = f_i(\theta_i, x_i)$, and the allocation x maximizes the utility of planner such that $\max_x u(\tau(\theta))$, where $\tau(\theta) = \sum_{i=1}^N t_i(\theta)$.

For any θ, assume the maximal utility of planner is $u^*(\theta) = \max_x u(\sum_{i=1}^N f_i(\theta_i, x_i))$ such that $\sum_{i=1}^N x_i = 1$. Assume $x^* : \mathbb{R}_+^N \mapsto \mathbb{R}_+^N$ is the allocation rule which maximizes the planner's utility given the profile of agents θ, that is, $x^*(\theta) = (x_1^*(\theta), \ldots, x_N^*(\theta)) = \arg\max_x u(\sum_{i=1}^N f_i(\theta_i, x_i))$. Assume $\bar{f}_{i\theta}(\theta_i, x_i(\theta)) > 0$ for any $x_i(\theta) > 0$, which means if agent i receives positive amount of resource, the aggregate production increases as his ability θ_i increases. Notice that FBE implies the planner allocates resource through agents optimally to achieve maximal utility with full-charge.

Proposition 1
Assume the preferences of the planner $\succeq$ is strongly monotonic, continuous, and there exists utility function $u : \mathbb{R}_+^M \mapsto \mathbb{R}$ representing the preferences.

i. There is no ex-ante SP and first-best efficient mechanism.

ii. For any FBE mechanism ϕ, the minimal punishment function $h^{\min}$ for ϕ to be h-SP satisfies $h_i^{\min}(\theta_i, \hat{\theta}_i) = \max\{\bar{f}_i(\theta_i, 1) - \bar{f}_i(\hat{\theta}_i, 1), 0\}$.

Proposition 1 shows that there does not exist any mechanism satisfying both ex-ante SP and FBE. It also provides the condition on the minimal punishment function for the existence of a mechanism to achieve both h-SP and FBE. The following examples show that the first-price mechanism is the first-best efficient when the production functions of agents are linear. In addition, for the linear case, the second-price mechanism is ex-ante SP, but not FBE. Thus, our work extends traditional traditional results in the auction literature to the multi-dimentional case.

Example 6 (Linear Production (Continued))
From Example 2, consider the agents with linear production function $f_i(\theta_i, x_i) = \theta_i x_i$*, and the punishment function is* $h_i(\theta_i, \hat{\theta}_i) = \pi|\theta_i - \hat{\theta}_i|^\gamma)$*. Without loss of generality, assume* $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_N$*. The utility of planner is increasing in total tax* τ*. The allocation* $x^*(\theta) = (1, 0, \ldots, 0)$ *maximizes utility of planner, and the maximal utility is* $u^*(\theta) = u(\theta_1)$*.*

In Example 5, the full-charge mechanism ϕ^F *is not ex-ante SP, but the truth-inducing mechanism* ϕ^T *is ex-ante SP.* ϕ^F *and* ϕ^T *allocate all resource to the agent with highest reported ability,* $x^F(\theta) = x^T(\theta) = (1, 0, \ldots, 0)$*.* ϕ^F *charges* $t_1^F(\hat{\theta}) = \hat{\theta}_1$*, and* ϕ^T *charges* $t_1^T(\hat{\theta}) = \hat{\theta}_2$*. The utility of planner from the full-charge mechanism* ϕ^F *is* $u(\hat{\theta}_1)$*, which is first-best efficiency, if the truthful report can be implemented. However, the truth-inducing mechanism* ϕ^T *charges* $\hat{\theta}_2$ *from agent 1 and leaves* $\theta_1 - \hat{\theta}_2$ *as information rent, utility of planner is* $u(\hat{\theta}_2) \leq u(\theta_1)$*. Thus,* ϕ^F *is FBE but not SP, and* ϕ^T *is ex-ante SP but not FBE.*

6 Discussion

We finalize the paper by noticing that the generality of our model encompasses a variety of applications previously discussed in the literature.

6.1 Procurement Auction

An agency (planner) asks a group of heterogeneous firms (agents) to report their abilities to produce goods/services at a set of different locations. Although the abilities of the firms are private information, the designer is able to elicit them and assign some resources to the firms based on them. The government is able to audit the firms after resources have been allocated and possibly punish those firms who lied about their ability. As such, understanding the class of incentive-compatible mechanisms in the presence of auditing and punishment is desirable.

Our study applies to the procurement auction, in which the planner has resources to allocate to agents for producing some goods. We focus on strategy-proof mechanisms with punishment for planner. Our results extend the widely studied VCG mechanisms in Vickrey[38], Clarke[13], Groves[16] to the case in which the planner could audit and punish the agents if agents are found lying. Theorem 1 shows that when the marginal punishment for misreporting is zero, the class of SP mechanisms does not expand, thus, these punishments are ineffective. If the planner chooses an allocation rule that allocates the objects to the agent with the highest valuation, the truth-inducing mechanism ϕ^T is the second-price mechanism, which is SP. The full-charge mechanism ϕ^F is the first-price mechanism, which charges the valuation of the agent who gets the objects.

Theorem 2 studies the effect of punishment on the welfare of the planner, and characterizes the optimal mechanism for the planner as a combination of the truth-inducing mechanism and the full-charge mechanism. The weight on the full-charge mechanism depends on the average marginal punishment.

Moreover, Carroll[11] shows that in many common preference domains a locally incentive-compatible mechanism is incentive-compatible. Lemma 1 shows that if a punishment is MNI, the locally h-strategy-proof mechanism is h-strategy-proof.

6.2 Private Equity Investment

Metrick and Yasuda[28], Nielsen[33] study the economics of investors in private equity funds empirically. Phalippou and Gottschalg[36] find that reports overestimate the performance of private equity funds. Our study applies to the problem of an investor endowed with a fixed level of capital that he can use to invest in projects. Each project can produce certain kinds of goods. The investor could elicit the of the projects based on reports, allocate capital to the projects and evaluate the performance over time, and punish the projects misreported their abilities by taking back the investment.

The model of strategy-proof mechanism with punishment provides a theory for the investor in private equity investment. Theorem 1 characterizes the ineffective punishment of the investor, which would not help the investor to eliminate the bias in reporting on the performance of private equity funds. Theorem 2 provides estimation of the best outcome for the investor, which depends on marginal punishment for deviation from truthful reporting.

6.3 Government Funding with Intermediation

The government is interested in delivering goods to recipients via a set of intermediaries (e.g., charities). The charities report their abilities to transmit a particular resource to recipients, and the government makes an allocation to the charities based on their reports. This ability is represented by the total amount of the resource that an intermediary sends to the recipients per unit of resource allocated, as well as by the proportions in which every recipient receives a resource relative to another from a given intermediary.

Galeotti and Condorelli[14], Choi, Galeotti and Goyal[12], Han and Juarez[17], Manea[27] study the strategic behavior of intermediaries in network. This paper complements the research about strategic behavior of intermediaries in network from the perspective of mechanism design, and studies the strategy-proof mechanism with punishment of the planner for resource transmission through intermediaries. Theorem 1 studies ex-ante strategy-proof mechanism, and characterizes ineffective punishments. Theorem 2 shows the optimal outcome for a planner to achieve with a given form of punishment.

7 Conclusion

This paper is a starting point in the study of strategy-proof mechanisms with punishments. The threat of ex-post punishment is shown to be able to expand traditional classes of incentive compatible mechanisms when punishment was not available. The h-optimal mechanism for the planner is discovered and characterized as a convex combination of full-charge mechanism and truth-inducing mechanism.

Not all ex-post punishments are equal, and some of them are ineffective, by providing no further mechanisms beyond the traditional ex-ante strategy-proof mechanisms. The paper characterizes the punishment functions that are ineffective as a first order condition of the punishment function.

Finally, when the planner is able to select the punishment function h, we provide the minimal punishment for a mechanism to be h-SP and achieve first-best efficiency.

Expansion of this work to other punishment functions, including random punishments, remain to be investigated.

Appendix: Proofs

Proof of Lemma 1

Proof.

Given the mechanism $\phi = (x(\cdot), t(\cdot))$, define $\Phi_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta)$. For this function, Part i is clearly satisfied. In order to prove that this satisfies the conditions of Part ii, note that for the mechanism $\phi = (x(\cdot), t(\cdot))$ to be h-SP, $\bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta) \geq \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{t}_i(\hat{\theta}_i, \theta_{-i}) - h_i(\theta_i, \hat{\theta}_i)$, $\forall \theta_i, \hat{\theta}_i, \theta_{-i}$.

$\Phi_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta)$ is the profit of agent i when he truthfully reports his ability θ_i, and the abilities of other agents are θ_{-i}. Then the condition for h-strategy-proofness is equivalent with $\Phi_i(\theta) \geq \Phi_i(\hat{\theta}_i, \theta_{-i}) + \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i})) - h_i(\theta_i, \hat{\theta}_i)$, $\forall \theta_i, \hat{\theta}_i, \theta_{-i}$.

For $\hat{\theta}_i > \theta_i$, we have $\frac{\Phi_i(\theta) - \Phi_i(\hat{\theta}_i, \theta_{-i})}{\theta_i - \hat{\theta}_i} \leq \frac{\bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i}))}{\theta_i - \hat{\theta}_i} + \frac{h_i(\theta_i, \theta_i) - h_i(\theta_i, \hat{\theta}_i)}{\theta_i - \hat{\theta}_i}$.

For $\hat{\theta}_i < \theta_i$, we have $\frac{\Phi_i(\theta) - \Phi_i(\hat{\theta}_i, \theta_{-i})}{\theta_i - \hat{\theta}_i} \geq \frac{\bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i}))}{\theta_i - \hat{\theta}_i} + \frac{h_i(\theta_i, \theta_i) - h_i(\theta_i, \hat{\theta}_i)}{\theta_i - \hat{\theta}_i}$.

Since the profit function Φ_i, production function $\bar{f}_i$ and punishment h_i are differentiable, take the limit $\hat{\theta}_i \to \theta_i$, there is $\frac{\partial \Phi_i(\theta_i, \theta_{-i})}{\partial \theta_i} = \bar{f}_{i\theta}(\theta_i, x_i(\theta)) + h'_{i2}(\theta_i, \theta_i)$.

For $h_i(\theta_i, \hat{\theta}_i) \geq 0$, assume $h'_{i2+}(\theta_i) = \lim_{\hat{\theta}_i \to \theta_i^+} \frac{h_i(\theta_i, \theta_i) - h_i(\theta_i, \hat{\theta}_i)}{\theta_i - \hat{\theta}_i} \geq 0$ with $\hat{\theta}_i > \theta_i$, and

$h'_{i2-}(\theta_i) = \lim_{\hat{\theta}_i \to \theta_i^-} \frac{h_i(\theta_i,\theta_i)-h_i(\theta_i,\hat{\theta}_i)}{\theta_i-\hat{\theta}_i} \leq 0$ with $\hat{\theta}_i < \theta_i$. Thus, $\bar{f}_{i\theta}(\theta_i, x_i(\theta)) + h'_{i2-}(\theta_i, \theta_i) \leq \frac{\partial \Phi_i(\theta_i,\theta_{-i})}{\partial \theta_i} \leq \bar{f}_{i\theta}(\theta_i, x_i(\theta)) + h'_{i2+}(\theta_i, \theta_i)$.

For the converse part, assume that the punishment function h satisfies the MNI condition. In addition, assume that there exists Φ that satisfies conditions i and ii.

From part ii, $\bar{f}_{i\theta}(\theta_i, x_i(\theta)) + h'_{i2-}(\theta_i, \theta_i) \leq \frac{\partial \Phi_i(\theta_i,\theta_{-i})}{\partial \theta_i} \leq \bar{f}_{i\theta}(\theta_i, x_i(\theta)) + h'_{i2+}(\theta_i, \theta_i)$.

The condition of h-SP requires $\bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta) \geq \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{t}_i(\hat{\theta}_i, \theta_{-i}) - h_i(\theta_i, \hat{\theta}_i), \forall \theta_i, \hat{\theta}_i, \theta_{-i}$. It is equivalent with $\Phi_i(\theta) \geq \Phi_i(\hat{\theta}_i, \theta_{-i}) + \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i})) - h_i(\theta_i, \hat{\theta}_i)$.

Note that $\Phi_i(\theta) - \Phi_i(\hat{\theta}_i, \theta_{-i}) = \int_{\hat{\theta}_i}^{\theta_i} \frac{\partial \Phi_i(q,\theta_{-i})}{\partial \theta_i} dq$, $\bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i})) = \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))dq$, $-h_i(\theta_i, \hat{\theta}_i) = -h_i(\theta_i, \hat{\theta}_i) + h_i(\theta_i, \theta_i) = \int_{\hat{\theta}_i}^{\theta_i} h'_{i2}(\theta_i, q)dq$.

Then condition of h-SP is equivalent to $\int_{\hat{\theta}_i}^{\theta_i} \frac{\partial \Phi_i(q,\theta_{-i})}{\partial \theta_i} dq \geq \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))dq + \int_{\hat{\theta}_i}^{\theta_i} h'_{i2}(\theta_i, q)dq$.

For $\hat{\theta}_i > \theta_i$, from condition ii, $\int_{\theta_i}^{\hat{\theta}_i} \frac{\partial \Phi_i(q,\theta_{-i})}{\partial \theta_i} dq \leq \int_{\theta_i}^{\hat{\theta}_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) + h'_{i2+}(q, q)dq$, and $x_i(q, \theta_{-i}) \leq x_i(\hat{\theta}_i, \theta_{-i})$ for $\hat{\theta}_i > q > \theta_i$. Since x_i and θ_i are complements, $\bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) \leq \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))$. The punishment h is marginal nondecreasing, $h'_{i2}(\theta_i, q) \geq h'_{i2+}(q, q) \geq 0$. Thus, $\int_{\theta_i}^{\hat{\theta}_i} \frac{\partial \Phi_i(q,\theta_{-i})}{\partial \theta_i} dq \leq \int_{\theta_i}^{\hat{\theta}_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) + h'_{i2+}(q, q)dq \leq \int_{\theta_i}^{\hat{\theta}_i} \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))dq + \int_{\theta_i}^{\hat{\theta}_i} h'_{i2}(\theta_i, q)dq$. Hence, the h-SP condition is satisfied.

For $\theta_i > \hat{\theta}_i$, from condition ii, $\int_{\hat{\theta}_i}^{\theta_i} \frac{\partial \Phi_i(q,\theta_{-i})}{\partial \theta_i} dq \geq \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) + h'_{i2-}(q, q)dq$, and $x_i(q, \theta_{-i}) \geq x_i(\hat{\theta}_i, \theta_{-i})$ for $\hat{\theta}_i < q < \theta_i$. Since x_i and θ_i are complements, $\bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) \geq \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))$. The punishment h is marginal nondecreasing, $0 \geq h'_{i2-}(q, q) \geq h'_{i2}(\theta_i, q)$. Thus, $\int_{\hat{\theta}_i}^{\theta_i} \frac{\partial \Phi_i(q,\theta_{-i})}{\partial \theta_i} dq \geq \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) + h'_{i2-}(q, q)dq \geq \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))dq + \int_{\hat{\theta}_i}^{\theta_i} h'_{i2}(\theta_i, q)dq$. Hence, the h-SP condition is satisfied. ∎

Proof of Theorem 1

Proof.

Recall that $h'_{i2}(\theta_i, \hat{\theta}_i) = \frac{\partial h_i(\theta_i,\hat{\theta}_i)}{\partial \hat{\theta}_i}$, $\theta = (\theta_i, \theta_{-i})$, $\bar{f}_{i\theta}(\theta_i, x_i(\theta)) = \frac{\partial \bar{f}_i(\theta_i,x_i(\theta))}{\partial \theta_i}$.

From Lemma 1, an h-SP mechanism satisfies $\bar{f}_{i\theta}(\theta_i, x_i(\theta)) + h'_{i2-}(\theta_i, \theta_i) \leq \frac{\partial \Phi_i(\theta_i,\theta_{-i})}{\partial \theta_i} \leq \bar{f}_{i\theta}(\theta_i, x_i(\theta)) + h'_{i2+}(\theta_i, \theta_i)$.

i. $\Rightarrow$ iii.

For the mechanism $\phi = (x(\cdot), t(\cdot))$ to be 0-SP, we have $\bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta) \geq \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{t}_i(\hat{\theta}_i, \theta_{-i})$. This is equivalent to $\Phi_i(\theta) \geq \Phi_i(\hat{\theta}_i, \theta_{-i}) + \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i}))$, $\forall \theta_i, \hat{\theta}_i, \theta_{-i}$.

For $\hat{\theta}_i > \theta_i$, $\frac{\Phi_i(\theta)-\Phi_i(\hat{\theta}_i,\theta_{-i})}{\hat{\theta}_i-\theta_i} \geq \frac{\bar{f}_i(\theta_i,x_i(\hat{\theta}_i,\theta_{-i}))-\bar{f}_i(\hat{\theta}_i,x_i(\hat{\theta}_i,\theta_{-i}))}{\hat{\theta}_i-\theta_i}$.

For $\hat{\theta}_i < \theta_i$, $\frac{\Phi_i(\theta)-\Phi_i(\hat{\theta}_i,\theta_{-i})}{\hat{\theta}_i-\theta_i} \leq \frac{\bar{f}_i(\theta_i,x_i(\hat{\theta}_i,\theta_{-i}))-\bar{f}_i(\hat{\theta}_i,x_i(\hat{\theta}_i,\theta_{-i}))}{\hat{\theta}_i-\theta_i}$.

As we take the limit of the two above inequalities when $\hat{\theta}_i \to \theta_i$, we have that $\frac{\partial \Phi_i(\theta_i,\theta_{-i})}{\partial \theta_i} = \bar{f}_{i\theta}(\theta_i, x_i(\theta))$.

ii. $\Rightarrow$ iii.

For the punishment function $h'_{i2}(\theta_i, \theta_i) = 0$, for any θ_i, then $h'_{i2-}(\theta_i, \theta_i) = h'_{i2+}(\theta_i, \theta_i) = 0$. Thus, from Lemma 1 (part ii), the h-SP mechanism satisfies $\frac{\partial \Phi_i(\theta_i, \theta_{-i})}{\partial \theta_i} = \bar{f}_{i\theta}(\theta_i, x_i(\theta))$.

iii. $\Rightarrow$ i.

For the mechanism ϕ that satisfies $\frac{\partial \Phi_i(\theta_i, \theta_{-i})}{\partial \theta_i} = \bar{f}_{i\theta}(\theta_i, x_i(\theta))$, we have $\Phi_i(\theta_i, \theta_{-i}) = \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq$, $\Phi_i(\hat{\theta}_i, \theta_{-i}) = \int_0^{\hat{\theta}_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq$. The planner charges agent i with $\bar{t}_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq$.

For any $\theta_i \geq q \geq \hat{\theta}_i$, $\Phi_i(\theta_i, \theta_{-i}) - \Phi_i(\hat{\theta}_i, \theta_{-i}) = \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq$, and $\bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i})) = \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))dq$. $x_i(\theta)$ is non-decreasing in θ_i, so $x_i(q, \theta_{-i}) \geq x_i(\hat{\theta}_i, \theta_{-i})$. Since $\frac{\partial^2 \bar{f}_i}{\partial \theta \partial x} \geq 0$, $\bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) \geq \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))$. Thus, $\int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq \geq \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))dq$. Then we have $\Phi_i(\theta_i, \theta_{-i}) - \Phi_i(\hat{\theta}_i, \theta_{-i}) \geq \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i}))$, which is equivalent with the 0-SP condition $\Phi_i(\theta_i, \theta_{-i}) \geq \Phi_i(\hat{\theta}_i, \theta_{-i}) + \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i}))$.

For any $\theta_i \leq q \leq \hat{\theta}_i$, $\Phi_i(\hat{\theta}_i, \theta_{-i}) - \Phi_i(\theta_i, \theta_{-i}) = \int_{\theta_i}^{\hat{\theta}_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq$, and $-\bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) + \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i})) = \int_{\theta_i}^{\hat{\theta}_i} \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))dq$. $x_i(\theta)$ is non-decreasing in θ_i, so $x_i(q, \theta_{-i}) \leq x_i(\hat{\theta}_i, \theta_{-i})$. Since $\frac{\partial^2 \bar{f}_i}{\partial \theta \partial x} \geq 0$, $\bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) \leq \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))$. Thus, $\int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq \leq \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))dq$. Then we have $-\Phi_i(\theta_i, \theta_{-i}) + \Phi_i(\hat{\theta}_i, \theta_{-i}) \leq -\bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) + \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i}))$, which is equivalent with the 0-SP condition $\Phi_i(\theta_i, \theta_{-i}) \geq \Phi_i(\hat{\theta}_i, \theta_{-i}) + \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i}))$.

iii. $\Rightarrow$ ii.

From above, any mechanism ϕ satisfying $\frac{\partial \Phi_i(\theta_i, \theta_{-i})}{\partial \theta_i} = \bar{f}_{i\theta}(\theta_i, x_i(\theta))$ is 0-SP. Thus, $\Phi_i(\theta) \geq \Phi_i(\hat{\theta}_i, \theta_{-i}) + \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i}))$, $\forall \theta_i, \hat{\theta}_i, \theta_{-i}$. Then $\Phi_i(\theta) \geq \Phi_i(\hat{\theta}_i, \theta_{-i}) + \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta}_i, \theta_{-i})) - h_i(\theta_i, \hat{\theta}_i)$, with $h_i(\theta_i, \hat{\theta}_i) \geq 0$. ■

Proof of Remark 1

Proof.

For any punishment function h and $\tilde{h}$ such that $h(\theta, \hat{\theta}) \leq \tilde{h}(\theta, \hat{\theta})$, for any θ and $\hat{\theta}$.

Any mechanism ϕ that is h-SP, satisfies $\bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta) \geq \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{t}_i(\hat{\theta}_i, \theta_{-i}) - h_i(\theta_i, \hat{\theta}_i)$, $\forall \theta, \hat{\theta}_i, i$.

Note that $\bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{t}_i(\hat{\theta}_i, \theta_{-i}) - h_i(\theta_i, \hat{\theta}_i) \geq \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{t}_i(\hat{\theta}_i, \theta_{-i}) - \tilde{h}_i(\theta_i, \hat{\theta}_i)$.

Thus, $\bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta) \geq \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{t}_i(\hat{\theta}_i, \theta_{-i}) - \tilde{h}_i(\theta_i, \hat{\theta}_i)$, $\forall \theta, \hat{\theta}_i, i$.

Hence, the mechanism $\phi = (x(\cdot), t(\cdot))$ is $\tilde{h}$-SP. ■

Proof of Theorem 2

Proof.

Proof of $\Leftarrow$. To prove any h-SP mechanism ϕ with allocation rule x charges not more than ϕ^h.

Consider any h-SP mechanism $\phi(\cdot) = (x(\cdot), t(\cdot))$, from Lemma 1, there is $\bar{f}_{i\theta}(\theta_i, x_i(\theta)) + h'_{i2-}(\theta_i, \theta_i) \le \frac{\partial \Phi_i(\theta_i,\theta_{-i})}{\partial \theta_i} \le \bar{f}_{i\theta}(\theta_i, x_i(\theta)) + h'_{i2+}(\theta_i, \theta_i)$.

Assume $\underline{\theta}_i(\theta_{-i}) = \sup\{\theta_i | x_i(\theta_i, \theta_{-i}) = 0\}$, for $\theta_i < \underline{\theta}_i(\theta_{-i})$, $x_i(\theta) = 0$, then $\bar{f}_i(\theta_i, x_i(\theta)) = 0$, $\bar{t}_i(\theta) \ge 0$, $\Phi_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta) \le 0$.

For $\theta_i < \underline{\theta}_i(\theta_{-i})$, the condition of individual rationality requires $\Phi_i(\theta) = 0$.

For $\theta_i \ge \underline{\theta}_i(\theta_{-i})$, integral the inequality $\bar{f}_{i\theta}(\theta_i, x_i(\theta)) + h'_{i2-}(\theta_i, \theta_i) \le \frac{\partial \Phi_i(\theta_i,\theta_{-i})}{\partial \theta_i}$ from $\underline{\theta}_i(\theta_{-i})$ to θ_i. $\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) + h'_{i2-}(q, q) dq \le \int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \frac{\partial \Phi_i(q,\theta_{-i})}{\partial \theta_i} dq = \Phi_i(\theta) - \Phi_i(\underline{\theta}_i(\theta_{-i}), \theta_{-i}) = \Phi_i(\theta)$. Substitute $\Phi_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta)$ into the inequality, $\bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta) \ge \int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) + h'_{i2-}(q, q) dq$.

Rearrange the inequality, $\bar{t}_i(\theta) \le \bar{f}_i(\theta_i, x_i(\theta)) - \int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} (\bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) + h'_{i2-}(q, q)) dq$.

From the condition of individual rationality, the maximal charge of planner is the production of agent, $\bar{t}_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta))$. From Theorem 1, the ex-ante SP mechanism charges $\bar{t}_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - \int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) dq$.

Assume the charge t in mechanism ϕ could be represented as weighted average of charge in the full-charge mechanism t^F and truth-inducing mechanism t^T, $\bar{t}_i(\theta) = \lambda_i(\theta)\bar{t}_i^F(\theta) + (1 - \lambda_i(\theta))\bar{t}_i^T(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - (1 - \lambda_i(\theta)) \int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) dq$.

Thus, h-SP requires $\bar{f}_i(\theta_i, x_i(\theta)) - (1 - \lambda_i(\theta)) \int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) dq \le \bar{f}_i(\theta_i, x_i(\theta)) - \int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} (\bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) + h'_{i2-}(q, q)) dq$, it is equivalent with $\lambda_i(\theta) \le \frac{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} -h'_{i2-}(q,q) dq}{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q,\theta_{-i})) dq}$.

For h-SP mechanism ϕ, the charge from agent i is not more than the weighted average of the full-charge mechanism and the ex-ante SP mechanism, with weight $\lambda_i(\theta)$ on the full-charge mechanism.

Proof of $\Rightarrow$.

Assume mechanism $\phi^h(\cdot) = (x(\cdot), t^h(\cdot))$, with charge $t_i^h(\theta) = \lambda_i(\theta) t_i^F(\theta) + (1 - \lambda_i(\theta)) t_i^T(\theta)$, with $\lambda_i(\theta) = \frac{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} -h'_{i2-}(q,q) dq}{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q,\theta_{-i})) dq}$. To prove ϕ^h is h-SP.

The charge of planner is $\bar{t}_i^h(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - (1 - \lambda_i(\theta)) \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) dq$.

The condition of h-SP requires $\bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i^h(\theta) \ge \bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{t}_i^h(\hat{\theta}_i, \theta_{-i}) - h_i(\theta_i, \hat{\theta}_i)$, $\forall \theta_i, \hat{\theta}_i, \theta_{-i}$. $-$ (1)

To simplify the notations, let $\hat{\theta} = (\hat{\theta}_i, \theta_{-i})$.

Substitute the charge $\bar{t}_i^h$ into agent i's profit $\bar{f}_i - \bar{t}_i^h$. Then we have $\bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i^h(\theta) = (1 - \lambda_i(\theta)) \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) dq$, and $\bar{f}_i(\theta_i, x_i(\hat{\theta}_i, \theta_{-i})) - \bar{t}_i^h(\hat{\theta}_i, \theta_{-i}) = \bar{f}_i(\theta_i, x_i(\hat{\theta})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta})) + (1 - \lambda_i(\hat{\theta})) \int_0^{\hat{\theta}_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) dq$. $-$ (2)

For $q < \underline{\theta}_i(\theta_{-i})$, $x_i(q, \theta_{-i}) = 0$, $\bar{f}_i(q, x_i(q, \theta_{-i})) = 0$, then $\bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) = 0$, and $\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) dq = \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) dq$.

$\lambda_i(\theta) = \frac{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} -h'_{i2-}(q,q)dq}{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q,x_i(q,\theta_{-i}))dq} = \frac{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} -h'_{i2-}(q,q)dq}{\int_0^{\theta_i} \bar{f}_{i\theta}(q,x_i(q,\theta_{-i}))dq}$, then we have $(1-\lambda_i(\theta)) \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq = \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq + \int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} h'_{i2-}(q, q)dq$. For $\hat{\theta}$, we have $(1-\lambda_i(\hat{\theta})) \int_0^{\hat{\theta}_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq = \int_0^{\hat{\theta}_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq + \int_{\underline{\theta}_i(\theta_{-i})}^{\hat{\theta}_i} h'_{i2-}(q, q)dq$. – (3)

Substitute (2) and (3) into (1), the h-SP condition is equivalent with $\int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq + \int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} h'_{i2-}(q, q)dq \geq \bar{f}_i(\theta_i, x_i(\hat{\theta})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta})) + \int_0^{\hat{\theta}_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq + \int_{\underline{\theta}_i(\theta_{-i})}^{\hat{\theta}_i} h'_{i2-}(q, q)dq - h_i(\theta_i, \hat{\theta}_i)$. Simplify the inequality, $\int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq + \int_{\hat{\theta}_i}^{\theta_i} h'_{i2-}(q, q)dq \geq \bar{f}_i(\theta_i, x_i(\hat{\theta})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta})) - h_i(\theta_i, \hat{\theta}_i)$. – (4)

Since $\bar{f}_i(\theta_i, x_i(\hat{\theta})) - \bar{f}_i(\hat{\theta}_i, x_i(\hat{\theta})) = \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))dq$, $h_i(\theta_i, \theta_i) - h_i(\theta_i, \hat{\theta}_i) = \int_{\hat{\theta}_i}^{\theta_i} h'_{i2}(\theta_i, q)dq$, the inequality (4) is equivalent with $\int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq + \int_{\hat{\theta}_i}^{\theta_i} h'_{i2-}(q, q)dq \geq \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))dq + \int_{\hat{\theta}_i}^{\theta_i} h'_{i2}(\theta_i, q)dq$.

If $\theta_i > \hat{\theta}_i$, since θ_i and x_i are complements, $\bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i})) \leq \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))$ for $\hat{\theta}_i < q < \theta_i$. The punishment function h_i is MNI, $h'_{i2-}(q, q) - h'_{i2}(\theta_i, q) \geq 0$ for $q < \theta_i$. Then $\bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) + h'_{i2-}(q, q) \geq \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i})) + h'_{i2}(\theta_i, q)$. Thus, integral the inequality, we have $\int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq + \int_{\hat{\theta}_i}^{\theta_i} h'_{i2-}(q, q)dq \geq \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))dq + \int_{\hat{\theta}_i}^{\theta_i} h'_{i2}(\theta_i, q)dq$, which is the inequality (4).

If $\theta_i < \hat{\theta}_i$, since θ_i and x_i are complements, $\bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i})) \geq \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))$, and $h'_{i2}(\theta_i, q) \geq 0 \geq h'_{i2-}(q, q)$ for $\theta_i < q < \hat{\theta}_i$. Then $\bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i})) + h'_{i2}(\theta_i, q) \geq \bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) + h'_{i2-}(q, q)$. Thus, integral the inequality, we have $\int_{\theta_i}^{\hat{\theta}_i} \bar{f}_{i\theta}(q, x_i(\hat{\theta}_i, \theta_{-i}))dq + \int_{\theta_i}^{\hat{\theta}_i} h'_{i2}(\theta_i, q)dq \geq \int_{\theta_i}^{\hat{\theta}_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq + \int_{\theta_i}^{\hat{\theta}_i} h'_{i2-}(q, q)dq$, which is the inequality (4).

From the proof above, the mechanism ϕ^h is h-SP, and any h-SP mechanism would not charge more than $\lambda_i(\theta)t_i^F(\theta) + (1 - \lambda_i(\theta))t_i^T(\theta)$. Thus, the mechanism ϕ^h is h-optimal. ■

Proof of Corollary 1

Proof.

For punishment function h with $h'_{i2}(\theta_i, \theta_i) = 0$, $\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} -h'_{i2-}(q, q)dq = 0$.

Then $\lambda_i^h(\theta) = \min\{\frac{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} -h'_{i2-}(q,q)dq}{\int_0^{\theta_i} \bar{f}_{i\theta}(q,x_i(q,\theta_{-i}))dq}, 1\} = \min\{0, 1\} = 0$, for any θ, i.

The mechanism ϕ^h satisfies $t_i^h(\theta) = \lambda_i^h(\theta)t_i^F(\theta) + (1 - \lambda_i^h(\theta))t_i^T(\theta) = t_i^T(\theta)$.

From Theorem 2, the h-optimal mechanism is the second-price mechanism ϕ^T. ■

Proof of Corollary 2

Proof.

The punishment function h satisfies $h_i(\theta_i, \hat{\theta}_i) \geq \bar{f}_i(\theta_i, 1) - \bar{f}_i(\hat{\theta}_i, 1)$, for any $\theta_i \geq \hat{\theta}_i$. Then $\frac{h_i(\theta_i,\hat{\theta}_i) - h_i(\theta_i,\theta_i)}{\theta_i - \hat{\theta}_i} \geq \frac{\bar{f}_i(\theta_i,1) - \bar{f}_i(\hat{\theta}_i,1)}{\theta_i - \hat{\theta}_i}$ for $\theta_i \geq \hat{\theta}_i$. Take the limit $\hat{\theta}_i \to \theta_i^-$, $-h'_{i2-}(\theta_i, \theta_i) \geq \bar{f}_{i\theta}(\theta_i, 1)$.

Since θ_i and x_i are complements for production, $\bar{f}_{i\theta}(\theta_i, 1) \geq \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))$ for any q.

Then $\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} -h'_{i2-}(q,q)dq \geq \int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q,1)dq \geq \int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q,x_i(q,\theta_{-i}))dq$.

Thus, $\lambda_i^h(\theta) = \min\{\frac{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} -h'_{i2-}(q,q)dq}{\int_{\underline{\theta}_i(\theta_{-i})}^{\theta_i} \bar{f}_{i\theta}(q,x_i(q,\theta_{-i}))dq}, 1\} = 1$.

The mechanism ϕ^h satisfies $t_i^h(\theta) = \lambda_i^h(\theta)t_i^F(\theta) + (1-\lambda_i^h(\theta))t_i^T(\theta) = t_i^F(\theta)$.

From Theorem 2, the h-optimal mechanism is the first-price mechanism ϕ^F. ■

Proof of Lemma 2

Proof.

i. Consider any punishment function h, such that mechanism ϕ is h-SP. The condition of h-SP requires $\bar{f}_i(\theta_i, x_i(\theta_i, \hat{\theta}_{-i})) - \bar{t}_i(\theta_i, \hat{\theta}_{-i}) \geq \bar{f}_i(\theta_i, x_i(\hat{\theta})) - \bar{t}_i(\hat{\theta}) - h_i(\theta_i, \hat{\theta}_i)$. By definition, $v_i(\theta_i, \hat{\theta}) = \bar{f}_i(\theta_i, x_i(\hat{\theta})) - \bar{t}_i(\hat{\theta})$, $v_i(\theta_i, \theta_i, \hat{\theta}_{-i}) = \bar{f}_i(\theta_i, x_i(\theta_i, \hat{\theta}_{-i})) - \bar{t}_i(\theta_i, \hat{\theta}_{-i})$. h-SP requires $v_i(\theta_i, \theta_i, \hat{\theta}_{-i}) \geq v_i(\theta_i, \hat{\theta}) - h_i(\theta_i, \hat{\theta}_i)$, equivalent with $h_i(\theta_i, \hat{\theta}_i) \geq v_i(\theta_i, \hat{\theta}) - v_i(\theta_i, \theta_i, \hat{\theta}_{-i})$, for any θ_i, $\hat{\theta}$.

To prove existence of $h_i^{\min}$, we have $v_i(\theta_i, \hat{\theta}) - v_i(\theta_i, \theta_i, \hat{\theta}_{-i}) \leq v_i(\theta_i, \hat{\theta}) \leq \bar{f}_i(\theta_i, x_i(\hat{\theta})) \leq \bar{f}_i(\theta_i, 1)$. The first inequality is from individual rationality of mechanism ϕ, $v_i(\theta_i, \theta_i, \hat{\theta}_{-i}) = \bar{f}_i(\theta_i, x_i(\theta_i, \hat{\theta}_{-i})) - \bar{t}_i(\theta_i, \hat{\theta}_{-i}) \geq 0$. The second inequality is from charge of planner is nonnegative, $\bar{t}_i(\hat{\theta}) \geq 0$. The third inequality is from monotonicity of production $\bar{f}_i(\theta_i, x_i(\hat{\theta}))$ on resource allocation $x_i(\hat{\theta})$ and $x_i(\hat{\theta}) \leq 1$. From the inequality above, the upper bound of $v_i(\theta_i, \hat{\theta}) - v_i(\theta_i, \theta_i, \hat{\theta}_{-i})$ exists. Thus, there exists function $h_i^{\min}(\theta_i, \hat{\theta}_i) = \max_{\hat{\theta}_{-i}}\{v_i(\theta_i, \hat{\theta}) - v_i(\theta_i, \theta_i, \hat{\theta}_{-i}), 0\}$.

Since punishment function h_i satisfies $h_i(\theta_i, \hat{\theta}_i) \geq v_i(\theta_i, \hat{\theta}) - v_i(\theta_i, \theta_i, \hat{\theta}_{-i})$, for any θ_i, $\hat{\theta}$, and punishment is nonnegative, $h_i(\theta_i, \hat{\theta}_i) \geq 0$. We have $h_i(\theta_i, \hat{\theta}_i) \geq h_i^{\min}(\theta_i, \hat{\theta}_i)$. From definition 6, $h^{\min}$ is the minimal punishment function for mechanism ϕ.

ii. Consider a mechanism $\phi = (x(\cdot), s(\cdot))$, which is ex-ante SP. The agents have incentive to report truthfully, $v_i(\theta_i, \hat{\theta}) \leq v_i(\theta_i, \theta_i, \hat{\theta}_{-i})$, for any $\hat{\theta}$, θ_i. So $\max_{\hat{\theta}_{-i}}[v_i(\theta_i, \hat{\theta}) - v_i(\theta_i, \theta_i, \hat{\theta}_{-i})] \leq 0$ for any $\hat{\theta}_i$, θ_i. Then the minimal punishment function $h_i^{\min}(\theta_i, \hat{\theta}_i) = 0$, for any $\hat{\theta}_i$, θ_i.

iii. Consider a mechanism $\phi = (x(\cdot), s(\cdot))$, which is not ex-ante SP, there exists θ_i, $\hat{\theta}$, such that $v_i(\theta_i, \theta_i, \hat{\theta}_{-i}) < v_i(\theta_i, \hat{\theta})$. Suppose $h_i^{\min}(\theta_i, \hat{\theta}_i) = 0$, for any $\hat{\theta}_i$, θ_i. Then $v_i(\theta_i, \theta_i, \hat{\theta}_{-i}) < v_i(\theta_i, \hat{\theta}) - h_i^{\min}(\theta_i, \hat{\theta}_i)$. If ability of agent i is θ_i and the reports of other agents are $\hat{\theta}_{-i}$, agent i has incentive to misreport $\hat{\theta}_i$ for higher profit. It is a contradiction with h-SP, so $h_i^{\min}$ is nonzero. ■

Proof of Proposition 1

Proof.

i. To prove there does not exist a mechanism, such that it is both SP and FBE.

Assume mechanism $\phi = (x(\cdot), t(\cdot))$ is SP. From Theorem 1, the profit Φ_i satisfies $\Phi_i(\theta_i, \theta_{-i}) = \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq$, and $\Phi_i(\theta_i, \theta_{-i}) = \bar{f}_i(\theta_i, x_i(\theta)) - \bar{t}_i(\theta)$. ϕ is SP, agents will report truthfully, $\hat{\theta} = \theta$. The aggregate charge of planner from agent i is $\bar{t}_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq$. Given mechanism ϕ, the total tax is $\tau = \sum_{i=1}^N t_i(\theta)$.

By definition of FBE, $u^*(\theta) = \max_x u(\sum_{i=1}^N f_i(\theta_i, x_i))$, such that $\sum_{i=1}^N x_i = 1$. Thus, $u^*(\theta) \geq u(\sum_{i=1}^N f_i(\theta_i, x_i))$ for any allocation rule $x(\cdot)$.

To prove there exists θ, such that $\sum_{i=1}^N \bar{f}_i(\theta_i, x_i(\theta)) > \sum_{i=1}^N \bar{t}_i(\theta)$, which is equivalent with $\sum_{i=1}^N \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq > 0$. Without loss of generality, assume for $\tilde{\theta}$, $x_i(\tilde{\theta}) > 0$, and allocation rule $x_i(q, \theta_{-i})$ is non-decreasing in q. Let $\theta_{-i} = \tilde{\theta}_{-i}$, $\theta_i > \tilde{\theta}_i$, then $x_i(q, \theta_{-i}) > 0$ for any $q > \tilde{\theta}_i$, $\bar{f}_{i\theta}(q, x_i(q, \theta_{-i})) > 0$. Thus, $\int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq \geq \int_{\tilde{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq > 0$. The aggregate charge from agent i is $\bar{t}_i(\theta) = \bar{f}_i(\theta_i, x_i(\theta)) - \int_0^{\theta_i} \bar{f}_{i\theta}(q, x_i(q, \theta_{-i}))dq < \bar{f}_i(\theta_i, x_i(\theta))$.

The preference of planner is strongly monotone, there is $u^*(\theta) \geq u(\sum_{i=1}^N f_i(\theta_i, x_i)) > u(\sum_{i=1}^N t_i(\theta))$. The SP mechanism could not be FBE, so there does not exist any mechanism, which is both SP and FBE.

ii. Assume ϕ is a FBE mechanism, the total tax τ maximizes the utility of the planner. The allocation rule of ϕ is $x^*(\theta)$, which solves the utility maximization problem $\max_x u(\sum_{i=1}^N f_i(\theta_i, x_i))$ for any θ, the charge of planner is $\tau^*(\theta) = \sum_{i=1}^N f_i(\theta_i, x_i^*)$. To prove the charge is $t_i(\theta) = f_i(\theta_i, x_i^*(\theta))$.

Suppose there exists agent i with $x_i^*(\theta) > 0$, and the outcome charged is $t_i(\theta) \leq f_i(\theta_i, x_i^*(\theta))$, $t_i(\theta) \neq f_i(\theta_i, x_i^*(\theta))$. Then total tax of the planner $\tau = \sum_{i=1}^N t_i(\theta) \leq \tau^*(\theta)$ and $\tau \neq \tau^*(\theta)$. If the preferences of planner is strongly monotone, then $u(\tau) < u(\tau^*(\theta))$, the fist best efficiency will not be achieved. So the charge of agent i is $t_i(\theta) = f_i(\theta_i, x_i^*(\theta))$, for $x_i^*(\theta) > 0$. If $x_i^*(\theta) = 0$, $f_i(\theta_i, x_i^*(\theta)) = \mathbf{0}$, $t_i(\theta) \leq f_i(\theta_i, x_i^*(\theta))$, the charge of agent i is $t_i(\theta) = f_i(\theta_i, x_i^*(\theta)) = \mathbf{0}$. Thus, FBE mechanism ϕ charges $t_i(\theta) = f_i(\theta_i, x_i^*(\theta))$, for any agent i.

From above, we have $v_i(\theta_i, \theta_i, \hat{\theta}_{-i}) = \bar{f}_i(\theta_i, x_i^*(\theta_i, \hat{\theta}_{-i})) - \bar{t}_i(\theta_i, \hat{\theta}_{-i}) = 0$, and $v_i(\theta_i, \hat{\theta}) = \bar{f}_i(\theta_i, x_i^*(\hat{\theta})) - \bar{t}_i(\hat{\theta}) = \bar{f}_i(\theta_i, x_i^*(\hat{\theta})) - \bar{f}_i(\hat{\theta}_i, x_i^*(\hat{\theta}))$. Then $v_i(\theta_i, \hat{\theta}) - v_i(\theta_i, \theta_i, \hat{\theta}_{-i}) = \bar{f}_i(\theta_i, x_i^*(\hat{\theta})) - \bar{f}_i(\hat{\theta}_i, x_i^*(\hat{\theta}))$.

From Lemma 2, $h_i^{\min}(\theta_i, \hat{\theta}_i) = \max_{\hat{\theta}_{-i}} \{v_i(\theta_i, \hat{\theta}) - v_i(\theta_i, \theta_i, \hat{\theta}_{-i}), 0\}$ is the minimal punishment function. Since $\bar{f}_{i\theta} \geq 0$, for $\hat{\theta}_i \geq \theta_i$, $\bar{f}_i(\theta_i, x_i^*(\hat{\theta})) - \bar{f}_i(\hat{\theta}_i, x_i^*(\hat{\theta})) \leq 0$, $h_i^{\min}(\theta_i, \hat{\theta}_i) = 0$. For $\hat{\theta}_i < \theta_i$, $h_i^{\min}(\theta_i, \hat{\theta}_i) = \max_{\hat{\theta}_{-i}} \bar{f}_i(\theta_i, x_i^*(\hat{\theta})) - \bar{f}_i(\hat{\theta}_i, x_i^*(\hat{\theta})) \geq 0$.

For $\hat{\theta}_i < \theta_i$, $\bar{f}_i(\theta_i, x_i^*(\hat{\theta})) - \bar{f}_i(\hat{\theta}_i, x_i^*(\hat{\theta})) = \int_{\hat{\theta}_i}^{\theta_i} \bar{f}_{i\theta}(q, x_i^*(\hat{\theta}))dq$, $x_i^*(\hat{\theta}) \in [0, 1]$. Since θ_i and x_i are complements, $\frac{\partial^2 f_{ij}(\theta_i, x_i)}{\partial \theta_i \partial x_i} \geq 0$ for all θ_i and x_i, $\bar{f}_i(\theta_i, x_i^*(\hat{\theta})) - \bar{f}_i(\hat{\theta}_i, x_i^*(\hat{\theta}))$ is increasing in $x_i^*(\hat{\theta})$, $\bar{f}_i(\theta_i, x_i^*(\hat{\theta})) - \bar{f}_i(\hat{\theta}_i, x_i^*(\hat{\theta})) \leq \bar{f}_i(\theta_i, 1) - \bar{f}_i(\hat{\theta}_i, 1)$. Let $\hat{\theta}_{-i} = (0, \ldots, 0)_{(M-1)\times 1}$, when $\hat{\theta}_i > 0$, it is optimal for planner to allocate all resource to agent i with strongly monotonic preference, so $\max_{\hat{\theta}_{-i}} x_i^*(\hat{\theta}) = 1$, $\max_{\hat{\theta}_{-i}} \bar{f}_i(\theta_i, x_i^*(\hat{\theta})) - \bar{f}_i(\hat{\theta}_i, x_i^*(\hat{\theta})) = \bar{f}_i(\theta_i, 1) - \bar{f}_i(\hat{\theta}_i, 1)$. Thus, the minimal punishment function is $h_i^{\min}(\theta_i, \hat{\theta}_i) = 0$ for $\hat{\theta}_i \geq \theta_i$, and $h_i^{\min}(\theta_i, \hat{\theta}_i) =$

$\bar{f}_i(\theta_i, 1) - \bar{f}_i(\hat{\theta}_i, 1)$ for $\hat{\theta}_i < \theta_i$. ■

References

[1] Kenneth Arrow. *The property rights doctrine and demand revelation under incomplete information*. Economics and human welfare. New York Academic Press, 1979.

[2] Salvador Barberà. An introduction to strategy-proof social choice functions. *Social Choice and Welfare*, 18(4):619–653, 2001.

[3] Salvador Barberà, Bhaskar Dutta, and Arunava Sen. Strategy-proof social choice correspondences. *Journal of Economic Theory*, 101(2):374–394, 2001.

[4] Salvador Barberà and Matthew O Jackson. Strategy-proof exchange. *Econometrica*, 63(1):51–87, 1995.

[5] Cesare Beccaria. *An Essay on Crimes and Punishments, Translated from the Italian; with a Commentary Attributed to Mons. De Voltaire, Translated from the French*. F. Newbery, at the corner of St. Paul's church-yard, 1769.

[6] Elchanan Ben-Porath, Eddie Dekel, and Barton L Lipman. Optimal allocation with costly verification. *American Economic Review*, 104(12):3779–3813, 2014.

[7] Jeremy Bentham. *An Introduction to the Principles of Morals and Legislation*. London: T. Payne, 1789.

[8] Eleanor Birrell and Rafael Pass. Approximately strategy-proof voting. In *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

[9] Walter Bossert and Yves Sprumont. Strategy-proof preference aggregation: Possibilities and characterizations. *Games and Economic Behavior*, 85:109–126, 2014.

[10] Paolo Buccirossi and Giancarlo Spagnolo. Optimal fines in the era of whistleblowers. should price fixers still go to prison? *Contributions to Economic Analysis*, 282:81–122, 2007.

[11] Gabriel Carroll. When are local incentive constraints sufficient? *Econometrica*, 80(2):661–686, 2012.

[12] Syngjoo Choi, Andrea Galeotti, and Sanjeev Goyal. Trading in networks: theory and experiments. *Journal of the European Economic Association*, 15(4):784–817, 2017.

[13] Edward H Clarke. Multipart pricing of public goods. *Public choice*, 11(1):17–33, 1971.

[14] Daniele Condorelli and Andrea Galeotti. Strategic models of intermediation networks. In *The Oxford Handbook of the Economics of Networks*. Oxford University Press, 2016.

[15] Allan Gibbard. Manipulation of voting schemes: a general result. *Econometrica*, 41(4):587–601, 1973.

[16] Theodore Groves. Incentives in teams. *Econometrica*, 41(4):617–631, 1973.

[17] Lining Han and Ruben Juarez. Free intermediation in resource transmission. *Games and Economic Behavior*, 111:75–84, 2018.

[18] Joseph E Harrington Jr and Myong-Hun Chang. When can we expect a corporate leniency program to result in fewer cartels? *The Journal of Law and Economics*, 58(2):417–449, 2015.

[19] Jason Hartline and Anna Karlin. Profit maximization in mechanism design. *Algorithmic Game Theory*, 331, 2007.

[20] John William Hatfield and Fuhito Kojima. Group incentive compatibility for matching with contracts. *Games and Economic Behavior*, 67(2):745–749, 2009.

[21] Bengt Holmström. Groves' scheme on restricted domains. *Econometrica*, 47:1137–1144, 1979.

[22] Ruben Juarez. Group strategyproof cost sharing: the role of indifferences. *Games and Economic Behavior*, 82:218–239, 2013.

[23] Ruben Juarez. *Optimal group strategyproof cost sharing*. PhD thesis, Working paper, 2013.

[24] Yannis Katsoulacos, Evgenia Motchenkova, and David Ulph. Penalizing cartels: The case for basing penalties on price overcharge. *International Journal of Industrial Organization*, 42:70–80, 2015.

[25] Shengwu Li. Obviously strategy-proof mechanisms. *American Economic Review*, 107(11):3257–87, 2017.

[26] Yunan Li. Mechanism design with costly verification and limited punishments. *Journal of Economic Theory*, 186:105000, 2020.

[27] Mihai Manea. Intermediation and resale in networks. *Journal of Political Economy*, 126(3):1250–1301, 2018.

[28] Andrew Metrick and Ayako Yasuda. The economics of private equity funds. *The Review of Financial Studies*, 23(6):2303–2341, 2010.

[29] Eiichi Miyagawa. House allocation with transfers. *Journal of Economic Theory*, 100(2):329–355, 2001.

[30] Hervé Moulin. Serial cost-sharing of excludable public goods. *The Review of Economic Studies*, 61(2):305–325, 1994.

[31] Herve Moulin and Scott Shenker. Serial cost sharing. *Econometrica*, 60(5):1009–1037, 1992.

[32] Tymofiy Mylovanov and Andriy Zapechelnyuk. Optimal allocation with ex post verification and limited penalties. *American Economic Review*, 107(9):2666–94, 2017.

[33] Kasper Meisner Nielsen. Institutional investors and private equity. *Review of Finance*, 12(1):185–219, 2008.

[34] Linda Nøstbakken. Fisheries law enforcement—a survey of the economic literature. *Marine Policy*, 32(3):293–300, 2008.

[35] Parag A Pathak and Tayfun Sönmez. School admissions reform in chicago and england: Comparing mechanisms by their vulnerability to manipulation. *The American Economic Review*, 103(1):80–106, 2013.

[36] Ludovic Phalippou and Oliver Gottschalg. The performance of private equity funds. *The Review of Financial Studies*, 22(4):1747–1776, 2009.

[37] Yves Sprumont. An axiomatization of the serial cost-sharing method. *Econometrica*, 78(5):1711–1748, 2010.

[38] William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961.

[39] Jung S You and Ruben Juarez. Incentive-compatible simple mechanisms. *Economic Theory*, 71(4):1569–1589, 2021.