

Mechanism Design for a Solution to the Tragedy of Commons*

Akira Yamada[†] Naoki Yoshihara[‡]

This version is from August 2006.

Abstract

We consider the problem of the tragedy of commons in cooperative production economies, and propose a mechanism to resolve this tragedy, taking into account that the coordinator cannot perfectly monitor each agent's labor skill and each agent may have an incentive to overstate as well as understate his own skill. Even in such a situation, the mechanism implements the *proportional solution* [Roemer and Silvestre (1989, 1993)] in Nash and strong equilibria when it is played as a normal form game. Moreover, the mechanism triply implements the solution in Nash, subgame-perfect, and strong equilibria when it is played as a two-stage extensive form game.

Journal of Economic Literature Classification Numbers: C72, D51, D78, D82

Keywords: triple implementation, proportional solution, unknown and possibly overstated labor skills, labor sovereignty

Corresponding Author: Naoki Yoshihara

*We are greatly thankful to Semi Koray, William Thomson, and an anonymous referee of this journal for their concrete and helpful comments on improving the paper. An earlier version of this paper was presented at the annual meeting of the Japanese Economic Association held at Hitotsubashi University in October 2001 and at the Conference on Economic Design held at NYU in July 2002. We are grateful to Takehiko Yamato for his useful comments in the former conference. We are also thankful to Tatsuyoshi Saijo, Kotaro Suzumura, and Yoshikatsu Tatamitani for their kind comments.

[†]Faculty of Economics, Sapporo University, 7-3-1 Nishioka 3-jo, Toyohira-ku, Sapporo, Hokkaido 062-8520, Japan. (e-mail: a-yamada@sapporo-u.ac.jp)

[‡]Institute of Economic Research, Hitotsubashi University, Naka 2-1, Kunitachi, Tokyo, 186-0004, Japan. (e-mail: yosihara@ier.hit-u.ac.jp)

1 Introduction

The fact that resource allocation under free access to technology results in “overproduction” and inefficient Nash equilibria in cooperative production economies is best known as the “*tragedy of commons*”. This paper provides a mechanism that can solve this tragedy. As a normative solution for the tragedy, we adopt the *proportional solution* [Roemer and Silvestre (1993)] under *joint ownership of the technology*, which assigns Pareto efficient allocations, in which each agent’s output consumption is proportional to his labor contribution. Then, we construct an incentive-compatible mechanism that implements the proportional solution.

There are some works on the implementation of the proportional solution, such as Suh (1995), Yoshihara (1999, 2000a), and Tian (2000), as well as of other social choice correspondences in production economies.¹ However, in most of the literature on implementation in production economies, a nonnegligible problem of asymmetric information in the production process appears to be treated as a “*black box*.” Under any mechanism, each agent is usually required to provide some information, and the outcome function assigns an allocation to each profile of agents’ strategies. This implicitly assumes, in production economies with labor input, that the mechanism coordinator is authorized to make agents supply their labor hours consistent with the assigned allocation.² This is because the original concern of implementation theory was in regard to *adverse selection problems*, and such a focus was valid whenever there was a decentralized resource allocation in exchange economies and/or production economies with no labor input. However, in production economies with labor input, this assumption is not realistic.

As an alternative, in this paper we assume that the coordinator is not authorized to make agents work as he pleases; the coordinator can monitor each agent’s labor hours, but the coordinator cannot perfectly monitor each agent’s *effective labor contribution measured in efficiency units*, since the coordinator is incapable of observing each agent’s *labor skill* exercised in the production process. Thus, there may be an incentive for each agent

¹For example, Hurwicz *et al.*(1995), Hong (1995), Tian (1999) for private ownership production economies with only private goods, Varian (1994) for production economies with the presence of an externality, and Kaplan and Wettstein (2000) and Corchón and Puy (2002) for cooperative production economies.

²Roemer (1989) pointed out this assumption explicitly.

to *overstate* or *understate* his own skill.³ Even under such a more realistic model of the tragedy of commons, the mechanism proposed in this paper can implement the solution.

This mechanism is a type of *sharing mechanism*: each agent can freely supply his labor hours, and the agent is asked to provide some information about his demand for the consumption good and his skill. Then, the outcome function only distributes the produced output to agents, according to the given information and the record of their supply of labor hours. In this mechanism, there is no restriction on the strategy spaces that prohibits agents from understating or overstating their skills. We will demonstrate that this mechanism *triply implements* the proportional solution *in Nash, strong Nash, and subgame-perfect equilibria*.

The basic model of economies and sharing mechanisms is defined in Section 2. Section 3 provides a sharing mechanism that implements the proportional solution. Concluding remarks are in Section 4. All proofs are provided in the Appendix.

2 The Basic Model

There are two goods, one of which is an input (labor time) $x \in \mathbb{R}_+$ to be used to produce the other good $y \in \mathbb{R}_+$.⁴ There is a set $N = \{1, \dots, n\}$ of agents, where $2 \leq n < +\infty$. Each agent i 's consumption is denoted by $z_i = (x_i, y_i)$, where x_i denotes his labor time, and y_i the amount of the output to be consumed by i . All agents face a common upper bound of labor time \bar{x} , where $0 < \bar{x} < +\infty$, and so have the same consumption set $Z \equiv [0, \bar{x}] \times \mathbb{R}_+$. Each agent i 's preference is defined on Z and represented by a utility function $u_i : Z \rightarrow \mathbb{R}$, which is continuous and quasi-concave on Z , and strictly monotonic (decreasing in labor time and increasing in the

³Tian (2000) constructed a mechanism that implements the proportional solution even if the coordinator does not know the agents' endowment vectors of commodities, under the assumption that agents cannot overstate their endowments. As Tian (2000) mentioned, such an assumption may be justified when endowments consist solely of material goods, since the coordinator can require agents to "place the claimed endowments on the table" (Hurwicz *et al.* (1995)). In our setting where endowments are labor skills, such a requirement is no longer valid, since the coordinator may not inspect the amount of skills in advance of production.

⁴The symbol \mathbb{R}_+ denotes the set of non-negative real numbers.

share of output) on $\overset{\circ}{Z} \equiv [0, \bar{x}] \times \mathbb{R}_{++}$.⁵ We use \mathcal{U} to denote the class of such utility functions. Each agent i also has a **labor skill** which is represented by a positive real number $s_i \in \mathbb{R}_{++}$. The universal set of skills for all agents is denoted by $\mathcal{S} = \mathbb{R}_{++}$. The labor skill $s_i \in \mathcal{S}$ implies i 's **effective labor supply** per hour measured in efficiency units. Thus, if the agent's **supply of labor time** is $x_i \in [0, \bar{x}]$ and the agent's skill is $s_i \in \mathcal{S}$, then $s_i x_i \in \mathbb{R}_+$ represents the agent's substantive contribution in labor supply to production. The production technology is a function $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, which is continuous, strictly increasing, concave with $f(0) = 0$. We choose an arbitrary such production technology function f and keep it fixed in the sequel. Thus, an economy is a pair of profiles $\mathbf{e} \equiv (\mathbf{u}, \mathbf{s})$ with $\mathbf{u} = (u_i)_{i \in N} \in \mathcal{U}^n$ and $\mathbf{s} = (s_i)_{i \in N} \in \mathcal{S}^n$. Denote the class of such economies by $\mathcal{E} \equiv \mathcal{U}^n \times \mathcal{S}^n$.

Given $\mathbf{s} \in \mathcal{S}^n$, an allocation $\mathbf{z} = (x_i, y_i)_{i \in N} \in Z^n$ is **feasible for \mathbf{s}** if $\sum y_i \leq f(\sum s_i x_i)$. We denote by $Z(\mathbf{s})$ the set of feasible allocations for $\mathbf{s} \in \mathcal{S}^n$. An allocation $\mathbf{z} = (z_i)_{i \in N} \in Z^n$ is **Pareto efficient for $\mathbf{e} = (\mathbf{u}, \mathbf{s}) \in \mathcal{E}$** if $\mathbf{z} \in Z(\mathbf{s})$ and there does not exist $\mathbf{z}' = (z'_i)_{i \in N} \in Z(\mathbf{s})$ such that for all $i \in N$, $u_i(z'_i) \geq u_i(z_i)$, and for some $i \in N$, $u_i(z'_i) > u_i(z_i)$. The **proportional solution** [Roemer and Silvestre (1993)] is a correspondence $PR : \mathcal{E} \rightrightarrows Z^n$ such that, for each $\mathbf{e} = (\mathbf{u}, \mathbf{s}) \in \mathcal{E}$, $PR(\mathbf{e})$ stands for the set of all allocations $\mathbf{z} = (x_i, y_i)_{i \in N} \in Z^n$ which are Pareto efficient for \mathbf{e} such that, for each $i \in N$, $y_i = \frac{s_i x_i}{\sum s_j x_j} f(\sum s_j x_j)$. An allocation $\mathbf{z} \in Z^n$ is **PR-optimal for $\mathbf{e} \in \mathcal{E}$** if $\mathbf{z} \in PR(\mathbf{e})$.

2.1 Sharing mechanisms

We are interested in mechanisms having the property of **labor sovereignty** [Kranich (1994); Yoshihara (2000b)],⁶ which says that every agent can choose freely his own labor time. As such, we focus on the following types of mechanisms. For each $i \in N$, let his strategy space be $A_i \equiv M_i \times [0, \bar{x}]$, with generic element (m_i, x_i) . Note that here M_i stands for an abstract general message space as in classical mechanisms, while the members of $[0, \bar{x}]$, which represent i 's choice of labor time as part of his observable action, are also considered as a strategic variable for i . Let $A \equiv \times_{i \in N} A_i$. Let $w \in \mathbb{R}_+$ be the *total output the coordinator observes after production*. Then, a **sharing mechanism** is a

⁵The symbol \mathbb{R}_{++} denotes the set of positive real numbers.

⁶The previous mechanisms such as Suh (1995), Yoshihara (1999, 2000a), Tian (2000) do not have this property.

function $g : A \times \mathbb{R}_+ \rightarrow \mathbb{R}_+^n$ such that for each $(\mathbf{m}, \mathbf{x}) \in A$ and each $w \in \mathbb{R}_+$, $g(\mathbf{m}, \mathbf{x}, w) = \mathbf{y}$ for some $\mathbf{y} \in \mathbb{R}_+^n$. A sharing mechanism g is **feasible** if for each $(\mathbf{m}, \mathbf{x}) \in A$ and each $w \in \mathbb{R}_+$, $\sum g_i(\mathbf{m}, \mathbf{x}, w) \leq w$. We denote by \mathcal{G} the class of all (feasible sharing) mechanisms. In the following discussion, we assume that the production technology function f is known and the total output after production is observable to the coordinator. Thus, for each $\mathbf{s} \in \mathcal{S}^n$ and each $\mathbf{x} \in [0, \bar{x}]^n$, $w = f(\sum s_j x_j)$ is known to the coordinator after production, without the true information about \mathbf{s} .⁷ Then, $g \in \mathcal{G}$ implies that for each $\mathbf{s} \in \mathcal{S}^n$ and each $(\mathbf{m}, \mathbf{x}) \in A$, $(\mathbf{x}, g(\mathbf{m}, \mathbf{x}, f(\sum s_j x_j))) \in Z(\mathbf{s})$. In the following discussion, for each $g \in \mathcal{G}$, we simply write a value of g as $g(\mathbf{m}, \mathbf{x})$ instead of $g(\mathbf{m}, \mathbf{x}, f(\sum s_j x_j))$ except for when we define new mechanisms in \mathcal{G} .

Given $g \in \mathcal{G}$, a **(feasible) sharing game** is defined for each economy $\mathbf{e} \in \mathcal{E}$ as a non-cooperative game (N, A, g, \mathbf{e}) . Fixing the set of players N and their strategy sets A , we simply denote a feasible sharing game (N, A, g, \mathbf{e}) by (g, \mathbf{e}) .

Given a profile $(\mathbf{m}, \mathbf{x}) \in A$, let $(m'_i, \mathbf{m}_{-i}, x'_i, \mathbf{x}_{-i}) \in A$ be another strategy profile that is obtained by replacing the i -th component (m_i, x_i) of (\mathbf{m}, \mathbf{x}) with (m'_i, x'_i) . A profile $(\mathbf{m}^*, \mathbf{x}^*) \in A$ is a **(pure-strategy) Nash equilibrium of (g, \mathbf{e})** if for each $i \in N$ and each $(m_i, x_i) \in A_i$, $u_i(x_i^*, g_i(\mathbf{m}^*, \mathbf{x}^*)) \geq u_i(x_i, g_i(m_i, \mathbf{m}_{-i}^*, x_i, \mathbf{x}_{-i}^*))$. Let $NE(g, \mathbf{e})$ denote the set of Nash equilibria of (g, \mathbf{e}) . An allocation $\mathbf{z} = (x_i, y_i)_{i \in N} \in Z^n$ is a **Nash equilibrium allocation of (g, \mathbf{e})** if there exists $\mathbf{m} \in M$ such that $(\mathbf{m}, \mathbf{x}) \in NE(g, \mathbf{e})$ and $\mathbf{y} = g(\mathbf{m}, \mathbf{x})$, where $\mathbf{x} = (x_i)_{i \in N}$ and $\mathbf{y} = (y_i)_{i \in N}$. Let $NA(g, \mathbf{e})$ denote the set of Nash equilibrium allocations of (g, \mathbf{e}) . A mechanism $g \in \mathcal{G}$ **implements PR in Nash equilibria** if for each $\mathbf{e} \in \mathcal{E}$, $NA(g, \mathbf{e}) = PR(\mathbf{e})$.

A profile $(\mathbf{m}^*, \mathbf{x}^*) \in A$ is a **(pure-strategy) strong (Nash) equilibrium of (g, \mathbf{e})** if for each $T \subseteq N$ and each $(m_i, x_i)_{i \in T} \in (A_i)_{i \in T}$, there exists $j \in T$ such that

$$u_j(x_j^*, g_j(\mathbf{m}^*, \mathbf{x}^*)) \geq u_j(x_j, g_j((m_i, x_i)_{i \in T}, (m_k^*, x_k^*)_{k \in T^c})) .^8$$

⁷Since the coordinator also knows f and \mathbf{x} , he can figure out that the true skill profile belongs to the hyperplane $\{\mathbf{s} \in \mathcal{S}^n \mid \mathbf{s} \cdot \mathbf{x} = f^{-1}(w)\}$. However, the exact location of the true skill profile in this hyperplane cannot be figured out. Note that, to see which of the feasible allocations are true PR -optimal allocations, one needs to know the information of the true skill profile.

⁸For each $T \subseteq N$, $\#T$ denotes the number of agents in T . For each $T \subseteq N$, T^c denotes the complement of T in N .

Let $SNE(g, \mathbf{e})$ denote the set of strong equilibria of (g, \mathbf{e}) . An allocation $\mathbf{z} = (x_i, y_i)_{i \in N} \in Z^n$ is a **strong equilibrium allocation** of (g, \mathbf{e}) if there exists $\mathbf{m} \in M$ such that $(\mathbf{m}, \mathbf{x}) \in SNE(g, \mathbf{e})$ and $\mathbf{y} = g(\mathbf{m}, \mathbf{x})$. Let $SNA(g, \mathbf{e})$ denote the set of strong equilibrium allocations of (g, \mathbf{e}) . A mechanism $g \in \mathcal{G}$ **implements PR in strong equilibria**, if for each $\mathbf{e} \in \mathcal{E}$, $SNA(g, \mathbf{e}) = PR(\mathbf{e})$. Moreover, a mechanism $g \in \mathcal{G}$ **doubly implements PR in Nash and strong equilibria** if for each $\mathbf{e} \in \mathcal{E}$, $NA(g, \mathbf{e}) = SNA(g, \mathbf{e}) = PR(\mathbf{e})$.

2.2 Timing problem in sharing mechanisms

Before discussing our own mechanism, we should mention the timing problem of strategy-decision in real applications of mechanisms, which is particularly relevant to the case of production economies. Note that \mathbf{m} and \mathbf{x} represent different kinds of strategic actions: \mathbf{m} is the list of agents' announcements of their private information, while \mathbf{x} is their production activity from supplying labor time. Thus, there may be a difference between the point in time when \mathbf{m} is announced and the time when \mathbf{x} is exercised. It implies that there may be at least *two polar cases of time sequence with regard to decision making*: the agents may announce \mathbf{m} before they engage in production, or they may announce \mathbf{m} after supplying \mathbf{x} . The former enables each agent i to choose his labor supply with knowledge of the announcements \mathbf{m} , whereas the latter enables each agent i to choose m_i with knowledge of the agents' actions \mathbf{x} in the production process.

Thus, we should consider at least two types of two-stage game forms: Given $g \in \mathcal{G}$, the first two-stage extensive game form derived from g is a feasible mechanism $\Gamma_g^{m \circ x}$ in which Stage 1 consists of selecting $\mathbf{m} \in M$, Stage 2 consists of selecting $\mathbf{x} \in [0, \bar{x}]^n$, and then $(\mathbf{x}, g(\mathbf{m}, \mathbf{x}))$ is the outcome. The second two-stage extensive game form derived from g is a feasible mechanism $\Gamma_g^{x \circ m}$ in which Stage 1 consists of selecting $\mathbf{x} \in [0, \bar{x}]^n$, Stage 2 consists of selecting $\mathbf{m} \in M$, and then $(\mathbf{x}, g(\mathbf{m}, \mathbf{x}))$ is the outcome.

Given a two-stage game $(\Gamma_g^{m \circ x}, \mathbf{e})$ and a strategy profile $\mathbf{m} \in M$ in Stage 1, let $(\Gamma_g^{m \circ x}(\mathbf{m}), \mathbf{e})$ be the corresponding Stage 2 subgame. A **strategy mapping** $\chi : M \rightarrow [0, \bar{x}]^n$ is a function such that for each $\mathbf{m} \in M$, $\chi(\mathbf{m})$ is a strategy profile of the subgame $(\Gamma_g^{m \circ x}(\mathbf{m}), \mathbf{e})$. Let \mathcal{X} be the set of such mappings. A profile $(\mathbf{m}^*, \chi^*) \in M \times \mathcal{X}$ is a **(pure-strategy) subgame perfect (Nash) equilibrium** of $(\Gamma_g^{m \circ x}, \mathbf{e})$ if for each $i \in N$, each $m_i \in M_i$,

each $\chi \in \mathcal{X}$ with $\chi = (\chi_i, \chi_{-i}^*)$, and each $\mathbf{m} \in M$,

$$u_i(\chi_i^*(\mathbf{m}^*), g_i(\mathbf{m}^*, \chi^*(\mathbf{m}^*))) \geq u_i(\chi_i^*(m_i, \mathbf{m}_{-i}^*), g_i(m_i, \mathbf{m}_{-i}^*, \chi^*(m_i, \mathbf{m}_{-i}^*)))$$

and $u_i(\chi_i^*(\mathbf{m}), g_i(\mathbf{m}, \chi^*(\mathbf{m}))) \geq u_i(\chi_i(\mathbf{m}), g_i(\mathbf{m}, \chi(\mathbf{m}))),$

where $\chi_i^*(\mathbf{m})$ (*resp.* $\chi_i(\mathbf{m})$) is the i -th component of $\chi^*(\mathbf{m})$ (*resp.* $\chi(\mathbf{m})$) in Stage 2 subgame induced by the choice \mathbf{m} in Stage 1.

Given a two-stage game $(\Gamma_g^{x^om}, \mathbf{e})$ and a strategy profile $\mathbf{x} \in [0, \bar{x}]^n$ in Stage 1, let $(\Gamma_g^{x^om}(\mathbf{x}), \mathbf{e})$ be the corresponding Stage 2 subgame. A **strategy mapping** $\mu : [0, \bar{x}]^n \rightarrow M$ is a function such that for each $\mathbf{x} \in [0, \bar{x}]^n$, $\mu(\mathbf{x})$ is a strategy profile of the subgame $(\Gamma_g^{x^om}(\mathbf{x}), \mathbf{e})$. Let \mathcal{M} be the set of such mappings. A profile $(\mu^*, \mathbf{x}^*) \in \mathcal{M} \times [0, \bar{x}]^n$ is a **(pure-strategy) subgame perfect (Nash) equilibrium of $(\Gamma_g^{x^om}, \mathbf{e})$** if for each $i \in N$, each $x_i \in [0, \bar{x}]$, each $\mu \in \mathcal{M}$ with $\mu = (\mu_i, \mu_{-i}^*)$, and each $\mathbf{x} \in [0, \bar{x}]^n$,

$$u_i(x_i^*, g_i(\mu^*(\mathbf{x}^*), \mathbf{x}^*)) \geq u_i(x_i, g_i(\mu^*(x_i, \mathbf{x}_{-i}^*), x_i, \mathbf{x}_{-i}^*)),$$

and $u_i(x_i, g_i(\mu^*(\mathbf{x}), \mathbf{x})) \geq u_i(x_i, g_i(\mu(\mathbf{x}), \mathbf{x})).$

Let $SPE(\Gamma_g^{m^ox}, \mathbf{e})$ be the set of subgame perfect equilibria of $(\Gamma_g^{m^ox}, \mathbf{e})$. An allocation $\mathbf{z} = (x_i, y_i)_{i \in N} \in Z^n$ is a **subgame perfect equilibrium allocation of $(\Gamma_g^{m^ox}, \mathbf{e})$** if there exists $(\mathbf{m}, \chi) \in SPE(\Gamma_g^{m^ox}, \mathbf{e})$ such that $\chi(\mathbf{m}) = \mathbf{x}$ and $\mathbf{y} = g(\mathbf{m}, \chi(\mathbf{m}))$. Let $SPA(\Gamma_g^{m^ox}, \mathbf{e})$ be the set of subgame perfect equilibrium allocations of $(\Gamma_g^{m^ox}, \mathbf{e})$. Given $g \in \mathcal{G}$, $\Gamma_g^{m^ox}$ **implements PR in subgame perfect equilibria** if for each $\mathbf{e} \in \mathcal{E}$, $SPA(\Gamma_g^{m^ox}, \mathbf{e}) = PR(\mathbf{e})$. Given $g \in \mathcal{G}$, $\Gamma_g^{m^ox}$ **triply implements PR in Nash, strong, and subgame perfect equilibria** if for each $\mathbf{e} \in \mathcal{E}$, $NA(g, \mathbf{e}) = SNA(g, \mathbf{e}) = SPA(\Gamma_g^{m^ox}, \mathbf{e}) = PR(\mathbf{e})$.⁹ Parallel definitions apply to $(\Gamma_g^{x^om}, \mathbf{e})$.

3 Implementation of the Proportional Solution

In the following, we impose two additional assumptions.

Assumption 1 (boundary condition): $\forall i \in N, \forall z_i \in \overset{\circ}{Z}, \forall z'_i \in [0, \bar{x}] \times \{0\}, u_i(z_i) > u_i(z'_i)$.

⁹This definition contains some abuse of language, as the implementation in Nash and strong equilibria is achieved by the mechanism (g, \mathbf{e}) , while it is $\Gamma_g^{m^ox}$ that implements PR in subgame perfect equilibria.

Assumption 2: *The production function f is continuously differentiable.*

We denote by $f'(x)$ the *derivative of f at x* .

The message space M of the mechanism in this paper is defined by $M \equiv \mathcal{S}^n \times \mathbb{R}_+^n$ with generic element $(\boldsymbol{\sigma}, \mathbf{y})$, where $\boldsymbol{\sigma} = (\sigma_i)_{i \in N}$, in which σ_i denotes i 's reported skill, and $\mathbf{y} = (y_i)_{i \in N}$, in which y_i denotes i 's demand for output.

3.1 Nash and strong implementability

In this subsection, we will set aside the timing problem of sharing mechanisms and propose a sharing mechanism as a normal form game form that implements PR in Nash and strong equilibria. To propose our mechanism, let us introduce two feasible sharing mechanisms, defined as follows:

- $g^\pi \in \mathcal{G}$ is such that for each $(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}) \in \mathcal{S}^n \times [0, \bar{x}]^n \times \mathbb{R}_+^n$, each $w \in \mathbb{R}_+$, and each $i \in N$,

$$g_i^\pi(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}, w) = \begin{cases} w & \text{if } \Pi(\mathbf{x}_{-i}) \neq \emptyset, x_i = \pi(\mathbf{x}_{-i}), \text{ and } y_i > \max\{y_j \mid j \neq i\} \\ 0 & \text{otherwise} \end{cases}$$
 where $\Pi(\mathbf{x}_{-i}) \equiv \{ \frac{x_j + \bar{x}}{2} \mid x_j < \bar{x} \text{ for } j \neq i \}$ and $\pi(\mathbf{x}_{-i}) \equiv \max \Pi(\mathbf{x}_{-i})$.
- $g^\sigma \in \mathcal{G}$ is such that for each $(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}) \in \mathcal{S}^n \times [0, \bar{x}]^n \times \mathbb{R}_+^n$, each $w \in \mathbb{R}_+$, and each $i \in N$,

$$g_i^\sigma(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}, w) = \begin{cases} w & \text{if } x_i = 0, \text{ and } \sigma_i > \sigma_j \text{ for all } j \neq i, \\ 0 & \text{otherwise.} \end{cases}$$

The mechanism $g^\pi \in \mathcal{G}$ assigns all of the produced output to only one agent who provides less than \bar{x} of labor time, but the maximal positive amount among those who provide less than \bar{x} of labor time, and reports a maximal amount of demand for the output, if there is such an agent at all. The mechanism $g^\sigma \in \mathcal{G}$ also assigns all of the produced output to only one agent who reports the highest labor skill and provides no labor time.

Given $(\mathbf{s}, \mathbf{x}, \mathbf{y}) = (s_i, x_i, y_i)_{i \in N} \in \mathcal{S}^n \times Z^n$, let $PR(\mathbf{s}, \mathbf{x}, \mathbf{y})^{-1} \equiv \{ \mathbf{u} \in \mathcal{U}^n \mid (\mathbf{x}, \mathbf{y}) \in PR(\mathbf{u}, \mathbf{s}) \}$. If $PR(\mathbf{s}, \mathbf{x}, \mathbf{y})^{-1} \neq \emptyset$, then (\mathbf{x}, \mathbf{y}) should be a **PR -optimal allocation for some economy with skill profile \mathbf{s}** . Let us call such an $(\mathbf{s}, \mathbf{x}, \mathbf{y})$ a **PR -consistent profile**. Note that if for all $i \in N$, $y_i = \frac{s_i x_i}{\sum_{s_k x_k} s_k x_k} f(\sum s_k x_k)$ and (\mathbf{x}, \mathbf{y}) is an interior allocation, then $PR(\mathbf{s}, \mathbf{x}, \mathbf{y})^{-1} \neq \emptyset$. Given $(\mathbf{s}, \mathbf{x}, \mathbf{y}) \in \mathcal{S}^n \times Z^n$, let $N(\mathbf{s}, \mathbf{x}, \mathbf{y}) \equiv \{ i \in N \mid \exists (s'_i, x'_i, y'_i) \in \mathcal{S} \times Z \text{ s.t. } PR(s'_i, x'_i, y'_i, \mathbf{s}_{-i}, \mathbf{x}_{-i}, \mathbf{y}_{-i})^{-1} \neq \emptyset \}$. This $N(\mathbf{s}, \mathbf{x}, \mathbf{y})$ is the set of **potential**

deviators under the profile $(\mathbf{s}, \mathbf{x}, \mathbf{y})$, since any $i \in N(\mathbf{s}, \mathbf{x}, \mathbf{y})$ can constitute a PR -consistent profile with the others' fixed strategies by switching his strategy from (s_i, x_i, y_i) . Given $(\mathbf{s}, \mathbf{x}, \mathbf{y}) \in \mathcal{S}^n \times Z^n$ and $i \in N(\mathbf{s}, \mathbf{x}, \mathbf{y})$, let $\mathcal{S}^i(\mathbf{s}, \mathbf{x}, \mathbf{y}) \equiv \{s'_i \in \mathcal{S} \mid \exists (x'_i, y'_i) \in Z \text{ s.t. } PR(s'_i, x'_i, y'_i, \mathbf{s}_{-i}, \mathbf{x}_{-i}, \mathbf{y}_{-i})^{-1} \neq \emptyset\}$. Note that $\mathcal{S}^i(\mathbf{s}, \mathbf{x}, \mathbf{y})$ is closed and bounded from below, or otherwise, $\mathcal{S}^i(\mathbf{s}, \mathbf{x}, \mathbf{y}) = \mathcal{S}$. The latter case occurs if and only if there exists $b \in \mathbb{R}_{++}$ such that $\sum_{k \neq i} s_k x_k < b$ and f is linear on $[0, b]$.

Given $(\mathbf{s}, \mathbf{x}, \mathbf{y}) \in \mathcal{S}^n \times Z^n$ and $i \in N(\mathbf{s}, \mathbf{x}, \mathbf{y})$, let $(\widehat{s}_i, \widehat{x}_i, \widehat{y}_i) \in \mathcal{S} \times Z$ be defined by $\widehat{s}_i = \arg \min_{s'_i \in \mathcal{S}^i(\mathbf{s}, \mathbf{x}, \mathbf{y})} |s'_i - s_i|$, $0 < \widehat{x}_i < \bar{x}$, and \widehat{y}_i with $\widehat{y}_i + \sum_{j \neq i} y_j = f\left(\sum_{j \neq i} s_j x_j + \widehat{s}_i \widehat{x}_i\right)$ and $\frac{\sum_{j \neq i} y_j}{\widehat{y}_i} = \frac{\sum_{j \neq i} s_j x_j}{\widehat{s}_i \widehat{x}_i}$. Then, $PR(\widehat{s}_i, \widehat{x}_i, \widehat{y}_i, \mathbf{s}_{-i}, \mathbf{x}_{-i}, \mathbf{y}_{-i})^{-1} \neq \emptyset$. Note that such $(\widehat{s}_i, \widehat{x}_i, \widehat{y}_i)$ is well-defined: first, \widehat{s}_i is uniquely determined, since if $s_i \in \mathcal{S}^i(\mathbf{s}, \mathbf{x}, \mathbf{y})$, then $\widehat{s}_i = s_i$, whereas if $s_i \notin \mathcal{S}^i(\mathbf{s}, \mathbf{x}, \mathbf{y})$, then $\mathcal{S}^i(\mathbf{s}, \mathbf{x}, \mathbf{y})$ is bounded from below and $\widehat{s}_i = \min \mathcal{S}^i(\mathbf{s}, \mathbf{x}, \mathbf{y})$. Second, once \widehat{s}_i is uniquely determined, then the other agents' strategies together with \widehat{s}_i give us the information about i 's *potential consumption vector* $(\widehat{x}_i, \widehat{y}_i)$ by the proportionality of the PR -optimal allocation.

We introduce $g^* \in \mathcal{G}$ which works in each given $w \in \mathbb{R}_+$ as follows:

Let any $(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}) = (\sigma_i, x_i, y_i)_{i \in N} \in \mathcal{S}^n \times [0, \bar{x}]^n \times \mathbb{R}_+^n$ be given.

Rule 1: If $f(\sum \sigma_k x_k) = w$, then

1-1: if $PR(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y})^{-1} \neq \emptyset$, then $g^*(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}, w) = \mathbf{y}$,

1-2: if $PR(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y})^{-1} = \emptyset$, and $N(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}) \neq \emptyset$, then

1-2-1: if $\#N(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}) > 1$, then $g^*(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}, w) = \mathbf{0}$,

1-2-2: if $N(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}) = \{j\}$ for some $j \in N$, then

$g_j^*(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}, w) = \min \left\{ \max \left\{ 0, \widehat{y}_j + \widehat{s}_j \cdot f' \left(\sum_{k \neq j} \sigma_k x_k + \widehat{s}_j \widehat{x}_j \right) \cdot (x_j - \widehat{x}_j) \right\}, w \right\}$
and $g_i^*(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}, w) = 0$ for any $i \neq j$, where $PR(\widehat{s}_j, \widehat{x}_j, \widehat{y}_j, \boldsymbol{\sigma}_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j})^{-1} \neq \emptyset$,

1-3: in any other case, $g^*(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}, w) = g^\pi(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}, w)$.

Rule 2: If $f(\sum \sigma_k x_k) \neq w$, then $g^*(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}, w) = g^\sigma(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}, w)$.

First, g^* computes the expected output $f(\sum \sigma_k x_k)$ from the data $(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y})$ and compares this with the real output w . In the case where these two values coincide, if $(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y})$ is PR -consistent, then g^* gives the agents their desired \mathbf{y} under **Rule 1-1**; if $(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y})$ is not PR -consistent, and there exists at least one potential deviator, say j , then g^* gives him at most a share of outcome available in the budget set with the supporting price $\widehat{s}_j \cdot f' \left(\sum_{k \neq j} \sigma_k x_k + \widehat{s}_j \widehat{x}_j \right)$,

while g^* gives nothing to any other agents under **Rule 1-2**; for any other case, g^* applies g^π under **Rule 1-3**. Finally, if $f(\sum \sigma_k x_k)$ and w are different, then g^* applies g^σ under **Rule 2**.

It is easy to see that g^* is *forthright* [Saijo *et al.* (1996)] and satisfies *best response property* [Jackson *et al.* (1994)]. Moreover, g^* is a *quantity* type, and so *self-relevant* [Hurwicz (1960)]. It is also easy to check that g^* is feasible.

Let us briefly explain how g^* induces true information about skills. This g^* has to distribute the total output $f(\sum s_k x_k)$ according to $(\sigma, \mathbf{x}, \mathbf{y})$, where $(s_k)_{k \in N}$ stands for the true skill profile and the coordinator cannot know whether $\sigma = \mathbf{s}$ or not. If $f(\sum \sigma_k x_k) \neq f(\sum s_k x_k)$, however, then clearly $\sigma \neq \mathbf{s}$, and there is at least one agent, say $j \in N$, such that $\sigma_j \neq s_j$ and $x_j > 0$. Then, this agent is definitely punished under **Rule 2**. Next, consider the case where $f(\sum \sigma_k x_k) = f(\sum s_k x_k)$ but $\sigma \neq \mathbf{s}$. Then, there are either at least two agents $i, j \in N$ such that $\sigma_i \neq s_i$, $\sigma_j \neq s_j$, $x_i > 0$, and $x_j > 0$, or else there exists an agent $j \in N$ such that $\sigma_j \neq s_j$ and $x_j = 0$. If the latter case is applied, then the agents such as j will be punished under **Rule 1-2** or **Rule 1-3**. If the former case is applied, then one of the agents, $j \in N$, with $\sigma_j \neq s_j$ can induce **Rule 2** by switching from $x_j > 0$ to $x'_j = 0$, together with announcing a higher number $\sigma'_j \neq s_j$ than any other σ_{-j} . Thus, this case may not correspond to an equilibrium. The following lemma actually confirms this insight:

Lemma 1: *Let $g^* \in \mathcal{G}$ be given. Given $(\mathbf{u}, \mathbf{s}) \in \mathcal{E}$, let $(\sigma, \mathbf{x}, \mathbf{y}) \in \mathcal{S}^n \times [0, \bar{x}]^n \times \mathbb{R}_+^n$ be a Nash equilibrium of $(g^*, \mathbf{u}, \mathbf{s})$ such that $f(\sum \sigma_k x_k) = f(\sum s_k x_k)$. Then, it follows that for all $i \in N$ with $x_i > 0$, $\sigma_i = s_i$.*

Now, we examine the performance of g^* .

Theorem 1: *Let Assumptions 1 and 2 hold. Now the mechanism g^* implements PR in Nash and strong equilibria.*

Note that g^* works even in economies of *two agents*.

3.2 Implementation of the proportional solution with the timing problem

Because of the timing problem discussed in Section 2.2, g^* may be played as $\Gamma_{g^*}^{(\sigma, \mathbf{y}) \circ \mathbf{x}}$ or $\Gamma_{g^*}^{\mathbf{x} \circ (\sigma, \mathbf{y})}$. In this situation, the coordinator may not know in ad-

vance the *information structure of the two-stage game* induced by $\Gamma_{g^*}^{(\sigma, \mathbf{y}) \circ \mathbf{x}}$ or $\Gamma_{g^*}^{\mathbf{x} \circ (\sigma, \mathbf{y})}$, even if the coordinator has control over the number of stages in the mechanism: this information structure among agents may be characterized as *perfect information*, or as *complete but imperfect information* about Stage 1. If the game is played as one with perfect information (*resp.* complete but imperfect information), we should consider subgame-perfect equilibria (*resp.* Nash equilibria) in the two-stage game. For instance, let us assume that the stage of announcing (σ, \mathbf{y}) to the coordinator is in advance of production, in which \mathbf{x} is supplied, and the information (σ, \mathbf{y}) is not made public by the coordinator until the end of the production process. If this is effectively enforced, then the coordinator may be concerned with Nash and strong implementation only. Such a scenario implicitly assumes that the coordinator can effectively obstruct any private communication regarding (σ, \mathbf{y}) among agents until the end of the production process. However, the coordinator may not be able to effectively wield this power, and agents could privately communicate with each other regarding (σ, \mathbf{y}) . This is one case that the coordinator cannot control the information structure among agents. In such a situation, the double implementability by $\Gamma_{g^*}^{(\sigma, \mathbf{y}) \circ \mathbf{x}}$ (*resp.* $\Gamma_{g^*}^{\mathbf{x} \circ (\sigma, \mathbf{y})}$) in Nash and subgame-perfect equilibria would be strongly attractive, since it keeps the desirable performance of the mechanism without relying on the information structure among agents:

Theorem 2: *Let Assumptions 1 and 2 hold. Now the mechanism $\Gamma_{g^*}^{(\sigma, \mathbf{y}) \circ \mathbf{x}}$ doubly implements PR in Nash and subgame-perfect equilibria.*

Theorem 3: *Let Assumptions 1 and 2 hold. Now the mechanism $\Gamma_{g^*}^{\mathbf{x} \circ (\sigma, \mathbf{y})}$ doubly implements PR in Nash and subgame-perfect equilibria.*

By the three theorems discussed above, we can summarize as follows:

Corollary: *Let Assumptions 1 and 2 hold. Now both of the mechanisms $\Gamma_{g^*}^{(\sigma, \mathbf{y}) \circ \mathbf{x}}$ and $\Gamma_{g^*}^{\mathbf{x} \circ (\sigma, \mathbf{y})}$ respectively triply implement PR in Nash, subgame-perfect, and strong equilibria.*

This result implies that g^* implements PR even if it permits each agent various kinds of freedom: the agent may choose freely his own supply of labor time; the agent is permitted to overstate his labor skill; the agent

can behave unilaterally or coalitionally; and the agent can behave strong-rationally, as in the subgame-perfect response, or weak-rationally, as in the Nash-like response.¹⁰

4 Concluding Remarks

We have proposed a feasible sharing mechanism that triply implements the proportional solution in Nash, subgame-perfect, and strong equilibria, even when agents can understate or overstate their labor skills. The performance of our mechanism is summarized in Table 1, which provides a comparison with other relevant mechanisms.

Insert Table 1 around here.

As revealed in Table 1, our mechanism has two undesirable features. First, it lacks continuity. Second, the mechanism fails to meet *balancedness* or *nonwastefulness*. One reason of the second undesirability is that the mechanism permits agents to both overstate and understate their labor skills. Thus, it is difficult to find the deviator when only aggregate information ($f(\sum \sigma_k x_k)$ and $f(\sum s_k x_k)$) is available. Therefore, the mechanism basically punishes all agents when there is a deviator. The other reason is that this mechanism is characterized by labor sovereignty. The labor sovereignty mechanism should accept a profile of the agents' choice of labor time as an outcome, even when it may constitute a nondesirable allocation. Thus, if the mechanism needs to punish potential deviators, this is only possible by reducing their share of output, which leads to the violation of balancedness. We surmise that there may be a trade-off between labor sovereignty and balancedness. However, it remains an open question whether or not there exists

¹⁰Let us point out this more precisely. Note that *intermediate* situations may happen where some agents behave unilaterally, and others coalitionally. For instance, let \mathcal{H} be a subset of the power set of N , which is the set of admissible coalitions, and consider an equilibrium relative to \mathcal{H} , i.e. message-action profiles upon which no coalition in \mathcal{H} can improve. When \mathcal{H} stands for the set of all singletons, the equilibrium relative to \mathcal{H} becomes the Nash equilibrium, while it becomes the strong equilibrium in case \mathcal{H} is the set of all coalitions. The problem is that the mechanism coordinator may not expect what equilibrium notions the agents choose in the play of the game, since he has no information about the structure of \mathcal{H} . However, once a doubly-implementing mechanism is provided, such a difficulty is resolved, which gives us a motivation for the double implementation in Nash and strong equilibria.

a mechanism that satisfies both labor sovereignty and balancedness, and at the same time implements the proportional solution.

5 Appendix

Proof of Lemma 1. Suppose there exists $j \in N$ with $\sigma_j \neq s_j$ and $x_j > 0$. Let $N_L(\boldsymbol{\sigma})$ be the set of such j . Since $f(\sum \sigma_k x_k) = f(\sum s_k x_k)$, $N_L(\boldsymbol{\sigma})$ is not a singleton. Moreover, for each $j \in N_L(\boldsymbol{\sigma})$, $y'_j = f\left(\sum_{i \neq j} s_i x_i\right) > 0$ by $\sigma'_j > \max\{\sigma_i \mid i \neq j\}$ and $x'_j = 0$ under Rule 2. Note that

$$\begin{aligned}
\sum_{j \in N_L(\boldsymbol{\sigma})} y'_j &= \sum_{j \in N_L(\boldsymbol{\sigma})} f\left(\sum_{i \neq j} s_i x_i\right) = \sum_{j \in N_L(\boldsymbol{\sigma})} f\left(\sum_{i \in N_L(\boldsymbol{\sigma}) \setminus \{j\}} s_i x_i + \sum_{i \notin N_L(\boldsymbol{\sigma})} s_i x_i\right) \\
&\geq \sum_{j \in N_L(\boldsymbol{\sigma})} f\left(s_j x_j + \sum_{i \notin N_L(\boldsymbol{\sigma})} s_i x_i\right) \quad (\text{since } N_L(\boldsymbol{\sigma}) \text{ is not a singleton}) \\
&\geq f\left(\sum_{j \in N_L(\boldsymbol{\sigma})} \left(s_j x_j + \sum_{i \notin N_L(\boldsymbol{\sigma})} s_i x_i\right)\right) \quad (\text{since } f \text{ is concave and } f(0) \geq 0) \\
&\geq f\left(\sum_{j \in N_L(\boldsymbol{\sigma})} s_j x_j + \sum_{i \notin N_L(\boldsymbol{\sigma})} s_i x_i\right) \geq \sum_{j \in N_L(\boldsymbol{\sigma})} y_j \equiv \sum_{j \in N_L(\boldsymbol{\sigma})} g_j^*(\boldsymbol{\sigma}, \boldsymbol{x}, \boldsymbol{y}).
\end{aligned}$$

If there exists $j \in N_L(\boldsymbol{s})$ such that $y'_j > y_j$, then j has an incentive to switch from x_j to $x'_j = 0$ and report $\sigma'_j > \max\{\sigma_i \mid i \neq j\}$. If $y'_j = y_j$ for each $j \in N_L(\boldsymbol{\sigma})$, then each $j \in N_L(\boldsymbol{\sigma})$ has an incentive to switch from x_j to $x'_j = 0$, since $u_j(x'_j, y'_j) = u_j(0, y_j) > u_j(x_j, y_j)$ by the strict monotonicity of utility functions. Thus, in any case, it contradicts the fact that $(\boldsymbol{\sigma}, \boldsymbol{x}, \boldsymbol{y})$ is a Nash equilibrium. ■

Proof of Theorem 1.

(1) Show $PR(\boldsymbol{u}, \boldsymbol{s}) \subseteq NA(g^*, \boldsymbol{u}, \boldsymbol{s})$ for each $(\boldsymbol{u}, \boldsymbol{s}) \in \mathcal{E}$. Let $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y}) \in PR(\boldsymbol{u}, \boldsymbol{s})$. If the strategy profile is $(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) \in \mathcal{S}^n \times [0, \bar{x}]^n \times \mathbb{R}_+^n$, then $g^*(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y}$ by **Rule 1-1**. Since $\boldsymbol{s} \gg \mathbf{0}$ and \boldsymbol{z} is an efficient proportional allocation, **Assumption 1** implies $\boldsymbol{x} \gg \mathbf{0}$ and $g_i^*(\boldsymbol{s}, \boldsymbol{x}, \boldsymbol{y}) > 0$ for each $i \in N$. Suppose that $j \in N$ deviates from (s_j, x_j, y_j) to $(\sigma'_j, x'_j, y'_j) \in \mathcal{S} \times [0, \bar{x}] \times \mathbb{R}_+$.

Note first that the deviation cannot induce **Rule 1-3**. If the deviation results in **Rule 1-2-1**, then $g_j^*(\sigma'_j, x'_j, y'_j, \mathbf{s}_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) = 0$. If the deviation induces **Rule 2**, then $x'_j > 0$. Hence, $g_j^*(\sigma'_j, x'_j, y'_j, \mathbf{s}_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) = 0$ under **Rule 2**.

If the deviation induces **Rule 1-2-2** with $\sigma'_j \neq s_j$, then $x'_j = 0$. Thus, $g_j^*(\sigma'_j, x'_j, y'_j, \mathbf{s}_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) \leq \hat{y}_j - \hat{s}_j \hat{x}_j \cdot f' \left(\sum_{i \neq j} s_i x_i + \hat{s}_j \hat{x}_j \right)$ by some $(\hat{s}_j, \hat{x}_j, \hat{y}_j) \in \mathcal{S} \times [0, \bar{x}] \times \mathbb{R}_+$ such that $PR(\hat{s}_j, \hat{x}_j, \hat{y}_j, \mathbf{s}_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j})^{-1} \neq \emptyset$. Suppose $f \left(\sum_{i \neq j} s_i x_i + \hat{s}_j \hat{x}_j \right) \neq f \left(\sum_{i \neq j} s_i x_i + s_j x_j \right)$. Since $PR(\mathbf{s}, \mathbf{x}, \mathbf{y})^{-1} \neq \emptyset$, $\frac{f(\sum_{i \neq j} s_i x_i + \hat{s}_j \hat{x}_j)}{\sum_{i \neq j} s_i x_i + \hat{s}_j \hat{x}_j} = \frac{\sum_{i \neq j} y_i}{\sum_{i \neq j} s_i x_i} = \frac{f(\sum_{i \neq j} s_i x_i + s_j x_j)}{\sum_{i \neq j} s_i x_i + s_j x_j}$. Since f is concave, f must be linear on a closed interval $\left[0, \max \left\{ \sum_{i \neq j} s_i x_i + \hat{s}_j \hat{x}_j, \sum_{i \neq j} s_i x_i + s_j x_j \right\} \right]$. Hence, $\hat{y}_j - \hat{s}_j \hat{x}_j \cdot f' \left(\sum_{i \neq j} s_i x_i + \hat{s}_j \hat{x}_j \right) = 0$. Thus, $g_j^*(\sigma'_j, x'_j, y'_j, \mathbf{s}_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) \leq 0$. Next, suppose that $f \left(\sum_{i \neq j} s_i x_i + \hat{s}_j \hat{x}_j \right) = f \left(\sum_{i \neq j} s_i x_i + s_j x_j \right)$. This implies $\hat{s}_j \hat{x}_j = s_j x_j$ and $\hat{y}_j = y_j$. Thus, $g_j^*(\sigma'_j, x'_j, y'_j, \mathbf{s}_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) \leq \hat{y}_j - \hat{s}_j \hat{x}_j \cdot f' \left(\sum_{i \neq j} s_i x_i + \hat{s}_j \hat{x}_j \right) = y_j - s_j x_j \cdot f' \left(\sum s_k x_k \right)$.

Consider the case where the deviation induces **Rule 1-2-2** with $\sigma'_j = s_j$. Since $\frac{f(\sum_{i \neq j} s_i x_i + s_j \hat{x}_j)}{\sum_{i \neq j} s_i x_i + s_j \hat{x}_j} = \frac{\sum_{i \neq j} y_i}{\sum_{i \neq j} s_i x_i} = \frac{f(\sum_{i \neq j} s_i x_i + s_j x_j)}{\sum_{i \neq j} s_i x_i + s_j x_j}$ by $PR(\mathbf{s}, \mathbf{x}, \mathbf{y})^{-1} \neq \emptyset$, $f' \left(\sum_{i \neq j} s_i x_i + s_j \hat{x}_j \right) = f' \left(\sum_{i \neq j} s_i x_i + s_j x_j \right)$. Thus, $g_j^*(\sigma'_j, x'_j, \mathbf{x}_{-j}, y'_j, \mathbf{y}_{-j}) \leq y_j + s_j \cdot f' \left(\sum s_k x_k \right) \cdot (x'_j - x_j)$.

Next, if the deviation induces **Rule 1-1**, then $\sigma'_j = s_j$. This is because $\sigma'_j \neq s_j$ implies $x'_j = 0$ under **Rule 1-1**, which contradicts $PR(\sigma'_j, x'_j, y'_j, \mathbf{s}_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j})^{-1} \neq \emptyset$ by **Assumption 1**. However, if the deviation induces **Rule 1-1** with $\sigma'_j = s_j$, then f must be linear on a closed interval $\left[0, \max \left\{ \sum s_k x_k, \sum_{i \neq j} s_i x_i + s_j x'_j \right\} \right]$, and we obtain $g_j^*(\sigma'_j, x'_j, \mathbf{x}_{-j}, y'_j, \mathbf{y}_{-j}) \leq y_j + s_j \cdot f' \left(\sum s_k x_k \right) \cdot (x'_j - x_j)$. In summary, by **Assumption 1** and Pareto efficiency of \mathbf{z} , no agent has an incentive to deviate from $(\mathbf{s}, \mathbf{x}, \mathbf{y})$.

(2) Show $NA(g^*, \mathbf{u}, \mathbf{s}) \subseteq PR(\mathbf{u}, \mathbf{s})$ for each $(\mathbf{u}, \mathbf{s}) \in \mathcal{E}$. Let $(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}) \in NE(g^*, \mathbf{u}, \mathbf{s})$.

Suppose that $(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y})$ corresponds to **Rule 2**. Let $N^0(\mathbf{x}) \equiv \{i \in N \mid x_i = 0\}$. If $N^0(\mathbf{x}) = \emptyset$, then $g_i^*(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}) = 0$ for all $i \in N$. Then, there exists $j \in N$ such that $\sum_{i \neq j} \sigma_i x_i \neq \sum_{i \neq j} s_i x_i$. Thus, $g_j^*(\sigma'_j, x'_j, y'_j, \boldsymbol{\sigma}_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) > 0$ with $\sigma'_j > \max \{\sigma_i \mid i \in N\}$ and $x'_j = 0$ under **Rule 2**. If $\#N^0(\mathbf{x}) \geq 2$, then for

any $j \in N^0(\mathbf{x})$, $g_j^*(\sigma'_j, x'_j, y'_j, \sigma_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) = f(\sum s_k x_k)$ by $x'_j = 0$ and $\sigma'_j > \max\{\sigma_i \mid i \neq j\}$ under **Rule 2**. If $\#N^0(\mathbf{x}) = 1$ and $\#N \setminus N^0(\mathbf{x}) \geq 2$, then there exists $j \in N \setminus N^0(\mathbf{x})$ such that $\sum_{i \in N \setminus (N^0(\mathbf{x}) \cup \{j\})} \sigma_i x_i \neq \sum_{i \in N \setminus (N^0(\mathbf{x}) \cup \{j\})} s_i x_i$. Thus, $g_j^*(\sigma'_j, x'_j, y'_j, \sigma_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) > 0$ by $\sigma'_j > \max\{\sigma_i \mid i \in N\}$ and $x'_j = 0$ under **Rule 2**. If $N^0(\mathbf{x}) = \{i\}$ and $N \setminus N^0(\mathbf{x}) = \{j\}$, then $g_j^*(\sigma'_j, x'_j, y'_j, \sigma_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) > 0$ by $\sigma'_j = s_j$, $x'_j = \frac{\bar{x}}{2}$, and $y'_j > \max\{y_i, f(s_j x'_j)\}$ under **Rule 1-3**. In summary, no profile of strategies can constitute a Nash equilibrium in **Rule 2**.

Suppose that $(\sigma, \mathbf{x}, \mathbf{y})$ corresponds to **Rule 1-3**. Then, $g_j^*(\sigma, \mathbf{x}, \mathbf{y}) = 0$ for some $j \in N$. If either $x_j = 0$ or $\sigma_j = s_j$, then $g_j^*(\sigma'_j, x'_j, y'_j, \sigma_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) > 0$ by $\sigma'_j = s_j$, $x'_j = \pi(\mathbf{x}_{-j})$, and $y'_j > \max\{f(s_j x'_j), \max\{y_i \mid i \neq j\}\}$ under **Rule 1-3**. If $x_j > 0$ and $\sigma_j \neq s_j$, then $g_j^*(\sigma'_j, x'_j, y'_j, \sigma_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) > 0$ by $\sigma'_j > \max\{\sigma_i \mid i \in N\}$ and $x'_j = 0$ under **Rule 2**.

Suppose that $(\sigma, \mathbf{x}, \mathbf{y})$ corresponds to **Rule 1-2-1**. Then, $g_i^*(\sigma, \mathbf{x}, \mathbf{y}) = 0$ for each $i \in N$. Note $x_j = 0$ for some $j \in N$ implies $N(\sigma, \mathbf{x}, \mathbf{y}) = \{j\}$ or $N(\sigma, \mathbf{x}, \mathbf{y}) = \emptyset$. Hence $\mathbf{x} \gg \mathbf{0}$. For each $j \in N$ with $\sigma_j = s_j$, $g_j^*(\sigma'_j, x'_j, y'_j, \sigma_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) > 0$ by $\sigma'_j = s_j$, $x'_j = \pi(\mathbf{x}_{-j})$, and $y'_j > \max\{f(s_j x'_j), \max\{y_i \mid i \neq j\}\}$ under **Rule 1-3** or **Rule 1-2-2**. Also, for each $j \in N$ with $\sigma_j \neq s_j$, $g_j^*(\sigma'_j, x'_j, y'_j, \sigma_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) > 0$ by $\sigma'_j > \max\{\sigma_i \mid i \in N\}$ and $x'_j = 0$ under **Rule 2**.

Suppose that $(\sigma, \mathbf{x}, \mathbf{y})$ corresponds to **Rule 1-2-2**. Then, $N \setminus N(\sigma, \mathbf{x}, \mathbf{y}) \neq \emptyset$, and $x_i > 0$ and $g_i^*(\sigma, \mathbf{x}, \mathbf{y}) = 0$ for each $i \in N \setminus N(\sigma, \mathbf{x}, \mathbf{y})$. For each $j \in N \setminus N(\sigma, \mathbf{x}, \mathbf{y})$ with $\sigma_j = s_j$, $g_j^*(\sigma'_j, x'_j, y'_j, \sigma_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) > 0$ by $\sigma'_j = s_j$, $x'_j = \pi(\mathbf{x}_{-j})$, and $y'_j > \max\{f(s_j x'_j), \max\{y_i \mid i \neq j\}\}$ under **Rule 1-3**. Also, for any $j \in N \setminus N(\sigma, \mathbf{x}, \mathbf{y})$ with $\sigma_j \neq s_j$, $g_j^*(\sigma'_j, x'_j, y'_j, \sigma_{-j}, \mathbf{x}_{-j}, \mathbf{y}_{-j}) > 0$ by $\sigma'_j > \max\{\sigma_i \mid i \in N\}$ and $x'_j = 0$ under **Rule 2**.

Thus, $(\sigma, \mathbf{x}, \mathbf{y})$ corresponds to **Rule 1-1**. Then, $g^*(\sigma, \mathbf{x}, \mathbf{y}) = \mathbf{y}$, which implies $g_i^*(\sigma, \mathbf{x}, \mathbf{y}) = \frac{\sigma_i x_i}{\sum \sigma_k x_k} f(\sum s_k x_k)$ for each $i \in N$. By **Assumption 1**, $\mathbf{x} \gg \mathbf{0}$. Moreover, $f(\sum \sigma_k x_k) = f(\sum s_k x_k)$. Therefore, by Lemma 1, $\sigma = \mathbf{s}$. Since $(\sigma, \mathbf{x}, \mathbf{y}) \in NE(g^*, \mathbf{u}, \mathbf{s})$, for each $i \in N$ and each $(\sigma'_i, x'_i, y'_i) \in \mathcal{S} \times [0, \bar{x}] \times \mathbb{R}_+$, $u_i(x_i, g_i^*(\sigma, \mathbf{x}, \mathbf{y})) \geq u_i(x'_i, g_i^*(\sigma'_i, x'_i, y'_i, \sigma_{-i}, \mathbf{x}_{-i}, \mathbf{y}_{-i}))$. For each $i \in N$, $g_i^*(\sigma'_i, x'_i, y'_i, \sigma_{-i}, \mathbf{x}_{-i}, \mathbf{y}_{-i}) \leq y_i + s_i \cdot f'(\sum s_k x_k) \cdot (x'_i - x_i)$ by $\sigma'_i = s_i$ and $y'_i = 0$ under **Rule 1-2-2**. Thus, (\mathbf{x}, \mathbf{y}) is Pareto efficient for (\mathbf{u}, \mathbf{s}) , so that $(\mathbf{x}, \mathbf{y}) \in PR(\mathbf{u}, \mathbf{s})$.

(3) Show $SNA(g^*, \mathbf{e}) = NA(g^*, \mathbf{e})$ for each $\mathbf{e} \in \mathcal{E}$. By definition, $SNA(g^*, \mathbf{e}) \subseteq NA(g^*, \mathbf{e})$. Suppose $SNA(g^*, \mathbf{e}) \subsetneq NA(g^*, \mathbf{e})$. Then, there exists $(\sigma, \mathbf{x}, \mathbf{y}) \in NE(g^*, \mathbf{e})$ such that for some $T \subsetneq N$ and some $(\sigma'_i, x'_i, y'_i)_{i \in T} \in \mathcal{S}^{\#T} \times$

$[0, \bar{x}]^{\#T} \times \mathbb{R}_+^{\#T}$, and each $j \in T$,

$$u_j(x_j, g_j^*(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y})) < u_j\left(x'_j, g_j^*\left((\sigma'_i, x'_i, y'_i)_{i \in T}, (\sigma_l, x_l, y_l)_{l \in N \setminus T}\right)\right).$$

Since $(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y}) \in NE(g^*, \mathbf{e})$ corresponds to **Rule 1-1** as shown above, $(\boldsymbol{\sigma}, \mathbf{x}, \mathbf{y})$ is *PR-consistent*, so that $\mathbf{x} \gg \mathbf{0}$ under **Assumption 1**. Hence, $\boldsymbol{\sigma} = \mathbf{s}$ by Lemma 1. Note also that $T = N$ is eliminated by Pareto efficiency of $NA(g^*, \mathbf{e})$. By construction of g^* , there is at most one agent who has a positive share of output under **Rules 1-2-1, 1-2-2, 1-3, and 2**. Since $g_i^*(\mathbf{s}, \mathbf{x}, \mathbf{y}) > 0$ for all $i \in N$, $\left((\sigma'_i, x'_i, y'_i)_{i \in T}, (\sigma_l, x_l, y_l)_{l \in N \setminus T}\right)$ corresponds to **Rule 1-1** by **Assumption 1**. Then, f must be linear on a closed interval $\left[0, \max\left\{\sum s_k x_k, \sum_{i \in T} \sigma'_i x'_i + \sum_{l \in N \setminus T} s_l x_l\right\}\right]$, and we obtain

$$g_j^*\left((\sigma'_i, x'_i, y'_i)_{i \in T}, (\sigma_l, x_l, y_l)_{l \in N \setminus T}\right) \leq w_j + s_j \cdot f'\left(\sum s_k x_k\right) \cdot (x'_j - x_j)$$

for some $j \in T$. Thus, by Pareto efficiency of $(\mathbf{x}, g^*(\mathbf{s}, \mathbf{x}, \mathbf{y}))$, $NA(g^*, \mathbf{e}) = SNA(g^*, \mathbf{e})$. ■

Proof of Theorem 2. Since $NA(g^*, \mathbf{e}) = PR(\mathbf{e})$ for each $\mathbf{e} \in \mathcal{E}$, we have only to show $PR(\mathbf{e}) \subseteq SPA\left(\Gamma_{g^*}^{(\boldsymbol{\sigma}, \mathbf{y})^{\circ \mathbf{x}}}, \mathbf{e}\right)$ for each $\mathbf{e} \in \mathcal{E}$. First, we show that in every Stage 2-subgame, there exists a Nash equilibrium. Let a strategy mapping $\boldsymbol{\chi}^* : \mathcal{S}^n \times \mathbb{R}_+^n \rightarrow [0, \bar{x}]^n$ be such that for each $\left(\Gamma_{g^*}^{(\boldsymbol{\sigma}, \mathbf{y})^{\circ \mathbf{x}}}(\boldsymbol{\sigma}, \mathbf{y}), \mathbf{e}\right)$:

$$\text{for all } i \in N, \chi_i^*(\boldsymbol{\sigma}, \mathbf{y}) = \begin{cases} \frac{\bar{x}}{2} & \text{if } y_i > \max\{y_k \mid k \neq i\}, \sigma_i = s_i, \text{ and} \\ \frac{\bar{x}}{2} & N(\boldsymbol{\sigma}, (\frac{\bar{x}}{2}, \mathbf{0}_{-i}), \mathbf{y}) = \emptyset \\ 0 & \text{otherwise} \end{cases}.$$

Note $g^*(\boldsymbol{\sigma}, \boldsymbol{\chi}^*(\boldsymbol{\sigma}, \mathbf{y}), \mathbf{y})$ corresponds to **Rule 1-3**. To simplify the notation, let us denote $\mathbf{x}^* = \boldsymbol{\chi}^*(\boldsymbol{\sigma}, \mathbf{y})$ in the following discussion.

Suppose that $i \in N$ switches from x_i^* to x'_i . Note that this deviation does not induce **Rule 1-1**. If x'_i induces **Rule 1-2-2**, then $g_i^*(\boldsymbol{\sigma}, x'_i, \mathbf{x}_{-i}^*, \mathbf{y}) \leq g_i^*(\boldsymbol{\sigma}, \mathbf{x}^*, \mathbf{y})$, because $i \notin N(\boldsymbol{\sigma}, (x'_i, \mathbf{x}_{-i}^*), \mathbf{y})$. If x'_i induces **Rule 2**, then $x'_i > 0$, so that $g_i^*(\boldsymbol{\sigma}, x'_i, \mathbf{x}_{-i}^*, \mathbf{y}) = 0$. If x'_i induces **Rule 1-3**, then $g_i^*(\boldsymbol{\sigma}, x'_i, \mathbf{x}_{-i}^*, \mathbf{y}) \leq g_i^*(\boldsymbol{\sigma}, \mathbf{x}^*, \mathbf{y})$. To see this, let us assume that an agent, say j , has a positive output in $g^*(\boldsymbol{\sigma}, \mathbf{x}^*, \mathbf{y})$. Then, $y_j > \max\{y_k \mid k \neq j\}$. Since $\mathbf{x}^* = (\frac{\bar{x}}{2}, \mathbf{0}_{-j})$, and $g^*(\boldsymbol{\sigma}, \mathbf{x}^*, \mathbf{y})$ corresponds to **Rule 1-3**, $N(\mathbf{s}, (\frac{\bar{x}}{2}, \mathbf{0}_{-j}), \mathbf{w}) = \emptyset$. Thus,

if $i = j$, then $x_i^* = \frac{\bar{x}}{2}$ is the best response to $\mathbf{0}_{-i}$ in $\left(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ x}}(\sigma, \mathbf{y}), \mathbf{e}\right)$. If $i \neq j$, then $g_i^*(\sigma, x'_i, \mathbf{x}_{-i}^*, \mathbf{y}) = 0$. Thus, $\mathbf{x}^* \in NE\left(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ x}}(\sigma, \mathbf{y}), \mathbf{e}\right)$.

Now, we will demonstrate that for each $\mathbf{e} = (\mathbf{u}, \mathbf{s}) \in \mathcal{E}$, if $\widehat{\mathbf{z}} = (\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in PR(\mathbf{e})$, then there exists a subgame-perfect equilibrium whose corresponding outcome is $\widehat{\mathbf{z}}$. Consider the following strategy profile of $\left(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ x}}, \mathbf{e}\right)$:

- (1) In Stage 1, $(\sigma, \mathbf{y}) = (\mathbf{s}, \widehat{\mathbf{y}})$.
- (2) In Stage 2, $\chi : \mathcal{S}^n \times \mathbb{R}_+^n \rightarrow [0, \bar{x}]^n$ is given as follows:
 - (2-1): if $(\sigma, \mathbf{y}) = (\mathbf{s}, \widehat{\mathbf{y}})$ in Stage 1, then $\chi(\sigma, \mathbf{y}) = \widehat{\mathbf{x}}$;
 - (2-2): if $(\sigma, \mathbf{y}) = ((s'_j, \mathbf{s}_{-j}), (y'_j, \widehat{\mathbf{y}}_{-j}))$ is such that $s'_j = s_j$, $y'_j \neq \widehat{y}_j$, and for each $i \neq j$, $y'_i \geq y_i$ in Stage 1, then for this $j \in N$,

$$\chi_j(\sigma, \mathbf{y}) = \arg \max_{x'_j \in [0, \bar{x}]} u_j(x'_j, g_j^*(\mathbf{s}, x'_j, \widehat{\mathbf{x}}_{-j}, y'_j, \widehat{\mathbf{y}}_{-j}))$$

and for all $i \neq j$, $\chi_i(\sigma, \mathbf{y}) = \widehat{x}_i$;

(2-3): in any other case, $\chi(\sigma, \mathbf{y}) = \chi^*(\sigma, \mathbf{y})$.

If the subgame $\left(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ x}}(\sigma, \mathbf{y}), \mathbf{e}\right)$ corresponds to (2-1), then $\chi(\sigma, \mathbf{y}) \in NE\left(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ x}}(\sigma, \mathbf{y}), \mathbf{e}\right)$. This is because $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in NA(g^*, \mathbf{e})$ by Theorem 1. Also, by the above argument, $\chi(\sigma, \mathbf{y}) \in NE\left(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ x}}(\sigma, \mathbf{y}), \mathbf{e}\right)$ in the subgame (2-3) of Stage 2.

Suppose that $\left(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ x}}(\sigma, \mathbf{y}), \mathbf{e}\right)$ corresponds to (2-2). Then, $g^*(\sigma, \chi(\sigma, \mathbf{y}), \mathbf{y})$ corresponds to **Rule 1-1** or **Rule 1-2**, since $j \in N(\mathbf{s}, x'_j, \widehat{\mathbf{x}}_{-j}, y'_j, \widehat{\mathbf{y}}_{-j})$ or $PR(\mathbf{s}, x'_j, \widehat{\mathbf{x}}_{-j}, y'_j, \widehat{\mathbf{y}}_{-j})^{-1} \neq \emptyset$ for any $x'_j \in [0, \bar{x}]$. If $g^*(\sigma, \chi(\sigma, \mathbf{y}), \mathbf{y})$ corresponds to **Rule 1-1**, then f must be linear on $\left[0, \max\left\{\sum s_k \widehat{x}_k, \sum_{i \neq j} s_i \widehat{x}_i + s_j x'_j\right\}\right]$, which implies the Pareto efficiency of $((x'_j, \widehat{\mathbf{x}}_{-j}), (y'_j, \widehat{\mathbf{y}}_{-j}))$. Thus, $\chi(\sigma, \mathbf{y}) \in NE\left(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ x}}(\sigma, \mathbf{y}), \mathbf{e}\right)$. If $g^*(\sigma, \chi(\sigma, \mathbf{y}), \mathbf{y})$ corresponds to **Rule 1-2**, then $\chi_j(\sigma, \mathbf{y}) = 0$ is a best response to $\widehat{\mathbf{x}}_{-j}$ in $\left(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ x}}(\sigma, \mathbf{y}), \mathbf{e}\right)$. This is because $N(\mathbf{s}, 0, \widehat{\mathbf{x}}_{-j}, y'_j, \widehat{\mathbf{y}}_{-j}) = \{j\}$, and so $g^*(\sigma, \chi(\sigma, \mathbf{y}), \mathbf{y})$ corresponds to **Rule 1-2-2**. For any other $i \neq j$, any deviation from \widehat{x}_i to x'_i results in $g_i^*(\sigma, x'_i, \chi_{-i}(\sigma, \mathbf{y}), \mathbf{y}) = 0$. Thus, $\chi(\sigma, \mathbf{y}) \in NE\left(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ x}}(\sigma, \mathbf{y}), \mathbf{e}\right)$.

Now, let us see that the above strategy profile (1)-(2) constitutes a subgame-perfect equilibrium of $\left(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ x}}, \mathbf{e}\right)$. By the strategy profile (1)-

(2) of $(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ \mathbf{x}}}, \mathbf{e})$, $g^*(\sigma, \chi(\sigma, \mathbf{y}), \mathbf{y}) = g^*(s, \widehat{\mathbf{x}}, \widehat{\mathbf{y}}) = \widehat{\mathbf{y}}$. Suppose that $j \in N$ switches from (s_j, \widehat{y}_j) to (s'_j, y'_j) in Stage 1. Then by (2-2) and (2-3), $g_j^*(s'_j, \mathbf{s}_{-j}, \chi(s'_j, \mathbf{s}_{-j}, y'_j, \widehat{\mathbf{y}}_{-j}), y'_j, \widehat{\mathbf{y}}_{-j}) \leq \widehat{y}_j + s_j \cdot f'(\sum s_k \widehat{x}_k) \cdot (x'_j - \widehat{x}_j)$. Thus, since $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in PR(\mathbf{e})$, $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in SPA(\Gamma_{g^*}^{(\sigma, \mathbf{y})^{\circ \mathbf{x}}}, \mathbf{e})$. ■

Proof of Theorem 3. Since $NA(g^*, \mathbf{e}) = PR(\mathbf{e})$ for each $\mathbf{e} \in \mathcal{E}$, we have only to show $PR(\mathbf{e}) \subseteq SPA(\Gamma_{g^*}^{\mathbf{x}^{\circ}(\sigma, \mathbf{y})}, \mathbf{e})$. First, we will show that in every Stage 2-subgame, there is a Nash equilibrium.

Given $\mathbf{e} = (\mathbf{u}, \mathbf{s}) \in \mathcal{E}$, let a strategy mapping $\boldsymbol{\mu}^* : [0, \bar{x}]^n \rightarrow \mathcal{S}^n \times \mathbb{R}_+^n$ be such that for each $(\Gamma_{g^*}^{\mathbf{x}^{\circ}(\sigma, \mathbf{y})}(\mathbf{x}), \mathbf{e})$, $\boldsymbol{\mu}^*(\mathbf{x}) = (\boldsymbol{\sigma}^*, \mathbf{y}^*)$, where for each $i \in N$:

$$(\sigma_i^*, y_i^*) = \begin{cases} (s_i, 0) & \text{if } x_i \neq \pi(\mathbf{x}_{-i}) \\ (s_i, f(s_i x_i) + 1) & \text{otherwise} \end{cases}.$$

Suppose that $i \in N$ switches from (σ_i^*, y_i^*) to (σ'_i, y'_i) . Note that $g^*(\boldsymbol{\sigma}^*, \mathbf{x}, \mathbf{y}^*)$ corresponds to **Rule 1-3**. Thus, i cannot induce **Rule 1-1** by changing his strategy. If (σ'_i, y'_i) induces **Rule 1-2-2**, then $g_i^*(\sigma'_i, \boldsymbol{\sigma}_{-i}^*, \mathbf{x}, y'_i, \mathbf{y}_{-i}^*) = 0$, because $i \notin N(\sigma'_i, \boldsymbol{\sigma}_{-i}^*, \mathbf{x}, y'_i, \mathbf{y}_{-i}^*)$. If (σ'_i, y'_i) induces **Rule 2**, then $x_i > 0$, which implies $g_i^*(\sigma'_i, \boldsymbol{\sigma}_{-i}^*, \mathbf{x}, y'_i, \mathbf{y}_{-i}^*) = 0$. If (σ'_i, y'_i) induces **Rule 1-3**, then $g_i^*(\sigma'_i, \boldsymbol{\sigma}_{-i}^*, \mathbf{x}, y'_i, \mathbf{y}_{-i}^*) \leq g_i^*(\boldsymbol{\sigma}^*, \mathbf{x}, \mathbf{y}^*)$, since whether $x_i = \pi(\mathbf{x}_{-i})$ or not is already fixed in Stage 1. Thus, $(\boldsymbol{\sigma}^*, \mathbf{y}^*) \in NE(\Gamma_{g^*}^{\mathbf{x}^{\circ}(\sigma, \mathbf{y})}(\mathbf{x}), \mathbf{e})$.

Now, we will show that if $\widehat{\mathbf{z}} = (\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in PR(\mathbf{e})$, then there exists a subgame-perfect equilibrium whose corresponding outcome is $\widehat{\mathbf{z}}$. Consider the following strategy profile of $(\Gamma_{g^*}^{\mathbf{x}^{\circ}(\sigma, \mathbf{y})}, \mathbf{e})$:

- (1) In Stage 1, each $i \in N$ supplies $\widehat{x}_i > 0$.
- (2) In Stage 2, $\boldsymbol{\mu} : [0, \bar{x}]^n \rightarrow \mathcal{S}^n \times \mathbb{R}_+^n$ is given as follows:
 - (2-1): if $\mathbf{x} = \widehat{\mathbf{x}}$ in Stage 1, then $\boldsymbol{\mu}(\mathbf{x}) = (\mathbf{s}, \widehat{\mathbf{y}})$;
 - (2-2): if $\mathbf{x} = (x'_j, \widehat{\mathbf{x}}_{-j}) \gg \mathbf{0}$, where $x'_j \neq \widehat{x}_j$ in Stage 1, then

$$\begin{aligned} & \text{for } j \in N, \mu_j(\mathbf{x}) = (s_j, f(s_j x'_j) + 1), \\ & \text{for all } i \neq j, \mu_i(\mathbf{x}) = \begin{cases} (s_i, \widehat{y}_i) & \text{if } x_i \neq \pi(\mathbf{x}_{-i}) \\ (s_i, f(\sum_{k \neq j} s_k \widehat{x}_k + s_j x'_j)) & \text{otherwise} \end{cases}; \end{aligned}$$

- (2-3): in any other case, $\boldsymbol{\mu}(\mathbf{x}) = \boldsymbol{\mu}^*(\mathbf{x})$.

If the subgame $\left(\Gamma_{g^*}^{x^o(\sigma, y)}(\mathbf{x}), \mathbf{e}\right)$ corresponds to (2-1), then $(\mathbf{s}, \widehat{\mathbf{y}}) \in NE\left(\Gamma_{g^*}^{x^o(\sigma, y)}(\mathbf{x}), \mathbf{e}\right)$. This is because $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in NA(g^*, \mathbf{e})$ by Theorem 1. Also, by the above argument, $\boldsymbol{\mu}^*(\mathbf{x}) \in NE\left(\Gamma_{g^*}^{x^o(\sigma, y)}(\mathbf{x}), \mathbf{e}\right)$ in the subgame (2-3) of Stage 2.

Suppose that $\left(\Gamma_{g^*}^{x^o(\sigma, y)}(\mathbf{x}), \mathbf{e}\right)$ corresponds to (2-2), and $h \in N$ switches from $\mu_h(\mathbf{x})$ to (s'_h, y'_h) . First, if $s'_h \neq s_h$, then $g_h^*(s'_h, y'_h, \boldsymbol{\mu}_{-h}(\mathbf{x}), \mathbf{x}) = 0$ under **Rule 2**. Secondly, consider the following two cases:

(i) For all $i \neq j$, $x_i \neq \pi(\mathbf{x}_{-i})$. Then, $g^*(\boldsymbol{\mu}(\mathbf{x}), \mathbf{x})$ corresponds to **Rule 1-2-2**. If $h \neq j$ and (s'_h, y'_h) induces **Rule 1-2-2** or **Rule 1-3**, then $g_h^*(s'_h, y'_h, \boldsymbol{\mu}_{-h}(\mathbf{x}), \mathbf{x}) = 0$. Note if $h \neq j$, then (s'_h, y'_h) cannot induce **Rule 1-1**. If $h = j$ and (s'_h, y'_h) induces **Rule 1-2-2** or **Rule 1-1**, then $g_h^*(s'_h, y'_h, \boldsymbol{\mu}_{-h}(\mathbf{x}), \mathbf{x}) \leq \widehat{y}_h + s_h \cdot f'(\sum s_k \widehat{x}_k) \cdot (x'_h - \widehat{x}_h)$. Note if $h = j$, then (s'_h, y'_h) cannot induce **Rule 1-3**. Thus, $\boldsymbol{\mu}(\mathbf{x}) \in NE\left(\Gamma_{g^*}^{x^o(\sigma, y)}(\mathbf{x}), \mathbf{e}\right)$.

(ii) There exists $i^* \neq j$ with $x_{i^*} = \pi(\mathbf{x}_{-i^*})$. Then, $g^*(\boldsymbol{\mu}(\mathbf{x}), \mathbf{x})$ corresponds to **Rule 1-3**. Then, (s'_h, y'_h) cannot induce **Rule 1-1**. If (s'_h, y'_h) induces **Rule 1-2-2**, this implies $h = i^*$ or $h = j$. Thus, since $h \notin N(s'_h, y'_h, \boldsymbol{\mu}_{-h}(\mathbf{x}), \mathbf{x})$, $g_h^*(s'_h, y'_h, \boldsymbol{\mu}_{-h}(\mathbf{x}), \mathbf{x}) = 0$. If (s'_h, y'_h) induces **Rule 1-3**, then $g_h^*(s'_h, y'_h, \boldsymbol{\mu}_{-h}(\mathbf{x}), \mathbf{x}) \leq g^*(\boldsymbol{\mu}(\mathbf{x}), \mathbf{x})$, since $x_{i^*} = \pi(\mathbf{x}_{-i^*})$ is already fixed in Stage 1. Thus, $\boldsymbol{\mu}(\mathbf{x}) \in NE\left(\Gamma_{g^*}^{x^o(\sigma, y)}(\mathbf{x}), \mathbf{e}\right)$.

Now, let us see that the above strategy profile (1)-(2) constitutes a subgame-perfect equilibrium of $\left(\Gamma_{g^*}^{x^o(\sigma, y)}, \mathbf{e}\right)$. By the strategy profile (1)-

(2) of $\left(\Gamma_{g^*}^{x^o(\sigma, y)}, \mathbf{e}\right)$, $g^*(\boldsymbol{\mu}(\widehat{\mathbf{x}}), \widehat{\mathbf{x}}) = g^*(\mathbf{s}, \widehat{\mathbf{x}}, \widehat{\mathbf{y}}) = \widehat{\mathbf{y}}$. Suppose that $j \in N$ deviates from \widehat{x}_j to $x'_j \neq \widehat{x}_j$ in Stage 1. If $x'_j = 0$, then by (2-3), $g_j^*(\boldsymbol{\mu}(x'_j, \widehat{\mathbf{x}}_{-j}), (x'_j, \widehat{\mathbf{x}}_{-j})) = 0$. If $x'_j > 0$, then by (2-2), $g_j^*(\boldsymbol{\mu}(x'_j, \widehat{\mathbf{x}}_{-j}), (x'_j, \widehat{\mathbf{x}}_{-j})) \leq \widehat{y}_j + s_j \cdot f'(\sum s_k \widehat{x}_k) \cdot (x'_j - \widehat{x}_j)$. Thus, since $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in PR(\mathbf{e})$, $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in SPA\left(\Gamma_{g^*}^{x^o(\sigma, y)}, \mathbf{e}\right)$. ■

6 References

Corchón, L. C. and Puy, M. S. (2002): “Existence and Nash implementation of efficient sharing rules for a commonly owned technology,” *Social Choice and Welfare* **19**, pp.369–379.

Hong, L. (1995): “Nash implementation in production economies,” *Economic Theory* **5**, pp.401–417.

- Hurwicz, L. (1960): “Optimality and informational efficiency in resource allocation processes,” in Arrow, K. J., Karlin, S. and Suppes, P. (eds.), *Mathematical methods in social sciences*, pp.27–46. Stanford: Stanford Univ. Press.
- Hurwicz, L., Maskin, E. and Postlewaite, A. (1995): “Feasible Nash implementation of social choice rules when the designer does not know endowments or production sets,” in Ledyard, J. (ed.), *The economics of informational decentralization: Complexity, efficiency, and stability*, Essays in Honor of Stanley Reiter, pp.367–433. Amsterdam: Kluwer Academic Publishers.
- Jackson, M. O., Palfrey, T. and Srivastava, S. (1994): “Undominated Nash implementation in bounded mechanisms,” *Games and Economic Behavior* **6**, pp. 474–501.
- Kaplan, T. and Wettstein, D. (2000): “Surplus sharing with a two-stage mechanism,” *International Economic Review* **41**, pp.399–409.
- Kranich, L. (1994): “Equal division, efficiency, and the sovereign supply of labor,” *American Economic Review* **84**, pp.178–189.
- Roemer, J. (1989): “A public ownership resolution of the tragedy of the commons,” *Social Philosophy & Policy* **6**, pp.74–92.
- Roemer, J. and Silvestre, J. (1993): “The proportional solution for economies with both private and public ownership,” *Journal of Economic Theory* **59**, pp.426–444.
- Saijo, T., Tatamitani, Y. and Yamato, T. (1996): “Toward natural implementation,” *International Economic Review* **37**, pp.949–980.
- Suh, S.-C. (1995): “A mechanism implementing the proportional solution,” *Economic Design* **1**, pp.301–317.
- Tian, G. (1999): “Double implementation in economies with production technologies unknown to the designer,” *Economic Theory* **13**, pp.689–707.
- Tian, G. (2000): “Incentive mechanism design for production economies with both private and public ownerships,” *Games and Economic Behavior* **33**, pp.294–320.
- Varian, H. (1994): “A solution to the problem of externalities when agents are well informed,” *American Economic Review* **84**, pp.1278–1293.

Yoshihara, N. (1999): “Natural and double implementation of public ownership solutions in differentiable production economies,” *Review of Economic Design* **4**, pp.127–151.

Yoshihara, N. (2000a): “A characterization of natural and double implementation in production economies,” *Social Choice and Welfare* **17**, pp.571–599.

Yoshihara, N. (2000b): “On efficient and procedurally-fair equilibrium allocations in sharing games,” Institute of Economic Research Discussion Paper Series A No. 397, Hitotsubashi University.

	Suh (1995)	Yoshihara (1999)	Yoshihara (2000)	Kaplan-Wettstein (2000)	Tian (2000)	Corchon-Puy (2002)	our mechanism
equilibrium notions*	NA SNA UNA	NA SNA UNA	NA SNA	SPA	NA SNA	NA	NA SNA SPA
# of goods	2	2	2 or more	2	2 or more	2	2
# of agents	2 or more	3 or more	3 or more	3 or more	2 or more	3 or more	2 or more
permission of overstatement of skills	no	no	no	no	no	no	yes
permission of understatement of skills	no	no	no	no	yes	no	yes
labor sovereignty	no	no	no	no	no	no	yes
feasibility	yes	yes	yes	yes	yes	yes	yes
self-relevancy	no	yes	yes	no	no	no	yes
best response property	no	yes	yes	no	no	yes	yes
forthrightness	no	yes	yes	yes	yes	yes	yes
balancedness	no	yes	yes	no	no	yes	no
continuity	no	no	no	no	yes	no	no

Table 1: Performance of mechanisms implementing solutions for the tragedy of commons

* NA, SNA, UNA, and SPA mean Nash implementability, strong Nash implementability, undominated Nash implementability, and subgame-perfect implementability respectively.