CENTER FOR ADVANCED RESEARCH IN FINANCE

CARF Working Paper

CARF-F-556

Free-Rider Problem and Commitment

Hitoshi Matsushima
University of Tokyo

First Version: January 28, 2022
Current Version: February 20, 2023

CARF is presently supported by Nomura Holdings, Inc., Sumitomo Mitsui Banking Corporation, The Dai-ichi Life Insurance Company, Limited, The Norinchukin Bank, MUFG Bank, Ltd. and Ernst & Young ShinNihon LLC. This financial support enables us to issue CARF Working Papers.

Free-Rider Problem and Commitment[1]

Hitoshi Matsushima[2]

University of Tokyo

First Version: January 28, 2022
Current Version: February 20, 2023

[1] Matsushima (2022a) is an earlier version of this study. The stepwise commitment rule in this study is a significant refinement of the cautious commitment rule presented in the earlier version. All the main theorems are new in the current version. This study was supported by a grant-in-aid for scientific research (KAKENHI 20H00070) from the Japan Society for the Promotion of Science (JSPS) and the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT). I am grateful to Professors Satoru Hibiki, Takeo Hoshi, and Tetsuji Okazaki for their useful comments and encouragement. All errors are mine.

[2] Department of Economics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: hitoshi@e.u-tokyo.ac.jp

Abstract

This study investigates a free-rider problem caused by externalities inherent in global commons, industrial pollution, and the hidden social costs of everyday activities. Within the range of sovereignty and privacy protection, to transform the social dilemma into an easier-to-solve coordination game format, a method of commitment rule design is adopted wherein each agent declares a scope of actions that they are willing to commit to, and a committee presents the agent with a promise to action within their scope. To ensure the fulfillment of each agent's promise, a social network is devised that links local trust to address the global problem. This study demonstrates the stepwise commitment rule, which incentivizes all agents to act cooperatively as a unique decent Nash equilibrium, and is sustainable, whereby cooperative relationships can be maintained even if a non-negligible number of agents adhere to uncooperative attitudes.

JEL Classification Numbers, C72, DD91, H41, H77, Q54

Keywords: Global Commons, Local Trust, Stepwise Commitment Rule, Uniqueness, Virtual Sustainability

1. Introduction

This study investigates a free-rider problem wherein each agent takes an action in a continuous interval that has a private cost but brings positive externalities to other agents. Unless an institutional device is used, every agent would predominantly utilize this strategy to act uncooperatively. This study proposes an institutional method to develop an ideal solution for the free-rider problem by introducing a commitment rule design.

Each agent declares a range of actions to which they are willing to commit (i.e., the upper limit of what they can commit) as their message. Thereafter, a committee presents each agent with a promise to action determined by a preset commitment rule and their announced upper limits. Each agent is assumed to fulfill their promise within certain limits. This study also assumes that each agent exhibits minimal prosociality as a lexicographic preference. With these assumptions, the free-rider problem is considered a commitment game associated with a commitment rule, where agents announce their respective upper limits as their maximal best-response messages. Accordingly, in this study's terminology, they will play a decent Nash equilibrium with respect to their message selections.

The central interest of this study is to develop a method of commitment rule design that would ideally solve free-rider problems that capture the common essence of externalities that make it difficult to maintain the global commons such as climate change, industrial pollution, and the hidden social costs of everyday economic activities.

For an ideal solution, a commitment rule should satisfy the following four requirements:

Sovereignty Protection: The commitment rule does not require each agent any promise that exceeds their upper limit.

Maintenance of Cooperation: The commitment rule requires each agent promises that are close to their upper limit.

Uniqueness: The maximal message profile is the unique decent Nash equilibrium in the associated commitment game, and it ensures the achievement of full cooperation.

Sustainability: Even if there are problematic agents who adhere to uncooperative attitudes, many of the remaining agents are still willing to act cooperatively in the associated commitment game.

Sustainability and maintenance of cooperation imply that even in unforeseen circumstances caused by some problematic agents, a high level of cooperation is being maintained by the remaining agents without changing the preset commitment rule to another rule that demands nothing from these problematic agents. Hence, it can remove the advantage of those who attempt to free-ride by taking advantage of non-excludability and deliberately staying uncooperative.

In this study, a characterization of free-rider problems is demonstrated wherein there exist commitment rules that satisfy these requirements (i.e., sovereignty protection, the maintenance of cooperation, uniqueness, and [a virtual version of] sustainability). Then, this study shows that the free-rider problem can be solved ideally by satisfying these requirements if the number of participants in the commitment game is sufficiently large.

To verify these results, the stepwise commitment rule is specified, which is designed such that the presence of agents with uncooperative attitudes lowers the promise of other agents who adopt more cooperative attitudes. This design acts as a deterrent to ensure that full cooperation is a decent Nash equilibrium in the associated commitment game.

The stepwise commitment rule is made depending on the minimal upper limit among all agents' announcements. Each agent's promise is always set to the minimal upper limit or greater. This dependence on the minimal upper limit is shown to have an irreplaceable role in balancing the uniqueness and maintenance of cooperation.

The stepwise commitment rule depends on the number of agents whose upper limits are not fully cooperative. The set of all message profiles is divided into a finite number of categories according to the number of agents whose messages are not fully cooperative.

According to the category the message profile belongs to, the range of discounts from each agent's upper limit to their actual promise gradually increases. Such category-based discounts, which are implemented in the stepwise commitment rule, have a central role in satisfying sustainability requirements.

The novelty of this study lies in the formulation of global commons as free-rider problems with many agents, the formulation of sustainability issue, and presenting concrete solutions to these problems. Specifically, this study clarifies the unique role of the minimal upper limit, and the design method of setting commitments on a category basis is the original proposal of this study, which is significant in solving various free-rider problems on a global scale.

Furthermore, the skillful use of local trust in communities is essential for solving the free-rider problem in the global commons. To facilitate the global sharing of local information, each community must be connected hierarchically with the committee at the top. Each community must be sized appropriately to avoid destabilizing the stepwise commitment rule by adopting another rule of its own. Privacy concerns regarding communication across communities are discussed and a trade-off between stability and sustainability is demonstrated.

Citizens seek to maintain trust within their communities by acting honestly and keeping promises to their fellow citizens within the bounds of privacy and sovereignty. However, such trust does not exist in a global society. Hence, another novelty of this study lies in relating a global problem, which is considered difficult to solve, to the potential problem-solving abilities of local communities.

2. Literature Review

This study is related to MacKay et al. (2015), which introduces commitment rule design for the free-rider problem in the global context of market failure and lack of coercive control by a centralized government. As a proposal for solving the climate change issue through international negotiations, they present the institutional framework wherein the

committee (i.e., the Conference of Parties [the COP]), would let each agent (i.e., each state) express the upper limit of what they can commit to, forcing them to commit within that scope. Under this framework, setting an appropriate commitment rule transforms the social dilemma into an easier-to-solve coordination game format, such as unanimity voting, wherein any agent promises nothing unless all agents are fully cooperative, and the Bertrand oligopoly with the lowest price guarantee clause, wherein each firm promises its customers a price below its rivals' prices.

MacKay et al. (2015) present the common commitment rule modeled on the lowest price guarantee clause, and suggest that it satisfies sovereignty protection and uniqueness. However, the common commitment rule does not ensure cooperation and sustainability. Moreover, MacKay et al. (2015) did not discuss why commitment rules that depend on the minimal upper limit are needed over a multitude of other rules. In fact, the unanimous commitment rule can be modeled on unanimous voting, which requires less information than the common commitment rule but satisfies sovereignty protection and uniqueness.

Addressing this gap, this study investigates the maintenance of cooperation and sustainability, as well as sovereignty protection and uniqueness. As a new rule design, the stepwise commitment rule is demonstrated, which satisfies all these requirements. This study also clarifies that the dependence on the minimal upper limit has a unique role in reconciling the maintenance of cooperation with uniqueness.

In the global commons, it has been difficult to achieve both the elimination of free-riding caused by non-excludability (i.e., sustainability in this study's terminology) and sovereignty protection. Nordhaus (2015) proposed a system design called the Climate Club, which relaxes the requirement to protect sovereignty and allows an external mechanism for retaliatory sanctions and ostracism against problematic states. In this study, such external measures are not relied upon; instead, we seek to integrate these requirements within the commitment rule design.

The difficulty of incorporating sustainability within the scope of commitment rule design is that as more people persist in being uncooperative, discounts on other people's promises accumulate cumulatively, which makes sustainability incompatible with the

maintenance of cooperation. To overcome this difficulty, we propose a virtualization of sustainability, which requires almost all, but not all, of the remaining agents to act cooperatively. Notably, the category-based design presented in this study is effective in achieving virtual sustainability.

The work of Eliaz (2002) is relevant to the issue of sustainability although out of our context. Eliaz (2002) considered a mechanism design in social choice theory, wherein some agents may behave irrationally, but the remaining rational agents are still willing to act cooperatively.

The framework of the commitment rule design, as explored by MacKay et al. (2015) and cultivated in this study, is supported by the assumption that each agent will keep their promise and that the necessary information, such as the minimal upper limit and the number of uncooperative agents, are shared with all the agents. However, in global commons, the validity of these needs must be carefully justified because there is no global trust or human bond.

Accordingly, Ostrom (1990, 2010) proposed polycentric governance, albeit informally, whereby local trust could be used to address the problem of the global commons. Inspired by Ostrom's proposal, this study introduced a hierarchical communication system with a tree structure as a formal model. Additional considerations that accompany this model include dealing with the potential of a community to unilaterally adopt commitment rules that are different from other communities (i.e., stability) and dealing with the potential of a community to lose local trust (i.e., sustainability). Particularly, this study argues for a trade-off between stability and sustainability.

Moreover, each agent is assumed to have ethical preferences such as honesty and observance. The COP, which continues to adopt a pledge-and-review approach, is largely premised on such ethical motives as wishful thinking. There exists a reputation theory in finitely repeated games, such as Kreps et al. (1982), showing that the existence of players who strictly adhere to cooperative behavior drives other selfish players to cooperative behavior. Roemer (2010) introduced the Kantian equilibrium as an alternative to the Nash

equilibrium, where ethical preferences are employed to justify cooperative behaviors that are difficult to explain by selfish motives alone.

In this study, no assumptions are made regarding the presence of the strong ethical preferences that these studies have in common. The reason for rejecting such strong ethics is that the COP has not been successful in applying the pledge-and-review approach (Victor, 2007; Cramton et al, 2017).

Instead, this research emphasizes that small ethical tendencies have a significant effect on achieving uniqueness, eliminating unwanted equilibria, while leaving only cooperative equilibria. Hence, this study is related to the recent progress in implementation theory, such as Matsushima (2022b), which shows that with a tiny possibility of preference for honesty, any social choice function that balances efficiency and equity is uniquely implementable.

The remainder of this paper is organized as follows. Section 3 defines the free-rider problem and introduces the commitment-rule design. Section 4 explains the sustainability requirement and presents an impossibility concerning an exact sense of sustainability. Section 5 defines virtual sustainability and describes the stepwise commitment rule. Afterward, a characterization of ideally solvable free-rider problems, together with a positive limit theorem in the global commons, is shown. Section 6 discusses the role of local trust in fostering global cooperation, and Section 7 concludes the paper.

3. Free-Rider Problem

Let $N = \{1,...,n\}$ denote the set of all agents. Each agent $i \in N$ has a set of actions $A_i = [0,1]$ and a utility function $u_i : A \to R$, which is specified as:

$$u_i(a) = \sum_{j \in N} a_j - c a_i \text{ for all } i \in N \text{ and } a \in A,$$

where $A \equiv \underset{i \in N}{\times} A_i$, $a \equiv (a_i)_{i \in N} \in A$, c is a positive integer, and

$$1 < c < n.$$

As $c > 1$, in the strategic game defined as a triple $(N, A, (u_i)_{i \in N})$, any agent $i \in N$ prefers to select the lowest action level 0 as a dominant strategy. However, because of $c < n$,

each agent prefers to increase their action level if the other agents simultaneously increase their action levels by the same amount. Hence:

$$u_i(\overline{a}) > u_i(\underline{a}) \quad \text{for all} \quad i \in N,$$

where it is denoted that $\overline{a} \equiv (\overline{a}_i)_{i \in N}$, $\underline{a} \equiv (\underline{a}_i)_{i \in N}$, and

$$\overline{a}_i = 1 \quad \text{and} \quad \underline{a}_i = 0 \quad \text{for each} \quad i \in N.$$

Therefore, agents experience a free-rider problem where even if the maximal action profile $\overline{a}$ is desirable, all agents are willing to select the minimal action profile $\underline{a}$ as a dominant strategy profile; however, it is Pareto-dominated by $\overline{a}$.

To overcome this free-rider problem, the following commitment device is explored. Each agent $i \in N$ contains a set of messages $M_i \in [0,1]$. A message $m_i \in M_i$ announced by agent i defines the upper limit of their action selection, that is, the upper limit of promises that they can tolerate. Let $M \equiv \underset{i \in N}{\times} M_i$.

A commitment rule $\alpha = (\alpha_i)_{i \in N}$ is introduced, where

$$\alpha_i : M \to A_i \quad \text{for each} \quad i \in N.$$

The action level $\alpha_i(m) \in A_i$ of agent i implies the promise of action selection that agent i must keep. It is assumed that every agent $i \in N$ will keep their promise $\alpha_i(m) \in A_i$. With this assumption, if agents announce a message profile $m \in M$, the resultant action profile is given by $\alpha(m) \in A$.

A commitment rule α must protect each agent's sovereignty, implying that each agent's promise should not exceed their announced upper limit.

Sovereignty Protection (SP): For every $i \in N$ and $m \in M$,

$$\alpha_i(m) \leq m_i.$$

A commitment rule α must help maintain cooperation whenever possible, implying that each agent's promise should be as close to their upper limit as possible. Fix an arbitrary positive integer $\varepsilon > 0$.

Maintenance of Cooperation (MC): For every $i \in N$ and $m \in M$,

$$\alpha_i(m) \geq m_i - \varepsilon .$$

A commitment game associated with a commitment rule α is defined as a triple $(N, M, (v_i(\cdot, \alpha))_{i \in N})$, where

$$v_i(\cdot, \alpha) = v_i(\cdot) : M \to R \text{ for each } i \in N ,$$

and

$$v_i(m) = u_i(\alpha(m)) = \sum_{j \in N} \alpha_j(m) - c\alpha_i(m) \text{ for all } i \in N \text{ and } m \in M .$$

A message profile $m \in M$ is said to be a decent Nash equilibrium in the commitment game associated with a commitment rule α if for every $i \in N$,

$$v_i(m) \geq v_i(m_i', m_{-i}) \text{ for all } m_i' \in M_i ,$$

and

$$v_i(m) > v_i(m_i', m_{-i}) \text{ for all } m_i' > m_i .$$

The decent Nash equilibrium is a refinement of the Nash equilibrium, where each agent has minimal prosociality in that it announces the maximal best-response message.

A commitment rule α must induce all agents to act cooperatively as the unique decent Nash equilibrium. The maximal message profile is denoted by $\bar{m} \equiv (\bar{m}_i)_{i \in N} \in M$, where

$$\bar{m}_i = 1 \text{ for each } i \in N .$$

Uniqueness (U): The maximal message profile $\bar{m}$ is the unique decent Nash equilibrium in the commitment game associated with the commitment rule α. It achieves full cooperation (i.e., $\alpha(\bar{m}) = \bar{a}$).

4. Sustainability

Furthermore, a commitment rule must be sustainable in that even if a non-negligible number of agents adhere to uncooperative attitudes, many of the remaining agents are still willing to act cooperatively. For each subset of agents $\tilde{N} \subset N$, we define a limited decent Nash equilibrium in the commitment game associated with a commitment rule α as a message profile $m \in M$, such that every agent belonging to $\tilde{N}$ selects the maximal best response to m, whereas any other agent adheres to the minimal message. Thus, a message profile $m \in M$ is said to be a limited decent Nash equilibrium for $\tilde{N}$ in the commitment game associated with a commitment rule α if for every $i \in \tilde{N}$,

$$v_i(m) \geq v_i(m_i', m_{-i}) \text{ for all } m_i' \in M_i,$$

and

$$v_i(m) > v_i(m_i', m_{-i}) \text{ for all } m_i' > m_i,$$

whereas for every $i \in N \setminus \tilde{N}$,

$$m_i = 0.$$

Fix an arbitrary positive real number $\lambda \in (0,1)$.

Exact Sustainability (ES): Consider an arbitrary subset $\tilde{N} \subset N$, where it is assumed that

$$\left|\tilde{N}\right| \geq \lambda n.$$

There exists a limited decent Nash equilibrium m for $\tilde{N}$ in the commitment game, such that

$$m_i = 1 \text{ for all } i \in \tilde{N}.$$

ES and MC imply that even if a non-negligible number of agents (i.e., $N \setminus \tilde{N}$) adhere to uncooperative attitudes, all the remaining agents (i.e., $\tilde{N}$) are still willing to act

cooperatively. The following theorem states that the requirement of ES is quite restrictive and that there is a serious trade-off between SP and ES.

Theorem 1: If $\varepsilon > 0$ is close to zero and n is sufficiently large, there is no commitment rule α that satisfies SP, MC, and ES. If $\lambda > 0$ is close to zero and n is sufficiently large, there is no commitment rule α that satisfies SP, MC, and ES. Moreover, if

(1) $$c < 1 - \frac{\varepsilon}{\ln \lambda},$$

then, for a sufficiently large n, there exists no commitment rule α that satisfies the SP, MC, and ES.

Proof: Suppose that a commitment rule α exists that satisfies the SP, MC, and ES. Note that any commitment rule derived from this commitment rule and a permutation on N also satisfies these requirements. Moreover, any commitment rule derived from a weighted sum of these commitment rules also satisfies these requirements. Thus, without loss of generality, it can be assumed that the commitment rule is symmetric in that for every permutation $\mu : N \to N$ and $m \in M$,

$$\alpha_i(m) = \alpha_{\mu(i)}(m') \text{ for all } i \in N,$$

where $m' = (m'_j)_{j \in N}$ and $m'_{\mu(i)} = m_i$ are denoted for all $i \in N$.

The first part of this theorem can be proved as follows. From the ES, $\bar{m}$ must be a (limited) decent Nash equilibrium, where we set $\tilde{N} = N$. Hence, if agent n announces 0 instead of 1, each of the other agents decreases their promises at least by $\frac{c-1}{n-1}$. Therefore, their promise must be at most $1 - \frac{c-1}{n-1}$. Next, consider $\tilde{N} = \{1, 2, ..., n-1\}$ and the limited decent Nash equilibrium for $\tilde{N}$, where every agent in $\tilde{N}$ selects the maximal message 1. If agent $n-1$ announces 0 instead of 1, each of the other agents in $\tilde{N}$ must decrease their promises at least by $\frac{c-1}{n-2}$ for their cooperative incentive. Thus, their promise must be at

most $1-\frac{c-1}{n-1}-\frac{c-1}{n-2}$. Recursively, for each $l \in \{2,...,n-1\}$, consider $\tilde{N}=\{1,2,...,n-l\}$ and the limited decent Nash equilibrium for $\tilde{N}$, where every agent in $\tilde{N}$ selects the maximal message 1. If agent $n-l$ announces 0 instead of 1, each of the other agents in $\tilde{N}$ must decrease their promises at least by $\frac{c-1}{n-l-1}$. Hence, their promise must be at most $1-\sum_{l'=0}^{l}\frac{c-1}{n-l'-1}$.

From the ES and MC, for each $l<(1-\lambda)n$, $\sum_{l'=0}^{l}\frac{c-1}{n-l'-1}\leq\varepsilon$ must hold. For a sufficiently large n, we can approximate $\sup_{l<(1-\lambda)n}\sum_{l'=0}^{l}\frac{c-1}{n-l'-1}$ using $(c-1)\ln\frac{1}{\lambda}$, which is greater than zero. Hence, if ε is close to zero, then for a sufficiently large n, $\sup_{l<(1-\lambda)n}\sum_{l'=0}^{l}\frac{c-1}{n-l'-1}$ is greater than ε, which contradicts the MC. Hence, the first part of the theorem has been proved.

Further, if λ is close to zero, then $(c-1)\ln\frac{1}{\lambda}$ is greater than 1, indicating that any commitment rule fails to satisfy MC regardless of the specification of ε, provided that n is sufficiently large. Therefore, the second part of this theorem has been proved.

From the above observations, given a sufficiently large n, $\max_{m\in M}\{m_1-\alpha_1(m)\}$ is approximated by $(c-1)\ln\frac{1}{\lambda}$ or more. Hence, $(c-1)\ln\frac{1}{\lambda}\leq\varepsilon$ must hold. However, this notion contradicts the inequality (1). Hence, we have proved the third part of this theorem.

Q.E.D.

ES requires a commitment rule to incentivize all rational agents to act cooperatively even if there exist multiple irrational agents who adhere to the minimal-message announcements. Thus, for each $l \in \{1,2,...,n\}$, a commitment rule must set a penalty against the $l-th$ deviant in an accumulative manner, whose value should be at least as

large as $\frac{c-1}{n-l-1}$. Therefore, the cumulative value of the penalties must be at least as large as $\sum_{l'=0}^{l}\frac{c-1}{n-l'-1}$, which, however, is greater than 1, provided that n and l are sufficiently large. It is imperative to weaken the requirement of ES for the cumulative penalties to be kept as low as possible.

5. Virtual Sustainability

To overcome the difficulty of the sustainable commitment rule implied by Theorem 1, we replace ES with a weaker requirement within a range that does not impair the original meaning. Fix an arbitrary positive integer $w>2$. For convenience, we assume that n is an integer multiple of w. The role of w corresponds to that of $\frac{1}{\lambda}$ in the definition of ES. Fix an arbitrary positive integer $z\in\{1,...,w-1\}$.

Virtual Sustainability (VS): Consider an arbitrary subset $\tilde{N}\subset N$, where

$$\left|\tilde{N}\right|\geq\frac{z+1}{w}n.$$

There exists a limited decent Nash equilibrium m for $\tilde{N}$ in the commitment game associated with the commitment rule $\boldsymbol{\alpha}$, which has a subset $\bar{N}\subset\tilde{N}$, such that

$$\left|\bar{N}\right|\geq\left|\tilde{N}\right|-\frac{n}{w},$$

and

$$m_i=1 \text{ for all } i\in\bar{N}.$$

VS along with MC implies that even if a non-negligible number of agents (i.e., $N\setminus\tilde{N}$) adhere to uncooperative attitudes, not all but a large proportion of the remaining agents (i.e., $\bar{N}$) are still willing to act cooperatively. For a sufficiently large w, VS can be

considered an approximation of the ES. For $z=1$, VS can be considered an approximation of the most restrictive sustainability requirement implied by ES associated with a sufficiently small λ.

The following theorem shows a necessary and sufficient condition for the existence of a commitment rule that satisfies SP, MC, U, and VS. The proof of the theorem is presented in the following subsections (i.e., Subsections 5. 1 and 5. 2).

Theorem 2: A commitment rule exists that satisfies SP, MC, U, and VS if and only if:

$$\sum_{x'=1}^{w-z}\frac{(c-1)w}{(w-x')n}\leq\varepsilon. \tag{2}$$

We assume that the private cost is dependent on n, which is denoted by $c=c(n)$, and that there exists $\rho\in[0,1]$ (called the limit relative private cost) such that

$$\lim_{n\to\infty}\frac{c(n)}{n}=\rho.$$

The following theorem characterizes the possibility of a commitment rule being sustainable within the limit with respect to n.

Theorem 3: Suppose that $\rho>0$. Then, for a sufficiently large n, there exists a commitment rule that satisfies SP, MC, U, and VS if

$$\sum_{x'=1}^{w-z}\frac{w}{w-x'}<\frac{\varepsilon}{\rho}. \tag{3}$$

For a sufficiently large n, there is no commitment rule that satisfies SP, MC, U, and VS if

$$\sum_{x'=1}^{w-z}\frac{w}{w-x'}>\frac{\varepsilon}{\rho}. \tag{4}$$

Proof: Since

$$\lim_{n\to\infty}\sum_{x'=1}^{w-z}\frac{\{c(n)-1\}w}{(w-x')n}=\rho\sum_{x'=1}^{w-z}\frac{w}{w-x'},$$

it follows that if the inequality (3) holds, then for a sufficiently large n, the inequality (2) also holds. Moreover, it follows that if the inequality (4) holds, then for a sufficiently large n, the inequality (2) does not hold. Hence, from Theorem 2, Theorem 3 can be proved.

Q.E.D.

The following theorem states that if the social benefit $(n-1)$ is much greater than the private cost saving $(c-1)$, i.e., $\frac{c}{n}$ is close to zero, then there exists a commitment rule that meets VS, as well as SP, MC, and U.

Theorem 4: Suppose $\rho = 0$. Then, for a sufficiently large n, there exists a commitment rule that satisfies SP, MC, U, and VS.

Proof: Since

$$\lim_{n\to\infty} \sum_{x'=1}^{w-z} \frac{(c-1)w}{(w-x')n} = 0,$$

it follows from Theorem 2 and the inequality (2) that Theorem 4 holds true.

Q.E.D.

5. 1. Stepwise Commitment Rule

The proof of theorem 2 is constructive. We define $\delta(\cdot, n) = \delta(\cdot) : \{0,...,w\} \to R$ as follows. Let $\delta(0) \equiv 0$. For each integer $x \in \{1,...,w-1\}$, let

$$\delta(x) \equiv \sum_{x'=1}^{x} \frac{(c-1)w}{(w-x')n}.$$

Let $\delta(w) \equiv \delta(w-1)$. For each $m \in M$, the number of agents whose messages are less than one is denoted by:

$$y(m) \equiv \left|\{i \in N \mid m_i < 1\}\right| \in \{0,...,n\}.$$

Then, we specify $x(m) \in \{0,...,w\}$ as follows:

$$x(m) = 0 \qquad \text{if } y(m) = 0,$$

and for each $x \in \{1,...,w\}$,

$$x(m) = x \qquad \text{if } \frac{x-1}{w} < \frac{y(m)}{n} \leq \frac{x}{w}.$$

We define the stepwise commitment rule $\alpha = \alpha^*$ as follows. For every $i \in N$ and $m \in M$, let

$$\alpha_i^*(m) = m_i - \min[\delta(x(m)), m_i - \min_{j \in N} m_j, \varepsilon],$$

that is,

$$\alpha_i^*(m) = m_i - \varepsilon \qquad \text{if } m_i - \varepsilon \geq \max[\min_{j \in N} m_j, m_i - \delta(x(m))],$$

$$\alpha_i^*(m) = \min_{j \in N} m_j \qquad \text{if } \min_{j \in N} m_j \geq m_i - \min[\delta(x(m)), \varepsilon],$$

and

$$\alpha_i^*(m) = m_i - \delta(x(m)) \quad \text{if } m_i - \delta(x(m)) \geq \max[\min_{j \in N} m_j, m_i - \varepsilon].$$

The stepwise commitment rule α^* can be interpreted as follows. We classify the message space M into $w+1$ categories. Each category $x \in \{0,...,w\}$ includes all the message profiles m, such that the number of agents whose messages are less than unity is between $\frac{(x-1)n}{w}$ and $\frac{xn}{w}$ (i.e., $\frac{x-1}{w} < \frac{y(m)}{n} \leq \frac{x}{w}$). If a message profile changes from category $x-1$ to category x, the discount amount of an agent's promise from their upper limit increases by

$$\delta(x) - \delta(x-1) = \frac{(c-1)w}{(w-x)n}.$$

If a message profile changes but the category remains unchanged, this discount remains the same. (For convenience, we ignored the minimal upper limit and ε in this explanation.)

5. 2. Proof of Theorem 2

We show that, with the inequality (2), the stepwise commitment rule α^* satisfies SP, MC, U, and VS. By definition, α^* satisfies SP and MC.

Moreover, α^* satisfies U as follows. Consider the maximal message profile $\bar{m}$. Suppose that agent 1 selects $m_1 < 1$ instead of $\bar{m}_1 = 1$. If $m_1 \geq 1 - \delta(m_1, \bar{m}_{-1}) = 1 - \frac{(c-1)w}{(w-1)n}$, then any agent's promise decreases from 1 to m_1. As $n > c$, we have $(n-1)(1-m_1) > (c-1)(1-m_1)$, which implies that agent 1 decreases its utility. If $m_1 < 1 - \delta(m_1, \bar{m}_{-1}) = 1 - \frac{(c-1)w}{(w-1)n}$, then any other agent's promise decreases from 1 to $1 - \frac{(c-1)w}{(w-1)n}$. As $n > w$, we have

$$(n-1)\frac{(c-1)w}{(w-1)n} > c - 1 \geq (c-1)(1-m_1),$$

This notion signifies that agent 1 decreases their utility. Therefore, we have proved that $\bar{m}$ is a decent Nash equilibrium. By definition, we obtain $\alpha^*(\bar{m}) = \bar{a}$.

Consider an arbitrary message profile $m \in M \setminus \{\bar{m}\}$. There exists an agent $i \in N$ such that $m_i = \min_{j \in N} m_j < 1$. Suppose that m is a decent Nash equilibrium. Due to minimal prosociality, any other agent's message is $\min[1, m_i + \delta(x(m)), m_i + \varepsilon] > m_i$. Accordingly, their promise is equal to m_i. Agent i has an incentive to increase their message in this case because any other agent's promise increases simultaneously. However, this notion is a contradiction; thus, we have proved that α^* satisfies U.

Thereafter, we show that α^* satisfies VS as follows. Consider an arbitrary subset $\tilde{N} \subset N$, where we assume $\left|\tilde{N}\right| \geq \frac{z+1}{w} n$. We can select a subset $\bar{N} \subset \tilde{N}$ where

$$x(\left|N \setminus \tilde{N}\right|) = x(\left|N \setminus \bar{N}\right|) \leq w - z - 1 \text{ and } x(\left|N \setminus \bar{N}\right| + 1) = x(\left|N \setminus \bar{N}\right|) + 1.$$

Notably, $\left|\bar{N}\right| \geq \left|\tilde{N}\right| - \frac{n}{w}$. Consider a message profile m that is specified as

$$m_i = 0 \text{ for all } i \in N \setminus \tilde{N},$$

$$m_i = 1 \text{ for all } i \in \bar{N},$$

and

$$m_i = \delta(x(m)) \text{ for all } i \in \tilde{N} \setminus \bar{N},$$

where, from the inequality (2), we have $\delta(x(m)) \leq \varepsilon$. We can prove that this message profile is a limited decent Nash equilibrium for $\tilde{N}$ as follows. Note that no agent $i \in \tilde{N} \setminus \bar{N}$ influences $x(m)$ through message selection. Hence, they prefer to set their promise equal to zero; therefore, because of (2), their maximal best response is $\delta(x(m))$. Next, any agent $i \in \bar{N}$ can influence $x(m)$. In other words, by selecting their message to be less than one, they can change the category from $x(m)$ to $x(m)+1$. This change decreases the promise of any agent in $\bar{N}$ from $1-\delta(x(m))$ to $1-\delta(x(m)+1)$, that is, by the amount of

$$\delta(x(m)+1) - \delta(x(m)) = \frac{(c-1)w}{(w-x(m)-1)n}.$$

As $\left|\bar{N}\right| = \frac{\{w-x(m)\}n}{w}$ and n are sufficiently large, this change decreases the agent $i's$ utility by the amount of

$$\begin{aligned}
&(\left|\bar{N}\right| - 1)\{\delta(x(m)+1) - \delta(x(m))\} \\
&= [\frac{\{w-x(m)\}n}{w} - 1]\frac{(c-1)w}{(w-x(m)-1)n} \\
&= \frac{\{w-x(m)\}n - w}{\{w-x(m)\}n - n}(c-1) \geq c-1.
\end{aligned}$$

This notion implies that they prefer the maximal message. Hence, we have proved that the specified message profile is a limited decent Nash equilibrium for $\tilde{N}$, and therefore, α^* satisfies VS.

From these observations, we have proved that α^* satisfies the SP, MC, U, and VS.

Similar to the proof of Theorem 1, it follows that if a commitment rule satisfies VS, the cumulative penalty when a message profile m belongs to category z must be at least as large as $\sum_{x'=1}^{w-z} \frac{(c-1)w}{(w-x')n}$. Therefore, for this rule to satisfy MC, the inequality (2) must hold. Hence, we have completed the proof of Theorem 2.

5. 3. Role of Minimal Upper Limit

The stepwise commitment rule makes each agent's promise depend on the minimal upper limit. Note that if we do not require a commitment rule to satisfy MC, then there is no point in making each agent's promise dependent on the minimal upper limit.

Consider a commitment rule $\tilde{\alpha}$ with SP, which is specified by $\tilde{\alpha}(\bar{m}) = \bar{a}$, and

$$\tilde{\alpha}(m) = \underline{a} \text{ for all } m \neq \bar{m}.$$

We call $\tilde{\alpha}$ the unanimous commitment rule. The unanimous commitment rule $\tilde{\alpha}$ does not depend on the minimal upper limit. Nevertheless, it satisfies U; that is, the maximal message profile $\bar{m}$ is the unique decent Nash equilibrium in the associated commitment game.

If we limit the scope to commitment rules that satisfy MC, the minimal upper limit will have a crucial role for U. Consider a commitment rule with SP and MC, $\hat{\alpha}$, which is specified by $\hat{\alpha}(\bar{m}) = \bar{a}$ and

$$\hat{\alpha}_i(m) = \max[0, m_i - \varepsilon] \text{ for all } m \neq \bar{m} \text{ and } i \in N.$$

This rule is a natural extension of the unanimous commitment rule $\tilde{\alpha}$. It is independent of the minimal upper limit, and the maximal message profile $\bar{m}$ is a decent Nash equilibrium in the associated commitment game. However, due to MC, it fails to satisfy U. In fact, the message profile $\hat{m}$ specified by $\hat{m}_i = \varepsilon$ for all $i \in N$ is another decent Nash equilibrium, but it fails to achieve cooperation.

Next, consider a commitment rule with SP and MC, $\hat{\alpha}^*$, which is specified by:

$$\hat{\alpha}_i^*(m) = \max[m_i - \varepsilon, \min_{j \in N} m_j] \text{ for all } m \in M \text{ and } i \in N.$$

We can consider this rule as a modification of $\hat{\alpha}$, which we call the cautious commitment rule. As the cautious commitment rule $\hat{\alpha}^*$ depends on the minimal upper limit, it satisfies U (i.e., the maximal message profile $\bar{m}$ is the unique decent Nash equilibrium) as follows. Due to minimal prosociality, many agents prefer to select messages that are greater than the minimal upper limit, which motivates the agent who announces the minimal upper limit to increase their message because many agents increase their promises simultaneously. Thus, agents can ascend the minimal upper limit like climbing stairs. Consequently, the maximal message profile is the only equilibrium that will survive through this stair-climbing procedure. However, the cautious commitment rule $\hat{\alpha}^*$ does not satisfy VS. Hence, the stepwise commitment rule α^* is proposed as a more elaborate extension of the cautious commitment rule $\hat{\alpha}^*$.

To better understand the role of the minimal upper limit, let us consider a commitment rule α^{**} as a modification of the stepwise commitment rule α^*, where for every $i \in N$ and $m \in M$,

$$\alpha_i^{**}(m) = \max[m_i - \delta(x(m)), 0].$$

Note that the α^{**} does not depend on the minimal upper limit. Like Theorem 2, we can show that α^{**} satisfies SP, MC, and VS, and the maximal message profile is a decent Nash equilibrium in the associated commitment game. However, this decent Nash equilibrium is not the only one. Consider a message profile m' where

$$m_i' = \delta(x(m)) = \delta(w) \text{ for all } i \in N.$$

Note that

$$\alpha_i^{**}(m') = 0 \text{ for all } i \in N,$$

and that if an agent i increases their message, then their promise also increases, but the other agents' promises remain at zero. Hence, the message profile m' is a decent Nash equilibrium but results in no cooperation. Due to the failure of stair-climbing procedure, α^{**} fails to satisfy U.

If there exists an agent who adheres to the minimal message, the role of the minimal upper limit will cease to function even in the commitment game associated with the stepwise commitment rule α^*. The dependence on the minimal upper limit becomes incapable of deriving uniqueness in this case. Consider a message profile m'', where $m_1'' = 0$, and

$$m_i'' = \delta(x(m'')) = \delta(w) \text{ for all } i \in N \setminus \{1\}.$$

Note that $\alpha_i^*(m'') = 0$ for all $i \in N$, that is, the message profile m'' makes all agents act uncooperatively. However, it is a limited decent Nash equilibrium for $\tilde{N} = \{2,...,n\}$ because any agent other than agent 1 has no incentive to change the message for the same reason as in the case of m'.

Remark: To prove Theorem 4, we only need to consider the following commitment rule, α^{***}, whose design is simpler than that of the stepwise commitment rule. That is, for every $i \in N$ and $m \in M$,

$$\alpha_i^{***}(m) = m_i - \min[\frac{x(m)}{w}\varepsilon, m_i - \min_{j \in N} m_j],$$

that is,

$$\alpha_i^{***}(m) = m_i - \frac{x(m)}{w}\varepsilon \quad \text{if } m_i - \frac{x(m)}{w}\varepsilon \geq \min_{j \in N} m_j,$$

and

$$\alpha_i^{***}(m) = \min_{j \in N} m_j \quad \text{if } m_i - \frac{x(m)}{w}\varepsilon < \min_{j \in N} m_j.$$

According to α^{***}, the $(\frac{x(m)n}{w}+1)-th$ deviant will be penalized from each of $\{w - x(m) - 1\}n$ agents by the amount of $\frac{\varepsilon}{w}$. For a sufficiently large n,

$$\frac{\varepsilon}{w} \geq \frac{(c-1)w}{(w-x-1)n} = \delta(x+1) - \delta(x) \text{ for all } x \in \{0,...,w-2\},$$

which, along with the argument in the proofs of Theorems 2 and 4, implies that α^{***} satisfies VS if $\rho = 0$.

6. Local Trust

In this study, we assumed that every agent would keep their promise as long as they do not outweigh their announced upper limit. To validate this assumption, local trust is utilized as follows. We assume that agents are locally linked to each other in a hierarchical manner, which is expressed by a tree $\tau : N \to N \bigcup \{0\}$, where $\tau(i) \neq i$ for all $i \in N$. A committee exists as a dummy agent 0 at the top of the tree. Every agent $i \in N$ is (indirectly) connected to this committee in that a positive integer K and a finite sequence $(i^1,...,i^K)$ exist such that $i^1 = i$, $i^K = 0$, and $i^k = \tau(i^{k-1})$ for all $k \in \{2,...,K\}$. We consider agent $\tau(i) \in N \bigcup \{0\}$ as the (direct) superior of agent $i \in N$, that is, we consider agent $i \in N$ as a (direct) subordinate of agent $\tau(i) \in N \bigcup \{0\}$.

A community is defined as (i, C), where $i \in N \bigcup \{0\}$, $C \subset N$, and $i \notin C$. Agent i is considered to be the common superior of all agents in C. Let

$$C(i,\tau) = C(i) \equiv \{j \in N \mid \tau(j) = i\}$$

denote the set of all the subordinates of agent i in the tree τ. We consider $(i, C(i,\tau))$ as the community composed of all agents whose superiors are agent i. Each member's message is observable to their superior and their colleagues. However, it is not observed by any other agent.

Within the scope of privacy protection, any agent will never tell lies to their superiors and subordinates, and within the scope of SP, they will keep the promise required by their superior.

6. 1. Communication

Associated with a tree τ, we introduce communication as the following bottom-up and top-down procedures. Let $N^0 \equiv \{0\}$. Recursively, for each $h \geq 1$, let $N^h \equiv \{i \in N \mid \tau(i) \in N^{h-1}\}$. $N^h = N^{h,\tau}$ expresses the h-th layer of tree τ. Consider $H = H^\tau$ denote the minimal integer h such that $N^{h+1,\tau}$ is empty. We consider N^H as the bottom layer.

First, any bottom-layer agent $i \in N^H$ sends a signal $l_i = m_i$ to their superior $\iota(i) \in N^{H-1}$, which is set equal to their own upper limit. Then, each agent $i \in N^{H-1}$ sends $l_i = \min[m_i, \min_{j \in C(i)} l_j]$, that is, the minimum between agent $i's$ own upper limit and their subordinates' signals, and also sends the number of their subordinates $j \in C(i) \bigcup \{i\}$ (including themselves) whose messages are less than unity (i.e., $m_j < 1$), which is denoted by y_i, to their superior $\iota(i) \in N^{H-2}$. Recursively, for each $h \in \{1,...,H-2\}$, each agent $i \in N^h$ sends $l_i = \min[m_i, \min_{j \in C(i)} l_j]$ and y_i to their superior $\iota(i) \in N^{h-1}$, where we define

$$y_i = \sum_{j \in C(i)} y_j + 1 \qquad \text{if } m_i < 1,$$

and

$$y_i = \sum_{j \in C(i)} y_j \qquad \text{if } m_i = 1.$$

At the end of this bottom-up procedure, the committee can know the minimal upper limit among all agents, i.e., $\min_{i \in N} m_i$, and the number of all agents whose messages are less than unity, i.e., $y(m)$.

After completing the bottom-up procedure, the committee informs their subordinates of $\min_{i \in N} m_i$ and $y(m)$. Recursively, for each $h \in \{1,...,H-1\}$, any agent $i \in N^h$ informs their subordinates of $\min_{j \in N} m_j$ and $y(m)$ like a bucket brigade. At the end of this top-down procedure, both $\min_{i \in N} m_i$ and $y(m)$ become common knowledge among all the agents.

Notably, $\alpha_i^*(m)$ only depends on $\min_{i\in N} m_i$, $y(m)$, and m_i. Therefore, we can protect the enforceability of the stepwise commitment rule α^* through local trust and communication.

6. 2. Stability

A community may benefit from adopting another commitment rule of its own, which destabilizes the stepwise commitment rule α^*. To eliminate this destabilization, we discuss the optimal community size in this subsection. Consider an arbitrary message profile $m \in M$ and an arbitrary superior $i \in N \bigcup \{0\}$, where we assume that $\min_{j\in C(i)} m_j - \varepsilon > \min_{j\in N} m_j$, that is,

$$\alpha_i^*(m) = m_j - \varepsilon \quad \text{for all} \quad j \in C(i).$$

Suppose that the size of agent $i's$ community is greater than c (i.e., $|C(i)| > c$). Then, all subordinates of agent i can raise their utilities if they uniformly increase their action levels to the same extent. Thus, all subordinates of agent i may prefer to replace the stepwise commitment rule for this community, $\alpha_{C(i)}^*$, with another rule $\alpha_{C(i)}$ such that $\alpha_j(m) = m_j - \tilde{\varepsilon}$ for all $j \in C(i)$, where $\tilde{\varepsilon}$ is set close to ε and $0 \le \tilde{\varepsilon} < \varepsilon$. This replacement raises all their promises to the same extent (i.e., $\varepsilon - \tilde{\varepsilon}$), which increases all their utilities. However, this replacement runs the risk of losing the incentives for all agents outside the community to act cooperatively.

Next, suppose that the size of agent $i's$ community is smaller than c (i.e., $|C(i)| < c$). Then, all subordinates of agent i can raise their utilities if they uniformly decrease their action levels to the same extent. Hence, all subordinates of agent i may prefer to replace $\alpha_{C(i)}^*$ with another rule $\alpha_{C(i)}$ such that $\alpha_j(m) = m_j - \tilde{\varepsilon}$ for all $j \in C(i)$, where $\tilde{\varepsilon}$ is set close to ε (i.e., $m_j - \tilde{\varepsilon} \ge \min_{j\in N} m_j$) and $\tilde{\varepsilon} > \varepsilon$. This replacement reduces all their promises to the same extent (i.e., $\tilde{\varepsilon} - \varepsilon$), which increases all their utilities. However, this replacement runs the risk of violating MC.

From these observations, it can be inferred that, to stabilize the stepwise commitment rule α^*, it is desirable for the tree τ to satisfy

$$\left|C(i)\right| \in \{0, c\} \text{ for all } i \in N \bigcup \{0\}.$$

Accordingly, we can consider c as the optimal community size.

6. 3. Tree Renovation

Assume that a tree τ has an agent $i \in N \bigcup \{0\}$ whose community size is not optimal (i.e., $\left|C(i,\tau)\right| \notin \{0, c\}$), and therefore, agent $i's$ community is a destabilizing factor to the stepwise commitment rule. Hence, we need to renovate this existing tree and make it into another tree $\hat{\tau} \neq \tau$ such that $\left|C(i,\hat{\tau})\right| \in \{0, c\}$ for all $i \in N \bigcup \{0\}$. In this subsection, we present a spontaneous-order method that changes a tree (i.e., communities and their linkage) over time to a desirable tree that can stabilize the stepwise commitment rule α^*.

We assume that if an agent $i \in N$ is a subordinate of an agent $j \in N \bigcup \{0\}$, then agent i can become a subordinate of agent $j's$ superior in the next step. We also assume that if an agent $i \in N$ is a subordinate of an agent $j \in N \bigcup \{0\}$ and an agent $j' \in N$ is another subordinate of agent j, then agent i can become a subordinate of agent j' in the next step. With these assumptions, a tree τ is said to be changeable to another tree τ' if for every $i \in N$, either $\tau'(i) = \tau(i)$, $\tau'(i) = \tau(\tau(i))$, or $\tau'(i) \in C(\tau(i), \tau)$. A tree τ is said to be procedurally changeable to another tree τ' if there exists a positive integer K and a finite sequence of trees $(\tau^k)_{k=1}^K$ such that $\tau^1 = \tau$, $\tau^K = \tau'$, and τ^k is changeable to τ^{k+1} for all $k \in \{1, ..., K-1\}$. Note that

$$[\, \tau'(i) = \tau(\tau(i)) \,] \Leftrightarrow [\, \tau(i) \in C(\tau'(i), \tau') \,].$$

Hence, we have reflexivity in that if τ is procedurally changeable to τ', τ' is procedurally changeable to τ. We have transitivity in that if τ is procedurally changeable to τ' and τ' is procedurally changeable to τ'', τ is procedurally changeable to τ''.

Proposition 1: Any tree is procedurally changeable to any other tree.

Proof: Consider an arbitrary tree τ. We specify a positive integer K and a finite sequence $(\tau^k)_{k=1}^K$ as follows. Let $\tau^1 = \tau$. Let

$$\tau^2(i) = \tau^1(i) \text{ for all } i \in N \text{ such that } \tau^1(i) = 0,$$

and

$$\tau^2(i) = \tau^1(\tau^1(i)) \text{ for all } i \in N \text{ such that } \tau^1(i) \neq 0.$$

Thus, the superior of agent i in the tree τ^2 is the superior of their superior in the tree τ^1. Recursively, for each integer $k \geq 3$, let

$$\tau^k(i) = \tau^{k-1}(i) \text{ for all } i \in N \text{ such that } \tau^{k-1}(i) = 0,$$

and

$$\tau^k(i) = \tau^{k-1}(\tau^{k-1}(i)) \text{ for all } i \in N \text{ such that } \tau^{k-1}(i) \neq 0.$$

We define a tree τ^0 by

$$\tau^0(i) = 0 \text{ for all } i \in N.$$

We eventually reach a positive integer K such that $\tau^K(i) = 0$ for all $i \in N$ (i.e., $\tau^K = \tau^0$). As τ^k is changeable to τ^{k+1} for all $k \in \{1,...,K-1\}$, it follows that τ is procedurally changeable to τ^0. From reflexivity, τ^0 is procedurally changeable to τ. These properties hold for all the trees. Hence, from transitivity, any two trees are procedurally changeable to each other.

Q.E.D.

We assume that n is an integer multiple of c. Note that there exists a tree τ that stabilizes the stepwise commitment rule, that is, $|C(i,\tau)| \in \{0,c\}$ for all $i \in N \cup \{0\}$. From Proposition 1, by renovating trees within the scope of changeability, we can transform the existing tree, which fails to stabilize the stepwise commitment rule, into any tree that can stabilize it.

6. 4. Privacy

We have assumed that each agent sent the correct information to their superior. To maintain local trust, it is better to limit the content of information leaked to the outside as much as possible. However, as discussed in Subsection 5.3, if agents communicate only the number of agents whose messages are less than 1 (i.e., x_i), the problem arises that a commitment rule cannot satisfy U.

In this subsection, we consider a situation wherein each agent $i \in N$ only conveys information about the minimal upper limit (i.e., l_i). In this case, the committee is informed of the number of their subordinates who have uncooperative (indirect) subordinates but is not informed of the exact number of these uncooperative agents. Hence, the committee cannot discriminate between the case wherein all agents are uncooperative and the case wherein any subordinate of the committee has just a single (indirect) subordinate who is uncooperative. The committee cannot construct any commitment rule that satisfies VS unless it has a discrimination ability. Therefore, each community should understand that some degree of privacy must be sacrificed to solve the free rider problem.

6. 5. Community-Based Sustainability

In this subsection, we investigate the case of unforeseen circumstances wherein a single community adheres to uncooperative attitudes because of the temporal collapse of its local trust. We require a commitment rule to be sustainable in a community-based sense, such that even if a single community adheres to uncooperative attitudes, the remaining communities are still willing to act cooperatively.

We assume that any (non-empty) community has the optimal size given by $c = c(n)$. The community-based sustainability corresponds to the VS requirement associated with

$$w = w(n) = \frac{n}{c(n)} \text{ and } z = z(n) = w(n) - 2.$$

For the stepwise commitment rule α^* to satisfy this requirement, we need ε should not be less than $\delta(1)=\frac{\{c(n)-1\}w(n)}{(w(n)-1)n}$. Notably,

$$\lim_{n\to\infty}\frac{\{c(n)-1\}w(n)}{\{w(n)-1\}n}=\lim_{n\to\infty}\frac{\frac{c(n)}{n}}{1-\frac{c(n)}{n}}=\frac{\rho}{1-\rho}.$$

To ensure stability, the size of each community must be equal to $c(n)$. Assume that the limit relative private cost ρ (i.e., $\lim_{n\to\infty}\frac{c(n)}{n}$) is positive. Note that $c(n)$ diverges to infinity as n increases, and that to meet MC, ρ must not be greater than $\frac{\varepsilon}{1+\varepsilon}$. Thus, unless the limit relative private cost ρ is zero, we must have a trade-off between stability and community-based sustainability.

7. Conclusion

We investigate the free-rider problem inherent in the global commons and various wide-range externalities with non-excludability. To convert social dilemmas into an easier-to-solve coordination game format, we adopted a commitment rule design. Instead of using external retaliatory sanctions and ostracism measures, we propose a new method to achieve full cooperation as a unique decent Nash equilibrium within the scope of commitment rule design, where agents have minimal prosociality as lexicographical preferences. Specifically, we propose the stepwise commitment rule as a sustainable rule such that cooperative relationships can be maintained even if there are problematic agents who adhere to uncooperative attitudes. We further discuss social networking to help local trust solve the global problem within the scope of protecting privacy and sovereignty. The novelty of this study lies in the formulation of the global commons as free-rider problems, the formulation of the sustainability issue, and presenting ideal solutions to these problems. The clarification of the unique role of the minimum upper limit and the design method of

setting commitments on a category basis are the original proposals of this study. Moreover, this study is novel in relating local trust to global commons problems.

In this study, we only consider symmetric models and assume that all agents agree on their common social goal and how their interests are opposed. Asymmetric models should be considered in future studies. In this case, it might be more difficult for agents to agree on a common social goal and shift their perception from social dilemmas to a coordination game format. It has been pointed out that one of the reasons why the COP has been in trouble for so long is that the target of the negotiations is to allocate the burden of reducing CO2 emissions among countries, which creates a long-running dilemma. Changing the negotiation target from such reduction quotas to a common carbon price will facilitate a game shift from the current social dilemma to a more hopeful coordinated game format.

A prompt shift to a coordination game format will greatly contribute to ameliorating conflicts regarding the North-South disparity. For instance, developed countries' incentives to promote innovation in CO2 emission reduction technologies are strengthened by the simultaneous promotion of such technologies in developing countries. Accordingly, the economic development and rectification of the North-South disparity are in line with each other. Thus, the stepwise commitment rule will serve as a guideline for developing a mechanism design for social common capitals that maintain cultural and human life around the world by harmonizing natural environments, social environments, and economic development.

Attempting to use local trust to help solve global problems is a new research direction. Many important arguments still remain open. For example, in the event of a temporary collapse of trust in one community, the information regarding lower-layer communities may become unavailable. When transitioning from an existing destabilizing tree to a more stabilizing tree, we may have to temporarily navigate through more destabilizing trees. Careful consideration of such concerns is beyond the scope of this paper and is left to future research.

References

Cramton, Peter, David J.C. MacKay, Axel Ockenfels, and Steven Stoft (2017): Global Carbon Pricing: The Path to Climate Cooperation. Cambridge, MA: MIT Press.

Eliaz, Kfir (2002): Fault Tolerant Implementation, Review of Economic Studies, 69, 589-610.

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson (1982): Rational Cooperation in the Finitely Repeated Prisoners' Dilemma, Journal of Economic Theory, 27, 245-252.

MacKay, David J.C., Peter Cramton, Axel Ockenfels, and Steven Stoft (2015): Price Carbon — I Will If You Will, Nature, 526, 315-316.

Matsushima, Hitoshi (2022a): Free-Rider Problem and Sovereignty Protection, Discussion Paper UTMD-024. https://www.mdc.e.u-tokyo.ac.jp/2022/01/28/free-rider-problem-and-sovereignty-protection/. University of Tokyo.

Matsushima, Hitoshi (2022b): Epistemological Implementation of Social Choice Functions, Games & Economic Behavior, 136, 389-402.

Nordhaus, William (2015): Climate Clubs: Overcoming Free-Riding in International Climate Policy, American Economic Review, 105, 1339-1370.

Ostrom, Elinor (1990): Governing the Commons: The Evolution of Institution for Collective Action. Cambridge: UK: Cambridge University Press.

Ostrom, Elinor (2010): Beyond Markets and States: Polycentric Governance of Complex Economic Systems, American Economic Review, 100, 641-672.

Roemer, John E. (2010): Kantian Equilibrium, Scandinavian Journal of Economics, 112, 1-24.

Victor, David (2007): The Collapse of the Kyoto Protocol and the Struggle to Slow Global Warming. Princeton, NJ: Princeton University Press.