Neutral Public Good Mechanisms

Jin Yeub Kim*

November 11, 2019

Abstract

In this paper, I characterize neutral mechanisms for the provision of a public good. I show that neutral mechanisms form a reasonable set of predictions for mechanism selection in public goods problems: Such predictions are sufficiently sharp, robust to a perturbation of the information structure at the time of selection, and invulnerable to the possibility of information leakage during the selection process. I also illustrate that neutral mechanisms have the desirable properties of both efficiency and equity, and can be conveniently computed by the tractable set of conditions. These results are shown to have interesting implications for the analysis of ex ante and interim incentive efficient mechanisms for public goods problems.

JEL classification: C71; C78; D74; D82; H41

Keywords: Public goods; Mechanism design; Neutral optimum; Interim efficiency; Ex ante efficiency; Almost ex ante stage

*School of Economics, Yonsei University, 50 Yonsei-ro Seodaemun-gu, Seoul 03722, Republic of Korea (email: shiningjin@gmail.com). I am deeply grateful to Roger Myerson and Lars Stole for their valuable advice and guidance on this paper. I received helpful feedback from Rabah Amir, Dongkyu Chang, Youngsub Chun, Kyungmin (Teddy) Kim, Wooyoung Lim, Qinggong Wu, Nicholas Yannelis, and the seminar participants at the University of Iowa, Korea University, Kyung Hee University, the University of Pittsburgh, the 2018 Nanjing Conference on Game Theory, the 2018 Asian Meeting of the Econometric Society, the 14th SSCW Meeting, the 76th MPSA Annual Conference, the 2017 SEA Annual Meetings, and the 2017 KAEA-KEA-APEA International Conference. Part of this paper was circulated under the title "Efficient and Neutral Mechanisms in Almost Ex Ante Bargaining Problems." The author declares that there is no relevant or material financial interests that relate to the research described in this paper.

1 Introduction

In the classic public goods problem, agents must decide whether or not to produce a public good and how to divide the cost of production. This paper considers incomplete information settings where the agents themselves may agree on some decision rule or mechanism to help them make decisions. Which mechanisms should we expect to be selected and used by the agents? That is, what would be considered a reasonable set of predictions for mechanism selection that has a strong predictive power as well as desirable properties?

If the agents can choose a mechanism before learning their private information, then they would agree on a mechanism that is ex ante incentive efficient.[1] Selecting an ex ante incentive efficient mechanism is appropriate for maximizing the probability of the public good being produced in situations where it appears to be worth more than it costs ex ante. If the mechanism is selected after the agents learn their private information, then chosen mechanisms would be, minimally, interim incentive efficient (Holmström and Myerson, 1983). Any selection from among the set of those mechanisms is reasonable when the only concern is achieving Pareto efficiency for public good provision.

In this paper, I am concerned with neutral mechanisms using a concept proposed by Myerson (1984*b*). Roughly speaking, a mechanism is neutral if it is incentive efficient and virtually equitable in terms of players' virtual preferences that incorporate what they would have wanted if they were of different types. While focusing on ex ante or interim incentive efficient mechanisms may be deemed appropriate on their own merits, my goal of this paper is to illustrate that characterizing neutral mecha-

[1] In this case, it is also assumed that agents cannot commit themselves ex ante to participate honestly and obediently in the chosen mechanism after they observe their private information (Holmström and Myerson, 1983).

nisms for public good provision is more theoretically appealing and reasonable from efficiency and equity standpoints.

My formal analysis begins by setting up a two-person Bayesian bargaining problem to represent public good environments with incomplete information.[2] In this problem, two players can jointly choose among possible decisions about whether a discrete public good should be produced, and if so, how much each player should pay for producing it. A player's type is a complete description of his private information about his valuation for the public good, and each player has beliefs over the other player's possible types.[3]

A mechanism for this problem chooses the decision as a function of the players' independently and confidentially reported types. By the revelation principle, attention should be restricted mechanisms that are feasible in the sense that the players are willing to participate and to reveal their types honestly in the mechanism. I assume that the players are able to negotiate with each other for which mechanism to implement among all feasible mechanisms.

Given the set of feasible mechanisms, the concept of Pareto efficiency is clearly a minimal requirement for defining reasonable selections by the players. An ex ante (resp. interim) incentive efficient mechanism incorporates efficient aggregation of players' ex ante (resp. interim) preferences over feasible mechanisms. Further, the concept of neutral optimum can be considered a reasonable requirement for fair bargaining solutions when the properties of both efficiency and equity are concerned. A useful characterization of neutral mechanisms is also provided in the context of public goods problem.

I first show that ex ante incentive efficient mechanisms are not robust to a pertur-

[2]I restrict attention to two players; this substantially simplifies the exposition while conveying all the main insights.

[3]For exposition, I use male pronouns for a player.

bation of the ex ante informational structure at the time of mechanism selection. To show this, I consider a perturbed setting where the players are not absolutely certain that nobody has any private information at the time when they meet initially to select a mechanism. I call this stage of mechanism selection *almost ex ante*, which indicates informational environments in between the ex ante and interim ones.[4] When the selection is made at the almost ex ante stage, the set of incentive efficient mechanisms should consist of those feasible mechanisms that incorporate efficient aggregation of both interim and ex ante preferences. I find that this set coincides with the set of interim incentive efficient mechanisms and thus is a superset of the set of ex ante incentive efficient mechanisms. Therefore, the focus on an ex ante incentive efficient mechanism as the most reasonable mechanism for players to choose is valid only when there is absolutely no doubt that all players do not know their types at the selection stage. If any doubt exists, then reasonable selections must be defined on a larger set of interim incentive efficient mechanisms.

I then discuss two possible cautions in focusing on the set of interim incentive efficient mechanisms as a reasonable set of predictions. First, the set of interim incentive efficient mechanisms can be quite large, so characterizing this set generates indefinite predictions of public good mechanisms that may arise. Second, selecting a particular interim incentive efficient mechanism is vulnerable to the possibility of information leakage, which will be discussed in detail in Section 3.2.

These results call for some other solution criterion to delimit reasonable predictions of mechanisms that are sufficiently sharp, robust to alternative specifications of the information structures at the time of mechanism selection, and invulnerable to the issue of information leakage from the selection of the mechanism. The neutral

[4] I thank Roger Myerson for suggesting this term. The probability of being informed can be any value between zero and one, so the almost ex ante stage can be alternatively called an almost interim stage.

mechanism not only satisfies those desiderata but also has the desirable properties of both efficiency and equity. That is, it achieves an efficient and equitable balance between different preferences of alternative types of players, and hence is inscrutable. Thus, I find the application of neutral mechanisms to public goods problems to be a theoretically and intuitively more appealing way to represent reasonable choices of public good mechanisms that may actually arise in practice.

1.1 Related literature

This paper connects with three lines of research. First, the most closely related literature is a series of papers by Ledyard and Palfrey (1994, 1999, 2002, 2007). Ledyard and Palfrey (1994, 1999) fully characterize interim incentive efficient mechanisms for the provision of public goods in Bayesian environments; Ledyard and Palfrey (2002) compare the performance of simple voting rules with that of interim incentive efficient public good mechanisms; and Ledyard and Palfrey (2007) provide a more general framework to study the properties of interim efficient mechanisms for the class of linear independent environments.[5]

In terms of the model and applications, my paper shares several common features with their paper. Ledyard and Palfrey (1994) consider a simple case with two types and a 0-1 public good decision; likewise, I work with finite type sets and a 0-1 public good decision. In the models of Ledyard and Palfrey (1999, 2002) with a continuum of types, the individuals decide on a level of a public good; but with the linear production technology, the optimal level of the public good will always be either 0 or 1, so this is essentially equivalent to making a 0-1 public good decision. The key difference is that my analysis makes use of the relatively stronger concept of neutral optimum

[5]See also Gresik (1996) and Wilson (1985) who explore interim incentive mechanisms for sealed-bid trading problems.

developed by Myerson (1984*b*). In this sense, this paper complements theirs by taking one step further and characterizing the set of neutral mechanisms in a Bayesian public good environment.

In a broader sense, this paper contributes to the literature on bargaining solution concepts and mechanism design problems for Bayesian environments. Harsanyi and Selten (1972) first explored the question of how to define reasonable bargaining solutions in games with incomplete information. Other attempts were made in the seminal works by Myerson (1983, 1984*a*,*b*) as well as in Maskin and Tirole (1990, 1992).[6] Several other authors have addressed the issue of information leakage in mechanism selection games and/or the robustness or stability of mechanisms (e.g., Celik and Peters, 2011; Cramton and Palfrey, 1995; Crawford, 1985; Holmström and Myerson, 1983; Laffont and Martimort, 2000; and Lagunoff, 1995; Liu et al., 2014; Pomatto, 2019, among many others).

The aforementioned papers share the assumption that the individuals already have their private information when the game begins. This paper considers a richer framework that allows the state of individuals' information at the mechanism selection stage to be different from that at the implementation stage. The exercise delivers the conclusion that ex ante solutions are sensitive to the specification of what information individuals possess at the time of bargaining, whereas the concept of neutral optimum is robust to a perturbation of information specification as well as to a possibility of information leakage.[7] This paper solidifies the justification for applying such concept

[6]Myerson (1983) proposes solutions for the problem of mechanism selection by an informed principal who has all of the negotiating ability. This problem is further analyzed as a noncooperative game by Maskin and Tirole (1990, 1992) for the cases of private and common values. Balkenborg and Makris (2015) make connections between Myerson (1983) and Maskin and Tirole (1992).

[7]Kim (2017) characterizes the sets of interim incentive efficient mechanisms and of neutral mechanisms for a class of examples. While the present paper does not directly tackle information leakage issues, it implicitly considers the possibility of information leakage when selecting among interim incentive efficient mechanisms. de Clippel and Minelli (2004) provide characterizations of neutral solutions under the additional assumption of verifiable types at the implementation stage.

to many bargaining situations with incomplete information. The applications may encompass pretrial negotiations, labor and employment disputes, selling or hiring situations, international conflicts, and bargaining in over-the-counter markets.[8]

Finally, this paper contributes to the conflict literature on institutional design (e.g., Bester and Wärneryd, 2006; Hörner, Morelli and Squintani, 2015; Kydd, 2003; Meirowitz et al., 2017, and others). The literature has addressed several critical questions about the effectiveness of institutions in preventing conflicts. When comparing the performance of different institutions, it may seem natural to focus on institutions that minimize the ex ante likelihood of conflict. Invoking such measure is valid for situations where a conflict-minimizing institution is the only one that would naturally arise, or such an institution is imposed exogenously. But if disputing parties themselves are able to choose among many institutions, a conflict-minimizing institution might not be chosen. The selection of an institution would depend crucially on the information structure at the time of selection. Hence, this paper extends the study of conflict and institutional design by suggesting that the informational environment faced by disputing parties at the time they select an institution should inform which performance measure to use for evaluating institutions.

The remainder of the paper is organized as follows. Section 2 outlines the model of public goods problems, summarizes the definition of ex ante and interim incentive efficiency, and provides the characterization of neutral mechanisms. Section 3 presents the main results on ex ante and interim incentive efficient mechanisms, and discusses the advantage of using neutral mechanisms as well as offering the implications for the analysis of ex ante and interim mechanism selections. Section 4 concludes.

[8]See, e.g., Kim (2017, 2019).

2 The model

2.1 Setup

There are two players who must decide whether or not to produce 1 unit of a discrete public good and how to divide the production cost that is equal to K. Each player has private information about its type, and has prior beliefs about the other player's type. Each type represents a player's valuation for the public good. Players cannot verify any claims that the other player might make about her type, and the two players do not expect to make any further joint decision in the future.

The public goods problem described above can be formulated as a Bayesian bargaining problem à la Myerson (1984*b*).[9] The set of feasible collective decisions (or bargaining outcomes) is $\mathcal{D} = \{(q, y_1, y_2) | 0 \leq q \leq 1, y_1 + y_2 = K, y_i \in \mathbb{R}\ \forall i\}$, where, for each $(q, y_1, y_2) \in \mathcal{D}$, q represents the probability that the public good is produced and y_i represents player i's share of the cost. Let $y = (y_1, y_2)$ denote the profile of cost shares. For each player i, T_i is the set of possible types t_i for player i. For simplicity, I assume that the players' types are independent random variables. Player $-i$ believes that the probability of player i being of type $t_i \in T_i$ is $p_i(t_i)$ such that $\sum_{t_i \in T_i} p_i(t_i) = 1$. As a regularity condition, I assume that all types have positive probability, so $p_i(t_i) > 0$ for all i and all t_i. Let $T = T_1 \times T_2$ denote the set of all possible type combinations $t = (t_1, t_2)$. For mathematical convenience, T is assumed to be a finite set.

Let u_i denote player i's utility function from $\mathcal{D} \times T$ into $\mathbb{R}$, such that $u_i((q, y), t)$ is the payoff which player i would get if $(q, y) \in \mathcal{D}$ were chosen and if t were the

[9] The concept of Bayesian bargaining problem was proposed by Harsanyi (1967-8) and further analyzed for the fixed-threats case by Harsanyi and Selten (1972) and Myerson (1979, 1984*b*). The present paper also considers the fixed-threats case.

vector of players' types. Let $d^* \equiv (q, y) = (0, \mathbf{0})$ represent the decision not to produce the public good, which is the natural conflict outcome for this problem because no production occurs if the players cannot agree on the division of the cost. Under this formulation, the utility functions are defined by the formula

$$u_i\left((q, y), t\right) = t_i q - y_i, \ \forall i, \ \forall t.$$

2.2 Feasible and efficient mechanisms

By the revelation principle, I can set up the bargaining problem as a direct-revelation mechanism, without loss of generality. That is, the players do not have to agree on a specific decision; instead they may agree on some mechanism.

Because of the linearity of the utility functions, I can restrict attention to deterministic mechanisms, mapping from T to $\mathcal{D}$, without loss of generality. So let $(Q(\cdot), Y(\cdot)) = (Q(\cdot), Y_1(\cdot), Y_2(\cdot))$ be a *mechanism* for determining the decision as a function of the players' reported types, where $Q(t)$ is the probability that the public good is produced and each $Y_i(t)$ is the expected share of the production cost to be made by player i if t is the profile of reported types. This mechanism must satisfy $0 \leq Q(t) \leq 1$ and $Y_1(t) + Y_2(t) = Q(t)K$ for all $t \in T$. If $Q(t) > 0$, then $P_i(t) \equiv Y_i(t)/Q(t)$ represents player i's expected payment per unit of the public good produced when the profile of reported types is t. If $Q(t) = 0$, any cost payment need not be specified because the public good would not be produced.

The interim expected utility to type t_i of player i, given that both players report their types honestly, if (Q, Y) is implemented is

$$U_i(Q, Y|t_i) = t_i \sum_{t_{-i} \in T_{-i}} p_{-i}(t_{-i}) Q(t_{-i}, t_i) - \sum_{t_{-i} \in T_{-i}} p_{-i}(t_{-i}) Y_i(t_{-i}, t_i).$$

The implementation of a mechanism is restricted by two constraints of incentive compatibility and individual rationality. If player i's type were t_i but he reports some other type s_i, while the other player remains honest, then the expected utility to player i is

$$U_i^*(Q, Y, s_i|t_i) = t_i \sum_{t_{-i} \in T_{-i}} p_{-i}(t_{-i}) Q(t_{-i}, s_i) - \sum_{t_{-i} \in T_{-i}} p_{-i}(t_{-i}) Y_i(t_{-i}, s_i).$$

A mechanism (Q, Y) is interim *incentive compatible* if and only if $U_i(Q, Y|t_i) \geq U_i^*(Q, Y, s_i|t_i)$ for all i, for all $t_i \in T_i$, and for all $s_i \in T_i$.

The conflict outcome d^* will occur if the players disagree, and each player has the right to refuse public good production. So no type of any player should expect to do worse under the mechanism, given that both players report their types honestly, than in the conflict outcome. So a mechanism (Q, Y) is interim *individually rational* if and only if $U_i(Q, Y|t_i) \geq 0$ for all i and for all $t_i \in T_i$. Then a mechanism is defined to be *feasible* for the players in this public goods problem if and only if it is both incentive compatible and individually rational. By the revelation principle, there is no loss of generality in focusing on feasible mechanisms.

Given the set of feasible mechanisms, the concept of Pareto efficiency can be applied to identify the entire set of efficient mechanisms among which the players would reasonably choose from. If the mechanism is selected by the players with asymmetric information, then the proper concept of Pareto efficiency is interim incentive efficiency. A mechanism (Q, Y) is *interim incentive efficient* (IIE) if and only if (Q, Y) is feasible and there does not exist another feasible mechanism (Q', Y') such that all types of all players would prefer (Q', Y') over (Q, Y), that is, $U_i(Q', Y'|t_i) \geq U_i(Q, Y|t_i)$, for all i and for all t_i with at least one strict inequality. If the mechanism can be selected before the players learn their private information, then the concept of ex ante incentive

efficiency should be applied. A mechanism (Q, Y) is *ex ante incentive efficient* if and only if (Q, Y) is feasible and there does not exist another feasible mechanism (Q', Y') such that all players would prefer (Q', Y') over (Q, Y) before learning their private types, that is, $\sum_{t_i \in T_i} p_i(t_i) U_i(Q', Y'|t_i) \geq \sum_{t_i \in T_i} p_i(t_i) U_i(Q, Y|t_i)$, for all i, with at least one strict inequality.

2.3 Neutral mechanisms: characterization

Myerson (1984*b*) developed a generalization of the Nash bargaining solution for Bayesian bargaining problems, called the *neutral bargaining solution*. This solution concept is axiomatically defined: A neutral bargaining solution is any mechanism such that it is contained in every solution correspondence that satisfies the probability-invariance, extension, and random-dictatorship axioms. Scrutinizing all of the axioms is not the primary goal of the present paper; so without loss of comprehension of the solution concept, I focus only on the random-dictatorship axiom. This axiom provides the key logic in understanding the neutral bargaining solution, which is stated below in the context of our public good problem.

Axiom 1 (Random dictatorship) *If there exist two interim incentive efficient mechanisms* (Q^1, Y^1) *and* (Q^2, Y^2) *such that* $U_2(Q^1, Y^1|t_2) = 0$*, for all* $t_2 \in T_2$*, and* $U_1(Q^2, Y^2|t_1) = 0$*, for all* $t_1 \in T_1$*; and if the mechanism* (Q^n, Y^n) *defined by* $Q^n(t) = 0.5Q^1(t) + 0.5Q^2(t)$ *and* $Y^n(t) = 0.5Y^1(t) + 0.5Y^2(t)$*, for all* $t \in T$*, is interim incentive efficient, then* (Q^n, Y^n) *is a neutral public good mechanism.*

The hypotheses of Axiom 1 are satisfied for public good problems in which there is a clear mechanism (or public good decision) that each player should demand if he could have all of the bargaining power. In the terminology of Myerson (1983, 1984*b*), (Q^1, Y^2) and (Q^2, Y^2) in Axiom 1 are strongly optimal decisions for players 1 and 2

respectively; that is, each of the two mechanisms is the most reasonable solution for the player if he could dictatorially choose the mechanism. For such problems, if the 50-50 randomization between the two best mechanisms is incentive efficient, then it is a neutral mechanism.

When the hypotheses of Axiom 1 are satisfied for a given public good problem, then the neutral mechanism is easy to compute. But those hypotheses may be restrictive for many public good problems. In such cases, one can appeal to the following well-known result that gives a complete characterization of neutral mechanisms as the solutions to a constrained optimization problem, stated below without proof.

Lemma 1 (Myerson (1984*b*)) *A mechanism (Q, Y) is neutral for a public good problem defined in this paper if and only if, for each positive number ε, there exist vectors $\lambda = \big((\lambda_i(t_i))_{t_i \in T_i}\big)_{i \in \{1,2\}}$, $\alpha = \big(\alpha_i(s_i|t_i)\big)_{i \in \{1,2\}, s_i \in T_i, t_i \in T_i}$, and $\omega = \big(\omega_i(t_i)\big)_{i \in \{1,2\}, t_i \in T_i}$ such that $\lambda_i(t_i) > 0$ and $\alpha_i(s_i|t_i) \geq 0$, $\forall i \in \{1,2\}$, $\forall s_i \in T_i$, $\forall t_i \in T_i$;*

$$\begin{aligned} &\Big(\big(\lambda_i(t_i) + \sum_{s_i \in T_i} \alpha_i(s_i|t_i)\big)\omega_i(t_i) - \sum_{s_i \in T_i} \alpha_i(t_i|s_i)\omega_i(s_i)\Big)/p_i(t_i) \\ &= \sum_{t_{-i} \in T_{-i}} p_{-i}(t_{-i}) \max_{(q,y) \in \mathcal{D}} \sum_{j \in \{1,2\}} \frac{v_j\big((q,y), t, \lambda, \alpha\big)}{2}, \ \forall i \in \{1,2\}, \forall t_i \in T_i; \end{aligned} \tag{1}$$

$$U_i(Q, Y|t_i) \geq \omega_i(t_i) - \varepsilon, \ \forall i \in \{1,2\}, \forall t_i \in T_i; \tag{2}$$

where

$$\begin{aligned} &v_i\big((q,y), t, \lambda, \alpha\big) \\ &= \Big(\big(\lambda_i(t_i) + \sum_{s_i \in T_i} \alpha_i(s_i|t_i)\big)u_i((q,y), t) - \sum_{s_i \in T_i} \alpha_i(t_i|s_i)u_i((q,y), (t_{-i}, s_i))\Big)/p_i(t_i) \end{aligned} \tag{3}$$

This lemma offers the tractable set of conditions for computing neutral public good mechanisms. A neutral mechanism can be characterized as an incentive-feasible

mechanism that is not only interim incentive efficient in terms of actual utility payoffs but also both efficient and equitable in terms of transferable virtual-utility payoffs; where a virtual-utility payoff v_i is defined in eq. (3) by taking into account the shadow price of the incentive constraints. So each v_i exaggerates the difference from the types that want to pretend to be player i's type. Conditions (1) and (2) establish that the neutral mechanism maximizes the sum of the players' transferable virtual-utility payoffs and allocates the total transferable payoff equally among the players in every state of types; and it gives each player a real expected utility that is at least as large as the limit of virtually equitable allocations for each type, where a virtually equitable allocation ω_i balances out conflicting goals of different possible types of player i.

3 Why neutral mechanisms?

In this section, I show that neutral mechanisms are both predictively and prescriptively appealing from efficiency and equity standpoints for the production of public goods, although ex ante or interim incentive efficient mechanisms may be deemed desirable on their own merits.

3.1 Ex ante efficient mechanisms

In many practical settings, not producing the public good can be inefficient in the sense that it leads to social welfare reduction. In terms of the model specification, such problems have utilities that satisfy $\sum_i u_i((q, y), t) \geq 0, \ \forall t \in T$. In this class of public good problems, the mechanism that minimizes the chance of no production appears to be a natural choice that the players can agree on.

Proposition 1 *For the class of public goods problems in which the provision of the public good is ex ante efficient, the mechanism that minimizes the ex ante probability of no production is equivalent to the ex ante incentive efficient mechanism.*

Proof. In my setting, the ex ante probability of no provision under mechanism (Q, Y) is $\sum_{t\in T} p(t)(1-Q(t))$, where $p(t) = \sum_i p_i(t_i)$. Player i's ex ante expected utility can be written as $\sum_{t\in T} p(t)\big(t_i Q(t) - Y_i(t)\big)$. Then the optimization problem of maximizing the ex ante expected utility differs from that of minimizing the ex ante probability of no provision only by a positive linear transformation. □

For the ex ante incentive efficient mechanism to be a reasonable prediction that the players would use to help them make public goods decisions, the implied assumption should be that the mechanism is selected at the ex ante stage to be implemented at the interim stage, or that the mechanism is chosen and enforced by an external planner whose goal is to minimize the ex ante probability of no production. That is, invoking the ex ante incentive efficient mechanism relies on the strong assumption that the players must be able to commit themselves to the chosen mechanism ex ante.

While this assumption may be valid in some practical settings, there exists another conceptual issue in assuming that players retreat behind the veil of ignorance to choose a mechanism. What if the players are no longer truly ignorant at the time when they meet to agree on a mechanism? Would they still select an ex ante incentive efficient mechanism? The ex ante incentive efficient mechanism may be sensitive to the assumption that the players are absolutely certain that nobody has any private information.

To investigate this, I consider a perturbation of the ex ante information structure at the mechanism selection stage. At the moment when players meet initially to decide on a mechanism, each player has already received his private information t_i with some

probability, independently of the other player. I say that mechanism selection is at the *almost ex ante* stage. Formally, I assume that at the almost ex ante stage of mechanism selection each player has probability $\varepsilon \in (0,1)$ of having learned his type, and a complementary probability, $1-\varepsilon$, of still waiting to learn his type. Then for any $t_i \in T_i$, $\varepsilon p_i(t_i)$ is the probability that player i already knows his type and the type is t_i, and $(1-\varepsilon)p_i(t_i)$ is the probability that player i does not know his type but is expected to be of type t_i, as would be assessed by player $-i$. This paper's results do not depend on the assumption of type-independent probability of being informed, which is only for simplicity.[10]

Implementation of the selected mechanism takes place at the standard interim stage, when every player has received his private information (but does not know the other's information). The players cannot pre-commit themselves to report their types honestly and not to force the conflict outcome in implementing the selected mechanism after every player has learned his type. Therefore, players should choose among the set of available mechanisms that are subject to the feasibility constraints, as is assumed in Myerson's works. I assume that all feasible mechanisms for a given public good problem are available to players at the selection stage.

In this perturbed setting, the public good mechanism that is expected to arise would be, minimally, incentive efficient in the appropriate sense. The proper concept of efficiency must be based on the players' evaluations of the anticipated effects of feasible mechanisms. How a player should evaluate a mechanism depends crucially on what information, if any, he possesses at the time of mechanism selection. In

[10]For example, let $\varepsilon_i(t_i)$ denote the conditional probability that player i will be informed of his type if he were of type t_i, for each t_i of player i. Then $\varepsilon_i(t_i)p_i(t_i)$ is the probability that player $-i$ would assign to the event that player i is informed and is type t_i, and $(1-\varepsilon_i(t_i))p_i(t_i)$ is the probability that player $-i$ would assign to the event that player i is uninformed but will be type t_i. Note that the marginal probabilities of player i being informed and uninformed are respectively $\sum_{t_i}\varepsilon_i(t_i)p_i(t_i)$ and $1-\sum_{t_i}\varepsilon_i(t_i)p_i(t_i)$. All of the results would hold under this specification.

my setting, each player may or may not have learned his private information at the almost ex ante stage of selection. For a player who has received private information about his type, mechanisms are evaluated according to his interim preferences. For a player who does not possess any private information, mechanisms are evaluated according his ex ante preferences. The efficient choice of a mechanism at the almost ex ante stage must then be characterized based on all levels of possible interim and ex ante expected utilities for players.

Definition 1. A mechanism (Q, Y) is *almost ex ante incentive efficient* (AAIE) if and only if (Q, Y) is feasible and there does not exist another feasible mechanism (Q', Y') such that $U_i(Q', Y'|t_i) \geq U_i(Q, Y|t_i)$, for all i and for all t_i, and $\sum_{t_i \in T_i} p_i(t_i) U_i(Q', Y'|t_i) \geq \sum_{t_i \in T_i} p_i(t_i) U_i(Q, Y|t_i)$, for all i, with at least one strict inequality.

The almost ex ante notion of incentive efficiency in Definition 1 is a version of Pareto efficiency concepts on the set of feasible mechanisms, the taxonomy for which is developed by Holmström and Myerson (1983). They let Δ_I^* denote the set of mechanisms that are interim incentive efficient (IIE). I similarly denote the set of AAIE mechanisms by Δ_{AA}^*, which delimits the set of mechanisms that the players could reasonably consider at the almost ex ante stage of mechanism selection. The following equivalence result entails a complete characterization of the set of AAIE mechanisms.[11]

Theorem 1 *The notion of almost ex ante incentive efficiency is equivalent to the notion of interim incentive efficiency:* $\Delta_{AA}^* = \Delta_I^*$.

Proof. Relative to the interim notion of incentive efficiency, Definition 1 has an additional inequality to be satisfied for mechanism (Q', Y') to dominate mechanism

[11] The equivalence holds true on any set of classically feasible mechanisms, not just on the set of incentive feasible ones.

(Q, Y) with respect to uninformed player i's expected utility. For any given mechanism, for each i, uninformed player i's expected utility is simply a weighted average of his interim utilities of all possible types. So $U_i(\mu'|t_i) \geq U_i(\mu|t_i),\ \forall t_i, \forall i$ implies $\sum_{t_i \in T_i} p_i(t_i) U_i(Q', Y'|t_i) \geq \sum_{t_i \in T_i} p_i(t_i) U_i(Q, Y|t_i),\ \forall i$. □

This result can be justified simply by an intuitive reasoning without making recourse to technical proofs. At the almost ex ante stage, each player privately knows his type with probability $\varepsilon \in (0, 1)$. An uninformed player knows that he has yet to learn his type, and his opponent would assign probability $1-\varepsilon$ to this event. Whether a player has observed private information about his type or not is also private information for the player. That is, there are effectively $|T_i|+1$ number of privately known types of player i at the time of mechanism selection: the t_i type for all $t_i \in T_i$ and the "uninformed" type. But it is common knowledge that at the implementation stage every player will exactly know his type t_i. Because any player's expected utility in implementing a mechanism depends on the players' true types, player $-i$ would assign probability $\varepsilon p_i(t_i) + (1-\varepsilon) p_i(t_i) = p_i(t_i)$ to the event that t_i is the true type of player i, regardless of whether player i is informed or not at the selection stage. Thus, the almost ex ante stage becomes essentially identical to the interim stage with an "extended" type set where players have the same probabilistic beliefs over t_i-types as they would have at the usual interim stage.

Holmström and Myerson (1983) show that ex ante incentive efficiency implies interim incentive efficiency. With Δ_A^* denoting the set of ex ante incentive efficient mechanisms, Theorem 1 has an immediate corollary.

Corollary 1 *Ex ante incentive efficiency implies almost ex ante incentive efficiency:* $\Delta_A^* \subseteq \Delta_{AA}^*$.

The equivalence result and the corollary hold for any $\varepsilon \in (0, 1)$. The case of $\varepsilon = 0$

corresponds to situations where the mechanism selection is made *ex ante*, before any player's type is specified. In this case, Holmström and Myerson (1983) suggest that the efficient choice of a mechanism will be from among the set Δ_A^*. If there were some chance that a player may have learned his type at the time of selection, even if that chance were vanishingly small, the set of incentive efficient mechanisms that are implementable and reasonable for the players to choose would be enlarged.

Although the demonstration of the results and the underlying intuition are quite simple, their economic significance is large. When the public good appears to be worth more than it costs, the ex ante incentive efficiency seems to be a desirable property that a public good mechanism should have from the perspective of a planner that is interested in minimizing the ex ante probability of production never occurring. In fact, an ex ante incentive efficient mechanism is interim incentive efficient; so from the perspective of players who themselves are choosing a mechanism that is to be implemented ex interim, the ex ante incentive efficient mechanism can be considered a reasonable selection. However, it is vulnerable to any perturbation of the ex ante informational structure at the mechanism selection stage. This result destroys the validity of ex ante incentive efficient mechanisms as the only reasonable choices for the players, and makes the interim notion of incentive efficiency the relevant solution concept.

3.2 Interim efficient mechanisms

Then what would be the reasonable mechanisms that one might expect to actually arise for the production of public goods and are insensitive to variations in the information structure? The simplest natural answer to this question is the whole set of interim incentive efficient mechanisms.

When the mechanism selection is made ex interim, the choice of a mechanism can be determined by an incomplete information bargaining solution (e.g., Harsanyi and Selten 1972; Myerson 1983, 1984*b*) applied to the set of mechanisms. Crawford (1985) shows one specification of the rules for bargaining over mechanisms that makes any IIE mechanism attainable when mechanism selection takes place at the interim stage. So as a minimal requirement, the players should be expected to choose from among the set Δ_I^*. Even if there is some chance that a player may not have learned his type at the time of selection, by Theorem 1, the players would still choose from the set Δ_I^*. That is, the set of IIE mechanisms is a set of mechanisms, which the players would reasonably consider, that is robust to any perturbation of the interim informational structure.

But one drawback of characterizing IIE mechanisms is that it identifies too large a set of attainable mechanisms in many settings. Hence, the consideration of the set of IIE mechanisms may not give practical or compelling predictions for public goods problems.

More importantly, when focusing on the set of IIE mechanisms, there is one evident informational issue that implicitly arises during the mechanism selection process. The feasible mechanism that is best for each player depends on whether he is informed or not, as well as on his type if he is informed. Therefore, when the players are discussing which mechanism to implement, demanding a particular IIE (or AAIE) mechanism might convey information about the player's type; even an uninformed player might be incorrectly identified as being of a certain type by his demand. In that case, the proposed mechanism may no longer be incentive compatible, or the players may refuse to participate. Hence, whether a player is informed or not and no matter what an informed player's type is, each player should maintain an inscrutable facade in the mechanism selection process (see Myerson (1983) for the inscrutability principle).

To do so, each player must make some sort of equitable compromise between what she really wants and what she might have wanted if her type had been different, due to the conflicting incentives of different types of the player. Even if a player is uninformed of his true type, he must also express an equitable compromise between all of his possible types.

We must then use an appropriate solution concept that captures the idea of this inscrutable intertype compromise, as well as to refine a possibly large set of IIE mechanisms and get a stronger prediction of public good mechanisms that one might expect to reasonably arise as an outcome of our public good problem. Fortunately, Myerson's neutral bargaining solution resolves both the multiplicity and informational issues that arise when using the concept of interim incentive efficiency.

3.3 Neutral mechanisms: implications

The previous results and discussions highlight several advantages of using neutral mechanisms rather than ex ante or interim incentive efficient mechanisms for the provision of the public good.

First, because neutral optimum by definition implies interim incentive efficiency, neutral mechanisms are also robust to alternative specifications of the information structure at the time of mechanism selection, as are interim incentive mechanisms;[12] whereas ex ante incentive efficient mechanisms are not robust.

[12]Note that the neutral bargaining solution Myerson (1984*b*) is defined for a class of problems where the information structures are the same at the selection stage as at the implementation stage. If the information structure at the selection stage is perturbed such that there is some chance that a player may not have learned his type, player i who is possibly uninformed can be treated as a player who is informed of being the "uninformed" type. Such uninformed-type player has probability $p_i(t_i)$ of being type t_i for each $t_i \in T_i$; so this player's deliberation of equitable intertype compromise subsumes t_i-player's intertype compromise deliberation for every $t_i \in T_i$. Therefore, the neutral bargaining solution's prescription should be the same for the perturbed settings as for the fully interim settings.

Second, unlike interim incentive efficient mechanisms, neutral mechanisms do not carry the issue of information leakage during the process of mechanism selection or bargaining. The player's demand of a neutral mechanism is accepted as independent of the player's type. The need for this inscrutability concern is taken care of by balancing out conflicting goals of different types, which is captured by the conditions in Lemma 1.

Further, relative to the concept of interim incentive efficiency, the concept of neutral optimum gives a stronger prediction of which mechanism should reasonably be chosen. While the set of interim incentive efficient mechanisms may be large, the neutral mechanism is essentially unique if Axiom 1 is satisfied. There is no general uniqueness theorem, but the neutral mechanism is shown to be unique for many class of symmetric problems given in Kim (2017, 2019) and Myerson (1984*b*, 1985, 1991).

Lastly, neutral mechanisms have the desirable properties of efficiency and equity that should be satisfied by a fair and reasonable bargaining solution. These properties, which incorporate the inscrutable intertype compromise concern, are demonstrated by the axioms that the neutral bargaining solution should satisfy or by the condition for virtual-equity in Lemma 1.

These features deliver important implications for the analysis of mechanism selection problems.

If ex ante mechanism selection is to be applied to real situations, then players must be absolutely certain that no one is informed of any relevant private information at the stage of mechanism selection. The ex ante incentive efficient mechanism is not robust to adding some uncertainty that some player may be informed of his type; and so it will lose its validity as a reasonable prediction when there is some possibility, even a very small one, that players are not truly ex ante with regard to their private information. Moreover, the players often seek the assistance of a mutually agreed-

upon mechanism to help reduce the risk of public good production never occurring that arises precisely because of information asymmetries. Thus, it is more plausible to assume that the players already have their private information at the time they make a decision about which mechanism to use.[13]

When mechanism selection takes place at the interim stage, and if the only concern is achieving Pareto efficiency, the players should be expected to reasonably choose from among the set of interim incentive efficient mechanisms. If achieving the property of fairness is also part of the concern, a neutral mechanism can be considered as a reasonable selection; such selection may not be an ex ante incentive efficient mechanism that would have been chosen had players selected at the ex ante stage.

Another set of implications concern the ex ante criterion that is used to evaluate the performance of mechanisms. The proper performance or welfare criterion to evaluate the selected mechanism depends on what information players possess at the time of selection. In terms of the ex ante measures, my result asserts that the almost ex ante solution may maximize neither the ex ante probability of provision nor the ex ante expected gains of the players. On one hand, the ex ante measures should be irrelevant when it comes to evaluating the performance of the interim (or almost ex ante) choice of mechanism. On the other hand, the result implies that when evaluating the performance of different mechanisms, it is important to distinguish between situations in which players are allowed to choose their mechanism and those in which they are not; also important is to carefully identify the informational environment that players face when they select a mechanism. Otherwise, ex ante efficiency can be seriously misleading as a welfare measure of the chosen mechanism even if uncertainty about whether players are informed or not is vanishingly small. Also, to

[13]Even if some player may still be waiting to learn his type at the time of selection, the interim incentive efficiency and neutral optimum concepts' prescriptions remain unchanged.

evaluate the mechanism's performance in terms of the ex ante probability of provision may understate the usefulness of the chosen mechanism. The selection of an interim incentive efficient or neutral mechanism may not maximize the ex ante probability of provision, yet it is Pareto efficient and will improve upon unmediated communication or no communication.

4 Conclusion

There is no generally accepted interim bargaining solution concept in the literature, but many bargaining situations take place under incomplete information such as mechanism selection problems for public goods environments. This paper provides a more solid grounding for the relevance of mechanism selection at the interim stage to actual public good problems and of Myerson's concept of neutral optimum to the process of mechanism selection in such problems.

The analysis of ex ante mechanism selection crucially depends on players having absolutely no doubt that all players are ignorant of their types. If that doubt exists, the players may play on each other's doubt. Hence, the result under the assumption of ex ante selection stage is not robust to a perturbation of the information structure at the selection stage. Further, the set of interim incentive efficient mechanisms can be large, generating infinite predictions of public good mechanisms; and any choice from that set is vulnerable to the possibility of information leakage. These results justify that neutral mechanisms are theoretically appealing and easy to use from a practical perspective in applications to public goods provision. They are computed by the tractable set of conditions with interpretations that are insightful, essentially admitting a unique prediction of which mechanism would reasonably arise.

References

Balkenborg, Dieter and Miltiadis Makris. 2015. "An Undominated Mechanism for a Class of Informed Principal Problems with Common Values." *Journal of Economic Theory* 157:918–958.

Bester, Helmut and Karl Wärneryd. 2006. "Conflict and the Social Contract." *Scandinavian Journal of Economics* 108(2):231–49.

Celik, Gorkem and Michael Peters. 2011. "Equilibrium Rejection of a Mechanism." *Games and Economic Behavior* 73(2):375–387.

Cramton, Peter C. and Thomas R. Palfrey. 1995. "Ratifiable Mechanisms: Learning from Disagreement." *Games and Economic Behavior* 10(2):255–283.

Crawford, Vincent P. 1985. "Efficient and Durable Decision Rules: A Reformulation." *Econometrica* 53(4):817–836.

de Clippel, Geoffroy and Enrico Minelli. 2004. "Two-Person Bargaining with Verifiable Information." *Journal of Mathematical Economics* 40:799–813.

Gresik, Thomas A. 1996. "Incentive-Efficient Equilibria of Two-Party Sealed-Bid Bargaining Games." *Journal of Economic Theory* 68:26–48.

Harsanyi, John C. 1967-8. "Games with Incomplete Information Played by 'Bayesian' Players." *Management Science* 14:159–189, 320–334, 348–502.

Harsanyi, John C and Reinhard Selten. 1972. "A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information." *Management Science* 18(5):80–106.

Holmström, Bengt and Roger B Myerson. 1983. "Efficient and Durable Decision Rules with Incomplete Information." *Econometrica* 51(6):1799–1819.

Hörner, Johannes, Massimo Morelli and Francesco Squintani. 2015. "Mediation and Peace." *Review of Economic Studies* 82(4):1483–1501.

Kim, Jin Yeub. 2017. "Interim Third-Party Selection in Bargaining." *Games and Economic Behavior* 102:645–665.

Kim, Jin Yeub. 2019. "Neutral Bargaining in Financial Over-The-Counter Markets." *AEA Papers and Proceedings* 109:539–544.

Kydd, Andrew. 2003. "Which Side Are You On? Bias, Credibility, and Mediation." *American Journal of Political Science* 47(4):597–611.

Laffont, Jean-Jacques and David Martimort. 2000. "Mechanism Design with Collusion and Correlation." *Econometrica* 68(2):309–342.

Lagunoff, Roger D. 1995. "Resilient Allocation Rules for Bilateral Trade." *Journal of Economic Theory* 66(2):463–487.

Ledyard, John O. and Thomas R. Palfrey. 1994. "Voting and Lottery Drafts as Efficient Public Goods Mechanisms." *Review of Economic Studies* 61:327–355.

Ledyard, John O. and Thomas R. Palfrey. 1999. "A Characterization of Interim Efficiency with Public Goods." *Econometrica* 67(2):435–448.

Ledyard, John O. and Thomas R. Palfrey. 2002. "The Approximation of Efficient Public Good Mechanisms by Simple Voting Schemes." *Journal of Public Economics* 83:153–171.

Ledyard, John O. and Thomas R. Palfrey. 2007. "A General Characterization of Interim Efficient Mechanisms for Indepedent Linear Environments." *Journal of Economic Theory* 133(1):441–466.

Liu, Qingmin, George J. Mailath, Andrew Postlewaite and Larry Samuelson. 2014. "Stable Matching with Incomplete Information." *Econometrica* 82(2):541–587.

Maskin, Eric and Jean Tirole. 1990. "The Principal-Agent Relationship with an Informed Principal: The Case of Private Values." *Econometrica* 58(2):379–409.

Maskin, Eric and Jean Tirole. 1992. "The Principal-Agent Relationship with an Informed Principal, II: Common Values." *Econometrica* 60(1):1–42.

Meirowitz, Adam, Massimo Morelli, Kristopher W. Ramsay and Francesco Squintani. 2017. "Dispute Resolution Institutions and Strategic Militarization." *Journal of Political Economy* Forthcoming.

Myerson, Roger B. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47(1):61–74.

Myerson, Roger B. 1983. "Mechanism Design by an Informed Principal." *Econometrica* 51(6):1767–1797.

Myerson, Roger B. 1984*a*. "Cooperative Games with Incomplete Information." *International Journal of Game Theory* 13(2):69–96.

Myerson, Roger B. 1984*b*. "Two-Person Bargaining Problems with Incomplete Information." *Econometrica* 52(2):461–488.

Myerson, Roger B. 1985. Analysis of Two Bargaining Problems with Incomplete Information. In *Game-theoretic Models of Bargaining*, ed. Alvin E. Roth. Cambridge University Press chapter 7, pp. 115–147.

Myerson, Roger B. 1991. *Game Theory: Analysis of Conflict*. Cambridge, M.A.: Harvard University Press.

Pomatto, Luciano. 2019. "Stable Matching under Forward-Induction Reasoning." Unpublished.

Wilson, Robert. 1985. "Incentive Efficiency of Double Auctions." *Econometrica* 53(5):1101–1115.