A Unique and Robust Social Contract: An Application to Negotiations with Probabilistic Conflicts

Jin Yeub Kim*

Abstract

This paper considers social contracts (or mechanisms) in negotiations with incomplete information in which an outside option is a probabilistic conflict and a peaceful agreement is ex ante efficient. Applications include partnership, labor-management bargaining, pretrial negotiations, and international negotiations. I compute the set of interim incentive efficient mechanisms, the ex ante incentive efficient mechanism, as well as the neutral bargaining solution. I numerically illustrate that the focus on the ex ante incentive efficient mechanism as the most reasonable prediction is not robust. This paper justifies the neutral bargaining solution as the unique, robust solution among all interim incentive efficient mechanisms.

JEL classification: C78; D74; D82; F51; J52.

Keywords: Negotiation; Social contracts; Incomplete information; Incentive efficiency; Neutral bargaining solution

Word Count: 6393

*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 also thank Dieter Balkenborg for helpful suggestions. This work was supported by the Yonsei University Future-Leading Research Grants (No. 2019-22-0187, 2020-22-0499).

1. Introduction

In many economic, political, or social interactions, agents bargain over the gains from some opportunity that they cannot exploit without reaching an agreement. For example, in a partnership, two parties bargain for the division of a given social surplus, otherwise resulting in partnership dissolution. A union and a firm negotiate the terms of wage contract; in case negotiations break down, a costly strike occurs that neither side want. A plaintiff and a defendant bargain over the settlement amount; if they fail to reach a private settlement, then they resort to expensive litigation. Two countries can reach a peaceful agreement about dividing a fixed amount of contestable resources, or else go to war.

In all those examples, information asymmetries typically exist and outside options are probabilistic conflicts. In such situations under incomplete information, agents may agree on some social contract, or bargaining mechanism, to help them reach agreements. To identify those mechanisms that are expected to reasonably arise in Bayesian environments, the concept of interim incentive efficiency in the sense of Holmström and Myerson (1983) can be applied. However, the set of interim incentive efficient mechanisms is large for many problems, and one might need a sharper solution concept to delimit reasonable predictions.

One natural concept might be ex ante incentive efficiency because the set of ex ante incentive efficient mechanisms is a subset of the set of interim incentive efficient mechanisms (Holmström and Myerson 1983). However, Kim (2020) shows that the (typically smaller) set of ex ante incentive efficient mechanisms is not robust to a perturbation of the information structure at the time of mechanism selection. Another possible concept is the neutral bargaining solution, proposed by Myerson (1984*b*),

which is shown to provide predictions that are sharper than interim incentive efficiency.

To illustrate those results and justify the neutral bargaining solution, this paper provides an example of conflict games with incomplete information in which a peaceful agreement is ex ante efficient. I explicitly characterize the set of interim incentive efficient mechanisms, and compute the unique ex ante incentive efficient mechanism and the unique neutral bargaining solution. In the example, the ex ante incentive efficient mechanism is associated with the highest ex ante probability of agreement, whereas the neutral bargaining solution has the lowest, among all interim incentive efficient mechanisms. I also illustrate that the ex ante incentive efficient mechanism is not robust by computing the enlarged set of incentive efficient mechanisms in a perturbed setting. Further, the neutral bargaining solution is the only one among all interim incentive efficient mechanisms that is invulnerable to the possibility of information leakage during the bargaining process.

This paper relates to the large body of literature on bargaining solution concepts and mechanism design problems for Bayesian environments; e.g., Harsanyi and Selten (1972), Maskin and Tirole (1990, 1992), Myerson (1983, 1984*a*,*b*), among many others. My paper also connects with the literature that characterizes interim incentive efficient mechanisms in Bayesian environments; e.g., Gresik (1996), Holmström and Myerson (1983), Wilson (1985), and a series of papers by Ledyard and Palfrey (1994, 1999, 2002, 2007). There are only a few papers that study Myerson's (1983; 1984*b*) neutral solutions (e.g., Balkenborg and Makris 2015; de Clippel and Minelli 2004); and the concept has seen few applications (e.g., Kim 2017, 2019, 2020).[1] In terms of the application to conflict games, this paper is most closely related to the works

[1]See Kim (2017) for a general application to conflict games and Kim (2019) for an application to financial over-the-counter markets. Kim (2020) investigates neutral public good mechanisms.

of Bester and Wärneryd (2006) and Hörner, Morelli and Squintani (2015). This paper has implications for the analysis of ex ante incentive efficient mechanisms, and adds to justifying the neutral bargaining solution as a reasonable solution concept for negotiations with incomplete information.

2. Bayesian Bargaining Problems

In this section, I recapitulate the general formulation of two-person Bayesian bargaining problem of Myerson (1984*b*).[2] Then I briefly review the concepts of incentive efficiency and neutral bargaining solution.

A two-person Bayesian bargaining problem Γ is defined as an object of the form

$$\Gamma = (D, d^*, T_1, T_2, u_1, u_2, p_1, p_2)\,.$$

The D is the set of collective decisions or feasible outcomes that the players can jointly choose among, and $d^* \in D$ is the conflict outcome that occurs in the absence of cooperation. For each $i \in \{1, 2\}$, T_i is the set of possible types for player i, u_i is player i's utility payoff function from $D \times T_1 \times T_2$ into $\mathbb{R}$, and p_i is the probability function that represents player i's beliefs about the other player's type as a function of his own type.

Let $T = T_1 \times T_2$ denote the set of all possible type combinations $t = (t_1, t_2)$. For mathematical convenience, D and T are assumed to be finite sets. Without loss of generality, utilities are normalized so that $u_i(d^*, t) = 0$ for all i and t. For simplification of formulas, I assume that the players' types are independent random variables under the common prior probability distribution $p \in \Delta(T)$. That is, if

[2]Starting with the seminal work of Harsanyi (1967-8), the concept of Bayesian bargaining problem was further developed by Harsanyi and Selten (1972), Myerson (1979, 1984*b*), and many others.

$\bar{p}_i(t_i)$ denotes the prior marginal probability that player i's type will be t_i, then the probability that some $t \in T$ will be the true combination of players' types is $p(t) = \prod_i \bar{p}_i(t_i)$ and the probability that player $-i$ would assign to the event that t_i is the actual type of player i is $\bar{p}_i(t_i)$. As a regularity condition, all types are assumed to have positive probability: $\bar{p}_i(t_i) > 0$ for all i and all $t_i \in T_i$.

A decision rule or mechanism for the Bayesian bargaining problem Γ specifies how the choice $d \in D$ should depend on the players' types $t \in T$. Formally, a mechanism is defined as a function $\mu : D \times T \to \mathbb{R}$ such that $\sum_{c \in D} \mu(c|t) = 1$ and $\mu(d|t) \geq 0$ for all $d \in D$, for all $t \in T$.

The implementation of a mechanism is restricted by two factors. First, the players' types are not verifiable, so any mechanism cannot be implemented unless the players are given incentives to reveal their types honestly. Let $U_i(\mu|t_i)$ denote the interim expected utility for player i in mechanism μ if his type is t_i and all players report their types honestly:

$$U_i(\mu|t_i) = \sum_{t_{-i} \in T_{-i}} \sum_{d \in D} \bar{p}_{-i}(t_{-i}) \mu(d|t) u_i(d, t).$$

Then a mechanism μ is incentive compatible if and only if

$$U_i(\mu|t_i) \geq \sum_{t_{-i} \in T_{-i}} \sum_{d \in D} \bar{p}_{-i}(t_{-i}) \mu(d|t_{-i}, s_i) u_i(d, t), \quad \forall i, \forall t_i \in T_i, \forall s_i \in T_i,$$

where the right-hand-side is the interim expected utility for player i in mechanism μ if his type were t_i but reported s_i while the other player remained honest.

Second, the conflict outcome occurs when the players fail to cooperate, and any player can force the conflict outcome whenever his expected utility in the mechanism is less than zero. So any mechanism cannot be implemented unless the players are given

incentives to participate obediently in the mechanism. A mechanism μ is individually rational if and only if

$$U_i(\mu|t_i) \geq 0, \quad \forall i, \forall t_i \in T_i.$$

Then a mechanism μ is defined to be *feasible* for the players in Γ if and only if μ is both incentive compatible and individually rational.

Given the set of feasible mechanisms, the concept of interim incentive efficiency (in the sense of Holmström and Myerson (1983)) can be applied to identify a set of mechanisms among which the players would reasonably choose from. A mechanism μ is *interim incentive efficient* (IIE) if μ is feasible and there does not exist another feasible mechanism μ' that interim (Pareto) dominates μ, i.e., $U_i(\mu'|t_i) \geq U_i(\mu|t_i)$ for all i and all t_i with at least one strict inequality. The set of IIE mechanisms is often quite large, so some other solution criterion may be used to refine a possibly large set of IIE mechanisms. In this paper, I consider the following two concepts: ex ante incentive efficiency and neutral bargaining solution.

A mechanism μ is *ex ante incentive efficient* (AIE) if μ is feasible and there does not exist another feasible mechanism μ' that ex ante (Pareto) dominates μ, i.e., $\sum_{t_i \in T_i} \bar{p}_i(t_i) U_i(\mu'|t_i) \geq \sum_{t_i \in T_i} \bar{p}_i(t_i) U_i(\mu|t_i)$ for all i with at least one strict inequality. Holmström and Myerson (1983) show that ex ante incentive efficiency implies interim incentive efficiency. With Δ_A^* and Δ_I^* denoting the sets of mechanisms that are respectively ex ante and interim incentive efficient, we have $\Delta_A^* \subseteq \Delta_I^*$.

Myerson (1984*b*) proposed the concept of neutral bargaining solution for two-person Bayesian bargaining problems. This concept is axiomatically derived; I omit the exposition of the axioms. Instead, the neutral bargaining solution can be characterized as an incentive-feasible mechanism that is not only efficient in terms of players' actual utilities but also both equitable and efficient in terms of players' virtual util-

ities. Importantly, the neutral bargaining solution captures the idea of equitable compromise between all possible types of each player, as well as between two players, in order for a player's true type not to be revealed during the bargaining process.

In this paper, I appeal to the well-known characterization theorem for computing neutral bargaining solutions, stated below without proof.

Theorem 1 (Myerson, 1991, Theorem 10.3.). *A mechanism μ is a neutral bargaining solution if and only if, for each positive number ε, there exist vectors λ, α, and ω (which may depend on ε) such that*

$$\begin{aligned}
&\Big(\big(\lambda_i(t_i) + \sum_{s_i \in T_i} \alpha_i(s_i|t_i)\big)\omega_i(t_i) - \sum_{s_i \in T_i} \alpha_i(t_i|s_i)\omega_i(s_i)\Big)/\bar{p}_i(t_i) \\
&= \sum_{t_{-i} \in T_{-i}} \bar{p}_{-i}(t_{-i}) \max_{d \in D} \sum_{j \in \{1,2\}} v_j(d, t, \lambda, \alpha)/2, \ \forall i \in N, \ \forall t_i \in T_i; \\
&\lambda_i(t_i) > 0 \text{ and } \alpha_i(s_i|t_i) \geq 0, \ \forall i \in N, \ \forall s_i \in T_i, \ \forall t_i \in T_i; \\
&\text{and } U_i(\mu|t_i) \geq \omega_i(t_i) - \varepsilon, \ \forall i \in N, \ \forall t_i \in T_i,
\end{aligned} \tag{1}$$

where

$$v_i(d, t, \lambda, \alpha) = \Big(\big(\lambda_i(t_i) + \sum_{s_i \in T_i} \alpha_i(s_i|t_i)\big)u_i(d, t) - \sum_{s_i \in T_i} \alpha_i(t_i|s_i)u_i(d, (t_{-i}, s_i))\Big)/\bar{p}_i(t_i). \tag{2}$$

I briefly explain the intuition behind the characterization. A virtual-utility payoff v_i, defined in (2) takes 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 in (1) guarantee that the neutral bargaining solution 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. Application: Conflict Games

In this section, I use the framework of Bayesian bargaining problems (with two bargaining outcomes and two private types) to analyse an application to negotiations, modelled as conflict games.

I consider a stylized conflict environment á la Bester and Wärneryd (2006) and Hörner, Morelli and Squintani (2015), which I review here. Two players (1 and 2) want as much as possible of a given surplus of size 1. Both players can agree to a peaceful split, or else an outright, probabilistic conflict occurs and shrinks the value of the surplus to $\theta < 1$.[3] Each player can be of type H or L, privately and independently drawn from the same distribution with probability q and $1 - q$ respectively. The player's type can be thought of as his resolve, potential, or strength in outright conflict, determining the probability of winning the fight for each player and thus the expected conflict payoffs. When the two players are of the same type, they have the same expected share of the remaining surplus in case of conflict, so each player's expected conflict payoff is $\theta/2$. When a type H player fights against an L type, the H-type player's expected share is $p > 1/2$ and the expected conflict payoff is $p\theta > 1/2$.

In this setting, Hörner, Morelli and Squintani (2015) consider (direct-revelation) mechanisms that determine the division of the surplus under a peaceful agreement and the probability of conflict, given type reports; and they compare the mediation and arbitration mechanisms that are feasible and that maximize the ex ante proba-

[3]Partnership dissolution, strikes in labor-management negotiations, litigations, legal disputes, and warfare are examples of outside options in negotiations that can be thought of as probabilistic conflicts.

bility of peaceful agreement. Because my focus is not on comparing mediation and arbitration, I simplify their model by considering mechanisms that recommend either an equal split or outright conflict; but I generalize by allowing two players to choose a mechanism from the set of feasible mechanisms.[4]

This bilateral conflict game can be formally represented as a Bayesian bargaining problem of the form Γ. Let $D = \{d_0, d_1\}$, $T_1 = T_2 = \{H, L\}$, $\bar{p}_i(H) = q$ and $\bar{p}_i(L) = 1 - q$ for all i, and the utility functions are given in Table 1. The outcomes in D are interpreted as follows: d_0 is the outcome of negotiation breakdown (or outright conflict), and d_1 is the outcome of agreement (or an equal split). The natural conflict outcome d^* for this problem is then d_0 because conflict occurs if the players cannot agree to a peaceful settlement.

Table 1: The players' utility payoffs (u_1, u_2) that depend on $d \in D$ and $t \in T_1 \times T_2$

	H, H	H, L	L, H	L, L
d_0	$(\theta/2, \theta/2)$	$(p\theta, (1-p)\theta)$	$((1-p)\theta, p\theta)$	$(\theta/2, \theta/2)$
d_1	$(1/2, 1/2)$	$(1/2, 1/2)$	$(1/2, 1/2)$	$(1/2, 1/2)$

I select the values for the parameters: $\theta = 8/10$, $q = 3/8$, $p = 6/8$. Then the utility payoffs are as follows.

Table 2: A numerical example

	H, H	H, L	L, H	L, L
d_0	$(2/5, 2/5)$	$(3/5, 1/5)$	$(1/5, 3/5)$	$(2/5, 2/5)$
d_1	$(1/2, 1/2)$	$(1/2, 1/2)$	$(1/2, 1/2)$	$(1/2, 1/2)$

Normalizing utilities so that $u_i(d_0, t) = 0$ for all i and t, the utility payoffs can be rewritten as in Table 3.

[4]My simplification of abstracting away from different split recommendations does not eliminate the informational incentives of the players that arise in Hörner, Morelli and Squintani's (2015) model.

Table 3: A numerical example (normalized)

	H, H	H, L	L, H	L, L
d_0	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$
d_1	$(1/10,1/10)$	$(-1/10,3/10)$	$(3/10,-1/10)$	$(1/10,1/10)$

A mechanism μ for this Bayesian bargaining problem specifies the probability of recommending an equal split (outcome d_1) or conflict (outcome d_0) given type reports. To simplify notation, I restrict attention to symmetric mechanisms and use the abbreviations

$$q_H = \mu(d_0|H,H),\ q_M = \mu(d_0|H,L) = \mu(d_0|L,H),\ \text{and}\ q_L = \mu(d_0|L,L)$$

where $0 \leq q_H, q_M, q_L \leq 1$ for a mechanism μ.

With this notation, the expected utility for a player of type L and H in mechanism $Q \equiv (q_H, q_M, q_L)$ are, respectively, written as:

$$\begin{aligned} U(Q|H) &= (3/8)(1-q_H)(1/10) + (5/8)(1-q_M)(-1/10), \\ U(Q|L) &= (3/8)(1-q_M)(3/10) + (5/8)(1-q_L)(1/10). \end{aligned} \tag{3}$$

Then feasible mechanisms are those that satisfy the following inequalities:

$$\begin{aligned} &(3/8)(1-q_H)(1/10) + (5/8)(1-q_M)(-1/10) \\ &\quad \geq (3/8)(1-q_M)(1/10) + (5/8)(1-q_L)(-1/10), \end{aligned} \tag{4}$$

$$\begin{aligned} &(3/8)(1-q_M)(3/10) + (5/8)(1-q_L)(1/10) \\ &\quad \geq (3/8)(1-q_H)(3/10) + (5/8)(1-q_M)(1/10), \end{aligned} \tag{5}$$

$$(3/8)(1-q_H)(1/10) + (5/8)(1-q_M)(-1/10) \geq 0, \tag{6}$$

$$(3/8)(1-q_M)(3/10) + (5/8)(1-q_L)(1/10) \geq 0. \tag{7}$$

The two inequalities in (4) and (5) are the type H incentive compatibility (H-IC) constraint and the type L incentive compatibility (L-IC) constraint, respectively; the two inequalities in (6) and (7) are the type H individual rationality (H-IR) constraint and the type L individual rationality (L-IR) constraint, respectively.

For this example, I first characterize the set of interim incentive efficient (IIE) mechanisms. Because the model is symmetric, I need not distinguish the identities of two players; it suffices to focus on the objective function and constraints for one player and thus omit the subscript i in what follows. A feasible mechanism (q_H, q_M, q_L) is IIE if and only if there exist some positive numbers $\lambda(H)$ and $\lambda(L)$ such that (q_H, q_M, q_L) is an optimal solution to the following primal problem:

$$\begin{aligned}\max_{(q_H,q_M,q_L)} \Big[&\lambda(H)\big((3/8)(1-q_H)(1/10)+(5/8)(1-q_M)(-1/10)\big)\\ &+\lambda(L)\big((3/8)(1-q_M)(3/10)+(5/8)(1-q_L)(1/10)\big)\Big] \qquad (8)\end{aligned}$$

subject to the constraints (4)–(7) where $0 \leq q_H, q_M, q_L \leq 1$.

The optimal solutions to (8) are characterized below.

Proposition 1. *The set of IIE mechanisms is*

$$\Delta_I^* = \big\{(q_H, q_M, q_L) | q_H = (4/9)q_M, q_M \in [6/11, 1], q_L = 0\big\}.$$

Proof. Let $\alpha(L|H)$ and $\alpha(H|L)$ denote the Lagrange multipliers for the H- and L-IC constraints respectively. First notice that setting $q_L = 0$ increases the value of the objective function only to relax the H-IC, L-IC, and L-IR constraints. Taking this into account, the Lagrangean function can be written as $1/10$ multiplied by the

following:

$$\begin{aligned}\lambda(H)\big[(3/8)(1-q_H)+(5/8)(1-q_M)(-1)\big]+\lambda(L)\big[(3/8)(1-q_M)(3)+(5/8)\big]\\ +\alpha(L|H)\big[(3/8)(1-q_H)+(5/8)(1-q_M)(-1)-\big((3/8)(1-q_M)+(5/8)(-1)\big)\big]\\ +\alpha(H|L)\big[(3/8)(1-q_M)(3)+(5/8)-\big((3/8)(1-q_H)(3)+(5/8)(1-q_M)\big)\big].\end{aligned}$$

This function can be simplified to:

$$(1-q_H)V_{HH}+(1-q_M)(V_{HL}+V_{LH})+V_{LL},$$

where

$$\begin{aligned}V_{HH}&\equiv(3/8)\big[\big(\lambda(H)+\alpha(L|H)\big)-\alpha(H|L)(3)\big],\\ V_{HL}&\equiv(5/8)\big[\big(\lambda(H)+\alpha(L|H)\big)(-1)-\alpha(H|L)\big],\\ V_{LH}&\equiv(3/8)\big[\big(\lambda(L)+\alpha(H|L)\big)(3)-\alpha(L|H)\big],\\ V_{LL}&\equiv(5/8)\big[\big(\lambda(L)+\alpha(H|L)\big)-\alpha(L|H)(-1)\big].\end{aligned}$$

Then the dual problem for λ can be written as:

$$\min_{\alpha}\big[\max\{V_{HH},0\}+\max\{V_{HL}+V_{LH},0\}+V_{LL}\big] \tag{9}$$

Applying Myerson's (1984*b*) Theorem 10.1 to my setting, a feasible mechanism is IIE iff there exist vectors λ and α such that $\lambda(H)>0$, $\lambda(L)>0$, $\alpha(L|H)\geq 0$, $\alpha(H|L)\geq 0$, $\alpha(L|H)\big[(3/8)(1-q_H)+(5/8)(1-q_M)(-1)-\big((3/8)(1-q_M)+(5/8)(-1)\big)\big]=0$, $\alpha(H|L)\big[(3/8)(1-q_M)(3)+(5/8)-\big((3/8)(1-q_H)(3)+(5/8)(1-q_M)\big)\big]=0$,

$$\begin{aligned}(1-q_H)V_{HH}&=\max\{V_{HH},0\},\\ (1-q_M)\big(V_{HL}+V_{LH}\big)&=\max\{V_{HL}+V_{LH},0\}.\end{aligned} \tag{10}$$

Without loss of generality, λ is normalized such that $\lambda(H) + \lambda(L) = 1$.

The H-IR constraint (6) can be rewritten as $5q_M \geq 2 + 3q_H$, which implies that $q_M > 0$. Also, the L-IC constraint (5) can be written as $9q_H \geq 4q_M$, which together with $q_M > 0$ implies that $q_H > 0$.

An IIE mechanism would set q_H as small as possible to bind the L-IC constraint. Thus it must be $q_H = (4/9)q_M < q_M$. Rewriting the H-IC constraint (4) gives $8q_M \geq 3q_H$, which is not binding because $q_M > q_H$. Thus, $\alpha(L|H) = 0$ and $\alpha(H|L) > 0$.

The lower bounds on q_M and q_H are simultaneously determined by the two binding H-IR and L-IC constraints: $5q_M = 2 + 3q_H$ and $9q_H = 4q_M$, which give $\underline{q}_M = 6/11$ and $\underline{q}_H = 8/33$. For any $q_M > 6/11$, the H-IR constraint is not binding, and the q_H is uniquely determined by the binding L-IC constraint $q_H = (4/9)q_M$ given q_M. For $q_M \in (6/11, 1)$ and $q_H \in (8/33, 1)$, to satisfy the conditions in (10), it must be $V_{HH} = 0$ and $V_{HL} + V_{LH} = 0$, which yield $\lambda(H) = 27/38$ and $\alpha(H|L) = (1/3)\lambda(H) = 9/38$. These parameters together with $\lambda(L) = 1 - \lambda(H)$ and $\alpha(L|H) = 0$ satisfy the conditions for mechanisms with $q_H = (4/9)q_M$, $q_M \in (6/11, 1)$, and $q_L = 0$ to be IIE. For $q_M = 1$, it is $q_H = 4/9$. The binding constraint is the same as before, so $\alpha(L|H) = 0$ and $\alpha(H|L) > 0$; any $\lambda(H) \geq 27/38$ so that $V_{HL} + V_{LH} \leq 0$, together with $\alpha(H|L) = (1/3)\lambda(H)$, satisfies the conditions in (10). For any $q_M < 6/11$, a mechanism is not feasible and thus not IIE. □

I now compute the set of ex ante incentive efficient (AIE) mechanisms. A feasible mechanism (q_H, q_M, q_L) is AIE if and only if it is an optimal solution to the problem

of maximizing the ex ante expected utility in mechanism (q_H, q_M, q_L):

$$\max_{(q_H,q_M,q_L)} \Big[(3/8)^2 (1-q_H)(1/10) + (3/8)(5/8)(1-q_M)(2/10) + (5/8)^2 (1-q_L)(1/10) \Big] \tag{11}$$

subject to the constraints (4)–(7) where $0 \leq q_H, q_M, q_L \leq 1$.

Proposition 2. *There is a unique AIE mechanism such that*

$$q_H = \frac{8}{33}, \;\; q_M = \frac{6}{11}, \;\; q_L = 0.$$

Let μ_A denote this solution, so $\Delta_A^* = \{\mu_A\}$.

Proof. First note that setting $q_L = 0$ increases the value of the objective function only to relax the H-IC, L-IC, and L-IR constraints. Then the L-IR constraint $(14 - 9q_M \geq 0)$ never binds for any q_M; the L-IC constraint $(9q_H - 4q_M \geq 0)$ must bind in the solution, or else one could decrease q_H thus increasing the value of the objective function without violating other constraints. Also, the H-IR constraint $(5q_M \geq 3q_H + 2)$ must bind in the solution, or else one could decrease q_M and make the L-IC constraint slack. Solving for q_H and q_M in the system defined by the binding L-IC and H-IR constraints yields a unique solution to the problem (11). □

Lastly, I apply Theorem 1 and obtain the following result.

Proposition 3. *There is a unique neutral bargaining solution such that*

$$q_H = \frac{4}{9}, \;\; q_M = 1, \;\; q_L = 0. \tag{12}$$

Let μ_N denote this solution, so $\Delta_N^* = \{\mu_N\}$.

Proof. Note that all IIE mechanisms satisfy $(27/38)U(\mu|H)+(11/38)U(\mu|L) = 5/152$. So all IIE mechanisms must be optimal solutions of the primal problem (8) for λ, where

$$\lambda(H) = 27/38,\ \lambda(L) = 11/38.$$

The optimal solution of the dual problem for λ is

$$\alpha(L|H) = 0,\ \alpha(H|L) = (1/3)\lambda(H).$$

So let us try the parameters $\lambda(H) = 27/38$ and so $\alpha(H|L) = 9/38$. With these parameters in (2), the virtual-utility payoffs are:

$$\begin{aligned}
v(d_1,(H,H),\lambda,\alpha) &= \big((27/38)u(d_1,(H,H)) - (9/38)u(d_1,(L,H))\big)/(3/8) = 0,\\
v(d_1,(H,L),\lambda,\alpha) &= \big((27/38)u(d_1,(H,L)) - (9/38)u(d_1,(L,L))\big)/(3/8) = -24/95,\\
v(d_1,(L,H),\lambda,\alpha) &= \big((11/38) + (9/38)\big)u(d_1,(L,H))/(5/8) = 24/95,\\
v(d_1,(L,L),\lambda,\alpha) &= \big((11/38) + (9/38)\big)u(d_1,(L,L))/(5/8) = 8/95,\\
v(d_0,t,\lambda,\alpha) &= 0,\ \forall t.
\end{aligned}$$

The "warrant" equations in (1) are:

$$\begin{aligned}
\big((27/38)\omega(H) - (9/38)\omega(L)\big)/(3/8) = 0,\\
\big((11/38 + 9/38)\omega(L)\big)/(5/8) = (5/8)(8/95),
\end{aligned}$$

which yield the unique solution: $\omega(H) = 1/48,\ \omega(L) = 1/16$. Among IIE mechanisms, $\mu_N \equiv (q_H = 4/9, q_M = 1, q_L = 0)$ is the only mechanism that gives $U(\cdot|H) \geq \omega(H)$ and $U(\cdot|L) \geq \omega(L)$. In fact, $U(\mu_N|H) = 1/48$ and $U(\mu_N|L) = 1/16$. By Theorem 1, μ_N is a unique neutral bargaining solution. □

4. Discussions

4.1. Comparison

An immediate observation of Propositions 1, 2, and 3 is that $\Delta_A^* = \{\mu_A\} \subset \Delta_I^*$ and $\Delta_N^* = \{\mu_N\} \subset \Delta_I^*$, while $\mu_A \neq \mu_N$. Table 4 summarizes the mechanism probabilities, the interim expected utilities for H and L types, the ex ante expected utility, as well as the ex ante probability of agreement in the two mechanisms μ_A and μ_N, between which there is a continuum of IIE mechanisms.

Table 4: The IIE Mechanisms

	q_H	q_M	q_L	$U(\cdot\|H)$	$U(\cdot\|L)$	$U(\cdot)$	$Pr(agreement)$
μ_A	8/33	6/11	0	0	5/44	25/352	500/704
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
μ_N	4/9	1	0	1/48	1/16	3/64	90/192

Note: For brevity, $U(\cdot|H)$ and $U(\cdot|L)$ denotes the interim expected utilities for H and L types, respectively; and $U(\cdot)$ denotes the ex ante expected utility.

Any mechanism with $q_M \in [6/11, 1]$, $q_H = (4/9)q_M$, and $q_L = 0$ is IIE. The set of IIE utility allocations is a line in $\mathbb{R}^2$ with end points $(U(\cdot|H), U(\cdot|L))$ as follows:

$$\left(0, \frac{5}{44}\right) \text{ and } \left(\frac{1}{48}, \frac{1}{16}\right). \tag{13}$$

The first of these allocations is implemented by using the unique AIE mechanism μ_A. The second of these allocations is implemented by using the unique neutral bargaining solution μ_N. One can see that the AIE mechanism μ_A is interim best for type L but interim worst for type H, and vice versa for the NBS μ_N, among all IIE mechanisms.

In terms of the mechanism probabilities as well as other properties, μ_A and μ_N

are the two extremes among all IIE mechanisms. All IIE mechanisms assign a strictly positive probability of conflict outcome for high-type dyads (H, H) and asymmetric type dyads (H, L) and (L, H). The μ_A has the lowest probability of recommending conflict to those dyads, whereas μ_N has the highest probability of recommending conflict. Therefore, the AIE mechanism μ_A is associated with the highest ex ante probability of agreement, whereas the neutral bargaining solution μ_N carries the lowest ex ante probability of agreement, among all IIE mechanisms.[5] The ex ante features are irrelevant for evaluating interim solution concepts, but μ_N happens to be ex ante worst for the players.

The comparison here does not depend on a specific numerical parameterization, but applies to any other example with the same payoff structure as the one considered here.

4.2. Robustness

When players are able to choose among feasible mechanisms, then the AIE mechanism may appear to be a natural choice that the players can agree on. For this mechanism to be the only reasonable selection, the players must be absolutely certain that nobody has any private information at the stage of mechanism selection. However, Kim (2020) shows that AIE mechanisms are not robust to a perturbation of such assumption in a class of problems in which an agreement is ex ante efficient. Here I illustrate this result in the example by explicitly characterizing the set of incentive efficient mechanisms in a perturbed setting.

[5]For the class of examples considered in this paper, the optimization problem of maximizing the ex ante probability of agreement differs from (11) only by a positive linear transformation. Hence μ_A is equivalent to the peace-maximizing mechanism, and μ_N is equivalent to the conflict-maximizing mechanism, among all IIE mechanisms. However, the maximization of the conflict probability or the minimization of the ex ante welfare is not the objective of the neutral bargaining solution, but rather an induced feature of the solution, as pointed out by Kim (2017).

Suppose that, at the moment when two players meet to decide on a mechanism, there is a 1 percent chance that each player has already received his private information, independently of the other player.[6] Kim (2020) calls this the almost ex ante stage of mechanism selection. At this stage, each player privately knows that his type is H with probability (1/100)(3/8) and L with probability (1/100)(5/8), as would be assessed by his opponent. An uninformed player knows that he has yet to learn his type, and his opponent would assign probability (99/100) to this event. Therefore, there are effectively three privately known "types" of each player at the time of mechanism selection; that is, type H, type L, and type U ("uninformed").

A set of incentive efficient mechanisms in this perturbed setting can be defined by applying the concept of Pareto efficiency. The proper concept of efficiency must be based on the players' evaluations of the anticipated effects of feasible mechanisms when implemented. For an informed player, mechanisms are evaluated according to his interim preferences, represented by the interim expected utilities in (3). For an uninformed player, mechanisms are evaluated according to his ex ante expected utility, $U^u(\cdot) \equiv (3/8)U(\cdot|H) + (5/8)U(\cdot|L)$.

Then a feasible mechanism (q_H, q_M, q_L) is almost ex ante incentive efficient (AAIE) if and only if there exist some positive numbers $\lambda(U)$ (independent of players' types),

[6]The illustration here is valid given any probability $\varepsilon \in (0, 1)$ of having learned the type. Also, the assumption of type-independent probability of being informed is only for simplicity.

$\lambda(H)$, and $\lambda(L)$ such that (q_H, q_M, q_L) is an optimal solution to the following problem:

$$\begin{aligned}\max_{(q_H,q_M,q_L)} \Bigg[& \lambda(U)\Big((3/8)^2 (1-q_H)(1/10) + (3/8)(5/8)(1-q_M)(2/10) \\ & + (5/8)^2 (1-q_L)(1/10)\Big) \\ & + \lambda(H)\Big((3/8)(1-q_H)(1/10) + (5/8)(1-q_M)(-1/10)\Big) \\ & + \lambda(L)\Big((3/8)(1-q_M)(3/10) + (5/8)(1-q_L)(1/10)\Big)\Bigg]\end{aligned} \tag{14}$$

subject to the constraints (4)–(7) where $0 \leq q_H, q_M, q_L \leq 1$.

By letting $(3/8)\lambda(U) + \lambda(H) = \hat{\lambda}(H)$ and $(5/8)\lambda(U) + \lambda(L) = \hat{\lambda}(L)$, the objective function in (14) is a linear transformation of the objective function in (8). Hence the solutions to (14) must be the same as the solutions to (8). So the set of AAIE mechanisms is equivalent to $\Delta_I^* = \big\{(q_H, q_M, q_L) | q_H = (4/9)q_M, q_M \in [6/11, 1], q_L = 0\big\}$. The set of AAIE utility allocations is a line in $\mathbb{R}^3$ with end points $(U(\cdot|H), U(\cdot|L), U^u(\cdot))$ as follows:

$$\left(0, \frac{5}{44}, \frac{25}{352}\right) \text{ and } \left(\frac{1}{48}, \frac{1}{16}, \frac{3}{64}\right). \tag{15}$$

This implies that if there were some chance that a player may have learned his type at the time of selection, even if that chance were small, the set of incentive efficient mechanisms that are implementable and reasonable for the players to choose would be enlarged. Hence, the focus on an ex ante incentive efficient mechanism as the most reasonable selection needs much justification.

Rather, the whole set of IIE mechanisms should be considered as reasonable predictions. However, a particular choice from that set might be vulnerable to the possibility of information leakage that implicitly arises during the mechanism selection stage. In fact, among the large set of IIE mechanisms, only μ_N survives the

informational leakage problem, which I explain below.

At the selection stage, each player may have uncertainty over whether the other player has private information or not. But the players know that μ_N is best for type H but worst for type L and an uninformed player among all of IIE mechanisms (see (15), where the second allocation is achieved by μ_N). So when the players are discussing which mechanism to implement, if a player insists on any IIE mechanism other than μ_N, it could be taken as a signal of being type L even if that player might have been uninformed. The other player, if informed and of type H, will then be convinced to force the conflict outcome. Therefore, no player—whether type H or L, or uninformed—wants the other player to infer via his mechanism choice that he is of type L. In some sense, both an L type and an uninformed player would have an incentive to conceal the (possible) state of their information. Accordingly, these players would mimic type H by choosing whatever an informed type H player would have chosen. Even if there is only a fairly small probability $(1/100)(3/8)$ that a player already knows that he is type H, the effect created by the early informed type H player who wants to break off from any IIE mechanism other than μ_N is influential on the players' behavior when they bargain over mechanisms. Thus each player would bargain for the mechanism that is most favorable to the H type, which is μ_N.

5. Conclusion

The concept of incentive efficiency is clearly a minimal requirement for defining reasonable selections by players in bargaining or negotiation situations with incomplete information. In this paper, I explicitly characterize the set of interim incentive efficient mechanisms for an example of standard conflict games where a peaceful agreement is ex ant efficient. Among those mechanisms are the unique ex ante incentive efficient

mechanism and the unique neutral bargaining solution. The focus on the ex ante incentive efficient mechanism may seem appealing because it maximizes the change of agreement; however, such focus is not robust to a perturbation of the information structure at the selection stage. Then focusing on any mechanism in the larger set of interim incentive efficient mechanisms is reasonable if the only concern is achieving Pareto efficiency. Yet, among those interim incentive efficient mechanisms, the neutral bargaining solution is the only one that is invulnerable to the information leakage problem during bargaining. Therefore, this paper justifies the concept of the neutral bargaining solution as a unique, robust solution concept that can be applied to negotiations under incomplete information.

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.

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.

Kim, Jin Yeub. 2020. "Neutral Public Good Mechanisms." Working paper.

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.

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.

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

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