Efficient and Neutral Mechanisms in Almost Ex Ante Bargaining Problems*

Jin Yeub Kim†

Initiated February 2016; Revised June 16, 2020

Abstract

I consider two-person bargaining problems in which mechanism is selected at the *almost ex ante* stage—when there is some positive probability that players may have learned their private types—and the chosen mechanism is implemented at the interim stage. For these problems, I define almost ex ante incentive efficient mechanisms and apply the concept of neutral optima (Myerson 1984*b*). I show that those mechanisms may not be ex ante incentive efficient. This note suggests that ex ante incentive efficient mechanisms are not robust to a perturbation of the ex ante informational structure at the time of mechanism selection.

JEL classification: C71; C78; D74; D82

Keywords: Bargaining; Incomplete information; Mechanism selection; Almost ex ante stage; Incentive efficiency; Neutral optima

1. Introduction

In bargaining situations with incomplete information, disputing parties may agree on some decision rule or mechanism to help them reach agreements. If the parties can choose a mechanism before observing their private information, then they would

*This manuscript is a research note providing the basis of my working paper titled "Neutral Public Good Mechanisms."

†School of Economics, Yonsei University, 50 Yonsei-ro Seodaemun-gu, Seoul 03722, Republic of Korea (email: shiningjin@gmail.com).

reasonably agree on a mechanism that is ex ante incentive efficient (Holmström and Myerson 1983). The ex ante choice of this mechanism may not be implementable if the parties can renegotiate their mechanism once they learn their private information.[1] This problem can be avoided by assuming either that the parties can commit themselves to the mechanism ex ante, or that the chosen mechanism can be enforced by an external actor.

But there exists another conceptual issue in the case of ex ante mechanism selection where parties retreat behind the veil of ignorance to choose a mechanism. What if the parties 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 problem of ex ante mechanism selection may be sensitive to the assumption that the parties are absolutely certain that nobody has any private information. The goal of this note is to examine the robustness of the bargaining solutions for ex ante mechanism selection to a perturbation of this assumption.

Here I only mention three related lines of research, the review of which can be found in Kim (2020). First, this note connects with the literature that characterizes interim incentive efficient mechanisms in Bayesian environments; e.g., Gresik (1996), Wilson (1985), and a series of papers by Ledyard and Palfrey (1994, 1999, 2002, 2007). Second, this note relates to the large body of literature on bargaining solution concepts and mechanism design problems for Bayesian environments; e.g., among many others, Balkenborg and Makris (2015), de Clippel and Minelli (2004), Harsanyi and Selten (1972), Kim (2017, 2019), Maskin and Tirole (1990, 1992), Myerson (1983,

[1] If (re)negotiation takes place when parties already know their private information, some of their private information may be leaked in the process of negotiation. This information leakage problem has received much attention in the literature on mechanism design (e.g., Celik and Peters 2011; Cramton and Palfrey 1995; Crawford 1985; Holmström and Myerson 1983; Lagunoff 1995; Myerson 1983, 1984*b*). This note does not directly tackle commitment and renegotiation issues, but implicitly considers the possibility of information leakage during bargaining in the sense of Myerson (1983, 1984*b*).

1984a,b); as well as several other papers addressing 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), Lagunoff (1995), Liu et al. (2014), Pomatto (2019). Finally, this note connects with the conflict literature on institutional design; e.g., Bester and Wärneryd (2006), Hörner, Morelli and Squintani (2015), Kydd (2003), and Meirowitz et al. (2017).

The remainder of the note is organized as follows. Section 2 introduces the model of almost ex ante bargaining problems. Section 3 defines almost ex ante incentive efficient mechanisms and compares the notions of ex ante, almost ex ante, and interim incentive efficiency. Section 4 discusses neutral mechanisms for almost ex ante bargaining problems. Section 5 discusses the implications of my results for the analysis of ex ante and interim mechanism selections. Section 6 provides an example to illustrate the results, and Section 7 concludes.

2. Definition of Almost Ex Ante Bargaining Problems

2.1. The Basic Structure of Bayesian Bargaining

To describe bargaining situations with incomplete information, I briefly review the formulation of two-person Bayesian bargaining problem à la Myerson (1984b).[2] A two-person Bayesian bargaining problem Γ is defined as an object of the form

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

[2]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, 1984b). The present note also considers the fixed-threats case.

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 throughout the note, I assume that the players' types are independent random variables under the common prior probability distribution $p \in \Delta(T)$. That is, if $\bar{p}_i(t_i)$ 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. Second, any player can force the conflict outcome whenever his expected utility in the mechanism is less than zero. Hence, I restrict attention to mechanisms that are incentive compatible and individually rational in the sense of the following constraints:

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}}\sum_{d\in D}\bar{p}_{-i}(t_{-i})\mu(d|t)u_i(d,t) \\ &\geq \sum_{t_{-i}\in T_{-i}}\sum_{d\in D}\bar{p}_{-i}(t_{-i})\mu(d|t_{-i},s_i)u_i(d,t), \quad \forall i, \forall t_i\in T_i, \forall s_i \in T_i; \end{aligned} \tag{1}$$

$$\sum_{t_{-i}\in T_{-i}}\sum_{d\in D}\bar{p}_{-i}(t_{-i})\mu(d|t)u_i(d,t) \geq 0, \quad \forall i, \forall t_i \in T_i. \tag{2}$$

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.

2.2. Information Structure at the Mechanism Selection Stage

In many bargaining situations, players are able to bargain over available mechanisms and to negotiate with each other for the mechanism they want implemented. The stage of this bargaining over mechanisms—i.e., mechanism selection—would then precede the stage of implementation of the mechanism in full account of the bargaining process. Myerson (1983, 1984*b*) consider the problem of mechanism selection in games with incomplete information.[3] In both papers, information structure at the mechanism selection stage is the same as that at the implementation stage. My innovation is to consider a more general class of information structures during the bargaining process in the sense that information structures differ in the two stages.

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.[4]

[3] The first paper deals with mechanism selection by a principal with all of the bargaining power; the second paper is on mechanism selection by two players with equal power.

[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*

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 $t_i \in T_i$, $\varepsilon \bar{p}_i(t_i)$ is the probability that player i already knows his type and the type is t_i, and $(1-\varepsilon)\bar{p}_i(t_i)$ 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 note's results do not depend on the assumption of type-independent probability of being informed, which is only for simplicity.[5]

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 bargaining problem Γ are available to players at the selection stage. I refer to the bargaining problem in which a mechanism is selected at the almost ex ante stage and is implemented at the interim stage as an *almost ex ante bargaining problem.*

In this note, I do not model the bargaining process as a noncooperative strategic game, leaving the actual structure of mechanism selection stage implicit. Rather, I take the cooperative approach to determine the mechanisms that the players should "reasonably" choose according to two solution criteria: incentive efficiency (Sect. 3)

stage.

[5] 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 $\varepsilon_i(t_i)\bar{p}_i(t_i)$ is the probability that player $-i$ would assign to the event that player i is informed and is type t_i, and $(1-\varepsilon_i(t_i))\bar{p}_i(t_i)$ 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 $\sum_{t_i} \varepsilon_i(t_i)\bar{p}_i(t_i)$ and $1-\sum_{t_i} \varepsilon_i(t_i)\bar{p}_i(t_i)$. All of the results would hold under this specification.

and neutral optima (Sect. 4).[6]

Before proceeding with the analysis, I define two benchmark cases. The case of $\varepsilon = 0$ characterizes mechanism selection at the ex ante stage, before players have received any private information. The case of $\varepsilon = 1$ describes the situation in which mechanism selection takes place at the interim stage. The almost ex ante stage, $\varepsilon \in (0, 1)$, can then be interpreted as a perturbation of mechanism selection taking place either "absolutely" ex ante or "absolutely" interim. The solutions for the two benchmark cases will be compared to the almost ex ante solutions in Sect. 3 and 4, and further illustrated in a simple example in Sect. 6.

3. Efficient Mechanisms in Almost Ex Ante Bargaining Problems

Given the set of feasible mechanisms, I can identify a set of mechanisms among which the players would choose from 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. How a player should evaluate a mechanism depends crucially on what information, if any, he possesses at the time of mechanism selection. In 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. The interim evaluation of a mechanism

[6] The concept of neutral optima or neutral bargaining solution is proposed by Myerson (1983, 1984*b*) for interim bargaining problems where the information structure is the same at both selection and implementation stages. Section 4 explains this solution concept in detail and delimits it for almost ex ante bargaining problems.

μ by informed player i given that he is of type t_i is

$$\sum_{t_{-i}\in T_{-i}} \varepsilon\bar{p}_{-i}(t_{-i}) \sum_{d\in D} \mu(d|t)u_i(d,t) + (1-\varepsilon)\Big[\sum_{t_{-i}\in T_{-i}} \bar{p}_{-i}(t_{-i}) \sum_{d\in D} \mu(d|t)u_i(d,t)\Big],$$

which reduces to

$$U_i(\mu|t_i) \equiv \sum_{t_{-i}\in T_{-i}} \sum_{d\in D} \bar{p}_{-i}(t_{-i})\mu(d|t)u_i(d,t) \tag{3}$$

where the expected utility that is conditioned on type t_i denotes that it is for player i who is informed and is of type t_i.

For a player who does not possess any private information, mechanisms are evaluated according his ex ante preferences. The ex ante evaluation of a mechanism μ by uninformed player i is

$$\begin{aligned}\sum_{t_i\in T_i} \bar{p}_i(t_i)\Big[&\sum_{t_{-i}\in T_{-i}} \varepsilon\bar{p}_{-i}(t_{-i}) \sum_{d\in D} \mu(d|t)u_i(d,t)\\ &+ (1-\varepsilon)\Big[\sum_{t_{-i}\in T_{-i}} \bar{p}_{-i}(t_{-i}) \sum_{d\in D} \mu(d|t)u_i(d,t)\Big]\Big],\end{aligned}$$

which reduces to

$$U_i^u(\mu) \equiv \sum_{t_i\in T_i} \bar{p}_i(t_i) \sum_{t_{-i}\in T_{-i}} \sum_{d\in D} \bar{p}_{-i}(t_{-i})\mu(d|t)u_i(d,t) \tag{4}$$

where the expected utility with superscript u denotes that it is for player i who is uninformed.

The efficient choice of a mechanism at the almost ex ante stage must then be characterized based on all levels of $U_i^u(\mu)$ and $U_i(\mu|t_i)$, for all t_i, for all i.

Definition 1. A mechanism μ is *almost ex ante incentive efficient* (AAIE) if and only if μ is feasible and there does not exist another feasible mechanism μ' such that

$$U_i^u(\mu') \geq U_i^u(\mu) \text{ and } U_i(\mu'|t_i) \geq U_i(\mu|t_i) \;\; \forall t_i \in T_i, \;\; \forall i$$

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.

Relative to the interim notion of incentive efficiency, Definition 1 has an additional inequality to be satisfied for mechanism μ' to dominate mechanism μ 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 $U_i(\mu'|t_i) \geq U_i(\mu|t_i)$, $\forall t_i \in T_i, \forall i$ implies $U_i^u(\mu') \geq U_i^u(\mu)$, $\forall i$. These observations establish the following equivalence result.[7]

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

The intuitive reason for the equivalence is as follows. 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 $|T_i|+1$

[7] The equivalence holds true on any set of classically feasible mechanisms, not just on the set of incentive feasible ones. The formal characterization theorem is referred to Appendix A.

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 $\varepsilon\bar{p}_i(t_i) + (1-\varepsilon)\bar{p}_i(t_i) = \bar{p}_i(t_i)$ 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 $\varepsilon \in (0,1)$. Although the demonstration of the two results is immediate and the underlying intuition is quite simple, their economic significance is large. Even when the players are very likely to be uninformed at the mechanism selection stage, an ε probability that a player has already observed his type makes the interim notion of incentive efficiency the relevant solution concept.

4. Neutral Mechanisms in Almost Ex Ante Bargaining Problems

The criterion of Pareto efficiency suggests, in a normative sense, that players in almost ex ante bargaining problems should bargain for mechanisms that incorporate efficient aggregation of their interim and ex ante preferences over possible feasible mechanisms. However, those AAIE mechanisms (or, equivalently, IIE mechanisms) are not generally unique, and the set of AAIE mechanisms may be quite large. We may then use some other solution criterion to refine a possibly large set of AAIE mechanisms, so as to get a stronger prediction of mechanism selection at the almost ex ante stage.

Importantly, any additional criterion shall deal with one evident informational issue that implicitly arises during the mechanism selection stage in almost ex ante bargaining problems. 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 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 whatever an informed player's type may be, each player should maintain an inscrutable facade in the mechanism selection process. To do so, each player must make some equitable compromise among the conflicting preferences of alternative types. Even if a player is uninformed of his true type, he must also express an equitable compromise between all of his possible types.

So my next task is to find an appropriate cooperative solution concept that captures the idea of this inscrutable intertype compromise. Fortunately, Myerson (1983) first raised the substantively same issue, and proposed the principle of inscrutability for Bayesian mechanism design and selection problems. In a closely related paper, Myerson (1984*b*) axiomatically defined the concept of neutral bargaining solution for two-person Bayesian bargaining problems to resolve the issue. Formally, 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; these axioms generalize the axioms of Nash bargaining solution for games with incomplete information.[8]

Scrutinizing the axioms is not the primary goal of the present note; so without loss of comprehension of the solution concept, I leave aside the axioms and instead appeal to a useful and tractable characterization in terms of efficiency and virtual-equity properties, developed by Myerson (1984*b*). A neutral bargaining solution can be characterized as an incentive-feasible mechanism that is not only IIE in terms of actual utility payoffs but also both efficient and equitable in terms of transferable virtual-utility payoffs. A player's virtual-utility payoff is defined by taking into account the shadow price of the incentive constraints, so it exaggerates the difference from the types that want to pretend to be the player's type. More precisely, 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. Such solution can be considered a fair bargaining solution. What is essential is that the notion of virtual equity captures exactly the inscrutable intertype compromise concern that is also present in my almost ex ante bargaining problems.

Note that the neutral bargaining solution is defined for a class of problems where the information structures are the same at the selection stage as at the implementation stage. In such fully interim bargaining problems, the consideration of informational issue with respect to an uninformed player is irrelevant because every player

[8]Detailed expositions of the axioms can be found in Myerson (1984*b*), which establishes the existence of the neutral bargaining solutions for any finite two-person Bayesian bargaining problem.

is informed of his own type. In my almost ex ante bargaining problems where the information structures are different at the two stages, player i who is uninformed at the almost ex ante selection stage can be treated as a player who is informed of being the uninformed type. Such uninformed-type player has probability $\bar{p}_i(t_i)$ 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 two bargaining problems have the substantively identical informational issue, and so the neutral bargaining solution's prescription should be the same for almost ex ante bargaining problems as for interim bargaining problems.

Without any need or imperative to formally define "almost ex ante neutrality," I can apply the solution concept of interim neutrality to characterize neutral mechanisms in almost ex ante bargaining problems.[9] Further, when mechanism selection takes place at the fully ex ante stage, a player's demand for a particular feasible mechanism does not convey anything about his private information because he has absolutely no information to reveal, misrepresent, or conceal in the first place. The informational issue that arises at the almost ex ante or interim selection stage does not exist at the ex ante selection stage; hence, the concept of neutrality is inapplicable and irrelevant to ex ante mechanism selection.

Neutrality by definition implies almost ex ante incentive efficiency. Hence, relative to the Pareto efficiency concept, the concept of neutrality gives a stronger prediction of which mechanisms should reasonably be chosen as cooperative solutions in a two-person bargaining problem Γ when mechanism selection is at the almost ex ante stage. Although there is no general uniqueness theorem, the neutral mechanism is computed to be unique in many examples of symmetric trading and bargaining problems given in Myerson (1984*b*, 1985, 1991) as well as in the example given in Sect. 6 of this note.

[9] The formal characterization is given in Appendix B.

5. Implications

The results in this note have two prescriptive implications for the benchmark cases of $\varepsilon = 0$ and $\varepsilon = 1$.

In the case of $\varepsilon = 0$, the selection is made *ex ante*, before any player's type is specified, and the implementation takes place ex interim, after each player is given his own type but not given the other's type. 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. Hence, any perturbation of the ex ante informational structure at the mechanism selection stage destroys the validity of ex ante incentive efficient mechanisms as the only reasonable choices for the players.

In the case of $\varepsilon = 1$ where the 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, 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.

Comparing the sets of almost ex ante incentive efficient and neutral mechanisms to the sets of ex ante and interim incentive efficient mechanisms deliver implications for the analysis of mechanism selection problems.

If the only concern is achieving Pareto efficiency, one can expect that when mechanism selection takes place at the almost ex ante stage, the players will choose from among the larger set of AAIE mechanisms Δ^*_{AA} than the set of ex ante incentive efficient mechanisms Δ^*_A. If one accepts the equity notion built into the bargaining solution, the players should be expected to reasonably consider the set of neutral mechanisms, denoted by Δ^*_N, that is contained in Δ^*_{AA}.

While $\Delta^*_A \subseteq \Delta^*_{AA}$ and $\Delta^*_N \subseteq \Delta^*_{AA}$, the two sets Δ^*_A and Δ^*_N are not generally or necessarily equivalent. Section 6 illustrates an example for which the sets Δ^*_A and Δ^*_N are singletons and $\Delta^*_A \cap \Delta^*_N$ is empty. When there is some uncertainty about whether players are truly ex ante with regard to their private information, 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. This questions the tenability of ex ante incentive efficient mechanisms as the solutions that are robust to a perturbation of the ex ante information structure at the mechanism selection stage by adding some possibility, even a very small one, that some player may be informed of his type.

Another set of implications concern the ex ante criterion that is used in several studies to evaluate the performance of conflict resolution institutions (e.g., Bester and Wärneryd 2006; Hörner, Morelli and Squintani 2015; Kydd 2003; Meirowitz et al. 2017). The proper performance or welfare criterion to evaluate the selected mechanism depends on what information players possess at the time of selection. If the mechanism is selected ex ante, then the performance of the mechanism in reducing the possibility of conflict can be measured by the ex ante probability of peace, and the appropriate welfare criterion is ex ante incentive efficiency, which uses the two components of $\big(U_i^u(\mu)\big)_{i\in\{1,2\}}$ as the relevant utility measures.[10] But if the mechanism is selected almost ex ante, then the right welfare criterion must be almost ex ante incentive efficiency, which is based on consideration of all $\sum_i(|T_i|+1)$ components of $\big(U_i^u(\mu),(U_i(\mu|t_i))_{t\in T_i}\big)_{i\in\{1,2\}}$ as the relevant utility measures.

In terms of the ex ante measures, my result asserts that the almost ex ante solution may maximize neither the ex ante probability of peace nor the ex ante expected gains of the bargaining parties. On one hand, the ex ante measures should be irrelevant when it comes to evaluating the performance of the almost ex ante choice of mechanism. On the other hand, the result implies that when evaluating the performance of different institutions, such as mediation and arbitration, it is important to distinguish between situations in which parties are allowed to choose their institution and those in which they are not; also important is to carefully identify the informational environment that parties face when they select an institution. Otherwise, ex ante efficiency can be seriously misleading as a welfare measure of the chosen institution even if uncertainty about whether parties are informed or not is vanishingly small. Also, to evaluate the institution's performance in terms of the ex ante probability of peace may understate the usefulness of the chosen institution. The selection of an almost ex ante efficient or neutral mechanism may not maximize the ex ante probability of peace, yet it is Pareto efficient and will improve upon unmediated communication or no communication.

[10]In my setting, the performance measure in terms of the ex ante probability of peace is $\sum_{t\in T}p(t)\sum_{d\in D\setminus\{d^*\}}\mu(d|t)$. Assuming for simplicity that players are symmetric, the welfare measure in terms of player 1's ex ante expected utility can be written as $\sum_{t\in T}p(t)\sum_{d\in D\setminus\{d^*\}}\mu(d|t)u_1(d,t)$. In this formulation, I have excluded d^* from the sum over all d because utilities are normalized so that $u_1(d^*,t)=0$ for all $t\in T$. Then for a class of standard bilateral conflict problems in which peaceful agreement is ex ante efficient, the problem of maximizing the ex ante expected utility differs from that of maximizing the ex ante probability of peace only by a positive linear transformation; hence, the ex ante peace-maximizing mechanism coincides with the ex ante incentive efficient mechanism.

Further implications pertain to the interim bargaining mechanisms. If ex ante mechanism selection is to be applied to real bargaining situations, parties must be absolutely certain that neither is informed of any relevant private information. However, the parties often seek the assistance of a mutually agreed-upon mechanism to help reduce conflicts that arise precisely because of information asymmetries. Thus, it is more reasonable to assume that the parties may already have their private information at the time they select a mechanism. That the almost ex ante selection and the interim selection are the same implies the robustness of the interim selection of mechanisms. This note provides a more solid grounding for the relevance of mechanism selection at the interim stage to real bargaining problems, and further justifies Myerson's neutral mechanism as a powerful interim bargaining solution concept.

6. Example

To illustrate these ideas, consider a numerical example of Hörner, Morelli and Squintani's (2015) model of conflict. Two players (1 and 2) want as much as possible of a given cake of size 10. War can be initiated unilaterally and shrinks the value of the cake to 8 unless both players accept a peaceful split. Each player can be of type H or L, privately and independently drawn from the same distribution with probability 3/8 and 5/8 respectively. When the two players are of the same type, they have the same expected share of the remaining cake in case of war, so each player's expected war payoff is 4. When a type H fights against an L type, H-player's expected share is 6/8, and hence H-player expects 6 from war while L-player expects 2.

In this setting, Hörner, Morelli and Squintani (2015) consider mechanisms that recommend splits $(x, 10 - x)$ with some probability and war with complementary probability, given type reports; and they compare the mediation and arbitration

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	$(4, 4)$	$(6, 2)$	$(2, 6)$	$(4, 4)$
d_1	$(5, 5)$	$(5, 5)$	$(5, 5)$	$(5, 5)$

Table 2: Normalized 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	$(0, 0)$	$(0, 0)$	$(0, 0)$	$(0, 0)$
d_1	$(1, 1)$	$(-1, 3)$	$(3, -1)$	$(1, 1)$

mechanisms that are feasible and that maximize the ex ante probability of peace. Because my focus is not on comparing mediation and arbitration, I simplify their model by considering mechanisms that recommend either an equal split $(5, 5)$ or war; but I generalize by allowing two players to choose a mechanism from the set of feasible mechanisms.[11]

To formally model this example as a Bayesian bargaining problem of the form Γ, let $D = \{d_0, d_1\}$, $T_1 = T_2 = \{H, L\}$, $p_1(H) = p_2(H) = 3/8$, $p_1(L) = p_2(L) = 5/8$, with the utility functions given in Table 1. The outcomes in D are interpreted as follows: d_0 is the outcome of war, and d_1 is the outcome of an equal split. The natural conflict outcome d^* for this problem is d_0 because war occurs if the players cannot agree to an equal split. Normalizing utilities so that $u_i(d_0, t) = 0$ for all i and t, the utility functions can be rewritten as in Table 2.

To simplify notation, I restrict attention to symmetric mechanisms and use the abbreviations

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

[11] 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.

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

With this notation, the players choose a mechanism among all feasible mechanisms, which are those satisfying the following inequalities:

$$8q_M - 3q_H \geq 5q_L, \qquad 9q_H - 4q_M \geq 5q_L, \tag{5}$$

$$5q_M \geq 3q_H + 2, \qquad 14 - 9q_M \geq 5q_L. \tag{6}$$

The two inequalities in (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) are the type H individual rationality (H-IR) constraint and the type L individual rationality (L-IR) constraint, respectively. Note that 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 in what follows.

6.1. Benchmark Case of $\varepsilon = 0$

I first compute the efficient choice of a mechanism in the benchmark case of mechanism selection at the ex ante stage.

For the example given above, a feasible mechanism (q_H, q_M, q_L) is ex ante incentive efficient if and only if it is an optimal solution to the problem of maximizing the ex ante expected utility:

$$\begin{aligned} &\max_{(q_H,q_M,q_L)} \left[(3/8)^2 (1-q_H) + 2(3/8)(5/8)(1-q_M) + (5/8)^2 (1-q_L) \right] \\ &\text{subject to the constraints (5) and (6) where } 0 \leq q_H, q_M, q_L \leq 1. \end{aligned} \tag{7}$$

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 (7), which is

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

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

This μ_A corresponds to the feasible mechanism that maximizes the ex ante probability of peace, the program of which can be written exactly as (7) for this example. For other examples with the same payoff structure as the one considered here, the optimization problem of maximizing the ex ante expected utility differs from (7) only by a positive linear transformation. Hence the ex ante incentive efficient mechanism coincides with the peace-maximizing mechanism for the class of those examples.

6.2. Benchmark Case of $\varepsilon = 1$

For the benchmark case of mechanism selection at the interim stage, the concept of interim efficiency can be represented through restrictions on λ-weights in the weighted sum of the expected utilities of all types.

A feasible mechanism (q_H, q_M, q_L) is IIE if and only if there exist some positive numbers $\hat{\lambda}(H)$ and $\hat{\lambda}(L)$ such that (q_H, q_M, q_L) is an optimal solution to the following problem:

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

subject to the constraints (5) and (6) where $0 \le q_H, q_M, q_L \le 1$.

The optimal solutions to (8) can be characterized by using Proposition 2 of Kim (2017). The set of IIE mechanisms is

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

For the set of neutral mechanisms in this example, it can also be shown by Proposition 3 of Kim (2017) that there is a unique neutral mechanism, which is

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

Let μ_N denote this solution, so $\Delta_N^* = \{\mu_N\}$.[12]

6.3. Almost Ex Ante Incentive Efficient and Neutral Mechanisms

I now assume that players at the stage of mechanism selection are not absolutely sure that everyone is uninformed or informed.

At this almost ex ante stage of mechanism selection, a feasible mechanism (q_H, q_M, q_L) is AAIE if and only if there exist some positive numbers λ^u, $\lambda(H)$, and $\lambda(L)$ such

[12]See Appendix C for verifying the solution.

that (q_H, q_M, q_L) is an optimal solution to the following problem:

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

subject to the constraints (5) and (6) 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 (10) is a linear transformation of the objective function in (8). Hence the solutions to (10) must be the same as the solutions to (8), verifying Theorem 1 that $\Delta^*_{AA} = \Delta^*_I$. So the set of AAIE mechanisms is also

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

The set of AAIE utility allocations satisfying feasibility is a line in $\mathbb{R}^3$ with end points $(U(\mu|H), U(\mu|L), U^u(\mu))$ as follows:

$$\left(0, \frac{25}{22}, \frac{125}{176}\right) \text{ and } \left(\frac{5}{24}, \frac{5}{8}, \frac{15}{32}\right). \tag{11}$$

The first of these allocations is implemented by using the unique ex ante incentive efficient mechanism μ_A. The second of these allocations is implemented by using the unique neutral mechanism μ_N. Any mechanism with $q_M \in [6/11, 1]$, $q_H = (4/9)q_M$, and $q_L = 0$ is AAIE. So there is a continuum of AAIE mechanisms that fall in between μ_A and μ_N in terms of q_M.

An immediate observation is that $\Delta^*_A = \{\mu_A\}$ is a proper subset of Δ^*_{AA}; that is

$\Delta^*_A \subset \Delta^*_{AA}$. This result is a stronger version of Corollary 1. I may refine the large set Δ^*_{AA} by characterizing the set of neutral mechanisms Δ^*_N, which is $\Delta^*_N = \{\mu_N\}$.

To understand the solution, I invoke the logic of inscrutable intertype compromise during the selection process in the context of almost ex ante bargaining. At the almost ex ante stage, each player has uncertainty over whether the other player has private information. But players know that μ_N is best for a type H player but worst for both a type L and an uninformed player among all of the AAIE mechanisms (see the two end points of AAIE utility allocations, given in (11), where the second allocation is achieved by μ_N). In such a case, if a player insists heavily on μ_A, for example, it could be taken as a signal of being type L, regardless of whether that player is informed of type L or actually uninformed. The other player, if informed and of type H, will then be convinced to force the conflict outcome. Therefore, no player—whether H, 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 player and an uninformed player would have an incentive to conceal the (possible) state of their information. Accordingly, these players would mimic the H type by choosing whatever an informed H type player would have chosen. Even if ε is fairly small so that there is only a small probability $(3/8)\varepsilon$ that a player already knows that he is type H, the effect created by the early informed H type player who wants to break off from μ_A (or any 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.

For the example here, when mechanism selection takes place at the ex ante stage, the two players should be able to agree on the unique ex ante incentive efficient mechanism μ_A. When mechanism selection takes place at the almost ex ante stage, the normative concepts of Pareto efficiency and virtual equity together suggest that the two players will agree on the unique neutral mechanism μ_N. This strong prediction of mechanism selection at the almost ex ante stage is different from the uniquely efficient mechanism selection at the ex ante stage; that is, $\Delta_N^* = \{\mu_N\} \neq \{\mu_A\} = \Delta_A^*$ regardless of the size of $\varepsilon \in (0,1)$. This implies that once there is some uncertainty about whether players are truly uninformed of their types, the players will not select the ex ante efficient choice. In terms of the mechanism probabilities, μ_N and μ_A are the two extremes among all AAIE mechanisms.

7. Conclusion

The concept of incentive efficiency is clearly a minimal requirement for defining reasonable selections by players in bargaining situations with incomplete information. I find that the set of almost ex ante incentive efficient mechanisms coincides with the set of interim incentive efficient mechanisms and thus is a superset of the set of ex ante incentive efficient mechanisms.

The analysis of ex ante mechanism selection in bargaining problems with incomplete information 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 mechanism selection stage is not robust to a perturbation of the information structure at the selection stage. This implies that 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; Reasonable selections must be defined on a larger set of almost ex ante (or equivalently, interim) incentive efficient mechanisms, which may not be ex ante incentive efficient. Further, ex ante efficiency can be seriously misleading as a solution concept for a theory of bargaining or as a welfare measure for evaluating mechanisms.

There is no generally accepted interim bargaining solution concept in the literature, but many bargaining situations take place under incomplete information such as those mentioned in the introduction. While the set of almost ex ante incentive efficient mechanisms may be large, Myerson's concept of neutral optima gives stronger predictions of mechanism selections. Such concept incorporates the consideration of an equitable balance between different preferences of players; in this sense, the neutrality concept can be considered a reasonable requirement for fair bargaining solutions when the properties of both efficiency and equity are concerned.[13] Further, the concept of neutral optima is a robust solution concept that can be applied to such situations and that admits a unique prediction of which mechanism would reasonably arise for a broad class of bargaining games. In this sense, this note reinforces the relevance of the interim bargaining solution suggested by Myerson (1983, 1984*b*) to models of the process of agreeing on a mechanism.

Appendix A. Characterization Theorem for AAIE Mechanisms

Holmström and Myerson (1983) give the representation of the concept of interim incentive efficiency through restrictions on individual weights in a social welfare function. Under the assumption that D and T are finite sets, the set of feasible mechanisms is defined by a finite number of linear constraints. Thus by the supporting hyperplane theorem, a feasible mechanism μ is interim incentive efficient (IIE) iff there exist some positive numbers $\hat{\lambda}_i(t_i)$ for each type t_i of each player i such that μ

[13] A neutral mechanism for any given almost ex ante bargaining problem can be interpreted as being both efficient and equitable in terms of players' virtual preferences that incorporate what they would have wanted if they were of different types or uninformed.

is an optimal solution to the optimization problem:

$$\max_{\mu:T\to\Delta(D)} \sum_i \sum_{t_i\in T_i} \hat{\lambda}_i(t_i)U_i(\mu|t_i) \tag{A.1}$$

subject to the feasibility constraints (1) and (2) (in the main text), where $U_i(\mu|t_i)$ is the conditional expected utility for player i in mechanism μ given that he is of type t_i. Solving this problem gives us the set of IIE mechanisms, Δ_I^*.

Myerson (1991, pp. 497-498) provides the characterization theorem for computing IIE mechanisms. The same idea behind the mathematical formulation of the conditions for characterizing incentive efficient mechanisms applies to characterizing the set of AAIE mechanisms, Δ_{AA}^*. So throughout this appendix, I keep the same notations when appropriate and use similar language as Myerson. A feasible mechanism μ is AAIE iff there exist some positive numbers λ_i^u independent of t_i and $\lambda_i(t_i)$ for each type t_i, for each player i, such that μ is an optimal solution to the optimization problem:

$$\max_{\mu:T\to\Delta(D)} \sum_i \left[\lambda_i^u U_i^u(\mu) + \sum_{t_i\in T_i} \lambda_i(t_i)U_i(\mu|t_i)\right]$$

subject to the feasibility constraints (1) and (2).

This optimization problem is a linear programming problem, so a Lagrangian function can be formed. Let $\alpha_i(s_i|t_i)$ denote the dual variables for the constraint (1), and let $\alpha = (\alpha_i(s_i|t_i))_{i\in\{1,2\},s_i\in T_i,t_i\in T_i}$. Because I can vary $\lambda = (\lambda_i^u, (\lambda_i(t_i))_{t_i\in T_i})_{i\in\{1,2\}}$ over the strictly positive orthants of $\mathbb{R}^{\{u\}\cup T_1}\times\mathbb{R}^{\{u\}\cup T_2}$, where $\{u\}$ denotes a singleton set of a player being uninformed, the dual variables for the constraint (2) can be suppressed. Then the Lagrangian function can be written as

$$\sum_i \Big[\lambda_i^u U_i^u(\mu) + \sum_{t_i \in T_i} \lambda_i(t_i) U_i(\mu|t_i)\Big]$$
$$+ \sum_i \sum_{t_i \in T_i} \sum_{s_i \in T_i} \alpha_i(s_i|t_i) \Big[\sum_{t_{-i} \in T_{-i}} \sum_{d \in D} \bar{p}_{-i}(t_{-i}) \mu(d|t) u_i(d,t)$$
$$- \sum_{t_{-i} \in T_{-i}} \sum_{d \in D} \bar{p}_{-i}(t_{-i}) \mu(d|t_{-i}, s_i) u_i(d,t)\Big].$$

This function can be simplified to

$$\sum_{t \in T} \sum_{d \in D} \sum_i \mu(d|t) V_i(d,t,\lambda,\alpha)$$

by letting

$$V_i(d,t,\lambda,\alpha) = \bar{p}_{-i}(t_{-i}) \bigg[\Big(\lambda_i^u \bar{p}_i(t_i) + \lambda_i(t_i) + \sum_{s_i \in T_i} \alpha_i(s_i|t_i) \Big) u_i(d,t)$$
$$- \sum_{s_i \in T_i} \alpha_i(t_i|s_i) u_i(d,(t_{-i}, s_i)) \bigg]. \tag{A.2}$$

I call this quantity $V_i(d,t,\lambda,\alpha)$ player i's *almost ex ante virtual evaluation* of decision d in expectation of state t with respect to λ and α. This expression slightly differs from the original expression of (interim) virtual evaluation introduced by Myerson (1984*b*, p. 473), in the sense that an extra term $\lambda_i^u \bar{p}_i(t_i) u_i(d,t)$ appears in brackets in (A.2). Myerson's virtual evaluation captures the idea that a player may want to distinguish himself from the other possible types even though he already knows his actual type. The interpretation of my expression (A.2) depends on whether a player is informed or uninformed: If a player is informed, his virtual evaluation incorporates what he would have wanted if he were uninformed, captured by the term $\lambda_i^u \bar{p}_i(t_i) u_i(d,t)$, in addition to what he would have wanted if he were of different types; if a player is uninformed, then the term $\lambda_i^u \bar{p}_i(t_i) u_i(d,t)$ captures his actual utility given his ex ante belief $\bar{p}_i(t_i)$ that he is of type t_i, while the remaining terms incorporate what he would have wanted if he were informed of different possible types.

By the duality theorem of linear programming, the dual problem for λ can be written as

$$\min_{\alpha} \sum_{t \in T} \max_{d \in D} \sum_i V_i(d, t, \lambda, \alpha).$$

These arguments lead to the following characterization theorem for AAIE mechanisms.

Theorem A.1. *For any two-person almost ex ante bargaining problem, a mechanism μ is AAIE if and only if μ is feasible and there exist vectors $\lambda = (\lambda_i^u, (\lambda_i(t_i))_{t_i \in T_i})_{i \in \{1,2\}}$ and $\alpha = (\alpha_i(s_i|t_i))_{i \in \{1,2\}, s_i \in T_i, t_i \in T_i}$ such that*

$$\lambda_i^u > 0,\ \lambda_i(t_i) > 0,\ \forall t_i \in T_i, \quad \forall i \in \{1,2\},$$

$$\alpha_i(s_i|t_i) \geq 0, \quad \forall i \in \{1,2\},\ \forall s_i \in T_i,\ \forall t_i \in T_i,$$

$$\sum_{d \in D} \mu(d|t) \sum_i V_i(d, t, \lambda, \alpha) = \max_{d \in D} \sum_i V_i(d, t, \lambda, \alpha), \quad \forall t \in T,$$

and complementary slackness conditions of dual optima are satisfied.

The AAIE mechanism essentially maximizes the sum of the players' almost ex ante virtual evaluations in every possible state. In a normative sense of Pareto efficiency, the players would reasonably choose from among the set of AAIE mechanisms at the almost ex ante stage of mechanism selection.

Theorem A.1 is exactly analogous to the characterization theorem for Δ_I^* except that, in the latter theorem, the utility weights $(\lambda_i^u)_{i \in \{1,2\}}$ are eliminated from the conditions of the theorem and the terms $\lambda_i^u \bar{p}_i(t_i)$ are deleted from the expression of virtual evaluations (A.2). The inclusion of additional utility weights λ_i^u in Theorem

A.1 arises from introducing uncertainty about whether players are informed or not, which constitutes an important point of departure of my note.

Suppose that the conditions for interim incentive efficiency are satisfied for a feasible mechanism μ with some vectors $\hat{\lambda} = (\hat{\lambda}(t_i))_{i\in\{1,2\},t_i\in T_i}$ and $\alpha = (\alpha_i(s_i|t_i))_{i\in\{1,2\},s_i\in T_i,t_i\in T_i}$. For any given $\bar{p}_i(t_i)$ for every type t_i of any player i, there exist strictly positive numbers λ_i^u and $\lambda_i(t_i)$ such that $\lambda_i^u\bar{p}_i(t_i) + \lambda_i(t_i) = \hat{\lambda}_i(t_i)$. Then the utility weights $(\lambda_i^u, (\lambda_i(t_i))_{t_i\in T_i})_{i\in\{1,2\}}$ together with $(\alpha_i(s_i|t_i))_{i\in\{1,2\},s_i\in T_i,t_i\in T_i}$ satisfy all the conditions in Theorem A.1 for the same feasible mechanism μ. The converse also holds by the same logic. This completes the technical proof of Theorem 1 in the main text.

Appendix B. Characterization Theorem for AAN Mechanisms

Myerson's (1984*b*) neutral bargaining solution is developed axiomatically and is proven to exist for any two-person Bayesian bargaining problem. Myerson (1984*b*) also offers the tractable set of conditions for characterizing interim neutral (IN) mechanisms that are not only IIE in terms of actual utility payoffs but also both efficient and equitable in terms of transferable virtual-utility payoffs. Given my definition of almost ex ante bargaining problems, the appropriately modified version of the characterization theorem for almost ex ante neutral (AAN) mechanisms can be formulated.

Theorem B.1. *For any two-person almost ex ante bargaining problem, a mechanism μ is AAN if and only if μ is AAIE and there exist sequences $\{\lambda^k\}_{k=1}^{\infty}$, $\{\alpha^k\}_{k=1}^{\infty}$, and $\{\omega^k\}_{k=1}^{\infty}$ such that*

$$\lambda_i^{u,k} > 0,\ \lambda_i^k(t_i) > 0,\ \forall t_i \in T_i, \quad \forall i \in \{1,2\},\ \forall k,$$

$$\alpha_i^k(s_i|t_i) \geq 0, \quad \forall i \in \{1,2\},\ \forall s_i \in T_i,\ \forall t_i \in T_i,\ \forall k,$$

$$\omega_i^k(t_i) \geq 0, \quad \forall i \in \{1,2\},\ \forall t_i \in T_i,\ \forall k,$$

$$
\begin{aligned}
&\Big(\lambda_i^{u,k}\bar{p}_i(t_i)+\lambda_i^k(t_i)+\sum_{s_i\in T_i}\alpha_i^k(s_i|t_i)\Big)\omega_i^k(t_i)-\sum_{s_i\in T_i}\alpha_i^k(t_i|s_i)\omega_i^k(s_i)\\
&\quad=\sum_{t_{-i}\in T_{-i}}\max_{d\in D}\sum_{j=1}^{2}V_j(d,t,\lambda^k,\alpha^k)/2,\quad \forall i\in\{1,2\},\ \forall t_i\in T_i,\ \forall k,
\end{aligned}
\tag{B.1}
$$

$$
\limsup_{k\to\infty}\omega_i^k(t_i)\le U_i(\mu|t_i),\quad \forall i\in\{1,2\},\ \forall t_i\in T_i, \tag{B.2}
$$

$$
\textit{and}\ \limsup_{k\to\infty}\sum_{t_i\in T_i}\bar{p}_i(t_i)\omega_i^k(t_i)\le U_i^u(\mu),\quad \forall i\in\{1,2\}, \tag{B.3}
$$

where $V_j(d,t,\lambda^k,\alpha^k)$ *is the almost ex ante virtual evaluation defined as in* (A.2) *with respect to* λ^k *and* α^k*.*

The condition (B.1) says that μ 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. The conditions (B.2) and (B.3) require that μ gives each player a real expected utility that is at least as large as the limit of virtually equitable allocations, $\omega_i(\cdot)$, for each type t_i and the limit of ex ante expected virtually equitable allocations, $\sum \bar{p}_i(\cdot)\omega_i(\cdot)$, for type "uninformed."

The equivalence between the sets of AN and IN mechanisms follows trivially by letting $\lambda_i^{u,k}\bar{p}_i(t_i)+\lambda_i^k(t_i)=\hat{\lambda}_i^k(t_i)$ for all k, for all $t_i\in T_i$ and i, where the sequences $\{\lambda^k\}_{k=1}^{\infty}$ and $\{\hat{\lambda}^k\}_{k=1}^{\infty}$ satisfy respectively the conditions of Theorem B.1 and those of Theorem 4 in Myerson (1984b). Further, the condition (B.2) immediately implies the condition (B.3).

The neutral solution concept captures the idea of inscrutable intertype compromise, which is to use a bargaining strategy that maintains an equitable balance between the conflicting goals of alternative types of the same player, so as not to reveal his true type during the selection process. Even if a player is uninformed of his true type, he must not be mistaken for some possible type, and so must also express an equitable compromise between alternative types as if he were informed. Hence, nothing essential is added to the concept of almost ex ante neutral mechanisms by the assumption of the almost ex ante stage, other than what the interim neutral concept already captures.

Appendix C. Verifying Neutral Mechanism in the Example

Consider the example from Section 6 in the main text. The interim neutral mechanism for this example is μ_N, which has the probabilities

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

I first verify that this mechanism satisfies the conditions of the characterization theorem for interim neutral mechanisms in Myerson (1984*b*).

Note that μ_N is an optimal solution of the primal problem (8) for $\hat{\lambda}$ (given in the main text), where

$$\hat{\lambda}(H) > 27/38, \; \hat{\lambda}(L) = 1 - \hat{\lambda}(H) > 0.$$

The optimal solution of the dual for $\hat{\lambda}$ is

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

Let $\hat{\lambda}(H) = 9/10$ and so $\alpha(H|L) = 3/10$. With these parameters, the interim virtual evaluations are:

$$
\begin{aligned}
V(d_1,(H,H),\lambda,\alpha) &= \tfrac{3}{8}\left[\tfrac{9}{10}u(d_1,H,H) - \tfrac{3}{10}u(d_1,L,H)\right] = 0,\\
V(d_1,(H,L),\lambda,\alpha) &= \tfrac{5}{8}\left[\tfrac{9}{10}u(d_1,H,L) - \tfrac{3}{10}u(d_1,L,L)\right] = -\tfrac{3}{4},\\
V(d_1,(L,H),\lambda,\alpha) &= \tfrac{3}{8}\left(\tfrac{1}{10} + \tfrac{3}{10}\right)u(d_1,L,H) = \tfrac{9}{20},\\
V(d_1,(L,L),\lambda,\alpha) &= \tfrac{5}{8}\left(\tfrac{1}{10} + \tfrac{3}{10}\right)u(d_1,L,L) = \tfrac{1}{4},\\
V(d_0,t,\lambda,\alpha) &= 0,\ \forall t \in T.
\end{aligned}
$$

Then $\omega(H)$ and $\omega(L)$ should satisfy $\frac{9}{10}\omega(H) - \frac{3}{10}\omega(L) = 0$ and $\frac{2}{5}\omega(L) = \frac{1}{4}$, and so $\omega(H) = \frac{5}{24} = U(\mu_N|H)$ and $\omega(L) = \frac{5}{8} = U(\mu_N|L)$. With these $\hat{\lambda}$, α, and ω, the conditions of Theorem 5 in Myerson (1984*b*) are satisfied. Then Myerson's Theorem 6 verifies that μ_N is an interim neutral mechanism.

I also verify that μ_N satisfies the conditions of Theorem B.1, checking that μ_N is indeed AAN. Using $\hat{\lambda}(H) = 9/10$ and $\hat{\lambda}(L) = 1/10$, I can find strictly positive numbers λ^u, $\lambda(H)$, and $\lambda(L)$ that satisfy $\hat{\lambda}(H) = (3/8)\lambda^u + \lambda(H)$ and $\hat{\lambda}(L) = (5/8)\lambda^u + \lambda(L)$. For example, let

$$\lambda^u = 2/15,\ \lambda(H) = 17/20,\ \lambda(L) = 1/60.$$

Because all components of $\lambda = (\lambda^u, \lambda(H), \lambda(L))$ are strictly positive, I can satisfy the conditions of Theorem B.1 with the constant sequences $(\lambda^k, \alpha^k, \omega^k) = (\lambda, \alpha, \omega)$ where $\alpha = (\alpha(L|H), \alpha(H|L)) = (0, 3/10)$ and $\omega = (\omega(H), \omega(L)) = (5/24, 5/8)$. Note that $(3/8)\omega(H) + (5/8)\omega(L) = \frac{15}{32} = U^u(\mu_N)$, satisfying (B.3). Thus, μ_N remains a neutral mechanism for the example of almost ex ante bargaining problem.

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.

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

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.