Conditional Random Dictatorship and the Value of a Bayesian Cooperative Game[☆]

Andrés Salamanca

Department of Economics, Ca' Foscari University of Venice, Italy

Abstract

We present and analyze a theory of cooperative bargaining under asymmetric information, based on the equity principles of the conditional random dictatorship—choosing with equal probability one of the individuals to act as a dictator under the participation constraint of the other players. Our approach leads to a unified bargaining solution that has as special cases the Shapley value in transferable utility games and the Maschler-Owen consistent value in non-transferable utility games. This solution is shown to be the smallest possible set satisfying two axioms: one of these axioms states what the solution should be for the class of problems where conditional random dictatorship yields an efficient outcome, and the other axiom relates the solutions of a problem to the solutions of its extensions.

Keywords: Cooperative games, incomplete information, strong solution, random dictatorship, virtual utility.

1. Introduction

In this paper, we address the problem of allocating the proceeds or costs of a cooperative endeavor among multiple participants with private information. Specifically, we consider situations in which, at the time when allocation decisions must be made, some individuals have information about their preferences and/or endowments (encoded in their *types*) that is not known by other individuals. In order to develop general principles for performing this allocation, we need to determine how equitable compromises are made, not just among the different individuals, but also among the distinct possible types of any one individual. To address this normative aspect of information, we adopt a procedural justice approach: individuals themselves select the social decision through a procedure that is fair in the sense that it gives each of them an equal chance to control the decision. A prominent example of such procedures is the *conditional random dictatorship*—every player gets the same chance at being allowed to demand any allocation he wants, provided that all individuals agree on the proposed division (hence the name conditional). A key insight from Myerson (1983, 1984a) is that conditional

[☆]This research benefited from funding provided by the European Union-Next Generation EU through the Italian National Recovery and Resilience Plan (NRRP)—Mission 4, Component 2, Investment 1.1: Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN)—Project "Political persuasion," grant number 2022WS8AJY–CUP H53D23002600006. This manuscript reflects only the author's views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Email address: `andres.salamanca@unive.it` (Andrés Salamanca)

dictatorship defines an endogenous inter-type compromise, insofar as the dictator needs to strike a balance between the goals of his actual type and the goals of the other possible types that the other individuals think he might be. Our contention is that conditional random dictatorship constitutes a good guiding principle to define a bargaining solution for general Bayesian collective-decision problems.

Consider the following simple allocation problem that will serve to draw a parallel between the complete information case and the new issues raised by the asymmetric information. Player 1 is the seller of some object and player 2 is the only potential buyer. The object is worth \$0 to the seller and \$10 to the buyer. Payoffs to each player are defined to be his/her monetary gains from trade (i.e., the seller gets p and the buyer obtains $\$10 - p$, whenever they trade at the price p). Thus, both individuals are risk-neutral in money. A trading plan specifies the probability that the object will be traded and (if so) at what price. The set of payoffs generated by all possible trading plans in this example is illustrated in Figure 1.1 (shaded area).

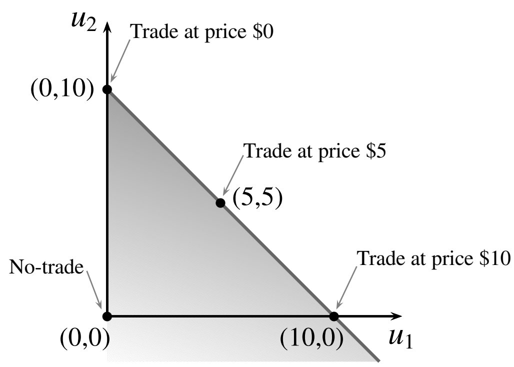

Figure 1.1: Feasible payoffs in the bilateral trading problem

An approach with a long tradition in economics would proceed by selecting an allocation that balances the marginal contributions of the various individuals. According to this *egalitarian principle*, trading (with probability 1) at the price \$5 is equitable in the sense that each player gains from trade as much as s/he is contributing to the other player. Conditional random dictatorship offers another alternative to the previous distributive approach. If the seller were a dictator, subject to the constraint that the buyer gets at least what he would obtain in the absence of trade, then she could make a take-it-or-leave-it offer to sell the good at the price \$10. After all, the buyer would be indifferent between paying \$10 and not buying the object. On the other hand, if the buyer were given the power of dictatorship, then he could make a first-and-final offer to buy the object for a price of \$0. When both players have equal bargaining ability, randomizing between the dictatorial outcomes with equal probabilities would be fair in the sense that both individuals are given equal chance to control the final outcome.[1] Because individuals are risk neutral, the *expected* outcome of this procedure is equivalent to setting a price equal to \$5, as in the initial distributive approach. However, this equivalence between the egalitarian

[1]Here *bargaining ability* means the ability to argue articulately and persuasively in the negotiation process. A related notion is that of *bargaining power*, which denotes the ability to help or hurt other players at will, and to defend against the threats of others.

principle and the conditional random dictatorship (in quasi-linear environments) does not survive the introduction of asymmetric information.

Let us now consider the conditional random dictatorship in the context of a Bayesian problem. For that, suppose now that the seller has private information related to the quality of the object, so that the value to the buyer is a function of the seller's valuation. Formally, if the quality of the good is high, then the object is worth \$5 to the seller, but if the quality is low instead, the object is worth only \$0 to her. On the other hand, if the value of the object for the seller is v, then it will be ultimately worth $v + \$5$ to the buyer. Only the seller knows the quality of the object, but the buyer thinks that the good is likely to be of high quality with a probability of $2/5$. A trading plan must now be contingent on the quality of the good. However, the seller is free to make any statement about the quality, and there is no way to verify whether those claims are true or false. Thus, a trading plan must satisfy some *incentive constraints* giving the seller an incentive to participate honestly in the trading plan. In this Bayesian problem it is not obvious what plan would the seller demand if she were a (conditional) dictator. Consider the plans that correspond to a take-it-or-leave-it offer to sell the object at a price p (regardless of the quality). Any such plan is incentive compatible as it does not depend on the quality of the object. Moreover, the gains from trade to each type of the seller are increasing in p, and the expected gains to the buyer are non-negative if and only if $p \leq \$7$. Therefore, if the seller were to implement a take-it-or-leave-it offer (as in the complete information case), her best choice would be to set a price equal to \$7.[2] However, there exists another plan that the seller would prefer if the quality were high: The buyer pays his full value $\tilde{v} + \$5$ depending on the seller's announced quality, $\tilde{v} \in \{\$0, \$5\}$, but to keep the seller truthful, trade only occurs with probability $1/2$ if the seller claims that the quality is high.[3] Thus, as an unavoidable cost of incentive compatibility, there must be a positive probability that no trade will occur in this plan. One theory that cannot be valid is to suggest that the seller would demand the buyer's full value if the quality were high and would make a take-it-or-leave-it offer to sell at \$7 if the quality were low. In this case, the selection of the trading plan itself will reveal the actual quality. Thus, the buyer would refuse to buy the object for \$7 if he believed that the seller would make this demand only when the quality is low, because the expected value of the object to him would only be \$5.

To resolve this dilemma, the seller would have to demand a trading plan that appears to be a fair *inscrutable compromise* between her actual preferences and what she would have preferred if her type had been different. In this example, there are compelling reasons to think that the conflict of interest between both seller's types should be resolved in favor of the high-quality type. In other words, both types should insist on asking for the buyer's full value. To see why, notice that this trading plan remains acceptable for the buyer no matter what he might infer about the seller's type

[2]Observe that the gains to the seller from a take-it-or-leave-it offer to sell at \$7 are $\$2 = \$7 - \$5$ for the high-quality type, and $\$7 = \$7 - \$0$ for the low-quality type.

[3]Under this trading plan the seller obtains expected gains equal to $\$5/2 = 1/2 \times (\$10 - \$5)$ in case she is a high-quality type, and $\$5 = \$5 - \$0$ in case she is a low-quality type.

from her choice (i.e., it is ex-post individually rational). Therefore, the seller can be confident that the buyer will accept the plan and trade accordingly regardless of his beliefs. We say then that such a plan is *safe* for the seller. On the other hand, this trading plan is not Pareto inferior for the seller to any other incentive-compatible plan. We say then that this trading plan is *undominated* for the seller. Myerson (1983) refers to a safe and undominated trading plan as a *strong solution*. Myerson (1983) argues that the seller should demand her strong solution, even though she might actually prefer some other incentive-compatible plan (given her true type): If the seller tried to demand the take-it-or-leave-it offer to sell at \$7, and if the buyer inferred from this demand that the seller's type is the one that strictly benefits from this "deviation" (i.e., the low-quality type), then such an alternative plan cannot be acceptable for the buyer given this new information. Thus, the seller cannot do better than to demand her strong solution. Unfortunately, a significant obstacle obstructs the use of strong solutions as a building block for extending the conditional random dictatorship procedure to general allocation problems with incomplete information. Specifically, a strong solution does not always exist in every allocation problem (see example 10.2 in Myerson, 1991b). However, whenever a strong solution exists, it is essentially unique, in the sense that even if there were many strong solution, they are all utility equivalent for the dictator. Before addressing the non-existence of strong solutions, we must first acknowledge a notable drawback with the conditional random dictatorship procedure. By resolving this issue, we will also offer a solution to the non-existence problem.

An "acceptable" allocation cannot always be obtained through the conditional random dictatorship procedure. This is due to the possibility that the expected outcome of the randomization may not be efficient. In such instances, we say that the game has *non-transferable utility* (*NTU*), that is, the players may not be able to make side payments to each other in such a way that the total utility gains are equal to the total utility losses. To illustrate this issue, let us return to our allocation problem with *complete* information. This time, we assume that the seller must pay a tax of 60% on the portion of her income that exceeds \$5. In other words, the seller loses \$0.6 for each additional unit of currency over \$5 that the buyer transfers to her. Thus, this modified allocation problem exhibits non-transferable utility. This restriction on the feasible utility set is equivalent to assuming that any trading plan with a price $p \geq \$5$ has a probability $\frac{3(p-5)}{3p+20}$ of being obstructed. Therefore, the seller may fail to sell the object if the price is $\$5 < p \leq \$10$, even though the object is actually worth more to the buyer. The set of feasible utility allocations looks like in the panel A of Figure 1.2. Now the conditional random dictatorship yields the expected allocation $\bar{u}$, which is Pareto dominated. In this modified problem, however, the allocation $u^* = (5, 5)$ is still a natural candidate for an equitable and efficient solution.

Similar to how the tax constrained the utility-possibility set, in the presence of asymmetric information, incentive constraints restrict what is feasible in such a way that the game may exhibit non-transferable utility even when the corresponding ex-post games have transferable utility (TU). For example, as illustrated in our previous allocation problem with incomplete information, when the quality of the object is indeed high, the seller may needed to accept a positive probability of no trade. This is needed to prove that she is not lying when she claims that the quality is high. Consequently,

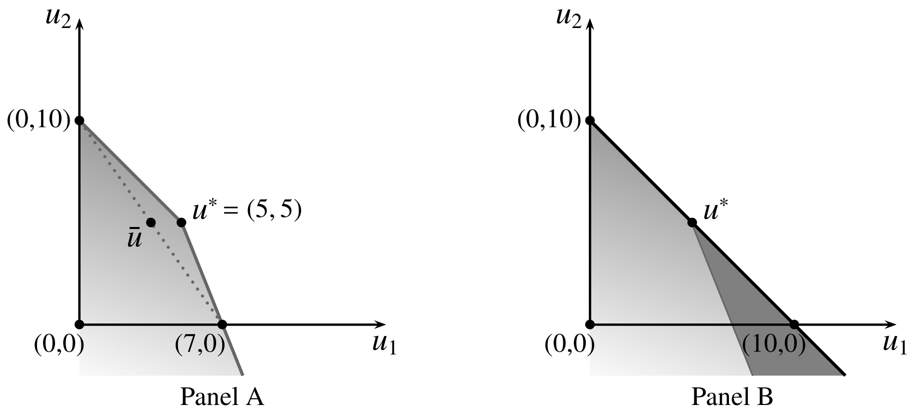

Figure 1.2: Feasible payoffs when utility is non-transferable

incentive constraints limit the likelihood of trade, resembling the obstruction effect observed with the tax. Hence, due to incentive compatibility, we cannot prevent the conditional random dictatorship procedure from producing Pareto-dominated allocations in Bayesian collective-decision problems, even if players have all quasi-linear utilities.

The question then arises as to how to reformulate the fairness principles of the conditional random dictatorship procedure to guarantee an efficient outcome. We resolve this dilemma by adopting the *independence of irrelevant alternatives* (*IIA*) axiom introduced by Nash (1950). This axiom states that the solution of the game should not change as the set of feasible outcomes is reduced, so long as the disagreement point remains unchanged, and the solution point originally selected remains feasible in the reduced game. Consider extending the NTU allocation problem in Panel A of Figure 1.2 by introducing additional decision options that expand the set of feasible utilities to include points as illustrated in the panel B of Figure 1.2. This extended problem is utility equivalent to the TU problem in Figure 1.1. Therefore, the application of conditional random dictatorship yields the expected allocation u^*, which is feasible in the original NTU problem. Thus, according to the IIA, u^* must also be regarded as the solution to the NTU problem. The additional decisions in the expanded problem may be considered as "irrelevant" in the sense that they are not actually required to achieve the final outcome. These decision are only important to the extent that they help to find out what an equitable allocation should look like in the NTU problem.

There is a substantial problem when we try to replicate the above reasoning under asymmetric information: adding new collective decisions may change the incentives structure of the game, which often modifies the efficient frontier in a way that makes the application of the IIA impossible. Indeed, we cannot arbitrarily extend the original game by introducing additional collective decisions (or equivalently ex-post utility vectors), while leaving the original utility allocation efficient in the expanded problem. The reason is that new decisions may be used to facilitate the fulfillment of incentive constraints (see Salamanca, 2020, Sec. 7.2 for an illustration of this phenomenon). Aware of this conceptual difficulty, Myerson (1983, 1984a,b) developed the *virtual utility approach* that, for each efficient allocation, identifies additional collective decisions that effectively linearize the set of feasible utilities (as if utility were transferable), while leaving the original allocation efficient in the enlarged problem (as required by the IIA). This extension has an important added feature—it can be constructed in a way that ensures every player has a strong solution (see Lemma 1). The virtual utility approach thus provides us with the necessary elements to address the non-existence of the strong solutions as well as the extendability of the IIA to environments with incomplete information.

The virtual utility approach tells us how to linearize the set of feasible utilities without removing a given efficient allocation, but it does not guarantee that the randomization between the dictatorial outcomes yields the candidate efficient allocation. Thus, the virtual utility approach is not sufficient to identify a solution for a given Bayesian collective-decision problem. To illustrate this issue, consider expanding the feasible utility set in panel A of Figure 1.2 as shown in Figure 1.3. This extension preserves the efficient allocation u^*. However, applying the conditional random dictatorship to the expanded problem now yields the allocation $\tilde{u}$, which not only does not coincide with u^*, but is also infeasible in the original problem. Hence, IIA cannot justify u^* as a solution to the original problem under this alternative extension.

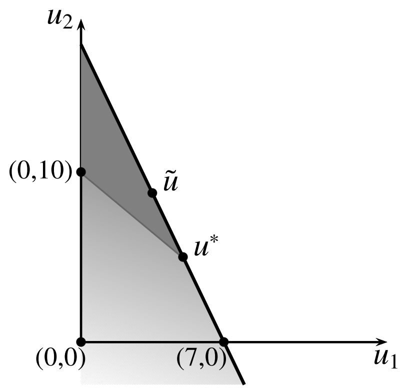

Figure 1.3: Infeasibility of conditional random dictatorship

A bargaining solution as defined here then solves a fixed-point problem: An allocation u^* is a *neutral procedural solution* iff u^* is efficient and there exists an extension of the original problem in which the conditional random dictatorship yields an expected outcome equal to u^*. For technical reasons, to guarantee the existence of a solution of this fixed-point problem, we slightly enlarge the above solution set to include allocations that emerge from some topological closure (see Theorem 1).

The neutral procedural solution is influenced by the outside options of the various players insofar as any random dictator must satisfy the participation constraints of all other individuals. In the examples above, we have assumed that the seller cannot trade the object with anyone else and that the buyer cannot purchase the object from another competing seller. Therefore, the outside option for both players corresponds to the no-trade decision. However, one may imagine other situations where, for instance, the buyer may acquire the object from a second seller. In this case, the buyer's outside option is determined by a preliminary agreement that he might reach with this second seller. Such a pre-agreement is the outcome of a Bayesian collective-choice problem similar to the one that the buyer faces with the first seller. Therefore, it would be natural to determine such a pre-agreement by applying the neutral procedural solution to this bilateral trading problem. More generally, we adopt the view that the final allocation in the grand coalition N will be influenced by partial cooperation in subcoalitions. For that, we assume that every smaller coalition S pre-commits to a *threat* that would actually be carried out if S were to form. Let S be a coalition with at least two members. Suppose that, for every $i \in S$, we have already identified $u^{S\setminus i}$ to be the threat from coalition $S \setminus i$. Then, the threat from S is the neutral procedural solution of the subgame restricted to the members of S, where a randomly chosen dictator $i \in S$ must satisfy the participation constraints of the other players in S, who have the outside option to form coalition $S \setminus i$, in which case they obtain $u^{S\setminus i}$. The neutral procedural solution for the grand coalition is then constructed inductively in the size of coalitions. This "recursive" application of the neutral procedural solution is somewhat similar to the idea of subgame perfection in extensive form games, where the same equilibrium conditions are applied in each subgame. Hart and Mas-Colell (1996, p. 366) refers to this property as "subcoalition perfectness".

The neutral procedural solution can be related to other cooperative solutions in the literature. When restricted to games with *complete* information, our solution concept coincides with the Shapley value in TU games. This can be easily seen from the fact that the Shapley value of a player is the average of his marginal contribution to the grand coalition and his Shapley values in the subgames with $|N| - 1$ players (see Hart, 2004, p. 39). In the case of NTU games, the neutral procedural solution generalizes the consistent NTU value introduced by Maschler and Owen (1989, 1992) (see Theorem 4). This should not be surprising, if we take into account that the consistent NTU value can be derived from a set of axioms that includes the IIA and the recursive conditional random dictatorship applied to the class of hyperplane games—games where the feasible utility set of each coalition is a half-space (see de Clippel et al., 2004).[4] More generally, in the family of games with *incomplete* information, our solution concept coincides with Myerson's (1984a) generalization of the Nash bargaining solution. This cooperative solution concept forms the smallest set of allocations satisfying two axioms. One of this axioms prescribes the use of the conditional random dictatorship for a class of bargaining problems in which there is a clear allocation that each player should demand, if he could act as a dictator. The other axiom defines a relationship between the solutions of one problem and the solutions of its extensions (as in the IIA axiom). Using appropriate generalizations of the previous two axioms we are able to provide an axiomatic characterization of the neutral procedural solution (see Theorem 6).

Alternative cooperative solutions for games with incomplete information have already been proposed in the literature. The first one is Myerson's (1984b) generalization of Shapley's (1969) NTU value (or M-solution for short).[5] In recent contributions, de Clippel (2005) and Salamanca (2020) have constructed two eloquent examples were the M-solution is shown to be insensitive to some informa-

[4]In the class of hyperplane games, the conditional random dictatorship always produces an expected outcome that is efficient. Moreover, a payoff configuration is *the* consistent NTU value of a hyperplane game iff it is the outcome of the recursive conditional random dictatorship.

[5]The Shapley NTU value is sometimes referred in the literature as the λ-transfer value.

tional externalities (see Sections 2 and 6). Building on other fairness criteria, these authors proposed alternative compelling outcomes for those games. De Clippel (2005) considers the perfect Bayesian equilibria (PBE) of an extensive form game inspired in the *random order arrival* that characterizes the consistent NTU value. Not surprisingly, the neutral procedural solution yields the same outcome as the unique PBE of this game. Thus, our cooperative solution prescribes an intuitively appealing outcome in de Clippel's example. On the other hand, Salamanca (2020) introduced a general cooperative solution (abbreviated as S-solution) that extends Harsanyi's (1963) NTU value to games with incomplete information. When applied to both examples, the S-solution solves the purported "difficulties" with the M-solution, but it also raises new issues in de Clippel's example. In contrast, the neutral procedural solution is subject to the same criticisms as the M-solution in Salamanca's example.

The rest of the paper is organized as follows. In Section 2, we use de Clippel's example to motivate our approach. The basic structure of a Bayesian cooperative game is presented in Section 3. Section 4 is devoted to a preliminary study of the strong solutions. In Section 5, we systematically approach the problem of developing a theory for allocating the benefits of cooperation in Bayesian cooperative games. The main results of this paper are also presented in this section. Section 6 concludes the paper with an analysis of the neutral procedural solution in Salamanca's example. In doing so, we illustrate an important property of the neutral procedural solutions called the *arrogance of strength*. Most of the technical proofs are deferred to the appendices. Appendix A contains the proof of existence of the neutral procedural solutions. Appendix B presents the proofs of various important lemmas supporting our main results.

2. Motivating Example: Bilateral Trade with a Broker

The following example was originally proposed by de Clippel (2005). The basic description of the game starts with a bilateral trading problem: Player 2 is the seller of a single indivisible commodity that has no value for himself. Player 1 is the only potential buyer. The good may be worth either \$30, with probability $1/5$, or \$90, with probability $4/5$ to the buyer. Only the buyer knows her own valuation of the good. We may say that the buyer's type is *weak* if the good is worth \$90 to her, since she may then be relatively more compelled to pay any higher price. Conversely, we refer to the buyer's type as *strong* if the good is worth \$30 to her. Payoffs to each player are defined to be his/her net monetary gains from trade, thus both individuals are risk neutral. Gains from no-trade are normalized to zero.

Individuals are trying to agree whether to trade and, if so, at what price. A *trading plan* (or *mechanism*) specifies the probability of trade and the price (if trade occurs), for each possible value that the buyer might have for the commodity. Because the buyer's valuation is not verifiable, we may anticipate that the buyer may try to convince the seller that she is strong, even when she is actually weak, to force the seller to accept a lower price. Therefore, for the players to trade according to a given trading plan, that plan must be *incentive compatible*—it must provide the incentives for the buyer to be truthful about her type.

Figure 2.1 depicts the set of incentive-efficient (i.e. second-best) allocations of expected gains for which every (type of each) individual obtains at least his or her reservation utility from no-trade.

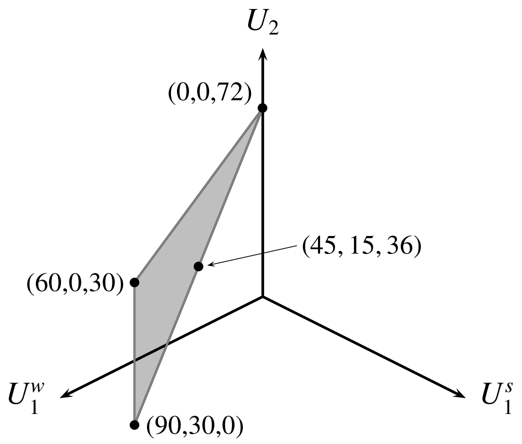

Figure 2.1: Incentive-efficient payoff allocations without brokerage

Suppose that the buyer is given the authority to demand any trading plan conditional on the participation of the seller. Then there is a clear inscrutable compromise for both buyer's types, namely, insisting on the plan that achieves the allocation $(U_1^w, U_1^s, U_2) = (90, 30, 0)$. Such a plan is equivalent to a take-it-or-leave-it price of \$0, and it is the buyer's strong solution. Hence, it gives the whole surplus from trade to the buyer. On the other hand, if the seller is given all the bargaining ability, he cannot do better than demanding the plan in Table 2.1,[6] which implements the allocation $(U_1^w, U_1^s, U_2) = (0, 0, 72)$.

Buyer's value	
\$30	\$90
(0, –)	(1, \$90)

Table 2.1: Seller's optimal plan for the bilateral problem

According to the plan 2.1, trade occurs only when the buyer claims to be weak and, in that case, the buyer pays her maximum reservation price. Although there exists a price range that would be acceptable to both parties when the buyer is strong (between \$0 and \$30), the seller needs to accept not to trade with the strong type to mitigate the incentive of the weak type to pretend to be strong. Incentive compatibility then limits the ability of the seller to extract the whole gains from trade when he is a "dictator". Indeed, the trading plan that gives the entire surplus to the seller in both states (i.e. give the good in exchange of \$30 or \$90 depending on the actual value that the buyer attributes to the good) is not incentive compatible.

Consider now a *conditional random dictatorship* in which each player gets a 50 per cent chance at being allowed to demand any incentive-compatible plan, constrained only by the requirement of giv-

[6] In Table 2.1, for each possible buyer's valuation, the probability of trade is shown first and the price of the good (if trade occurs) is shown second. The en-dash indicates that the price-if-trade cannot be defined when the probability of trade is zero.

ing the other player at least what he or she would get if cooperation breaks down (i.e., in the absence of trade). Such a random dictatorship yields the expected allocation in (2.1), which is generated by a $\frac{1}{2}-\frac{1}{2}$ randomization between the two dictatorial outcomes. From a procedural perspective, conditional random dictatorship is equitable (since each player gets equal chance to control the decision), thus the allocation (2.1) should be considered a fair outcome.

$$(U_1^w, U_1^s, U_2) = \tfrac{1}{2} \times (90, 30, 0) + \tfrac{1}{2} \times (0, 0, 72) = (45, 15, 36). \tag{2.1}$$

The allocation (2.1) is also incentive efficient (as evidenced in Figure 2.1), thus it can be regarded as a bargaining solution of the bilateral problem.

A third player, called the *broker*, can facilitate the transaction between the buyer and seller by releasing them from the incentive constraints they face. Specifically, the broker charges a commission for verifying the buyer's type, so that when a trading plan is carried out with the broker's help (i.e., in the grand coalition), the buyer and seller can proceed to trade as if there were no incentive constraints. Of course, when cooperation in the grand coalition breaks down into a coalition that does not contain both the buyer and the seller, its members cannot produce any surplus. A *brokerage plan* (or trading plan assisted by a broker) specifies, for each possible buyer's type, the probability of trade, the price (if trade occurs), the brokerage fee, and how the fee is distributed between the buyer and seller. To show the positive effect that the broker's assistance brings to the main parties, Figure 2.2 compares the set of incentive-efficient allocations of expected gains without brokerage (thin light gray contour) and with brokerage when the broker is rewarded \$0 (thick dark gray contour). We observe that, with the help of the broker, the seller may now achieve the allocation $(U_1^w, U_1^s, U_2, U_3) = (0, 0, 78, 0)$ corresponding to the situation in which the whole surplus of trade is given to him.

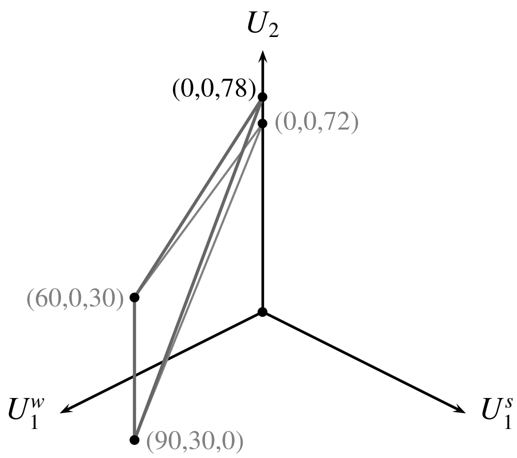

Figure 2.2: Incentive-efficient allocations with (dark gray) and without (light gray) brokerage

Before proceeding to calculate our bargaining solution under brokerage, it is instructive to first look at some other outcomes generated by alternative cooperative solutions. Let us consider first the M-solution. The unique payoff allocation that can be supported by some M-solution of the game with brokerage (see de Clippel, 2005) is

$$(U_1^w, U_1^s, U_2, U_3) = (45, 15, 39, 0). \tag{2.2}$$

This allocation considers the broker as a null player. Thus, the M-solution is insensitive to the broker's positive informational externality. Even though the broker does not create any additional surplus, it would be fair to give him a positive reward, as the buyer and seller have to rely on him in order to weaken the incentive constraints they face when they cooperate in coalition $\{1, 2\}$.

Consider now the S-solution. The unique payoff allocation that can be supported by some S-solution of this game (see Salamanca, 2020) is

$$(U_1^w, U_1^s, U_2, U_3) = (45, 13, 38.6, 0.8). \tag{2.3}$$

This allocation can be achieved by the following brokerage plan in Table 2.2.[7]

Buyer's value	
\$30	\$90
(1, \$15, \$4, \$2)	(1, \$45, \$0, –)

Table 2.2: S-solution with brokerage

Because only the buyer's weak type has incentives to impersonate the strong type, but not the opposite, the broker's service is only worthwhile when the buyer claims to be strong. Thus, the S-solution rewards the broker only when the buyer's value is \$30. Unlike the M-solution, the S-solution deems that, by releasing the traders from their incentive constraints, the broker brings a benefit to *both* of them. Thus, the burden of the commission is born by both traders—the brokerage plan in 2.2 splits the commission equally between the seller and the buyer. Although this outcome better reflects the situation, one can still argue against it. After all, only the seller needs the broker's help to extract all the surplus from the trade, and therefore the entire commission should be paid by the seller.

Let us now return to our procedural solution. In order to determine the fairness of a potential agreement, we need to assess the benefits that the broker brings to every trader. To do this, we need to determine what is thought to happen to the buyer and seller in the absence of brokerage. We adopt the view that when cooperation with the broker breaks down, the buyer and the seller form a coalition and agree on their bargaining solution in (2.1). The value of brokerage can then be assessed by applying the conditional random dictatorship to the grand coalition: each player gets $1/3$ of chance at being allowed to demand an incentive-compatible plan under the participation constraint of the other players, who have the outside option to form a two-person coalition and obtain the value of their bargaining solution.

[7]In Tables 2.2 and 2.3, for each buyer's value, the probability of trade is shown first, secondly the price (if trade occurs), thirdly the broker's commission, and lastly the commission incidence on the seller. The en-dash indicates that the incidence is not defined when the commission is zero.

Suppose the buyer is given the power of dictatorship. Since the seller and the broker cannot produce any surplus themselves, the buyer only needs to guarantee a non-negative payoff to the other players. It is clear from Figure 2.2 (thick dark gray contour) that the best inscrutable compromise for both buyer's types is to insist on the allocation $(U_1^w, U_1^s, U_2, U_3) = (90, 30, 0, 0)$. Similarly, when the seller is chosen to be a dictator, the best he can do is to demand the allocation $(U_1^w, U_1^s, U_2, U_3) = (0, 0, 78, 0)$. Assume now that the broker is given the power of dictatorship. Any incentive-compatible brokerage plan that satisfies the equation

$$\tfrac{4}{5}U_1^w + \tfrac{1}{5}U_1^s + U_2 + U_3 = 78 \tag{2.4}$$

must be incentive efficient (for the grand coalition). This equation asserts how the expected surplus ($\$78 = \frac{4}{5}\times\$90+\frac{1}{5}\times\$30$) can be efficiently distributed among the players. Because the broker needs to give the buyer and the seller at least what they would get with the allocation (2.1), the largest amount of expected surplus that the broker can extract is $U_3 = \$78 - \left(\frac{4}{5} \times \$45 + \frac{1}{5} \times \$15 + \$36\right) = \$3$.

Applying the conditional random dictatorship procedure to the grand coalition, we obtain the equitable allocation

$$(U_1^w, U_1^s, U_2, U_3) = \tfrac{1}{3}(90, 30, 0, 0) + \tfrac{1}{3}(0, 0, 78, 0) + \tfrac{1}{3}(45, 15, 36, 3) = (45, 15, 38, 1) \tag{2.5}$$

This allocation can be achieved by the brokerage plan in Table 2.3.

Buyer's value	
\$30	\$90
(1, \$15, \$5, \$5)	(1, \$45, \$0, –)

Table 2.3: Bargaining solution with brokerage

Clearly, the allocation (2.5) satisfies equation (2.4), thus it is incentive efficient and the trading plan in Table 2.3 can be regarded as a bargaining solution. Just like the S-solution, our procedural bargaining solution rewards the broker when his services are most valuable, that is, when the buyer's type is actually strong but the seller cannot believe in any such claim. Here, however, the burden of the commission rests solely on the seller. Thus, the neutral procedural solution seems to reflect the structure of the game better than the other cooperative solutions.

3. Model

3.1. Bayesian Cooperative Game

The model of a cooperative game with incomplete information is as follows. Let $N = \{1, 2, ..., n\}$ denote the set of players. For each (non-empty) coalition $S \subseteq N$, D_S denotes the set of feasible collective decisions for the members of S. For instance, D_S may represent the set of reallocations of individual endowments of goods between the members of S in an exchange economy. It may also constitute a set of possible trade deals as in the examples of the previous section. We assume that the sets of collective decisions are finite and *superadditive*, that is, for any two disjoint coalitions[8] S and R,

$$D_R \times D_S \subseteq D_{R \cup S}.$$

A player's private information is represented by a random privately known type. For any player $i \in N$, we let T_i denote the (finite) set of possible types of player i. We use the notation[9] $t_S = (t_i)_{i \in S} \in T_S = \prod_{i \in S} T_i$, $t_{-i} = t_{N \setminus i} \in T_{-i} = T_{N \setminus i}$, and $t_{-S} = t_{N \setminus S} \in T_{-S} = T_{N \setminus S}$. Types are assumed to be independently distributed random variables, so that for each $i \in N$, there exists a probability distribution $p_i \in \Delta(T_i)$.[10] Then we may write

$$p(t_S) := \prod_{i \in S} p_i(t_i), \quad \forall S \subseteq N, \quad \forall t_S \in T_S.$$

This assumption simplifies the algebraic expression of various formulas, although our theory may be developed without it (cf. equation (3.6) and Myerson, 1984b, eq. (3.6)). In addition, our theory satisfies the *invariance probability axiom* in Myerson (1984a); thus, for any game with correlated types, prior probabilities and utilities can be jointly modified in a way that the new game has independent types and both games impute probability and utility functions that are decision-theoretically equivalent (see Myerson, 1984a, p. 466). Without loss of generality, the probability of any type is assumed to be strictly positive, that is, $p_i(t_i) > 0$ for each $t_i \in T_i$ and each $i \in N$. Otherwise, we may simply dispose of all types that occur with zero probability. Each player knows his own type $t_i \in T_i$ and computes his beliefs using the Bayes rule: $p(t_{-i} \mid t_i) := \frac{p(t_N)}{p_i(t_i)} = p(t_{N \setminus i})$.

The utility function of player $i \in N$ is $u_i : D_N \times T_N \to \mathbb{R}$. We rule out strategic externalities across coalitions. Therefore, we assume that coalitions are *orthogonal* w.r.t. decisions. Formally,

$$u_i((d_S, d_{N \setminus S}), t_N) = u_i((d_S, d'_{N \setminus S}), t_N),$$

for all $d_S \in D_S$, $d_{N \setminus S}, d'_{N \setminus S} \in D_{N \setminus S}$, $t_N \in T_N$, $i \in S$, and $S \subseteq N$.

Orthogonality w.r.t. decisions is a standard hypothesis in cooperative game theory. It states that when a coalition $S \subseteq N$ chooses an action which is feasible for it, the payoffs to the members of S do not depend on the actions of the players in $N \setminus S$. This hypothesis is satisfied, for instance, in exchange economies. Also, the example in Section 2 satisfies this orthogonality requirement.

Under the orthogonality assumption, we can let $u_i^S(d_S, t_S)$ denote the expected utility to player $i \in S$ when $d_S \in D_S$ is carried out in coalition S in state $t_S \in T_S$. That is,

$$u_i^S(d_S, t_S) := \sum_{t_{-S} \in T_{-S}} p(t_{-S}) u_i((d_S, d_{N \setminus S}), (t_S, t_{-S})),$$

for some $d_{N \setminus S} \in D_{N \setminus S}$ (recall that $D_S \times D_{N \setminus S} \subseteq D_N$).

[8]For any two sets A and B, $A \subseteq B$ denotes *weak* inclusion (i.e., possibly $A = B$), and $A \subset B$ denotes strict inclusion.
[9]For simplicity we write $S \setminus i$, $S \cup i$, and D_i instead of the more cumbersome $S \setminus \{i\}$, $S \cup \{i\}$, and $D_{\{i\}}$.
[10]For any (finite) set A, $\Delta(A)$ denotes the set of probability distributions over A.

A *cooperative game with incomplete information* is defined by

$$\Gamma_N = \{N, (D_S)_{S\subseteq N}, (T_i, u_i, p_i)_{i\in N}\}.$$

For any coalition $S \subseteq N$, we denote Γ_S the game obtained by restricting Γ_N to the subcoalitions of S. The orthogonality assumption guarantees that Γ_S is a well defined game.

Players can use any communication mechanism to implement a state-contingent contract. Because information is not verifiable, the only feasible contracts are those which are induced by Bayesian Nash equilibria of the corresponding communication game. By the Revelation Principle (see Myerson, 1982), we can restrict attention to (Bayesian) incentive-compatible direct mechanisms. Formally, a *(direct) mechanism* for coalition S is a mapping $\mu_S : T_S \to \Delta(D_S)$. The interpretation is that if S forms, it makes a decision randomly as a function of its members' information. Let the set of mechanisms for S be denoted $\mathcal{M}_S$.

The (interim) expected utility of player $i \in S$ of type t_i under the mechanism μ_S when he pretends to be of type τ_i (while all other players in S are truthful) is

$$U_i^S(\mu_S, \tau_i \mid t_i) := \sum_{t_{S\setminus i}\in T_{S\setminus i}} p(t_{S\setminus i}) \sum_{d_S\in D_S} \mu_S(d_S \mid \tau_i, t_{S\setminus i}) u_i^S(d_S, t_S). \tag{3.1}$$

As is standard, we denote $U_i^S(\mu_S \mid t_i) := U_i^S(\mu_S, t_i \mid t_i)$.

A mechanism $\mu_S \in \mathcal{M}_S$ is *incentive compatible* for coalition S if and only if

$$U_i^S(\mu_S \mid t_i) \geq U_i^S(\mu_S, \tau_i \mid t_i), \qquad \forall t_i, \tau_i \in T_i, \quad \forall i \in S. \tag{3.2}$$

We denote as $\mathcal{M}_S^*$ the set of incentive-compatible mechanisms for coalition S.

A mechanism $\mu_S \in \mathcal{M}_S$ is *(interim) individually rational* for coalition S if and only if

$$U_i^S(\mu_S \mid t_i) \geq \max_{d_i\in D_i} u_i^{\{i\}}(d_i, t_i), \qquad \forall t_i \in T_i, \quad \forall i \in S. \tag{3.3}$$

3.2. Incentive Efficiency and the Virtual Utility

Let $S \subseteq N$ be a coalition. A mechanism $\mu_S \in \mathcal{M}_S$ is *(interim) incentive efficient* for S iff μ_S is incentive compatible for S and there does not exist any other incentive-compatible mechanism $\bar{\mu}_S$ such that

$$U_i^S(\bar{\mu}_S \mid t_i) \geq U_i^S(\mu_S \mid t_i), \qquad \forall t_i \in T_i, \quad \forall i \in S, \tag{3.4}$$

with strict inequality for at least one type t_i of some player $i \in S$.

To simplify our notation, we define

$$\Lambda_+^S := \left\{\lambda^S \in \prod_{i\in S} \mathbb{R}^{T_i} \,\middle|\, \sum_{i\in S}\sum_{t_i\in T_i} \lambda_i^S(t_i) = 1,\ \lambda_i^S(t_i) \geq 0,\ \forall i \in S,\ \forall t_i \in T_i\right\}$$

and denote its (relative) interior by Λ^S_{++}. Given any vector $\lambda^S \in \Lambda_+$, we define the *primal problem for S w.r.t. λ^S* to be the linear programming problem

$$\max_{\mu_S \in \mathcal{M}^*_S} \sum_{i \in S} \sum_{t_i \in T_i} \lambda^S_i(t_i) U^S_i(\mu_S \mid t_i) \tag{3.5}$$

That is, the primal problem is to find an incentive-compatible mechanism that maximizes the λ^S-weighted sum of the expected utilities of all types of players in S. When we vary λ^S as a free parameter over Λ^S_{++}, the optimal solutions to the primal problem (3.5) cover the entire set of incentive-efficient mechanisms for S.

Associated to any linear-programming problem there is a (Lagrangian) "dual problem" where each constraint in the primal problem becomes a variable in the dual problem. To formulate the dual problem of (3.5), we let $\alpha^S_i(\tau_i \mid t_i) \geq 0$ denote the dual variable (or Lagrange multiplier) for the constraint that the type t_i of player i should not gain by reporting τ_i. We let

$$A^S := \left\{ \alpha^S \in \prod_{i \in S} \mathbb{R}^{T_i \times T_i}_+ \mid \alpha^S_i(t_i \mid t_i) = 0,\ \forall t_i \in T_i,\ \forall i \in S \right\}.$$

Given any vectors $\lambda^S \in \Lambda^S_+$ and $\alpha^S \in A^S$, we define

$$\begin{aligned} v^S_i(d_S, t_S, \lambda^S_i, \alpha^S_i) := \frac{1}{p_i(t_i)} \Bigg[& \left(\lambda^S_i(t_i) + \sum_{\tau_i \in T_i} \alpha^S_i(\tau_i \mid t_i) \right) u^S_i(d_S, t_S) \\ & - \sum_{\tau_i \in T_i} \alpha^S_i(t_i \mid \tau_i) u^S_i(d_S, (\tau_i, t_{S \setminus i})) \Bigg] \end{aligned} \tag{3.6}$$

The quantity $v^S_i(d_S, t_i, \lambda^S_i, \alpha^S_i)$ is called the *virtual utility* of player $i \in S$ from decision $d_S \in D_S$, when he is type $t_i \in T_i$ (w.r.t. the utility weights λ^S_i and the dual variables α^S_i).

With this definition, the *dual problem for S w.r.t. λ^S* (i.e., the dual to (3.5)) may be written as

$$\min_{\alpha^S \in A^S} \sum_{t_S \in T_S} p(t_S) \left(\max_{d_S \in D_S} \sum_{i \in S} v^S_i(d_S, t_S, \lambda^S_i, \alpha^S_i) \right) \tag{3.7}$$

The virtual utilities are linear in the dual variables α^S, so this dual problem is a linear-programming problem. Strong duality of linear programming implies that the optimal value of the optimization problems (3.5) and (3.7) is the same.

Let $v^S_i(\mu_S, t_S, \lambda^S_i, \alpha^S_i)$ denote the linear extension of the virtual utility over μ_S:

$$v^S_i(\mu_S, t_S, \lambda^S_i, \alpha^S_i) := \sum_{d_S \in D_S} \mu_S(d_S \mid t_S) v^S_i(d_S, t_S, \lambda^S_i, \alpha^S_i).$$

Proposition 1 (Incentive Efficiency).
A mechanism $\mu_S \in \mathcal{M}_S$ is incentive efficient for $S \subseteq N$ iff it is incentive compatible for S and there exist vectors $\lambda^S \in \Lambda^S_{++}$ and $\alpha^S \in A^S$, such that

$$\alpha_i^S(\tau_i \mid t_i)\left[U_i^S(\mu_S \mid t_i) - U_i^S(\mu_S, \tau_i \mid t_i)\right] = 0, \quad \forall i \in S,\ \forall t_i \in T_i,\ \forall \tau_i \in T_i, \tag{3.8a}$$

$$\sum_{i \in S} v_i^S(\mu_S, t_S, \lambda_i^S, \alpha_i^S) = \max_{d_S \in D_S} \sum_{i \in S} v_i^S(d_S, t_S, \lambda_i^S, \alpha_i^S), \quad \forall t_S \in T_S. \tag{3.8b}$$

Equation (3.8a) is the usual *dual complementary slackness* condition. The appropriate vector α^S in Proposition 1 is any vector that solves the dual problem (3.7).

4. Strong solutions

This section is devoted to a preliminary study of the *strong solutions* introduced by Myerson (1983). This solution concept constitutes the basic building block of our recursive conditional random dictatorship procedure. It will allow us to predict which mechanisms an individual with conditional dictatorship power should select.

Let $S \subseteq N$ be a coalition (with $|S| \geq 2$) and i be a fixed player in S. Assume that player i is chosen to be a dictator in S, that is, he is given all the bargaining ability to determine a coordination mechanism for the members of S. However, he has to deal with the threats from the players in $S \setminus i$, each one of them having veto power to refuse i's proposition in order to enforce the *status quo* option, which is assumed to be a given mechanism $\mu_{S\setminus i} \in \mathcal{M}_{S\setminus i}$ implemented in coalition $S \setminus i$. The mechanism $\mu_{S\setminus i}$ determines the outside option available to the players in $S \setminus i$. Throughout this section we will assume that the mechanism $\mu_{S\setminus i}$ is incentive efficient for coalition $S \setminus i$. We denote this mechanism selection problem by $\Gamma_S^i(\mu_{S\setminus i})$.

Definition 1 (Feasibility).
A mechanism $\mu_S \in \mathcal{M}_S$ is feasible in $\Gamma_S^i(\mu_{S\setminus i})$ if it is incentive compatible for S, and it satisfies the following participation constraints:

$$U_j^S(\mu_S \mid t_j) \geq U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j), \quad \forall j \in S \setminus i, \quad \forall t_j \in T_j. \tag{4.1}$$

The problem $\Gamma_S^i(\mu_{S\setminus i})$ is an *informed-principal problem* as described and studied by Myerson (1983). Because player i already knows his type at the time when he selects the mechanism, the choice of the mechanism itself may convey information about his type to the other players in S. With this new information, the players in $S \setminus i$ may find new opportunities to gain by dishonesty or rejection of the proposed mechanism. So, the selected mechanism might not be feasible in practice, even though it satisfies (3.2) and (4.1). For inscrutability, all types of player i must choose the same mechanism. Fortunately, player i may never need to communicate any information to the other players in S by his choice of the mechanism, because he can always build such communication into the mechanism itself (in that $\mu_S(\cdot \mid t_S)$ depends on t_i). Myerson (1983) refers to this claim as the *inscrutability principle*.[11] However, because i's true type might actually prefer a different feasible mechanism, then the predicted mechanism must be a fair compromise between the alternative goals of all i's possible types. Myerson (1983) develops a formal theory of what such a "fair compromise" should be. In the following we summarize and adapt the main aspects of Myerson's theory in the context of the problem $\Gamma^i_S(\mu_{S\setminus i})$.

4.1. Undominated Mechanisms

We say that a mechanism $\mu_S \in \mathcal{M}_S$ is *dominated* in $\Gamma^i_S(\mu_{S\setminus i})$ if and only if there is another mechanism $\bar{\mu}_S \in \mathcal{M}_S$ for which

$$U_i^S(\bar{\mu}_S \mid t_i) \geq U_i^S(\mu_S \mid t_i), \qquad \forall t_i \in T_i, \tag{4.2}$$

with strict inequality for at least one type $t_i \in T_i$. If all inequalities in (4.2) are satisfied as strict inequalities, then $\bar{\mu}_S$ is said to be *strictly* dominated. Player i should never be expected to select a mechanism that is strictly dominated for him (see Myerson, 1983, p. 1775).

Definition 2 (**Undominated mechanisms).**
A mechanism $\mu_S \in \mathcal{M}_S$ is undominated in $\Gamma^i_S(\mu_{S\setminus i})$ if it is feasible in $\Gamma^i_S(\mu_{S\setminus i})$ and is not dominated by any other feasible mechanism.

As with the concept of incentive efficiency, undominated mechanisms can be equivalently characterized by the solutions of a weighted-utility maximization problem. A mechanism $\bar{\mu}_S \in \mathcal{M}_S$ is undominated in $\Gamma^i_S(\mu_{S\setminus i})$ iff there exists strictly positive numbers $\lambda_i^S = (\lambda_i^S(t_i))_{t_i\in T_i}$ such that $\bar{\mu}_S$ is a solution to

$$\begin{aligned} \max_{\mu_S\in\mathcal{M}_S^*} \quad & \sum_{t_i\in T_i} \lambda_i^S(t_i) U_i^S(\mu_S \mid t_i) \\ \text{s.t.} \quad & U_j^S(\mu_S \mid t_j) \geq U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j), \quad \forall j \in S \setminus i,\ t_j \in T_j. \end{aligned} \tag{4.3}$$

This linear-programming problem will be called the *primal problem for $\Gamma^i_S(\mu_{S\setminus i})$ w.r.t. λ_i^S*.

Remark 1. Because $\mu_{S\setminus i}$ is incentive compatible for $S \setminus i$, the optimization problem (4.3) is feasible. Indeed, let $\hat{\mu}_i \in \mathcal{M}_i$ be any mechanism defined by:

$$U_i^{\{i\}}(\hat{\mu}_i \mid t_i) = \max_{d_i\in D_i} u_i^{\{i\}}(d_i, t_i), \quad \forall t_i \in T_i. \tag{4.4}$$

Define the mechanism $\hat{\mu}_S \in \mathcal{M}_S$ by

$$\begin{aligned} \hat{\mu}_S([d_i, d_{S\setminus i}] \mid t_S) &= \hat{\mu}_i(d_i \mid t_i)\mu_{S\setminus i}(d_{S\setminus i} \mid t_{S\setminus i}), \quad \text{if } [d_i, d_{S\setminus i}] \in D_i \times D_{S\setminus i} \subseteq D_S \\ \hat{\mu}_S(d_S \mid t_S) &= 0, \quad \text{otherwise.} \end{aligned} \tag{4.5}$$

It can be easily checked that $\hat{\mu}_S$ is feasible in $\Gamma^i_S(\mu_{S\setminus i})$ whenever $\mu_{S\setminus i}$ is incentive compatible for $S \setminus i$.

[11] A formal justification for the inscrutability principle can be found in Myerson (1983, p. 1774).

Let $\alpha_j^S(\tau_j \mid t_j) \geq 0$ be the dual variable for the constraint that type t_j of player $j \in S$ should not gain by reporting τ_j in problem (4.3). Let also $\lambda_j^S(t_j) \geq 0$ denote the dual variable for the constraint that μ_S must give at least $U_j^S(\mu_{S \setminus i} \mid t_j)$ to type t_j of player $j \in S \setminus i$. Then using the concept of virtual utility, the Lagrangian for the primal problem for $\Gamma_S^i(\mu_{S \setminus i})$ can be written as

$$\mathcal{L}_S^i(\mu_S, \mu_{S\setminus i}, \lambda^S, \alpha^S) = \sum_{t_S \in T_S} p(t_S) \sum_{j \in S} v_j^S(\mu_S, t_S, \lambda_j^S, \alpha_j^S) - \sum_{j \in S \setminus i} \sum_{t_j \in T_j} \lambda_j^S(t_j) U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j). \quad (4.6)$$

The subtrahend in the above Lagrangian can also be conveniently expressed in terms of the virtual utilities. To do this, note that, because $\mu_{S\setminus i}$ is incentive efficient for $S \setminus i$, then by Proposition 1, there exists a vector $\alpha^{S\setminus i} \in A^{S\setminus i}$ such that

$$\alpha_j^{S\setminus i}(\tau_j \mid t_j)\left[U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j) - U_j^{S\setminus i}(\mu_{S\setminus i}, \tau_j \mid t_j)\right] = 0, \quad \forall j \in S \setminus i,\ \forall t_j, \tau_j \in T_j. \quad (4.7)$$

Hence, the following chain of equalities holds:

$$\begin{aligned}
&\sum_{j \in S \setminus i} \sum_{t_j \in T_j} \lambda_j^S(t_j) U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j) \\
&= \sum_{j \in S \setminus i} \sum_{t_j \in T_j} \lambda_j^S(t_j) U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j) \\
&\quad + \sum_{j \in S \setminus i} \sum_{t_j \in T_j} \sum_{\tau_j \in T_j} \alpha_j^{S\setminus i}(\tau_j \mid t_j)\left[U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j) - U_j^{S\setminus i}(\mu_{S\setminus i}, \tau_j \mid t_j)\right] \\
&= \sum_{t_{S\setminus i} \in T_{S\setminus i}} p(t_{S\setminus i}) \sum_{j \in S \setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i}). \qquad (4.8)
\end{aligned}$$

The equality (4.8) asserts that the λ^S-weighted sum of the players' reservation utilities in the game $\Gamma_S^i(\mu_{S\setminus i})$ equals the virtual worth that they expect from $\mu_{S\setminus i}$ given λ^S and $\alpha^{S\setminus i}$. Therefore, for any given vector $\alpha^{S\setminus i}$ satisfying (4.7), the Lagrangian in (4.6) can alternatively be formulated as:

$$\mathcal{L}_S^i(\mu_S, \mu_{S\setminus i}, \lambda^S, \alpha^S, \alpha^{S\setminus i}) = \sum_{t_S \in T_S} p(t_S) \sum_{j \in S} v_j^S(\mu_S, t_S, \lambda_j^S, \alpha_j^S) - \sum_{t_{S\setminus i} \in T_{S\setminus i}} p(t_{S\setminus i}) \sum_{j \in S \setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i}). \quad (4.9)$$

Proposition 2 (**Characterizing undominated mechanisms**).
A mechanism $\mu_S \in \mathcal{M}_S$ is undominated in $\Gamma_S^i(\mu_{S\setminus i})$ iff it is feasible in $\Gamma_S^i(\mu_{S\setminus i})$ and there exist vectors $\lambda^S \in \Lambda_+^S$ and $\alpha^S \in A^S$, such that

$$\lambda_i^S > 0, \quad (4.10a)$$

$$\alpha_j^S(\tau_j \mid t_j)\left[U_j^S(\mu_S \mid t_j) - U_j^S(\mu_S, \tau_j \mid t_j)\right] = 0, \quad \forall j \in S,\ \forall t_j, \tau_j \in T_j, \quad (4.10b)$$

$$\lambda_j^S(t_j)\left[U_j^S(\mu_S \mid t_j) - U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j)\right] = 0, \quad \forall j \in S \setminus i,\ \forall t_j \in T_j, \tag{4.10c}$$

and μ_S *maximizes the Lagrangian in* (4.9) *over all mechanisms in* $\mathcal{M}_S$*, namely,*

$$\sum_{j\in S} v_j^S(\mu_S, t_S, \lambda_j^S, \alpha_j^S) = \max_{d_S\in D_S} \sum_{j\in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S), \quad \forall t_S \in T_S. \tag{4.10d}$$

Remark 2. Propositions 1 and 2 imply that, whenever $\lambda_j^S > 0$ for all $j \in S \setminus i$, then an undominated mechanism is incentive efficient. More generally, an undominated mechanisms is weakly incentive efficient.

This Lagrangian analysis asserts that, for any combination of types, player i should select the decision in D_S that maximizes the virtual worth of coalition S, and then transfer virtual utility to the other players in $S \setminus i$ to compensate them according to $\mu_{S\setminus i}$ for their participation. The residual expected virtual utility for type t_i of player i is

$$\sum_{t_{S\setminus i}\in T_{S\setminus i}} p(t_{S\setminus i})\left(\max_{d_S\in D_S}\sum_{j\in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S) - \sum_{j\in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i})\right)$$

We may now ask, what allocations of real utility could correspond to this residual virtual payoff? We say that a utility allocation $\omega_i^S \in \mathbb{R}^{T_i}$ is *warranted* by λ^S, α^S, $\alpha^{S\setminus i}$ and $\mu_{S\setminus i}$ iff

$$\begin{aligned}\frac{1}{p(t_i)}&\left[\left(\lambda_i^S(t_i) + \sum_{\tau_i\in T_i}\alpha_i^S(\tau_i \mid t_i)\right)\omega_i^S(t_i) - \sum_{\tau_i\in T_i}\alpha_i^S(t_i \mid \tau_i)\omega_i^S(\tau_i)\right] \\ &= \sum_{t_{S\setminus i}\in T_{S\setminus i}} p(t_{S\setminus i})\left(\max_{d_S\in D_S}\sum_{j\in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S) - \sum_{j\in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i})\right), \quad \forall t_i \in T_i.\end{aligned} \tag{4.11}$$

The quantity $\omega_i^S(t_i)$ is called the *warranted claim* of type t_i of player i. Equations (4.11) implicitly define ω_i^S to be the vector of expected utilities for the possible types of player i that correspond to the residual expected virtual worth of coalition S once the players other than i have been rewarded according to their virtual reservation utilities from $\mu_{S\setminus i}$. The warrant equations are solvable in ω_i^S, and the solution is unique, provided that $\lambda^S \in \Lambda_{++}^S$ (see Lemma 1 in Myerson (1983)).

4.2. *Safe Mechanisms*

We say that a mechanism μ_S is *feasible in* $\Gamma_S^i(\mu_{S\setminus i})$ *given* t_i iff μ_S is incentive compatible for i (that is, it satisfies (3.2) for player i), and would be feasible for all players in $S \setminus i$ *after* they inferred that i's type is t_i. Formally, for all $j \in S \setminus i$, and every $t_j \in T_j$,

$$\begin{aligned}\sum_{t_{S\setminus\{i,j\}}\in T_{S\setminus\{i,j\}}} p(t_{S\setminus\{i,j\}}) &\sum_{d_S\in D_S}\mu_S(d_S \mid t_S)u_j^S(d_S, t_S) \\ &\geq \sum_{t_{S\setminus\{i,j\}}\in T_{S\setminus\{i,j\}}} p(t_{S\setminus\{i,j\}})\sum_{d_S\in D_S}\mu_S(d_S \mid t_i, \tau_j, t_{S\setminus\{i,j\}})u_j^S(d_S, t_S), \quad \forall \tau_j \in T_j,\end{aligned} \tag{4.12a}$$

and

$$\sum_{t_{S\setminus\{i,j\}}\in T_{S\setminus\{i,j\}}} p(t_{S\setminus\{i,j\}}) \sum_{d_S\in D_S} \mu_S(d_S \mid t_S) u_j^S(d_S, t_S) = U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j). \tag{4.12b}$$

Definition 3 (**Safe Mechanisms**).
A mechanism $\mu_S \in \mathcal{M}_S$ is safe in $\Gamma_S^i(\mu_{S\setminus i})$ iff, for every type $t_i \in T_i$, μ_S is feasible given t_i.

That is, a safe mechanism is one which would be implementable by player i no matter what the players in $S \setminus i$ might infer about i's type from his selection. Safe mechanisms do not necessarily exist for any given mechanism selection problem $\Gamma_S^i(\mu_{S\setminus i})$.

4.3. Strong Solutions

A safe and undominated mechanism is called a *strong solution*.[12] Two important facts justify why insisting on his strong solution is the most reasonable solution to player i, even when he might actually prefer some other mechanism (given his true type). On the one hand, no matter what the other players in S might infer about i's type from his selection of a strong solution, they would still be willing to participate honestly in the mechanism (because it is safe). On the other hand, if player i tried to demand some other feasible mechanism μ_S, and if the players in $S \setminus i$ inferred from this demand that player i must be in the set of types that prefer μ_S over the strong solution, then μ_S could not be feasible given this information—some type of a player $j \in S \setminus i$ would have incentives to lie or to insist in the status quo. Therefore, if player i has a strong solution, he cannot do better than to demand his strong solution. Any other demand would become infeasible as soon as it is selected, because of the information that it reveals.

Though compelling as a solution concept, a strong solution may not always exist. Yet, when it does, it is essentially unique, in the sense that, if μ_S and $\tilde{\mu}_S$ are both strong solutions for player i, then $U_i^S(\mu_S \mid t_i) = U_i^S(\tilde{\mu}_S \mid t_i)$ for all $t_i \in T_i$. Therefore, whenever a strong solution exists, there is no ambiguity about what a solution of $\Gamma_S^i(\mu_{S\setminus i})$ should be.

5. The General Bargaining Solution

The idea in defining our bargaining solution was already anticipated in the introduction and illustrated in the example of Section 2. The solution is constructed by induction on the size of coalitions, recursively applying the conditional random dictatorship procedure to growing coalitions. For any player i, let $\mu_i \in \mathcal{M}_i$ be a mechanism satisfying (4.4). Clearly, this is the best player i can do without any other player's help. Let $S \subseteq N$ be a coalition and assume that, for each $i \in S$, the members of $S \setminus i$ have

[12] Our definition of a safe mechanism differs slightly from the one originally introduced by Myerson (1983) in that we require exact feasibility in (4.12b). Therefore, our definition of a strong solution is closer to the "strong optima" defined in Myerson (1984a). As in Myerson (1984a), the use of this stronger class of safe mechanisms allows us to obtain an axiomatic characterization of our bargaining solution—permitting the possibility of inequalities in (4.12b) would invalidate our Theorem 6 (see footnote 18). However, this restriction does not modify the properties of a strong solution.

already agreed on the mechanism $\mu_{S\setminus i}$ that they would use as a threat in case negotiations with player i break down. The threat of coalition S is then computed applying the conditional random dictatorship procedure: a player $i \in S$ is picked at random, with all players in S having equal probability. Player i then faces the mechanism selection problem $\Gamma_S^i(\mu_{S\setminus i})$—he has the authority to demand any incentive-compatible mechanism constrained only by the participation of the other players in S, who have the outside option to form coalition $S \setminus i$, in which case they carry out $\mu_{S\setminus i}$ (by induction). Assume that every $i \in S$ has a strong solution μ_S^i that the he will select in $\Gamma_S^i(\mu_{S\setminus i})$. The expected outcome of such a conditional random dictatorship is equivalent to the outcome of the "average" mechanism $\mu_S = \frac{1}{|S|}\sum_{i\in S}\mu_S^i$. From the point of view of procedural justice, the mechanism μ_S can be regarded as an "equitable" agreement for coalition S, since it reproduces the same *expected* outcome that could be reached when each player gets equal opportunity to control the final decision. If μ_S happens to be incentive efficient for S, then it is defined to be the threat that S would carry out in the event that S forms. A vector of threats $\eta_N = (\mu_S)_{S\subseteq N}$ constructed following this *recursive conditional random dictatorship procedure* is then called a bargaining solution of the game Γ_N.

The requirement that, for each coalition S, the average mechanism μ_S be incentive efficient is quite restrictive, and is likely to be met by relatively few games. In fact, this condition will only be satisfied by coalitions in which the payoff allocations implemented by the various strong solutions $(\mu_S^i)_{i\in S}$ lie on a segment of the incentive-efficient frontier that coincides with a hyperplane. To appreciate the implications of this restriction, note that, being undominated, each strong solution is characterized by vectors $\lambda^S \in \Lambda_+^S$ and $\alpha^S \in A^S$ for which conditions (4.10a)–(4.10d) are satisfied. The various virtual scales (λ^S, α^S) may well depend on i; but μ_S can only be incentive efficient if the mechanisms $(\mu_S^i)_{i\in S}$ are all supported by the *same* virtual scales. However, if we insist on requiring these virtual scales to be the same across all mechanisms $(\mu_S^i)_{i\in S}$, then some of them may not be safe.

On the other hand, the construction in the recursive conditional random dictatorship procedure requires that every player has a strong solution in every possible coalition he may belong to. However, as previously noted, many mechanism selection problems have no strong solution. As a result, our bargaining solution, as it has been defined above, may not exist for a large number of games. We need then to reformulate our bargaining solution if we want to prove a general existence theorem. An alternative definition must, however, still yield the same expected outcome as the recursive conditional random dictatorship procedure, provided that all players have a strong solution in every coalition and the randomizations result in vectors of threats that are incentive efficient. The idea to solve this predicament is motivated by the use of the *independence of irrelevant alternatives* (*IIA*) axiom introduced by Nash (1950). This axiom says that the bargaining solution of the game should not change as the set of feasible outcomes is reduced, so long as the disagreement point remains unchanged, and the original solution (in the larger problem) remains feasible in the reduced problem.

Lemma 1 (Restricted extension property).
Given a coalition $S \subseteq N$, let $(\mu_{S\setminus i})_{i\in S}$ be a vector of incentive-compatible threats. Suppose that, for each $i \in S$, the vector ω_i^S is warranted by $\lambda^S \in \Lambda_{++}^S$, $\alpha^S \in A^S$, $\alpha^{S\setminus i} \in A^{S\setminus i}$, and $\mu_{S\setminus i}$. Then there exists

an extended set of decision options $\tilde{D}_S \supseteq D_S$ such that, for each $i \in S$, there is a strong solution $\tilde{\mu}^i_S$ of the corresponding extended mechanism-selection problem $\tilde{\Gamma}^i_S(\mu_{S\setminus i})$ satisfying that $\tilde{U}^S_i(\tilde{\mu}^i_S \mid t_i) = \omega^S_i(t_i)$ for all $t_i \in T_i$.[13] *Moreover, the mechanism $\tilde{\mu}_S = \frac{1}{|S|}\sum_{i\in S}\tilde{\mu}^i_S$ is incentive-efficient for S in the extended problem.*

The proof of Lemma 1 is deferred to Appendix B. The restricted extension property says that, so long as the "disagreement" threats $(\mu_{S\setminus i})_{i\in S}$ remain fixed (hence the name "restricted"), it is possible to extend the set of feasible decision options for coalition S in a way that each member $i \in S$ has a strong solution, $\tilde{\mu}^i_S$, in the extended problem, that gives each of his types its warranted claim. Thus, given that $\tilde{\mu}_S = \frac{1}{|S|}\sum_{i\in S}\tilde{\mu}^i_S$ is incentive efficient, it must be considered the threat that coalition S would carry out if it were to form *in the extended problem*. This mechanism $\tilde{\mu}_S$ gives every type t_i of a player $i \in S$ an expected payoff that equals the average between what he could get from his strong solution when he is a dictator for S (i.e., $\omega^S_i(t_i)$), and from the various disagreement threats, $(\mu_{S\setminus j})_{j\in S\setminus i}$, when any other player $j \in S \setminus i$ is a dictator for S. Clearly, the mechanism $\tilde{\mu}_S$ may not be feasible in the original problem, since it may require to put positive probability weight on decisions in $\tilde{D}_S \setminus D_S$. However, suppose that there exists an incentive-efficient mechanism μ_S *in the original problem* that is (interim) utility equivalent to $\tilde{\mu}_S$, that is,

$$U^S_i(\mu_S \mid t_i) = \frac{1}{|S|}\left[\sum_{j\in S\setminus i} U^{S\setminus j}_i(\mu_{S\setminus j} \mid t_i) + \omega^S_i(t_i)\right], \quad \forall i \in S, \quad \forall t_i \in T_i. \tag{5.1}$$

Thus, by a use of the IIA axiom, the mechanism μ_S should be regarded as a reasonable threat for coalition S *in the original problem*. Indeed, all decisions in $\tilde{D}_S \setminus D_S$ may be seen as irrelevant to the extent that they are not actually required to achieve the final outcome. Thus, eliminating those unchosen alternatives should not affect the selection of μ_S as a threat for S. The decisions in $\tilde{D}_S \setminus D_S$ are only important in sofar as they help to figure out what a fair and efficient agreement should look like for coalition S in the original problem.

A question still remains: What ensures that such a utility-equivalent mechanism actually exists in the original problem? The existence can be guaranteed by a standard fixed-point argument adjusting the virtual scales (λ^S, α^S). However, notice that the restricted extension property holds provided that *all* utility weights in λ^S are strictly positive. This is so because we only know the warrant equations (4.11) to be solvable when $\lambda^S \in \Lambda^S_{++}$ (see Lemma 1 in Myerson, 1983). But the Kakutani fixed-point theorem cannot be applied to the interior of a simplex.[14] To remedy this problem, we follow Myerson's idea to slightly enlarge the solution set by adding points that are reasonable as emerging from some topological closure. Thus, we define a *neutral procedural solution* to be a vector of

[13] Here $\tilde{U}^S_i(\cdot)$ is defined by the analog of (3.1) for the problem $\tilde{\Gamma}^i_S$ instead of Γ^i_S.

[14] This same dilemma also appears in the case of complete information, where the solution adopted by Shapley (1969) was to admit all utility-weight vectors in $\mathbb{R}^N_+ \setminus \{0\}$. This resolution has been shown to be unsatisfactory, insofar as it is possible to construct games where the λ-transfer value includes too many points, many of them not being efficient and/or individually rational.

incentive-efficient threats $(\mu_S)_{S\subseteq N}$ such that, for each coalition S, there exists a restricted extension (given the threats $(\mu_{S\setminus i})_{i\in S}$) in which an equitable and incentive-efficient mechanism gives every type of a player in S an expected payoff that does not exceed what it gets from μ_S by more than an arbitrarily small amount.[15]

Definition 4 (**Neutral Procedural Solution**).

A vector of threats $\eta_N = (\mu_S)_{S\subseteq N}$ is a neutral procedural solution of the game Γ_N if, for every coalition $S \subseteq N$, μ_S is incentive efficient for S, and for each $\epsilon > 0$ there exist vectors $\lambda^S \in \Lambda^S_{++}$, $\alpha^S \in A^S$, and $\omega^S \in \prod_{i\in S}\mathbb{R}^{T_i}$ such that:

NPS1. For every $i \in S$, ω^S_i is warranted by λ^S, α^S, $\alpha^{S\setminus i}$, and $\mu_{S\setminus i}$,

NPS2. $U^S_i(\mu_S \mid t_i) \geq \frac{1}{|S|}\left[\sum_{j\in S\setminus i} U^{S\setminus j}_i(\mu_{S\setminus j} \mid t_i) + \omega^S_i(t_i)\right] - \epsilon, \quad \forall t_i \in T_i,\ \forall i \in S.$

The vector of expected utilities $U^N(\mu_N) = (U^N_i(\mu_N \mid t_i))_{i\in N,\, t_i\in T_i}$ is called a neutral procedural value of the game Γ_N.

Remark 3. Note that the vectors λ^S, α^S, $\alpha^{S\setminus i}$, and ω^S_i may depend on ϵ. Even though for reasons of notational simplicity, this dependency has not been made explicit.

Remark 4. The vector $\alpha^{S\setminus i}$ can always be taken to satisfy (4.7), which is possible since $\mu_{S\setminus i}$ is incentive efficient for S. Therefore, $\alpha^{S\setminus i}$ can be set independent of ϵ.

Theorem 1 (**Existence**).

For any game Γ_N, there exists at least one neutral procedural solution.

We defer the proof of Theorem 1 to Appendix A.

Note that if the conditions *NPS1–NPS2* can be satisfied by some vectors λ^S, α^S, $\alpha^{S\setminus i}$, and ω^S_i with $\epsilon = 0$, then the same vectors will also satisfy these conditions for any $\epsilon > 0$. In this case, we can dispense with the semi-continuity requirement. We say that $(\mu_S)_{S\subseteq N}$ is a *non-degenerate* solution if we can find $\lambda^S \in \Lambda^S_{++}$, $\alpha^S \in A^S$, and $\omega^S \in \prod_{i\in S}\mathbb{R}^{T_i}$ that satisfy *NPS1–NPS2* with $\epsilon = 0$ for each $S \subseteq N$ (as it is possible in the examples of this paper).

To appreciate the implications of conditions *NPS1–NPS2*, let $(\mu_S)_{S\subseteq N}$ be a non-degenerate solution. Then, for each $S \subseteq N$, the mechanism μ_S is an optimal solution of the primal (3.5) w.r.t. λ^S, and α^S is an optimal solutions of the dual (3.7) w.r.t. λ^S. In particular, μ_S satisfies the conditions in Proposition 1 for the virtual scales (λ^S, α^S). To see this, we proceed by induction on the size of S. Assume that the

[15]This same idea also appears more explicitly in Myerson's (1984a) characterization of his two-person bargaining solution (see Theorem 4 in that paper). Likewise, it is used by Myerson (1984b) in his generalization of the λ-transfer value, and by Myerson (1991a, sec. 9.8) to define the inner core.

statement above is true for all smaller coalitions $R \subset S$. Consider the following chain of inequalities:

$$\begin{aligned}
&|S| \sum_{i \in S} \sum_{t_i \in T_i} \lambda_i^S(t_i) U_i^S(\mu_S \mid t_i) \\
&\quad \geq \sum_{i \in S} \sum_{t_i \in T_i} \lambda_i^S(t_i) \omega_i^S(t_i) + \sum_{i \in S} \sum_{j \in S \setminus i} \sum_{t_i \in T_i} \lambda_i^S(t_i) U_i^{S \setminus j}(\mu_{S \setminus j} \mid t_i) \\
&\quad = \sum_{i \in S} \sum_{t_i \in T_i} \lambda_i^S(t_i) \omega_i^S(t_i) + \sum_{i \in S} \sum_{j \in S \setminus i} \sum_{t_i \in T_i} \lambda_j^S(t_j) U_j^{S \setminus i}(\mu_{S \setminus i} \mid t_j) \\
&\quad = \sum_{i \in S} \sum_{t_S \in T_S} p(t_S) \left[\max_{d_S \in D_S} \sum_{j \in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S) - \sum_{j \in S \setminus i} v_j^{S \setminus i}(\mu_{S \setminus i}, t_{S \setminus i}, \lambda_j^S, \alpha_j^{S \setminus i}) \right]. \\
&\qquad + \sum_{i \in S} \sum_{j \in S \setminus i} \sum_{t_j \in T_j} \lambda_j^S(t_j) U_j^{S \setminus i}(\mu_{S \setminus i} \mid t_j) \\
&\quad = |S| \sum_{t_S \in T_S} p(t_S) \max_{d_S \in D_S} \sum_{j \in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S),
\end{aligned} \tag{5.2}$$

where the inequality in the first line follows from condition *NPS2*; the equality in the second line rearranges the terms in the second summation; the equality in the third line is obtained after summing the warrant equations (i.e., *NPS1*); and finally, the equality in the fourth line uses (4.8), since $\alpha^{S \setminus i}$ satisfies (4.7) by the induction hypothesis.

On the other hand, since μ_S is feasible in the primal (3.5) w.r.t. λ^S, and $\alpha^S \geq 0$ is feasible in the dual (3.7) w.r.t. λ^S, weak duality implies that

$$\sum_{i \in S} \sum_{t_i \in T_i} \lambda_i^S(t_i) U_i^S(\mu_S \mid t_i) \leq \sum_{t_S \in T_S} p(t_S) \max_{d_S \in D_S} \sum_{j \in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S).$$

Therefore, the inequalities in (5.2) hold as equality. Strong duality then yields the desired result.

Another interesting implication from the equality in (the first line of) (5.2) is that non-degenerate solutions satisfy (5.1), that is, condition *NPS2* holds with equality. Note that

$$\sum_{i \in S} \sum_{t_i \in T_i} \lambda_i^S(t_i) \left[U_i^S(\mu_S \mid t_i) - \frac{1}{|S|} \left(\sum_{j \in S \setminus i} U_i^{S \setminus j}(\mu_{S \setminus j} \mid t_i) + \omega_i^S(t_i) \right) \right] = 0.$$

Since $\lambda^S > 0$, then condition *NPS2* implies that all terms inside the bracket in the above summation must be zero.

More generally, the definition of a neutral procedural solution allows the possibility that some utility weights vanish in the limit when ϵ goes to zero. Therefore, we cannot exclude that eventually only "degenerate" solutions exist. However, letting $\epsilon \to 0$ and taking a convergent subsequence if necessary, it is always possible to find vectors λ^S, α^S, and ω^S for which conditions *NPS1*–*NPS2* are satisfied for $\epsilon = 0$ with $\lambda^S \in \Lambda_+^S$. Hence, using the same reasoning as above we can establish the following general result.

Theorem 2 (**Necessary conditions**).
Let $\eta_N = (\mu_S)_{S\subseteq N}$ be a neutral procedural solution of Γ_N. Then for each $S \subseteq N$, there exist vectors $\lambda^S \in \Lambda_+^S$, $\alpha^S \in A^S$, and $\omega^S \in \prod_{i\in S} \mathbb{R}^{T_i}$ such that

(i) *μ_S is an optimal solution of the primal (3.5) w.r.t. λ^S,*
(ii) *α^S is an optimal solutions of the dual (3.7) w.r.t. λ^S,*
(iii) *for each $t_i \in T_i$ of every $i \in S$, either (5.1) holds or $\lambda_i^S(t_i) = 0$.*

Non-degenerate solutions are also individually rational in the sense of (3.3). This result readily follows from Lemma 2 in Appendix B.

Theorem 3 (**Individual rationality**).
Let $\eta_N = (\mu_S)_{S\subseteq N}$ be a non-degenerate neutral procedural solution of Γ_N. Then for each $S \subseteq N$, μ_S is individually rational., i.e., it satisfies (3.3).

Even though we want to determine how the proceeds of cooperation within the grand coalition should be shared, fairness in N is affected by what is thought to happen if a smaller coalition S would form instead of N. We call the allocation achieved by S the *threat* from S. Our bargaining solution concept determines the threats μ_S for coalitions $S \subset N$ exactly in the same way μ_N is determined for N; this property—that μ_S is the bargaining solution of the subgame Γ_S for each $S \subseteq N$—is called "subcoalition perfectness" by Hart and Mas-Colell (1996). In this sense, coalitional threats constitute a "credible" pre-commitment similar to the notion of credibility in subgame perfection. In contrast, Myerson's (1984b) bargaining solution specifies "rational threats" that may be manifestly unfair. Indeed, for the M-solution, only the final agreement in the grand coalition is required to be equitable. Consider for instance the example in Section 2: the mechanism that gives the whole gains from trade in both states to the buyer is a rational threat for coalition $\{1, 2\}$. However, such a threat cannot be considered credible, since the broker could hardly believe that the seller would agree to accept an offer of a take-it-or-leave-it price of \$0 in case he (the broker) refuses to cooperate. Just as our credible threats express fairness using the same random dictatorship procedure in all coalitions, Salamanca's (2020) "egalitarian threats" apply a same equity principle to all coalitions. This principle requires that, for any two members of a coalition S, the average amount that each player would expect to gain from the cooperation of all other members of S should be equal (when utility comparisons are made in the virtual utility scales). However, unlike our bargaining solution, the S-solution does not satisfy subcoalition perfectness.

Subcoalition perfectness also characterizes another cooperative solution concept for cooperative games with *complete* information: the *consistent NTU value* introduced by Maschler and Owen (1989, 1992). A cooperative game Γ_N is said to have *complete information* if T_i is a singleton for every $i \in N$. For such games there are no incentive constraints and, therefore, we can set all dual variables $(\alpha^S)_{S\subseteq N}$ to zero. The virtual utility formula then reduces to λ^S-weighted utility and the conditions for a non-degenerate neutral procedural solution become

$$\lambda_i^S U_i^S(\mu_S) = \frac{1}{|S|}\left[W_S(\lambda^S) - \sum_{j\in S\setminus i} \lambda_j^S U_j^{S\setminus i}(\mu_{S\setminus i}) + \sum_{j\in S\setminus i} \lambda_i^S U_i^{S\setminus j}(\mu_{S\setminus j})\right], \tag{5.3}$$

where $W_S(\lambda^S) := \max_{d_S\in D_S} \sum_{j\in S} \lambda_j^S u^S(d_S)$. Equation (5.3) recursively characterizes the consistent NTU value (cf. equation (4) in Hart, 2004). We summarize this result in the following theorem.

Theorem 4 (Games with complete information).
Let Γ_N be a cooperative game with complete information. The vector of threats $\eta_N = (\mu_S)_{S\subseteq N}$ is a non-degenerate neutral procedural solution of Γ_N iff the vector payoff configuration $(U^S(\mu_S))_{S\subseteq N}$ is a consistent NTU value of Γ_N.

When defining the consistent NTU value, Maschler and Owen (1992) explicitly took $\lambda^S > 0$ for all coalitions S. Because utility weights are endogenous, to guarantee that whatever emerges from their definition is always associated with strictly positive utility weights, they restricted attention to a domain of games for which the feasible set of each coalition is "positively smooth" (i.e., there is a unique supporting hyperplane at each point of the Pareto frontier, with a normal vector that has all coordinates positive). Although their definition naturally extends to non-smooth games, their existence result does not. Thus, when specialized to the case of complete information, the neutral procedural solution also appears to be an appropriate extension of the consistent NTU value to non-smooth games. Obviously, to the extent that the consistent NTU value extends the Shapley value to general games where the players may not be able to make side payments to each other, the neutral procedural value constitutes a valid generalization of the Shapley value to Bayesian cooperative games.

5.1. Axiomatic Characterization

Myerson (1984a) derived a concept of *neutral bargaining solution* that extends the Nash bargaining solution to two-person bargaining problems with incomplete information. His solution forms the smallest set satisfying two postulates: a random-dictatorship axiom and an extension axiom. The first axiom prescribes what the solution should look like for the class of problems where choosing with equal probability one individual to act as a dictator yields an expected outcome that is incentive efficient. The second axiom relates the solutions of a bargaining problem to the solutions of its extensions (like the IIA).

A *two-person bargaining problem with incomplete information* is a Bayesian cooperative game satisfying: $n = 2$, $D_i = \{d_i\}$ for all $i \in N$, and $u_i(d^*, t_N) = 0$ for all $i \in N$ and $t_N \in T_N$, where $d^* := [d_1, d_2]$ is the disagreement outcome. That is, in the absence of agreement, a player can only enforce the disagreement outcome, which leaves both players with a normalized utility equal to zero in every state.

Theorem 5 (Two-person bargaining problems).
Let Γ_N be a two-person bargaining problem with incomplete information. The mechanism μ_N is a

neutral bargaining solution of Γ_N *iff it is a neutral procedural solution of* Γ_N.

Proof. The result is obtained immediately from Theorem 4 in Myerson (1984a) after dividing both sides of the warrant equations by two and redefining ω_i^N to be $\frac{\omega_i^N}{2}$. □

According to Theorem 5, our solution concept can be seen as a valid generalization of the neutral bargaining solution to games with more than two players. More importantly, this result suggests a way to obtain an axiomatic characterization of the neutral procedural solutions. For that, we first need to appropriately extend Myerson's (1984a) axioms to games involving more than two players.

A *bargaining solution correspondence* for Bayesian cooperative games is a set-valued mapping $BS(\cdot)$ that assigns to each Bayesian cooperative game a set of vectors $\eta_N = (\mu_S)_{S\subseteq N}$ of coalitional threats.[16]

Axiom RCRD (Restricted conditional random dictatorship). Let $\eta_N = (\mu_S)_{S\subseteq N}$ be a vector of threats in Γ_N satisfying that, for each $i \in N$, $\eta_{N\setminus i} = (\mu_{S\setminus i})_{S\subseteq N} \in BS(\Gamma_{N\setminus i})$. Suppose that each player $i \in N$ has a strong solution, μ_N^i, of the mechanism-selection problem $\Gamma_N^i(\mu_{N\setminus i})$. Suppose also that the mechanism $\mu_N := \frac{1}{|N|}\sum_{i\in N}\mu_N^i$ is incentive efficient for N. Then $\eta_N \in BS(\Gamma_N)$.

As it should be clear, *RCRD* is justified by the conditional random dictatorship procedure (hence its name). The hypotheses of this axiom are quite restrictive, which means that this axiom is very weak. On the one hand, many mechanism-selection problems do not have a strong solution. On the other hand, randomization between strong solutions may not be incentive efficient. When restricted to two-person bargaining problems, *RCRD* reduces to Myerson's (1984a) random dictatorship axiom. However, in larger games with $|N| > 2$, *RCRD* is even weaker, as it additionally requires the "disagreement" threats to be bargaining solutions of the corresponding subgames.

When applied recursively, *RCRD* can be used to construct a bargaining solution. Define the vector of threats $\eta_N^D = (\mu_S^D)_{S\subseteq N}$ (the superscript stands for "dictator") recursively as follows:

(*RCRD*.a) For all $i \in N$, μ_i^D satisfies (4.4).

(*RCRD*.b) For all $S \subseteq N$ with $|S| \geq 2$, $\mu_S^D := \frac{1}{|S|}\sum_{i\in S}\mu_S^i$, where for each $i \in S$, μ_S^i is a strong solution of the mechanism design problem $\Gamma_S^i(\mu_{S\setminus i}^D)$.

The vector of threats η_N^D is well defined whenever all strong solutions in (b) exist. According to *RCRD*, if μ_S^D is incentive efficient for all $S \subseteq N$, then η_N^D should be considered a bargaining solution of Γ_N. This construction suggests defining the following weaker axiom.

Axiom RCRD' (Recursive conditional random dictatorship). Suppose that the vector of threats $\eta_N^D = (\mu_S^D)_{S\subseteq N}$ is well defined in the game Γ_N. If μ_S^D is incentive efficient for each $S \subseteq N$, then $\eta_N^D \in BS(\Gamma_N)$.

The *RCRD'* is akin to axiom 7 in de Clippel et al. (2004), which is used by the authors to axiomatize the consistent NTU value. Clearly, a bargaining solution correspondence that satisfies *RCRD* will also satisfy *RCRD'*.

[16]When defining a bargaining solution correspondence, we consider vectors of threats and not just the mechanism of the grand coalition. In contrast, Myerson (1984b) and Salamanca (2020) deal explicitly only with μ_N, the mechanisms μ_S for all coalitions $S \neq N$ are, however, implicitly defined in their bargaining solutions.

Given two Bayesian cooperative games Γ_N and $\tilde{\Gamma}_N$, we say that $\tilde{\Gamma}_N$ is a *restricted extension* of Γ_N iff both games can be written in the form

$$\begin{aligned} \Gamma_N &= \{N, (D_S)_{S \subseteq N}, (T_i, u_i, p_i)_{i \in N}\}, \\ \tilde{\Gamma}_N &= \{N, (\tilde{D}_S)_{S \subseteq N}, (T_i, \tilde{u}_i, p_i)_{i \in N}\}, \end{aligned}$$

where $D_N \subseteq \tilde{D}_N$, $D_S = \tilde{D}_S$ for all $S \subset N$, and $u_i(d,t) = \tilde{u}_i(d,t)$ for each $(d,t) \in D_N \times T_N$ and every $i \in N$. That is, $\tilde{\Gamma}_N$ differs from Γ_N only in that more decision options are available for N in $\tilde{\Gamma}_N$.

Axiom RE (Restricted extension). Let $\eta_N = (\mu_S)_{S \subseteq N}$ be a vector of threats in Γ_N such that, for each $i \in N$, $\eta_{N \setminus i} = (\mu_{S \setminus i})_{S \subseteq N} \in BS(\Gamma_{N \setminus i})$, and μ_N is incentive efficient for N. Suppose that, for each positive number ϵ, there exists a restricted extension of Γ_N, denoted $\tilde{\Gamma}^\epsilon_N$, and an incentive-compatible mechanism $\tilde{\mu}^\epsilon_N$ for $\tilde{\Gamma}^\epsilon_N$ such that

(RE.a) $\tilde{\eta}^\epsilon_N = ((\mu_S)_{S \subset N}, \tilde{\mu}^\epsilon_N) \in BS(\tilde{\Gamma}^\epsilon_N)$,
(RE.b) $U^N_i(\mu_N \mid t_i) \geq \tilde{U}^N_i(\tilde{\mu}^\epsilon_N \mid t_i) - \epsilon, \quad \forall i \in N,\ \forall t_i \in T_i$.[17]

Then the vector of threats $\eta_N \in BS(\Gamma_N)$.

Axiom *RE* combines the arguments of the IIA axiom together with a kind of upper-semicontinuity condition. When the hypothesis of the axiom are satisfied, it is possible to increase the set of decision options available to the grand coalition (without changing the decision options of subcoalitions) in such a way that all players would be willing to settle on an allocation that is *almost* (interim) Pareto dominated by μ_N. Thus, *Axiom RE* asserts that the players ought to be willing to settle on μ_N when the extra options are not available. When restricted to the family of two-person bargaining problem, *RE* coincides with Myerson's (1984a) extension axiom.

Theorem 6 (**Axiomatic characterization**).
The neutral procedural solution is the minimal (relative to set-inclusion) bargaining solution correspondence satisfying the axioms RCRD and RE. In other words,

- *The neutral procedural solution satisfies axioms RCRD and RE.*
- *If another bargaining solution correspondence satisfies these two axioms, then it must contain the neutral procedural solutions.*

Unfortunately, we are not able to provide an exact axiomatic characterization. The works of de Clippel et al. (2004) and de Clippel and Minelli (2004) suggest that to obtain the maximality of the neutral procedural solution, regularity conditions need to be imposed on the set of feasible interim utility allocations. However, the way incentive constraints restrict what is feasible at the interim stage makes it considerably more difficult to derive conditions on the primitives of the game guaranteeing

[17] Here $\tilde{U}^N_i(\tilde{\mu}^\epsilon_N \mid t_i)$ denotes the expected utility of type t_i of player i from the mechanism $\tilde{\mu}^\epsilon_N$ in the game $\tilde{\Gamma}^\epsilon_N$.

the desired properties of the feasible utility sets. For instance, even if the ex-post games are well behaved, for example TU, the set of feasible interim utilities need not be positively smooth.

Proof. We show first that the neutral procedural solution satisfies the *RCRD* axiom. Let $\eta_N = (\mu_S)_{S \subseteq N}$ be a vector of threats in Γ_N satisfying that, for each $i \in N$, $\eta_{N\setminus i} = (\mu_{S\setminus i})_{S\subseteq N}$ is a neutral procedural solution of the subgame $\Gamma_{N\setminus i}$. Thus, for every $S \subset N$ and every $\epsilon > 0$ there exist vectors λ^S, α^S, and ω^S satisfying conditions *NPS1–NPS2* for μ_S. For every $i \in N$, let μ^i_N be a strong solution of $\Gamma^i_N(\mu_{N\setminus i})$. Define $\mu_N := \frac{1}{|N|} \sum_{i\in N} \mu^i_N$, and assume that μ_N is incentive efficient for N. We need to verify that for every $\epsilon > 0$ there exist vectors λ^N, α^N, and ω^N satisfying conditions *NPS1–NPS2* for μ_N.

Because μ_N is incentive efficient for N, there exist vectors $\lambda^N \in \Lambda^N_{++}$ and $\alpha^N \in A^N$ such that μ_N is an optimal solution of the primal (3.5) w.r.t. λ^N, and α^N is an optimal solution of the dual (3.7) w.r.t. λ^N. Let ω^N be such that, for every $i \in N$, ω^N_i is warranted by λ^N, α^N, $\alpha^{N\setminus i}$, and $\mu_{N\setminus i}$, where $\alpha^{N\setminus i} \in A^{N\setminus i}$ is any vector satisfying (4.7) (i.e., condition *NPS1*). Lemma 1 in Myerson (1983) guarantees that ω^N_i exists and is uniquely determined. Let $\tilde{\Gamma}_N$ be the restricted extension constructed in Lemma 1.

By giving zero probability to the new outcomes in $\tilde{D}_N \setminus D_N$, the strong solutions $(\mu^i_N)_{i\in N}$ can be made feasible in $\tilde{\Gamma}_N$. Therefore, for every $i \in N$, we must have that μ^i_N is also incentive efficient for N in $\tilde{\Gamma}_N$. This is so because μ^i_N is still an optimal solution of the primal problem (3.5) for λ^N in the context of $\tilde{\Gamma}_N$ (cf. (B.1c)).

Fix a player $i \in N$ and let $\tilde{\Gamma}^i_N(\mu_{N\setminus i})$ be player i's mechanism selection problem in $\tilde{\Gamma}_N$. Then by Lemma 1, there is a strong solution, $\tilde{\mu}^i_N$, of $\tilde{\Gamma}^i_N(\mu_{N\setminus i})$ that gives player i the vector of expected utilities ω^N_i. Clearly, μ^i_N is safe in the extended problem $\tilde{\Gamma}^i_N(\mu_{N\setminus i})$. Then we must have that

$$\omega^N_i(t_i) = \tilde{U}^N_i(\tilde{\mu}^i_N \mid t_i) \geq U^N_i(\mu^i_N \mid t_i), \qquad \forall t_i \in T_i. \tag{5.4a}$$

To see this, suppose that $P = \{t_i \in T_i \mid \tilde{U}^N_i(\tilde{\mu}^i_N \mid t_i) < U^N_i(\mu^i_N \mid t_i)\} \neq \emptyset$. Consider the mechanism $\bar{\mu}_N$ defined by

$$\bar{\mu}^i_N(d_N \mid t_N) = \begin{cases} \mu^i_N(d_N \mid t_N), & \text{if } t_i \in P; \\ \tilde{\mu}^i_N(d_N \mid t_N), & \text{if } t_i \notin P. \end{cases}$$

Then $\tilde{\mu}^i_N$ is dominated by $\bar{\mu}^i_N$ in $\tilde{\Gamma}^i_N(\mu_{N\setminus i})$. Because μ^i_N and $\tilde{\mu}^i_N$ are both feasible in $\tilde{\Gamma}^i_N(\mu_{N\setminus i})$ *given* every $t_i \in T_i$ (as they are both safe), then $\bar{\mu}^i_N$ is feasible in $\tilde{\Gamma}^i_N(\mu_{N\setminus i})$; but this contradicts the fact that $\tilde{\mu}^i_N$ is undominated in $\tilde{\Gamma}^i_N(\mu_{N\setminus i})$. Hence, $P = \emptyset$.

On the other hand, notice that $\tilde{U}^N_j(\tilde{\mu}^i_N \mid t_j) = U^N_j(\mu^i_N \mid t_j) = U^{N\setminus i}_j(\mu_{N\setminus i} \mid t_j)$ for all $j \in N \setminus i$ and $t_j \in T_j$, since both $\tilde{\mu}^i_N$ and μ^i_N are safe in $\tilde{\Gamma}^i_N(\mu_{N\setminus i})$. Suppose that (5.4a) holds as strict inequality for some $t_i \in T_i$. Then $\tilde{\mu}^i_N$ would give a higher expected utility to at least one type of player i without giving a lower expected utility to some other type of any other player. However, this contradicts the fact that μ^i_N is incentive efficient for N in $\tilde{\Gamma}_N(\mu_{N\setminus i})$.[18] Then we must conclude that all inequalities in (5.4a) must hold as equality and therefore, μ^i_N and $\tilde{\mu}^i_N$ are utility equivalent. That is, μ^i_N is also a strong solution

[18]This contradiction rests upon the fact that we have defined a safe mechanism using equalities in (4.12b) (i.e., binding

of $\tilde{\Gamma}_N(\mu_{N\setminus i})$. Because the same is true for all $i \in N$, then μ_N verifies (5.1) (i.e., condition *NPS2* with $\epsilon = 0$).

To check the restricted extension axiom, let $\eta_N = (\mu_S)_{S\subseteq N}$, $\tilde{\Gamma}^\epsilon_N$, and $\tilde{\eta}^\epsilon_N$ satisfy the hypotheses of the axiom *RE* with the neutral procedural solution as a bargaining solution correspondence. Thus, for every $S \subset N$ and every $\epsilon > 0$ there exist vectors λ^S, α^S, and ω^S satisfying conditions *NPS1–NPS2* for μ_S. Furthermore, for every $\epsilon > 0$, $\tilde{\eta}^\epsilon_N$ is a neutral procedural solution of $\tilde{\Gamma}^\epsilon_N$. Thus, for each $\epsilon > 0$, we can select $\tilde{\lambda}^{N,\epsilon} \in \Lambda^N_{++}$, $\tilde{\alpha}^{N,\epsilon} \in A^N$, $\tilde{\alpha}^{N\setminus i} \in A^{N\setminus i}$, and $\tilde{\omega}^{N,\epsilon} \in \prod_{i\in N}\mathbb{R}^{T_i}$ such that

(a) For every $i \in N$, $\tilde{\omega}^{N,\epsilon}_i$ is warranted by $\tilde{\lambda}^{N,\epsilon}$, $\tilde{\alpha}^{N,\epsilon}$, $\tilde{\alpha}^{N\setminus i}$, and $\mu_{N\setminus i}$ in $\tilde{\Gamma}^\epsilon_N$ (i.e., using the maximum over $\tilde{D}^\epsilon_N$, instead of D_N).

(b) $\tilde{U}^N_i(\tilde{\mu}^\epsilon_N \mid t_i) \geq \frac{1}{|N|}\left[\sum_{j\in N\setminus i} U^{N\setminus j}_i(\mu_{N\setminus j} \mid t_i) + \tilde{\omega}^{N,\epsilon}_i(t_i)\right] - \frac{\epsilon}{2}, \quad \forall t_i \in T_i,\ \forall i \in N.$

Now let $\omega^{N,\epsilon}$ satisfy the same warrant equations in (a) but in the game Γ_N; that is, with the maximum over D_N instead of $\tilde{D}^\epsilon_N$. Lemma 1 in Myerson (1983) implies that $\omega^{N,\epsilon}_i(t_i) \leq \tilde{\omega}^{N,\epsilon}_i(t_i)$ for every $i \in N$ and $t_i \in T_i$ (since $D_N \subseteq \tilde{D}^\epsilon_N$). Because $U^N_i(\mu_N \mid t_i) \geq \tilde{U}^N_i(\tilde{\mu}^\epsilon_N \mid t_i) - \frac{\epsilon}{2}$ for all $i \in N$ and $t_i \in T_i$, then we have that

$$U^N_i(\mu_N \mid t_i) \geq \frac{1}{|N|}\left[\sum_{j\in N\setminus i} U^{N\setminus j}_i(\mu_{N\setminus j} \mid t_i) + \omega^{N,\epsilon}_i(t_i)\right] - \epsilon, \quad \forall t_i \in T_i,\ \forall i \in N.$$

In summary, for every $\epsilon > 0$ there exist vectors $\tilde{\lambda}^{N,\epsilon} \in \Lambda^N_{++}$, $\tilde{\alpha}^{N,\epsilon} \in A^N$, and $\omega^{N,\epsilon} \in \prod_{i\in N}\mathbb{R}^{T_i}$ satisfying conditions *NPS1–NPS2* for μ_N. So the neutral procedural solutions do obey the restricted extension axiom.

Let $\eta_N = (\mu_S)_{S\subseteq N}$ be a neutral procedural solution of Γ_N. We shall see now that any alternative bargaining solution correspondence satisfying axioms *RCRD* and *RE* must contain η_N in its solution set for the game Γ_N. Let $BS(\cdot)$ be any bargaining solution correspondence satisfying *RCRD* and *RE*. We proceed by induction. Suppose that $(\mu_R)_{R\subseteq S} \in BS(\Gamma_S)$ for every subgame Γ_S with $|S| = |N| - 1$. Because η_N is a neutral procedural solution, then μ_N is incentive-efficient, and for each $\epsilon > 0$, there exist vectors $\lambda^{N,\epsilon} \in \Lambda^N_{++}$, $\alpha^{N,\epsilon} \in A^N$, and $\omega^{N,\epsilon} \in \prod_{i\in N}\mathbb{R}^{T_i}$ for which μ_N satisfies conditions *NPS1–NPS2*. Thus, by Lemma 1, for every ϵ, there exists a restricted extension of the game Γ_N, denoted $\tilde{\Gamma}^\epsilon_N$, such that, for each $i \in N$, there is a strong solution $\tilde{\mu}^{i,\epsilon}_S$ such that $\tilde{U}^N_i(\tilde{\mu}^{i,\epsilon}_N \mid t_i) = \omega^{N,\epsilon}_i(t_i)$ for all $t_i \in T_i$, and $\tilde{U}^N_j(\tilde{\mu}^{i,\epsilon}_N \mid t_j) = U^{N\setminus i}_j(\mu_{N\setminus i} \mid t_j)$ for all $j \in N \setminus i$ and every $t_j \in T_j$. Moreover, the mechanism $\tilde{\mu}^\epsilon_N = \frac{1}{|N|}\sum_{i\in N}\tilde{\mu}^{i,\epsilon}_N$ is incentive efficient (and a fortiori also incentive compatible) for N in $\tilde{\Gamma}^\epsilon_N$. Therefore, for every $i \in N$ and $t_i \in T_i$, we have that

$$U^N_i(\mu_N \mid t_i) \geq \frac{1}{|N|}\left[\sum_{j\in N\setminus i} U^{N\setminus j}_i(\mu_{N\setminus j} \mid t_i) + \omega^{N,\epsilon}_i(t_i)\right] - \epsilon = \tilde{U}^N_i(\tilde{\mu}^\epsilon_N \mid t_i) - \epsilon$$

By the restricted conditional random dictatorship axiom, $\tilde{\eta}^\epsilon_N = ((\mu_S)_{S\subset N}, \tilde{\mu}^\epsilon_N) \in BS(\tilde{\Gamma}^\epsilon_N)$ for every ϵ. Hence, by the restricted extension axiom, $\eta_N \in BS(\Gamma_N)$. □

participation constraints given t_i). If we allow for the possibility of strict inequalities in (4.12b), then it might be the case that $U^N_j(\mu^i_N \mid t_j) > \tilde{U}^N_j(\tilde{\mu}^i_N \mid t_j)$, which would invalidate our argument to absurdity.

6. Arrogance of Strength

We shall conclude the paper with an analysis of the eloquent example proposed by ourselves in a preceding contribution (see Salamanca, 2020). This example illustrates an important property of the neutral procedural solution, which Myerson (1985, p. 128) denominates *arrogance of strength*: "If two individuals of symmetric bargaining ability negotiate with each other, but one individual has a surprisingly strong bargaining position (i.e., the range of agreements that would be better for him than the disagreement outcome is smaller than the other individual expects), then the outcome of the [neutral procedural solution] tends to be similar to what would have been the outcome if the strong individual had had all of the bargaining ability, except that the probability of disagreement [...] is higher." As a result of this property, the neutral procedural solutions may be insensitive to some "negative" externalities generated by adverse selection.

The Bayesian cooperative game is as follows: $N = \{1, 2, 3\}$, $T_3 = \{w, s\}$, $p(w) = 1 - p(s) = 9/10$. Decision options for coalitions are $D_i = \{d_i\}$ $(i \in N)$, $D_{\{1,2\}} = (D_1 \times D_2) \cup \{d_{12}\} = \{[d_1, d_2], d_{12}\}$, $D_{\{i,3\}} = (D_i \times D_3) \cup \{d^i_{i3}, d^3_{i3}\} = \{[d_i, d_3], d^i_{i3}, d^3_{i3}\}$ $(i = 1, 2)$, and $D_N = (D_{\{1,2\}} \times D_3) \cup (D_{\{1,3\}} \times D_2) \cup (D_{\{2,3\}} \times D_1)$. Utility functions are described in Table 6.1.

(u_1, u_2, u_3)	s	w
$[d_1, d_2, d_3]$	$(0, 0, 0)$	$(0, 0, 0)$
$[d_{12}, d_3]$	$(5, 5, 0)$	$(5, 5, 0)$
$[d^1_{13}, d_2]$	$(0, 0, 5)$	$(0, 0, 10)$
$[d^3_{13}, d_2]$	$(10, 0, -5)$	$(10, 0, 0)$
$[d^2_{23}, d_1]$	$(0, 0, 5)$	$(0, 0, 10)$
$[d^3_{23}, d_1]$	$(0, 10, -5)$	$(0, 10, 0)$

Table 6.1: Utility functions for Example 2

The game situation is interpreted as the following collective choice problem. Three players may invest in a work project which would cost \$10. The project is worth \$10 to player 1 as well as to player 2; but its value to player 3 may be \$10 with probability 9/10 or \$5 with probability 1/10. We may say that player 3's type is *strong* if the project is worth \$5 to her, since she has relatively little to lose by refusing to cooperate than might otherwise be expected. On the other hand, we say that player 3's type is *weak* if the project is worth \$10 to her, since she would then be relatively more willing to bear any given cost. Every player $i \in N$ may decide not to cooperate (decision d_i), in which case he gets a reservation utility normalized to zero. If coalition $\{1, 2\}$ forms, its members may agree on the option d_{12} which carries out the project dividing the cost on equal parts. If players 1 and 3 form a coalition, decision d^j_{13} $(j = 1, 3)$ denotes the option to undertake the project at j's expense. Any other financing option may be represented by a lottery on $\{d^1_{13}, d^3_{13}\}$. Players 1 and 2 are symmetric, then decision options for coalition $\{2, 3\}$ are similarly interpreted. If all three players form a coalition, they may use a random device to pick a two-person coalition which must then make a decision as above.

We begin the analysis of this example by studying the different subgames. The subgame $\Gamma_{\{1,2\}}$ is a two-person bargaining problem with *complete* information. Clearly, the unique neutral procedural value of this subgame is its Nash bargaining solution $(U_1, U_2) = (5, 5)$ achieved by the decision d_{12}. Let $i = 1, 2$ be a fixed player and consider the subgame $\Gamma_{\{i,3\}}$. The unique neutral procedural value of this subgame is the allocation $(U_i, U_3^w, U_3^s) = (9/2, 5, 5/2)$, which corresponds to the expected outcome of the conditional random dictatorship procedure in coalition $\{i, 3\}$ (see Figure 6.1).

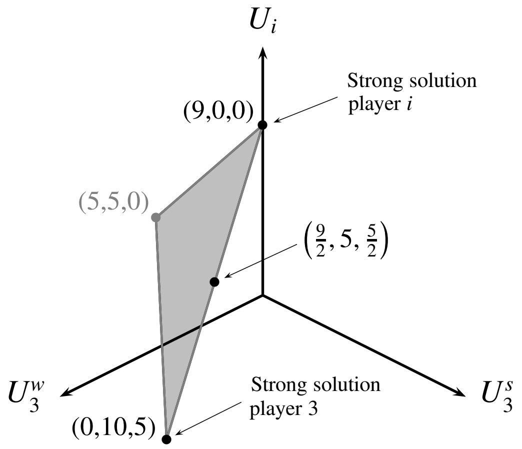

Figure 6.1: Incentive-efficient payoff allocations for coalition $\{i, 3\}$

It is noteworthy that player 3's strong solution in coalition $\{i, 3\}$ executes the project (with a probability of 1 in both states) at the expense of player i. This is equivalent to player 3 making a take-it-or-leave-it offer to pay \$0.

If the players abide by the equity principles underlying the neutral procedural solution, we observe that neither player 1 nor 2 can expect to get more than \$9/2 by forming a two-person coalition with player 3. Therefore, players 1 and 2 are better off in coalition $\{1, 2\}$, in which case they both get \$5 each. Moreover, by acting together, players 1 and 2 face no uncertainty at all. Indeed, it is commonly known that the project is worth \$10 to each of them. Salamanca (2020) then argues that coalition $\{1, 2\}$ should be "much" more likely to form, hence leaving *both* types of player 3 with a lower expected payoff.

To compute the neutral procedural value of the game Γ_N, it is convenient to express the various neutral procedural values, U^S, for each coalition $S \neq N$ as a four-dimensional vector (U_1, U_2, U_3^w, U_3^s), using "–" to denote the (types of the) players outside of S:

$$\begin{aligned} U^{\{1,2\}} &= (5, 5, -, -), \\ U^{\{1,3\}} &= \left(\tfrac{9}{2}, -, 5, \tfrac{5}{2}\right), \\ U^{\{2,3\}} &= \left(-, \tfrac{9}{2}, 5, \tfrac{5}{2}\right). \end{aligned}$$

Any incentive-compatible mechanism must satisfy the following equation:

$$U_1 + U_2 + \tfrac{9}{10} U_3^w + \tfrac{1}{5} U_3^s = 10. \tag{6.1}$$

For each coalition $S \neq N$, we can now adjoin to U^S a payoff for the missing player (in boldface below) so that the resulting payoff vector satisfies the equation (6.1):

$$U^{\{1,2\}} = (5, 5, \mathbf{0}, \mathbf{0}),$$
$$U^{\{1,3\}} = \left(\tfrac{9}{2}, \tfrac{\mathbf{1}}{\mathbf{2}}, 5, \tfrac{5}{2}\right),$$
$$U^{\{2,3\}} = \left(\tfrac{\mathbf{1}}{\mathbf{2}}, \tfrac{9}{2}, 5, \tfrac{5}{2}\right).$$

The payoff vectors above are the resulting payoffs from the strong solution in coalition N of the corresponding missing player. The unique neutral procedural value of the game Γ_N is the allocation that equally randomizes between the strong solutions of the various players in coalition N:

$$(U_1, U_2, U_3^w, U_3^s) = \left(\tfrac{10}{3}, \tfrac{10}{3}, \tfrac{10}{3}, \tfrac{5}{3}\right). \tag{6.2}$$

Any neutral procedural solution of Γ_N is *ex-post* efficient, that is, it executes the project with probability one in both states. Furthermore, the probability that a two-person coalition $\{i, 3\}$ is formed in state s is always $1/3$, and in this case the entire burden of costs falls on player i alone. Consider for instance the mechanism in Table 6.2.[19] This mechanism is a neutral procedural solution of Γ_N exhibiting the arrogance of strength described above—the outcome of the negotiation is similar to what would have been the outcome if player 3 had been a dictator on each two-person coalition, except that the probability of disagreement is positive, because coalition $\{1, 2\}$ can still form.

Type w	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
Prob($\{i, j\}$)	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$
Player 1	\$5	\$10	–
Player 2	\$5	–	\$10
Player 3	–	\$0	\$0

Type s	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
Prob($\{i, j\}$)	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$
Player 1	\$5	\$10	–
Player 2	\$5	–	\$10
Player 3	–	\$0	\$0

Table 6.2: Neutral procedural solution for N

In the mechanism 6.2 players 1 and 2 have nothing to gain by forming a coalition with player 3. Thus, they may simply spurn player 3's offer and form coalition $\{1, 2\}$ instead. Yet, the probability that a coalition containing player 3 will be formed in this mechanism is strikingly large (i.e., $1/3$). This opens the question of why players 1 and 2 would agree to sign such a contract. Salamanca (2020) already pointed out this difficulty in the context of the M-solution, which incidentally in this example, also yields the allocation (6.2). The S-solution offers an alternative outcome for this game. The unique utility allocation supported by some S-solution of Γ_N (see Salamanca, 2020) is

$$(U_1, U_2, U_3^w, U_3^s) = \left(\tfrac{41}{12}, \tfrac{41}{12}, \tfrac{40}{12}, \tfrac{10}{12}\right). \tag{6.3}$$

[19]Here and in Table 6.3, for a given type of player 3, each matrix describes the probability of coalition $\{i, j\}$ and the distribution of the costs among coalition members. The en-dash indicates that a player's cost cannot be defined if s/he does not belong to the corresponding coalition.

The allocation (6.3) gives less to the strong type of player 3 than (6.2). This is because, in the S-solution, the members of $\{i,3\}$ $(i = 1,2)$ have to settle for an "egalitarian" threat giving payoffs $(U_i, U_3^H, U_3^L) = (19/4, 5, 5/4)$. This payoff vector may be considered "more" equitable than the neutral procedural value of $\Gamma_{\{i,3\}}$ in the sense that type s of player 3 bears the efficiency losses originated on the adverse selection problem. As a result, the strong position of player 3 in coalition $\{i,3\}$ is weakened, which decreases the probability that player 3 be part of a coalition. To illustrate this argument, consider the mechanism in Table 6.3, where $0 \le \alpha \le 1/2$ denotes the probability that a coalition containing player 3 forms in state s. In case $\alpha = 1/6$, this mechanism is a neutral procedural solution of Γ_N. If $\alpha = 1/12$, this same mechanism is an S-solution instead.

	Type w				Type s		
	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$		$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
Prob($\{i,j\}$)	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	Prob($\{i,j\}$)	$1-2\alpha$	α	α
Player 1	\$5	\$5	–	Player 1	\$5	\$10	–
Player 2	\$5	–	\$5	Player 2	\$5	–	\$10
Player 3	–	\$5	\$5	Player 3	–	\$0	\$0

Table 6.3: S-solution vs. Neutral procedural solution

Which approach is correct? There cannot be a definite answer. It may depend on the way the negotiations are conducted, which may affect the coordination possibilities of the various players. For instance, if players expect the others to accept the first offer received, then safety issues may be perceived as preponderant, and the neutral procedural solution is justified. If players 1 and 2 have few coordination issues, then they may easily forego player 3's offer to form coalition $\{1,2\}$, and the S-solution would seem well-founded.

Appendix A. Proof of Theorem 1

Let $k \ge \sum_{i\in N} |T_i|$. For each $S \subseteq N$ we define

$$\Lambda_k^S = \left\{ \lambda \in \prod_{i\in S} \mathbb{R}^{T_i} \middle| \sum_{i\in S} \sum_{t_i \in T_i} \lambda_i(t_i) = 1,\ \lambda_i(t_i) \ge \tfrac{1}{k},\ \forall i \in S,\ \forall t_i \in T_i \right\}.$$

For each $S \subseteq N$, there exists a compact and convex set $\hat{A}^S \subseteq A^S$ such that, for each $\lambda^S \in \Lambda_+^S$, $\hat{A}_S$ contains at least one optimal solution of the dual problem (3.7). To prove this fact, notice that the feasible set in the primal problem (3.5) (i.e., $\mathcal{M}_S^*$) is compact and independent of λ^S. So the unit simplex Λ_+^S can be covered by a finite collections of sets (each corresponding to the range of optimality of one basic feasible solution in the primal) such that, within each set, an optimal solution of the dual can be given as a linear function of λ^S. Each of these linear functions is bounded on the compact unit simplex Λ_+^S, so we can choose $\hat{A}^S$ to be the convex hull of the union of the ranges of these linear functions on Λ_+^S.

For each k, we define a correspondence

$$\Phi_k : \prod_{S\subseteq N} \Lambda_k^S \times \prod_{S\subseteq N} \hat{A}_S \times \prod_{S\subseteq N} \mathcal{M}_S \rightrightarrows \prod_{S\subseteq N} \Lambda_k^S \times \prod_{S\subseteq N} \hat{A}_S \times \prod_{S\subseteq N} \mathcal{M}_S$$

so that $((\lambda^S)_{S\subseteq N}, (\alpha^S)_{S\subseteq N}, (\mu_S)_{S\subseteq N}) \in \Phi_k((\hat{\lambda}^S)_{S\subseteq N}, (\hat{\alpha}^S)_{S\subseteq N}, (\hat{\mu}_S)_{S\subseteq N})$ iff for each $S \subseteq N$:

$$\lambda^S \in \arg\min_{\gamma^S\in\Lambda_k^S} \sum_{i\in S}\sum_{t_i\in T_i} \gamma_i^S(t_i)\left[|S|U_i(\hat{\mu}_S \mid t_i) - \sum_{j\in S\setminus j} U_i(\hat{\mu}_{S\setminus j} \mid t_i) - \hat{\omega}_i^S(t_i)\right] \tag{A.1a}$$

$$\alpha^S \in \arg\min_{a_S\in\hat{A}_S} \sum_{t_S\in T_S} p(t_S) \max_{d_S\in D_S} \sum_{i\in S} v_i^S(d_S, t_S, \hat{\lambda}_i^S, a_i^S) \tag{A.1b}$$

$$\mu_S \in \arg\max_{m_S\in\mathcal{M}_S^*} \sum_{i\in S}\sum_{t_i\in T_i} \hat{\lambda}_i^S(t_i) U_i^S(m_S \mid t_i), \tag{A.1c}$$

where $\hat{\omega}_i^S = (\hat{\omega}_i^S(t_i))_{t_i\in T_i}$ is the unique allocation warranted by $\hat{\lambda}^S$, $\hat{\alpha}^S$, $\hat{\alpha}^{S\setminus i}$ and $\hat{\mu}_{S\setminus i}$ (Lemma 1 in Myerson (1983) guarantees that $\hat{\omega}_i^S$ is uniquely determined).

For any value of k, the correspondence Φ_k is non-empty convex valued and upper-hemicontinuous. Then by the Kakutani fixed point theorem, for each k there exits some $((\lambda^{S,k})_{S\subseteq N}, (\alpha^{S,k})_{S\subseteq N}, (\mu_S^k)_{S\subseteq N})$ such that

$$((\lambda^{S,k})_{S\subseteq N}, (\alpha^{S,k})_{S\subseteq N}, (\mu_S^k)_{S\subseteq N}) \in \Phi_k((\lambda^{S,k})_{S\subseteq N}, (\alpha^{S,k})_{S\subseteq N}, (\mu_S^k)_{S\subseteq N})$$

Thus, for every $S \subseteq N$ and each k, we have that μ_S^k is an optimal solution of the primal (3.5) w.r.t. $\lambda^{S,k}$ and $\alpha^{S,k}$ is an optimal solution of the dual problem (3.7) w.r.t. $\lambda^{S,k}$. Moreover, this sequence of fixed points lies on a compact domain, so we may assume w.l.g. that it converges to some $((\bar{\lambda}^S)_{S\subseteq N}, (\bar{\alpha}^S)_{S\subseteq N}, (\bar{\mu}_S)_{S\subseteq N})$. We shall see that $(\bar{\mu}_S)_{S\subseteq N}$ is a neutral procedural solution.

Let $S \subseteq N$ be a fixed coalition. For any k and for each $i \in S$, let $\omega_i^{S,k}$ denote the unique allocation warranted by $\lambda^{S,k}$, $\alpha^{S,k}$, $\alpha^{S\setminus i,k}$ and $\mu_{S\setminus i}^k$. Then, for any $\lambda^S \in \Lambda_k^S$ we have that

$$\begin{aligned}
&\sum_{i\in S}\sum_{t_i\in T_i} \lambda_i^S(t_i)\left[|S|U_i^S(\mu_S^k \mid t_i) - \sum_{j\in S\setminus i} U_i^{S\setminus j}(\mu_{S\setminus j}^k \mid t_i) - \omega_i^{S,k}(t_i)\right]\\
&\geq \sum_{i\in S}\sum_{t_i\in T_i} \lambda_i^{S,k}(t_i)\left[|S|U_i^S(\mu_S^k \mid t_i) - \sum_{j\in S\setminus i} U_i^{S\setminus j}(\mu_{S\setminus j}^k \mid t_i) - \omega_i^{S,k}(t_i)\right]\\
&= |S|\sum_{i\in S}\sum_{t_i\in T_i} \lambda_i^{S,k}(t_i)U_i^S(\mu_S^k \mid t_i) - \sum_{i\in S}\sum_{j\in S\setminus i}\sum_{t_i\in T_i} \lambda_i^{S,k}(t_i)U_i^{S\setminus j}(\mu_{S\setminus j}^k \mid t_i)\\
&\quad - \left[|S|\sum_{t_S\in T_S} p(t_S)\max_{d_S\in D_S}\sum_{i\in S} v_i^S(d_S, t_S, \lambda_i^{S,k}, \alpha_i^{S,k}) - \sum_{i\in S}\sum_{j\in S\setminus i}\sum_{t_j\in T_j} \lambda_j^{S,k}(t_j)U_j^{S\setminus i}(\mu_{S\setminus i}^k \mid t_j)\right]\\
&= |S|\left[\sum_{i\in S}\sum_{t_i\in T_i} \lambda_i^{S,k}(t_i)U_i^S(\mu_S^k \mid t_i) - \sum_{t_S\in T_S} p(t_S)\max_{d_S\in D_S}\sum_{i\in S} v_i^S(d_S, t_S, \lambda_i^{S,k}, \alpha_i^{S,k})\right]\\
&= 0,
\end{aligned} \tag{A.1d}$$

where the inequality in the first line is due to the fixed point condition together with (A.1a); the equality in the second line follows from summing the warrant equations and using the fact that $\mu^k_{S\setminus i}$ satisfies (4.8) for $\lambda^{S,k}$ and $\alpha^{S\setminus i,k}$; the equality in the third line holds because $\sum_{i\in S}\sum_{j\in S\setminus i}\sum_{t_i\in T_i}\lambda^{S,k}_i(t_i)U^{S\setminus j}_i(\mu^k_{S\setminus j}\mid t_i) = \sum_{i\in S}\sum_{j\in S\setminus i}\sum_{t_j\in T_j}\lambda^{S,k}_j(t_j)U^{S\setminus i}_j(\mu^k_{S\setminus i}\mid t_j)$; and finally, the equality in the fourth line follows from strong duality.

We shall now see that the sequence $\{(\omega^{S,k})_{S\subseteq N}\}_k$ is contained in a compact set. By Lemma 2,

$$\omega^{S,k}_i(t_i) \geq \max_{d_i\in D_i} u^{\{i\}}_i(d_i,t_i), \qquad \forall i\in S,\ \forall t_i\in T_i.$$

Thus, $\{\omega^{S,k}\}_k$ is bounded below. Now define $M := \max_{i\in N}\max_{t_i\in T_i}\max_{d\in D}|u_i(d,t_i)|$. For any $\lambda\in\Lambda^S_k$, we have that

$$\sum_{i\in S}\sum_{t_i\in T_i}\lambda_i(t_i)|S|U^S_i(\mu^k_S\mid t_i) \leq |S|M,$$

and

$$\sum_{i\in S}\sum_{t_i\in T_i}\lambda_i(t_i)\sum_{j\in S\setminus i}U^{S\setminus j}_i(\mu^k_{S\setminus j}\mid t_i) \geq -M(|S|-1).$$

Hence, (A.1d) implies that, for each k,

$$\sum_{i\in S}\sum_{t_i\in T_i}\lambda_i(t_i)\omega^{S,k}_i(t_i) \leq M(2|S|-1), \quad \forall \lambda\in\Lambda^S_k.$$

We conclude that $\{\omega^{S,k}\}_k$ is bounded above. Then, we may assume w.l.g. that $\{\omega^{S,k}\}_k$ converges to some $(\bar\omega^S)_{S\subseteq N}$.

For every $i\in S$ and $t_i\in T_i$, let δ^k_{i,t_i} be the probability vector in Λ^S_k that puts weight $\frac{1}{k}$ on all players' types except on player i's type t_i. Then $\{\delta^k_{i,t_i}\}_k$ converges to the Dirac measure that puts all probability weight on type t_i. Therefore, using δ^k_{i,t_i} in (A.1d) and taking the limit we obtain

$$|S|U_i(\bar\mu_S\mid t_i) - \sum_{j\in S\setminus i}U_i(\bar\mu_{S\setminus j}\mid t_i) \geq \bar\omega^S_i(t_i).$$

Thus, for every $\epsilon>0$, there exists k^* (sufficiently large) such that

$$\begin{aligned}\omega^{S,k^*}_i(t_i) &< \bar\omega^S_i(t_i)+|S|\epsilon\\ &\leq |S|U_i(\bar\mu_S\mid t_i) - \sum_{j\in S\setminus i}U_i(\bar\mu_{S\setminus j}\mid t_i)+|S|\epsilon.\end{aligned}$$

Or alternatively,

$$U_i(\bar\mu_S\mid t_i) > \frac{1}{|S|}\left[\sum_{j\in S\setminus i}U_i(\bar\mu_{S\setminus j}\mid t_i)+\omega^{S,k^*}_i(t_i)\right]-\epsilon,$$

and so $(\bar\mu_S)_{S\subseteq N}$ satisfies the conditions *NPS1–NPS2*. For each $S\subseteq N$, the mechanism $\bar\mu_S$ is incentive compatible, but not necessarily incentive efficient. However, there must exist an incentive-efficient mechanism $\tilde{\mu}_S$ such that $U_i^S(\tilde{\mu}_S \mid t_i) \geq U_i^S(\bar{\mu}_S \mid t_i)$ for all $i \in S$ and $t_i \in T_i$. Then conditions *NPS1–NPS2* are also satisfied for $\tilde{\mu}_S$. $\square$

Appendix B. Auxiliary Results

Proof of Lemma 1. Let $S \subseteq N$ be a coalition. Let $(\mu_{S\setminus i})_{i\in S}$ be a given vector of threats such that, for each $i \in S$, $\mu_{S\setminus i}$ is incentive-compatible for $S \setminus i$. Let $\lambda^S \in \Lambda^S_{++}$, $\alpha^S \in A^S$, $\alpha^{S\setminus i} \in A^{S\setminus i}$, and $\omega^S \in \prod_{i\in S} \mathbb{R}^{T_i}$ be such that, for every $i \in S$, ω_i^S is warranted by λ^S, α^S, $\alpha^{S\setminus i}$, and $\mu_{S\setminus i}$.

We define the quantities $w_i^S(t_S)$ so that

$$\frac{1}{p(t_S)}\left[\left(\lambda_i^S(t_i) + \sum_{\tau_i\in T_i} \alpha_i^S(\tau_i \mid t_i)\right) w_i^S(t_S) - \sum_{\tau_i\in T_i} \alpha_i^S(t_i \mid \tau_i) w^S(\tau_i, t_{S\setminus i})\right]$$
$$= \max_{d_S\in D_S} \sum_{j\in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S) - \sum_{j\in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i}), \quad \forall t_S \in T_S,\ \forall i \in S. \quad \text{(B.1a)}$$

As long as $\lambda^S \in \Lambda^S_{++}$, Lemma 1 in Myerson (1983) guarantees that these equations have a unique solution, so that the quantities $w_i^S(t_S)$ are well defined.

Now we let $\tilde{D}_S = D_S \cup \{\tilde{d}_S^i \mid i \in S\}$, and

$$\tilde{u}_j^S(d_S, t_S) = \begin{cases} \frac{w_j^S(t_S)}{p(t_{S\setminus j})}, & \text{if } d_S = \tilde{d}_S^j; \\ u_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}), & \text{if } d_S = \tilde{d}_S^i\ (i \neq j); \\ u_j^S(d_S, t_S), & \text{if } d_S \in D_S. \end{cases}$$

For every $i \in S$, let $\tilde{\Gamma}_S^i(\mu_{S\setminus i})$ be the mechanism selection game that differs from $\Gamma_S^i(\mu_{S\setminus i})$ only in that the set of decision options available to player i is $\tilde{D}_S$ (instead of D_S). Define $\tilde{\mu}_S^i$ to be the mechanism that always implements $\tilde{d}_S^i$ with probability 1.

Fix a player $i \in S$ and a type $t_i \in T_i$. For every $j \in S \setminus i$ and $t_j \in T_j$ we have that

$$\sum_{t_{S\setminus\{i,j\}}\in T_{S\setminus\{i,j\}}} p(t_{S\setminus\{i,j\}}) \sum_{d_S\in\tilde{D}_S} \tilde{\mu}_S^i(d_S \mid t_S)\tilde{u}_j^S(d_S, t_S) = \tilde{U}_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j) = U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j).$$

Moreover, since $\tilde{\mu}_S^i$ does not depend on t_S, then it is incentive compatible for i, and would also be incentive compatible for all players in $S \setminus i$ after they inferred that i's type is t_i. Therefore, $\tilde{\mu}_S^i$ is safe for player i in $\tilde{\Gamma}_S^i(\mu_{S\setminus i})$.

We shall now show that $\tilde{\mu}_S^i$ is undominated for player i in $\tilde{\Gamma}_S^i(\mu_{S\setminus i})$. For that, we first notice that

$$\sum_{t_S\in T_S} p(t_S) \sum_{j\in S} \tilde{v}_j^S(\tilde{d}_S^i, t_S, \lambda_j^S, \alpha_j^S)$$
$$= \sum_{j\in S} \sum_{t_S\in T_S} p(t_{S\setminus j}) \left[\left(\lambda_j^S(t_j) + \sum_{\tau_j\in T_j} \alpha_j^S(\tau_j \mid t_j)\right) \tilde{u}_j^S(\tilde{d}_S^i, t_S) - \sum_{\tau_j\in T_j} \alpha_j^S(t_j \mid \tau_j) \tilde{u}_j^S(\tilde{d}_S^i, \tau_j, t_{S\setminus j})\right]$$

$$
\begin{aligned}
&= \sum_{t_S \in T_S} \left[\left(\lambda_i^S(t_i) + \sum_{\tau_i \in T_i} \alpha_i^S(\tau_i \mid t_i) \right) w_i^S(t_S) - \sum_{\tau_i \in T_i} \alpha_i^S(t_i \mid \tau_i) w_i^S(\tau_i, t_{S \setminus i}) \right] \\
&\quad + \sum_{j \in S \setminus i} \sum_{t_j \in T_j} \sum_{t_{S \setminus j} \in T_{S \setminus j}} p(t_{S \setminus j}) \left[\left(\lambda_j^S(t_j) + \sum_{\tau_j \in T_j} \alpha_j^S(\tau_j \mid t_j) \right) u_j^{S \setminus i}(\mu_{S \setminus i}, t_{S \setminus i}) \right. \\
&\quad \left. - \sum_{\tau_j \in T_j} \alpha_j^S(t_j \mid \tau_j) u_j^{S \setminus i}(\mu_{S \setminus i}, \tau_j, t_{S \setminus \{i,j\}}) \right] \\
&= \sum_{t_S \in T_S} p(t_S) \max_{d_S \in D_S} \sum_{j \in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S) - \sum_{t_{S \setminus i} \in T_{S \setminus i}} p(t_{S \setminus i}) \sum_{j \in S \setminus i} v_j^{S \setminus i}(\mu_{S \setminus i}, t_{S \setminus i}, \lambda_j^S, \alpha_j^{S \setminus i}) \\
&\quad + \sum_{j \in S \setminus i} \sum_{t_j \in T_j} \left[\left(\lambda_j^S(t_j) + \sum_{\tau_j \in T_j} \alpha_j^S(\tau_j \mid t_j) \right) U_j^{S \setminus i}(\mu_{S \setminus i} \mid t_j) - \sum_{\tau_j \in T_j} \alpha_j^S(t_j \mid \tau_j) U_j^{S \setminus i}(\mu_{S \setminus i} \mid \tau_j) \right] \\
&\leq \sum_{t_S \in T_S} p(t_S) \max_{d_S \in D_S} \sum_{j \in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S) - \sum_{j \in S \setminus i} \sum_{t_j \in T_j} \lambda_j^S(t_j) U_j^{S \setminus i}(\mu_{S \setminus i} \mid t_j) \\
&\quad + \sum_{j \in S \setminus i} \sum_{t_j \in T_j} \left[\left(\lambda_j^S(t_j) + \sum_{\tau_j \in T_j} \alpha_j^S(\tau_j \mid t_j) \right) U_j^{S \setminus i}(\mu_{S \setminus i} \mid t_j) - \sum_{\tau_j \in T_j} \alpha_j^S(t_j \mid \tau_j) U_j^{S \setminus i}(\mu_{S \setminus i} \mid \tau_j) \right] \\
&= \sum_{t_S \in T_S} p(t_S) \max_{d_S \in D_S} \sum_{j \in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S), \qquad \text{(B.1b)}
\end{aligned}
$$

where the equality in the second line uses the definition of $\tilde{u}_j^S$; the equality in the third line follows from (B.1a); and the inequality in the fourth line is due to the fact that

$$
\begin{aligned}
&\sum_{j \in S \setminus i} \sum_{t_j \in T_j} \lambda_j^S(t_j) U_j^{S \setminus i}(\mu_{S \setminus i} \mid t_j) \\
&\quad \leq \sum_{j \in S \setminus i} \sum_{t_j \in T_j} \lambda_j^S(t_j) U_j^{S \setminus i}(\mu_{S \setminus i} \mid t_j) + \sum_{j \in S \setminus i} \sum_{t_j \in T_j} \sum_{\tau_j \in T_j} \alpha_j^{S \setminus i}(\tau_j \mid t_j) \left[U_j^{S \setminus i}(\mu_{S \setminus i} \mid t_j) - U_j^{S \setminus i}(\mu_{S \setminus i}, \tau_j \mid t_j) \right] \\
&\quad = \sum_{t_{S \setminus i} \in T_{S \setminus i}} p(t_{S \setminus i}) \sum_{j \in S \setminus i} v_j^{S \setminus i}(\mu_{S \setminus i}, t_{S \setminus i}, \lambda_j^S, \alpha_j^{S \setminus i}),
\end{aligned}
$$

since $\alpha^{S \setminus i} \geq 0$ and $\mu_{S \setminus i}$ is incentive compatible for $S \setminus i$. Because (B.1b) holds for every $\tilde{d}_S^i$, and $\tilde{D}_S = D_S \cup \{\tilde{d}_S^j \mid j \in S\}$, (B.1b) implies that

$$
\sum_{t_S \in T_S} p(t_S) \max_{d_S \in \tilde{D}_S} \sum_{j \in S} \tilde{v}_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S) = \sum_{t_S \in T_S} p(t_S) \max_{d_S \in D_S} \sum_{j \in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S). \qquad \text{(B.1c)}
$$

On the other hand, using (B.1a) we have that

$$
\begin{aligned}
&\sum_{t_S \in T_S} p(t_S) \max_{d_S \in D_S} \sum_{j \in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S) \\
&\quad = \sum_{t_{S \setminus i} \in T_{S \setminus i}} \sum_{t_i \in T_i} \left[\left(\lambda_i^S(t_i) + \sum_{\tau_i \in T_i} \alpha_i^S(\tau_i \mid t_i) \right) w_i^S(t_S) - \sum_{\tau_i \in T_i} \alpha_i^S(t_i \mid \tau_i) w_i^S(\tau_i, t_{S \setminus i}) \right]
\end{aligned}
$$

$$+\sum_{t_{S\setminus i}\in T_{S\setminus i}} p(t_{S\setminus i}) \sum_{j\in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i})$$
$$= \sum_{t_i\in T_i} \lambda_i^S(t_i) \sum_{t_{S\setminus i}\in T_{S\setminus i}} w_i^S(t_S) + \sum_{t_{S\setminus i}\in T_{S\setminus i}} p(t_{S\setminus i}) \sum_{j\in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i})$$
$$= \sum_{t_i\in T_i} \lambda_i^S(t_i) \sum_{t_{S\setminus i}\in T_{S\setminus i}} p(t_{S\setminus i})\tilde{u}_i^S(\tilde{d}_S^i, t_S) + \sum_{t_{S\setminus i}\in T_{S\setminus i}} p(t_{S\setminus i}) \sum_{j\in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i})$$
$$= \sum_{t_i\in T_i} \lambda_i^S(t_i)\tilde{U}_j^S(\tilde{\mu}_S^i \mid t_i) + \sum_{t_{S\setminus i}\in T_{S\setminus i}} p(t_{S\setminus i}) \sum_{j\in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i}). \tag{B.1d}$$

From (B.1c)–(B.1d) we conclude that

$$\sum_{t_i\in T_i} \lambda_i^S(t_i)\tilde{U}_j^S(\tilde{\mu}_S^i \mid t_i)$$
$$= \sum_{t_S\in T_S} p(t_S) \max_{d_S\in \tilde{D}_S} \sum_{j\in S} \tilde{v}_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S) - \sum_{t_{S\setminus i}\in T_{S\setminus i}} p(t_{S\setminus i}) \sum_{j\in S\setminus i} \tilde{v}_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i}). \tag{B.1e}$$

Therefore, by duality theory, $\tilde{\mu}_S^i$ and α^S respectively are optimal solutions of the primal (4.3) and its corresponding dual problem for λ_i^S, in the context of the problem $\tilde{\Gamma}_S^i(\mu_{S\setminus i})$, because they are feasible in their respective programs and give equal value to their corresponding objective functions. Therefore, $\tilde{\mu}_S^i$ is undominated in $\tilde{\Gamma}_S^i(\mu_{S\setminus i})$, and consequently, a strong solution of $\tilde{\Gamma}_S^i(\mu_{S\setminus i})$. It therefore follows that $\tilde{\mu}_S^i$ is incentive efficient and satisfies (3.8a)–(3.8b) in Proposition 1 for λ^S and α^S.

On the other hand, from the definition of $\tilde{u}_j^S$, it follows that, for every $t_i \in T_i$,

$$\left(\lambda_i^S(t_i) + \sum_{\tau_i\in T_i} \alpha_i^S(\tau_i \mid t_i)\right)\tilde{U}_i^S(\tilde{\mu}_S^i \mid t_i) - \sum_{\tau_i\in T_i} \alpha_i^S(t_i \mid \tau_i)\tilde{U}_i^S(\tilde{\mu}_S^i \mid \tau_i)$$
$$= \sum_{t_{S\setminus i}\in T_{S\setminus i}} \left[\left(\lambda_i^S(t_i) + \sum_{\tau_i\in T_i} \alpha_i^S(\tau_i \mid t_i)\right) w_i^S(t_S) - \sum_{\tau_i\in T_i} \alpha_i^S(t_i \mid \tau_i) w_i^S(\tau_i, t_{S\setminus i})\right]$$
$$= \sum_{t_{S\setminus i}\in T_{S\setminus i}} p(t_S)\left[\max_{d_S\in D_S} \sum_{j\in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S) - \sum_{j\in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i})\right],$$

that is, the allocation $\tilde{U}_i(\tilde{\mu}_S^i)$ is warranted by λ^S, α^S, $\alpha^{S\setminus i}$, and $\mu_{S\setminus i}$. Hence, $\tilde{U}_i^S(\tilde{\mu}_S^i \mid t_i) = \omega_i^S(t_i)$ for every $t_i \in T_i$, since ω_i^S is the unique allocation satisfying the warrant equations (4.11).

Let $\tilde{\mu}_S := \frac{1}{|S|}\sum_{i\in S} \tilde{\mu}_S^i$ be the mechanism that randomizes equally between all strong solutions $(\tilde{\mu}_S^i)_{i\in S}$. Because each $\tilde{\mu}_S^i$ satisfies (3.8a)–(3.8b) in Proposition 1 for λ^S and α^S, then by the linearity of all formulas involved, $\tilde{\mu}_S$ also satisfies (3.8a)–(3.8b) in Proposition 1. Since $\lambda^S \in \Lambda_{++}^S$, it follows that $\tilde{\mu}_S$ is incentive efficient. □

Lemma 2.

Let $S \subseteq N$ be a coalition and $i \in S$. Suppose that ω_i^S is warranted by λ^S, α^S, $\alpha^{S\setminus i}$ and $\mu_{S\setminus i}$, where the *mechanism $\mu_{S\setminus i}$ is incentive compatible for $S \setminus i$, $\alpha^{S\setminus i}$ satisfies (4.7), and $\lambda_i^S > 0$. Then*

$$\omega_i^S(t_i) \geq \max_{d_i \in D_i} u_i^{\{i\}}(d_i, t_i), \quad \forall t_i \in T_i. \tag{B.2}$$

Proof. Since $\alpha^S \geq 0$ and $\mu_{S\setminus i}$ is incentive compatible for $S \setminus i$, then for all $j \in S \setminus i$ we have that

$$\alpha_j^S(\tau_j \mid t_j)\left[U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j) - U_j^{S\setminus i}(\mu_{S\setminus i}, \tau_j \mid t_j)\right] \geq 0, \quad \forall t_j, \tau_j \in T_j. \tag{B.3a}$$

Therefore, the following chain of inequalities hold:

$$\begin{aligned}
&\sum_{t_{S\setminus i} \in T_{S\setminus i}} p(t_{S\setminus i}) \sum_{j \in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i}) \\
&\quad = \sum_{j \in S\setminus i} \sum_{t_j \in T_j} \lambda_j^S(t_j) U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j) \\
&\quad \leq \sum_{j \in S\setminus i} \sum_{t_j \in T_j} \lambda_j^S(t_j) U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j) + \sum_{j \in S\setminus i} \sum_{t_j \in T_j} \sum_{\tau_j \in T_j} \alpha_j^S(\tau_j \mid t_j)\left[U_j^{S\setminus i}(\mu_{S\setminus i} \mid t_j) - U_j^{S\setminus i}(\mu_{S\setminus i}, \tau_j \mid t_j)\right] \\
&\quad = \sum_{t_{S\setminus i} \in T_{S\setminus i}} p(t_{S\setminus i}) \sum_{j \in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^S),
\end{aligned} \tag{B.3b}$$

where the first equality is due to (4.8), the inequality in the second line follows from (B.3a), and finally the last equality is obtained by rearranging terms.

On the other hand, let $\hat{\mu}_S$ be defined as in (4.5). Then, for any $t_i \in T_i$ we have

$$\begin{aligned}
&\left(\lambda_i^S(t_i) + \sum_{\tau_i \in T_i} \alpha_i^S(\tau_i \mid t_i)\right)\omega_i^S(t_i) - \sum_{\tau_i \in T_i} \alpha_i^S(t_i \mid \tau_i)\omega_i^S(\tau_i) \\
&\quad = p_i(t_i) \sum_{t_{S\setminus i} \in T_{S\setminus i}} p(t_{S\setminus i})\left(\max_{d_S \in D_S} \sum_{j \in S} v_j^S(d_S, t_S, \lambda_j^S, \alpha_j^S) - \sum_{j \in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i})\right) \\
&\quad \geq p_i(t_i) \sum_{t_{S\setminus i} \in T_{S\setminus i}} p(t_{S\setminus i})\left(\sum_{j \in S} v_j^S(\hat{\mu}_S, t_S, \lambda_j^S, \alpha_j^S) - \sum_{j \in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i})\right) \\
&\quad = p_i(t_i) v_i^{\{i\}}(\hat{\mu}_i, t_i, \lambda_i^S, \alpha_i^S) \\
&\qquad\quad + \sum_{t_{S\setminus i} \in T_{S\setminus i}} p(t_{S\setminus i})\left(\sum_{j \in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^S) - \sum_{j \in S\setminus i} v_j^{S\setminus i}(\mu_{S\setminus i}, t_{S\setminus i}, \lambda_j^S, \alpha_j^{S\setminus i})\right) \\
&\quad \geq p_i(t_i) v_i^{\{i\}}(\hat{\mu}_i, t_i, \lambda_i^S, \alpha_i^S) \\
&\quad = \left(\lambda_i^S(t_i) + \sum_{\tau_i \in T_i} \alpha_i^S(\tau_i \mid t_i)\right) \sum_{d_i \in D_i} \hat{\mu}_i(d_i \mid t_i) u_i^{\{i\}}(d_i, t_i) - \sum_{\tau_i \in T_i} \alpha_i^S(t_i \mid \tau_i) \sum_{d_i \in D_i} \hat{\mu}_i(d_i \mid t_i) u_i^{\{i\}}(d_i, \tau_i), \\
&\quad \geq \left(\lambda_i^S(t_i) + \sum_{\tau_i \in T_i} \alpha_i^S(\tau_i \mid t_i)\right) \max_{d_i \in D_i} u_i^{\{i\}}(d_i, t_i) - \sum_{\tau_i \in T_i} \alpha_i^S(t_i \mid \tau_i) \max_{d_i \in D_i} u_i^{\{i\}}(d_i, \tau_i).
\end{aligned}$$

In this chain, the equality in the first line follows from the fact that ω_i^S is warranted by λ^S, α^S, $\alpha^{S\setminus i}$

and $\mu_{S\setminus i}$; the inequality in the second line is due to the max operator in the first line and the fact that $\hat{\mu}_S \in \mathcal{M}_S$; the equality in the third line uses the definition of $\hat{\mu}_S$ together with the orthogonality assumption; the inequality in the forth line is due to (B.3b); the equality in the fifth line uses the definition of virtual utility; and finally, the inequality in the sixth line follows from the definition of $\hat{\mu}_i$. The desired conclusion thus follows from Lemma 1 in Myerson (1983). □

References

de Clippel, Geoffroy, "Values for cooperative games with incomplete information: An eloquent example," *Games and Economic Behavior*, 2005, *53* (1), 73–82.

_ **and Enrico Minelli**, "Two-person bargaining with verifiable information," *Journal of Mathematical Economics*, 2004, *40* (7), 799–813.

_ **, Hans Peters, and Horst Zank**, "Axiomatizing the Harsanyi solution, the symmetric egalitarian solution and the consistent solution for NTU-games," *International Journal of Game Theory*, 2004, *33*, 145–158.

Harsanyi, John C., "A Simplified Bargaining Model for the n-Person Cooperative Game," *International Economic Review*, 1963, *4* (2), 194–220.

Hart, Sergiu, "A comparison of non-transferable utility values," *Theory and Decision*, 2004, *56*, 35–46.

_ **and Andreu Mas-Colell**, "Bargaining and Value," *Econometrica*, 1996, *64* (2), 357–380.

Maschler, Michael and Guillermo Owen, "The consistent Shapley value for hyperplane games," *International Journal of Game Theory*, 1989, *18*, 389–407.

_ **and** _ , *Rational Interaction*, Berlin Heidelberg: Springer-Verlag,

Myerson, Roger B., "Optimal coordination mechanisms in generalized principal-agent problems," *Journal of Mathemtical Economics*, 1982, *10* (1), 67–81.

_ , "Mechanism design by an informed principal," *Econometrica*, 1983, *51* (6), 1767–1797.

_ , "Two-person bargaining problems with incomplete information," *Econometrica*, 1984a, *52* (2), 461–488.

_ , "Cooperative Games with Incomplete Information," *International Journal of Game Theory*, 1984b, *13* (2), 69–96.

_ , "Analysis of two bargaining problems with incomplete information," in Alvin E. Roth, ed., *Game-theoretic Models of Bargaining*, Cambridge: Cambridge University Press, 1985, chapter 7, pp. 115–147.

_ , “Analysis of incentives in bargaining and mediation,” in Peyton Young, ed., *Negotiation Analysis*, Ann Arbor: The University of Michigan Press, 1991, chapter 4, pp. 67–85.

_ , *Game Theory: Analysis of Conflict*, Harvard University Press, 1991.

Nash, John F., “The Bargaining Problem,” *Econometrica*, 1950, *18* (2), 155–162.

Salamanca, Andrés, “A generalization of the Harsanyi NTU value to games with incomplete information,” *International Journal of Game Theory*, 2020, *49* (1), 195–225.

Shapley, Lloyd S., “Utility comparisons and the theory of games,” in “La Décision: Agrégation et Dynamique des Ordres de Préférence,” Paris: Centre National de la Recherche Scientifique, 1969, pp. 251–263.