A Noncooperative Reformulation of the Core

Charles Z. Zheng*

July 10, 2009

Abstract

The concept of the core is reformulated to handle externality problems where the total payoff for a coalition depends upon the actions of the players outside the coalition. A deviating coalition has rational expectations of the outsiders' individual attempts to stop the deviation and their coalitional response if the deviaiton cannot be stopped, with the coalitional response itself a core solution among the outsiders. A noncooperative game of competing principals is designed and the set of its subgame perfect equilibria, with two refinement conditions, is equal to the reformulated core. Applied to a problem of pollution externality, the new concept of the core prescribes a way to maintain Pareto optimal cooperation without repeated games or a central planner.

*Department of Economics, Iowa State University, 260 Heady Hall, Ames, IA 50011, czheng@iastate.edu, http://www.econ.iastate.edu/people/faculty/CV/669czheng.html.

1 Introduction

Human beings are often trapped in a suboptimal situation just because everyone expects everyone else to stick to it. Describing such situations is exactly what equilibrium concepts in noncooperative game theory are good at, as an equilibrium is required to be immune to only unilateral deviations. However, human beings do hope and sometimes even try to escape from a suboptimal situation through joint deviations. The theory of the core has a potential to provide a rigorous prescription, or even description, for such coalitional endeavors, as a core solution is required to be immune to both unilateral and coalitional deviations.

However, such a potential so far has been limited, because the concept of the core has not been fully developed to incorporate the possibility that a coalitional deviation may be countered by those outside the coalition. Traditionally, the core was formulated based on the assumption that the payoff for a coalition is independent of the actions of the outsiders. That makes the core inapplicable to problems involving externalities, where the independence assumption does not hold in general. Yet intuitively one would think that it is in such externality problems that coalitional cooperation could be particularly helpful.

In this paper, the concept of the core is reformulated to take such interdependency into account. A solution in the core corresponds to a mapping that maps any joint action of any coalition to a response from the complement of the coalition. The response consists of two parts, *preemptive* and *reactive*. If the coalition succeeds in committing to the joint action, the complementary coalition takes a joint action according to the reactive response which is itself a core solution within the complementary coalition given the previous committed joint action. To gain the commitment power, however, the previous coalition needs to compete with others. Before it can commit, the players outside the coalition try to stop the commitment by acting according to the preemptive response, which constitutes a Nash equilibrium given the expectation described by the aforementioned mapping. To *block* the solution, everyone in a coalition needs to profit from a joint action based on the rational expectation that the action will be responded by the outsiders according to the above mapping. A solution belongs to the core if no coalition can block it in such a manner.

Applied to a typical problem of pollution externality, the new concept of the core demonstrates how Pareto optimal cooperation among self-interest players can be maintained without repeated games or a central mediator.

The reformulated core can be viewed as a set of equilibria of a noncooperative game of competing principals, which is constructed in this paper. It is proved that the core is equal to the set of subgame perfect equilibria subject to two refinement conditions in this competing principals game. A blocking action of a coalition corresponds to a principal's deviation to propose an off-path contract to the players in that coalition. In organizing joint actions among the players, the principals hope to pocket some surpluses, which are competed away on equilibrium path.

Instead of trying to disentangle the interdependency among coalitions, an alternative solution idea is to propose a social choice for the entire economy that cannot be blocked by the objection from any coalition. Examples are the bargaining set introduced by Aumann and Maschler [3], whose main interest is the stability of the objections themselves, and the core formulated by Myerson [7] to incorporate incomplete information. Different from my approach is their centralization interpretation. While the core formulated in this paper is equivalent to a set of equilibrium outcomes of a game that has no central mediator, the proposal-objection process in their approaches seems to allow only centralized mechanisms.

Ray [11] has recognized the difficulty that the cooperative game theory has in dealing with externality problems. To solve such a problem, Ray and Vohra [12] have contributed an idea of separating a coalition's formation from its choice of actions. Specifically, the players in a coalition commit to act together without specifying the joint action that they will take. I doubt that such a separation could avoid the complexity dealt with in my paper, where one coalition tries to commit to a joint action while another coalition tries to thwart that commitment due to its externality. Even after the players have been partitioned into coalitions as described by Ray and Vohra, the coalitions would still need to best reply one another with certain actions. In their paper, such complexity is absent due to two implicit assumptions. First, there is an a priori sequencing of moves according to which players take turn to decide on forming coalitions (p1365). Second, in choosing an action, a coalition takes the actions of its outsiders as given (their Eq. (2) and the top of p1367). These assumptions allow them to pin down every coalition's action, as well as the formation of coalitions, in a Stackelberg manner.

My construction of a noncooperative foundation for the reformulated core is similar in spirit to Perry and Reny [10], Serrano [13], and especially Perez-Castrillo [9]. Perez-Castrillo has formulated a noncooperative game of competing principals where a principal's unilateral deviation corresponds to the coalitional deviation among the agents whom the principal's deviant contract intends to cover. A major difference is that his model, as well as those of Perry and Reny and Serrano, does not allow for the interdependency between a coalition and its outsiders, which is the focus of this paper. Much of the complications of the noncooperative game formulated in this paper is due to the interactive reasoning among the principals and agents. The noncooperative game needs to be complicated enough to allow them to preempt and react to one another's actions.

The idea of coalitional deviations has been used in the noncooperative game theory to formulate solution concepts such as strong Nash equilibrium and coalition-proof correlated equilibrium. See Aumann [1], Bernheim, Peleg and Whinston [4], and Moreno and Wooders [6]. In these equilibrium concepts, a deviating coalition assumes that the outsiders of the coalition stick to the status quo. Whereas, in the core reformulated in this paper, the outsiders may adjust their actions in response to the deviating coalition. In situations involving externality, it may be physically infeasible for the outsiders to stick to the status quo when a coalition deviates from it. That is true even in an Arrow-Debreu economy with jointly owned firms, as noted by Xiong and Zheng [14]. In addition, when a coalition deviates, it may be natural for its outsiders to work out a response through coalitional cooperation.

Section 2 explains the idea of the reformulated core through a pollution problem. Sections 3–4 formalize. Section 5 presents a noncooperative foundation for the core. Section 6 characterizes the core in a general problem with pollution externality.

2 A Motivating Example of Pollution

To explain the idea thoroughly, let us consider the following pollution problem distilled from Osborne and Rubinstein [8, p261]. Three factories share a lake, each drawing water from and discharging waste into it. Each factory incurs a cost kc to purify its water supply if k is the number of factories that take the action *pollute*, i.e., not treating one's waste. It costs any factory b to *clean* its waste before discharging it. Thus, if a factory takes the action Clean, its payoff is equal to $-b - kc$; else the payoff is $-kc$. Assume that

$$c < b < 2c. \tag{1}$$

If this situation is analyzed as a noncooperative game where the factories independently choose between Pollute and Clean, then Clean is strictly dominated by Pollute due to $c < b$.

Thus, the unique Nash equilibrium is that every factory pollutes; furthermore, the strong Nash and coalition-proof correlated equilibrium concepts cannot help either.

The question that we shall focus on is Can the factories improve upon their welfare through cooperation without resorting to repeated games or a central planner? One would naturally think of the cooperative game theory.

The sticking point of applying the cooperative game theory, however, is that the total payoff for a coalition depends upon the actions of those outside the coalition, as the cost kc of pollution enters the payoff function for every factory. To apply the cooperative theory, we need to handle a problem: when a coalition is deviating from a status quo, what would the complementary coalition do?

It would not do to assume that the outsiders stick to the status quo. As mentioned previously, given any actions that the outsiders would take, Clean is dominated by Pollute.

Aumann's [2] β-core could be an answer, which requires that a blocking plan should benefit the coalition no matter what the complementary coalition does. Then it would be a core solution that every factory takes the action Clean. If a single factory deviates, its payoff is reduced from $-b$ to $-2c < -b$ by (1) in the worst-case scenario where the other two factories both play Pollute. If two factories deviate, their total payoff reduce from $-2b$ to at most $2(-2c) = -4c$ no matter what the other factory does. The grand coalition certainly does not gain from deviation.

The problem of the above β-core solution is that the worst-case scenario is not credible. In fact, if it has been given that a factory plays Pollute, the best that the other two factories can do is for each of them to play Clean. To see that, first note that the total payoff for these two factories is maximized when they both play Clean. If both play Clean, then each of the two factories gets $-b - c$. If they both play Pollute, then each gets $-3c < -c - b$ by (1). If one plays Clean and the other plays Pollute, then the total payoff for the two of them is equal to $(-b - 2c) + (-2c) < 2(-b - c)$ by (1). Thus, given that a factory plays Pollute, it is uniquely Pareto optimal for the other two factories to both play Clean. Second, note that neither of these two factories can profit from deviating from Clean. If one of them plays Pollute, then the other factory would play Pollute in response because Pollute dominates Clean when the actions of the other two factories have been given (to be Pollute). Hence the deviator gets $-3c$, which by (1) is less than the payoff $-c - b$ from playing Clean.

Therefore, not only do we see that the worst-case scenario is too incredible to support the β-core solution, we also see that if one of the three factories can commit to Pollute, the other two factories would react with Clean. Furthermore, this reaction itself has the core-like property that it is Pareto optimal within the two factories and neither of the two can profit from a unilateral deviation from the reaction.

That motivates an idea: if a coalition manages to commit to its deviating action, the reaction from the complementary coalition should satisfy a core-like condition within the complementary coalition. This is somewhat similar to the idea of subgame perfection.

Let us consider the decision of a factory according to this idea. If a factory can commit to Pollute before anyone else does, then by the above deduction the other factories will play Clean and hence its payoff is equal to $-c$. Whereas, if the factory abides by Clean, its payoff is a lower $-b$ if the others all play Clean and is even lower if the others do not all play Clean. But then every factory would like to grab this first-mover advantage.

The question becomes Who gets to be the first to commit? To prescribe a self-supported mechanism of cooperation without resorting to an exogenous hierarchy among the players, let us not assume that there is an a priori sequential protocol according to which the factories take turn to make commitments. Even if there were such a hierarchy, either given exogenously or drawn randomly "by nature," one would still need an exogenous enforcer to prevent the second- and third-movers from disobeying the protocol.

In a decentralized environment without exogenous hierarchy among the players, let us assume that every coalition who wants to gain the commitment power has a chance to win it. Specifically, suppose that anyone who wants to make a commitment needs to use a *commitment device*, say signing a binding contract with an attorney. To use the device, a coalition needs to submit a request. If it is not challenged by others, the request is granted. However, any individual outside the coalition has the option to challenge it by submitting a competing request. An interpretation is that the individual, expecting the externality caused by the coalition, petitions the court to nullify the contract with which the coalition commits to its action. Suppose further that, when different parties have submitted competing requests for the commitment device, the winner of the device is randomly selected from one of the contenders, with probability proportional to the number of the individuals in each party. An interpretation is that the arbitration in the court involves randomness so that each side could win. Assuming the winning probabilities to be proportional to the size of the parties is merely for simplification. In general, the winning probabilities may depend on the resources initially endowed by the members of the contending parties such as political clout and resources for violence.

Based on such an institutional background, when a factory wants to be the first to commit to Pollute, it is a Nash equilibrium for each of the other two factories to challenge the first factory by submitting a competing request to commit to Pollute. To see that, consider any of the two factories, say j. If j abides by this proposed equilibrium, then all the three factories are competing for the commitment device, and each factory has probability 1/3 to commit to Pollute. Once the winner commits to Pollute, the other two will play Clean, as deduced previously. Thus, the payoff for factory j is equal to $-c - \frac{2}{3}b$. If j does not compete for commitment, then only two factories are competing. No matter which one wins, factory j and the loser will play Clean. Hence j's payoff is equal to $-c - b$. If j competes for commitment but tries to commit to Clean instead of Pollute, then its expected payoff is $-b - \frac{2}{3}c$, which is less than $-c - \frac{2}{3}b$ by (1). Thus, it is the best response for j to submit a competing request to commit to Pollute.

It follows that no single factory can profit from a deviation from the cooperative arrangement that everyone plays Clean. If a factory abides by the arrangement, its payoff is equal to $-b$. If it deviates, i.e., tries to commit to Pollute, then the factory knows that the other two will respond by submitting competing requests for the commitment device. Hence the expected payoff for the deviant is equal to

$$\frac{1}{3}(-c) + \frac{2}{3}(-c - b) = -c - \frac{2}{3}b,$$

which is less than $-b$ since $b < 3c$ by (1).

We are ready to spell out a solution for the pollution problem:

1. All three factories together commit to Clean.
2. If any factory deviates from the above step, then
   a. each of the other two factories submits a request to commit to Pollute;
   b. if a factory has committed to Pollute, the other two factories both play Clean.
3. If any two factories deviate, the third factory submits a request to commit to Pollute;
   a. if the 2-factory coalition wins the commitment device, the third factory reacts with Pollute;

b. if the third factory wins, the 2-factory coalition reacts with (Clean, Clean).

By previous deductions, step 2a. constitutes a Nash equilibrium for the two factories who are not the deviator, and steps 2b. and 3b. are each the unique core-like reaction for the two factories who lose the competition for the commitment device. Step 3a. is uniquely optimal for the third factory who loses the competition for commitment, because the actions of the other two factories have been given and so Clean is dominated by Pollute. In step 3, when two factories are deviating, it is uniquely optimal for the third factory to submit a competing request to commit to Pollute. In doing so, its expected payoff is equal to

$$\frac{1}{3}(-c) + \frac{2}{3}(-c - xc) = -c - \frac{2}{3}xc,$$

where $x \in \{1, 2\}$ denotes the number of polluters in the two-firm deviating coalition. In the above calculation we have used steps 3a. and 3b.. By contrast, if the factory does not compete for the commitment device, its payoff is equal to $-c - xc$ by step 3a.. Hence the third factory is better-off competing for the commitment device.

As calculated previously, no single factory can profit from deviating from the cooperative arrangement step 1, expecting the response in steps 2a. and 2b..

Can a two-firm coalition profit from deviation? Say they plan to have $x \in \{1, 2\}$ firms in their coalition to play Pollute. Then by step 3, the total expected payoff for the coalition is equal to

$$\begin{aligned}
& \frac{2}{3}(2(-c - xc) + (2 - x)(-b)) + \frac{1}{3}(2(-c - b)) \\
< \; & \frac{2}{3}(2(-c - xc) + (2 - x)(-c)) + \frac{1}{3}(2(-c - b)) \quad \text{since } b > c \text{ by } (1) \\
< \; & -\frac{7}{3}b \quad \text{since } b < 2c \text{ by } (1).
\end{aligned}$$

By contrast, if the two firms do not deviate, their total payoff is equal to $-2b$. Thus, the two-firm coalition cannot profit from deviation.

Thus, steps 1–3 constitutes a core-like solution. It supports the Pareto optimal arrangement that every factory plays Clean, and no coalition among the three factories can profit from deviation. Furthermore, the *preemptive* steps 2a. and 3 each constitute a Nash equilibrium among those outside the deviating coalition, and the *reactive* steps 2b., 3a. and 3b. each have the core-like property among those who have not committed an action.

Why do we require that a preemptive step constitutes a Nash equilibrium instead of a core among those outside the deviating coalition? That is to capture the surprise generated by the coalition's deviation from what has been expected. Surprised, those outside the deviating coalition may have little time to organize among themselves while the deviating coalition is about to get the commitment device unless the outsiders take immediate actions.

3 The Primitives

Let I be the set of players. Let $\mathbb{R}^l$ be this set of consumption bundles. When $l = 1$, a bundle means a quantity of payoff; when $l > 1$, the payoff is multidimensional. Every player $i \in I$ has a preference relation $\succeq_i$ on $\mathbb{R}^l$, is von Neumann Morgenstern rational, and $\succeq_i$ is represented by a utility function u_i. A profile $(z_i)_{i \in I} \in \left(\mathbb{R}^l_+\right)^I$ of consumption bundles across players is called an *allocation*.

Every player $i \in I$ has a set A_i of feasible actions. If $\varnothing \neq S \subseteq I$, denote

$$a_S := (a_i)_{i \in S} \in \Pi_{i \in S} A_i =: A_S$$

and likewise for $a_{\neg S}$ and $A_{\neg S}$, with $\neg S := I \setminus S$, $a_{\neg I} = a_{\varnothing} = \text{nil}$ and $\Pi_{i \in \neg I} A_i := \{\text{nil}\}$. Hence a_S is a profile of actions of the players in S, which we may regard as a joint action, or simply *action*, of S as a coalition.

The relationship between actions and outcomes is dictated by a family $(F_S)_{\varnothing \neq S \subseteq I}$ of *action-outcome correspondences*

$$F_S : \Pi_{i \in I} A_i \twoheadrightarrow \left(\mathbb{R}^l\right)^S$$

such that, for every $a \in A_I$, $F_S(a)$ is a subset of $\left(\mathbb{R}^l\right)^S$. The interpretation is that any element z_S of $F_S(a)$ is a feasible allocation among the players in S given the action profile a. Interdependency between a coalition S and its complement $\neg S$ is incorporated, as $F_S(a)$ may vary with the component $a_{\neg S}$ in a.

For example, in the pollution problem studied in §2, with $|T|$ denoting the size of set T,

$$F_S(a) = \left\{ (z_i)_{i \in S} \in \mathbb{R}^S : \sum_{i \in S} z_i \leq \ -b \left| \{ i \in S : a_i = \text{Clean} \} \right| - c|S| \left| \{ j \in I : a_j = \text{Pollute} \} \right| \ \right\}. \tag{2}$$

4 The Core

4.1 Solutions and Coalitional Responses

We are interested in predictions of the actions and allocations for the players, which we call solutions. If $T \subsetneq I$ is the set of players whose action $a_T \in A_T$ has been determined, a *solution* for the set $\neg T$ of players consists of a lottery $\alpha_{\neg T} \in \Delta A_{\neg T}$ of the actions for $\neg T$ and a *contingent allocation* $x_{\neg T}$, which is a function

$$x_{\neg T} : \operatorname{supp} \alpha_{\neg T} \to \left(\mathbb{R}^l\right)^{\neg T},$$

meaning that if $a_{\neg T}$ is in the support of $\alpha_{\neg T}$ and is selected as the action for $\neg T$ then $x_{\neg T}(a_{\neg T})$ is the allocation among the players in $\neg T$.

A solution (α_S, x_S) is *feasible* given $a_{\neg S} \in A_{\neg S}$ iff

$$\forall a_S \in \operatorname{supp} \alpha_S : x_S(a_S) \in F_S(a_S, a_{\neg S}).$$

The core is a set of feasible solutions subject to certain conditions that we shall introduce.

To formalize the core conditions, it is helpful to imagine the following institutional background motivated in §2. Any coalition S of players who have not committed to certain actions may submit a request for commitment, and the players in $\neg S$ who have not committed may respond. Specifically, the commitment request from S is in the form of a lottery $\alpha_S \in \Delta A_S$ that determines an action $a_S \in A_S$ if S wins the commitment device. Seeing α_S, the players $i \in \neg S$ independently submit competing commitment requests $\alpha_i \in (\Delta A_i) \cup \{\text{nil}\}$, where nil means that i does not submit a competing request. This is the *preemptive* response from $\neg S$. The usage of the commitment device is assigned to one of the competing parties with probabilities proportion to the sizes of the parties. If any party say S' wins the commitment device, the action $a_{S'} \in A_{S'}$ is selected according to the lottery $\alpha_{S'}$ and is committed to by S', and then the players in $\neg S'$ who have not committed react with an action $a_{\neg S'} \in A_{\neg S'}$. This is part of the *reactive* response from $\neg S'$.

Such "strategies" played by coalitions are formalized into the notion *coalitional response*. Suppose that a set $T \subsetneq I$ of players has committed to an action $a_T \in A_T$. For the rest of the economy, a coalitional response is a pair of two functions. The first is a function

$$\mathscr{P}_{T,a_T} : (S, \alpha_S) \longmapsto \left(\mathscr{P}^i_{T,a_T}(S, \alpha_S)\right)_{i \in \neg(S \cup T)}$$

that maps any coalition $S \subsetneq \neg T$ and its lottery α_S of actions to a preemptive response $\left(\mathscr{P}^i_{T,a_T}(S,\alpha_S)\right)_{i\in\neg(S\cup T)}$ for the players in $\neg(S\cup T)$ such that $\mathscr{P}^i_{T,a_T}(S,\alpha_S)\in(\Delta A_i)\cup\{\text{nil}\}$ for all players $i\in\neg(S\cup T)$. The second is a function

$$\mathscr{R}_{T,a_T} : (S,a_S) \longmapsto \mathscr{R}_{T,a_T}(S,a_S)$$

that maps any coalition $S \subsetneq \neg T$ and its committed action a_S to a reactive response $\mathscr{R}_{T,a_T}(S,a_S)$ such that $\mathscr{R}_{T,a_T}(S,a_S)$ is a feasible solution for $\neg(S\cup T)$ given (a_T,a_S).

If $\sigma_S := (\alpha_S, x_S)$ is a solution for a set S of players, denote

$$\pi^\alpha(\sigma_S) := \alpha_S,$$

and for any $i\in S$ denote $\pi_i(\sigma_S)$ for the lottery on player i's consumption bundles induced by σ_S (so $\pi_i(\sigma_S)\in\Delta\mathbb{R}^l$).

If $i\notin S\subsetneq I$, $\alpha_S\in\Delta A_S$, $a_{\neg S}\in A_{\neg S}$, denote

$$u_i(\alpha_S,a_{\neg S}) := \sum_{a_S\in A_S}\alpha_S(a_S)\max_{z_i\in F_{\{i\}}(a_S,a_{\neg S})}u_i(z_i). \tag{3}$$

Hence $u_i(\alpha_S,a_{\neg S})$ is the maximum expected utility for player i when i is free to choose any consumption bundle that is feasible given the committed actions and the lottery α_S of the actions of the set S (which does not include i).

4.2 The Definition of Blocking

Suppose a set $T\subsetneq I$ of players has committed to an action $a_T\in A_T$, a solution $\sigma_{\neg T}$ for $\neg T$ is *blocked* by a coalition $S\subseteq\neg T$ with respect to a coalitional response $(\mathscr{P}_{T,a_T},\mathscr{R}_{T,a_T})$ iff there exists a solution σ'_S for S such that

1. σ'_S is feasible if S gets to commit, i.e., for any $a_{\neg(S\cup T)}\in\operatorname{supp}\pi^\alpha\left(\mathscr{R}_{T,a_T}(S,a'_S)\right)$, σ'_S is feasible given $(a_{\neg T},a_{\neg(S\cup T)})$, and

2. for every $i\in S$, σ'_S makes player i better-off than $\sigma_{\neg T}$, i.e.,

$$\begin{aligned}&\frac{|S|}{|\neg T|}u_i\left(\pi_i(\sigma'_S)\right)+\sum_{j\in\neg(S\cup T)}\frac{1}{|\neg T|}\sum_{a_j\in A_j}\left(\mathscr{P}^j_{T,a_T}(S,\pi^\alpha(\sigma'_S))(a_j)\right)u_i\left(\pi_i\left(\mathscr{R}_{T,a_T}\left(\{j\},a_j\right)\right)\right)\\ >\;& u_i\left(\pi_i(\sigma_{\neg T})\right).\end{aligned}\tag{4}$$

Ineq. (4) reflects the institutional background described previously. With probability $\frac{|S|}{|\neg T|}$ the coalition S wins the commitment device and so player i gets the lottery $\pi_i(\sigma'_S)$ of bundles according to the alternative solution σ'_S. But the players in $\neg(S \cup T)$ challenge S according to the preemptive response $\mathscr{P}_{T,a_T}(S, \pi^\alpha(\sigma'_S))$. With probability $\frac{1}{|\neg T|}$ the commitment opportunity is given to a $j \in \neg(S \cup T)$ outside S, who will then commit to an action a_j with probability $\mathscr{P}^j_{T,a_T}(S, \pi^\alpha(\sigma'_S))(a_j)$. If j commits to an action a_j, then the rest of the economy will react with a solution $\mathscr{R}_{T,a_T}(\{j\}, a_j)$, which in turns induces a lottery $\pi_i(\mathscr{R}_{T,a_T}(\{j\}, a_j))$ of the consumption bundles for player i.

When $\neg T = \{i\}$ for some player i, the above notion that a solution σ_i for player i is blocked becomes equivalent to the notion that σ_i is not a best response to $a_{\neg\{i\}}$ for player i. That is because if $\varnothing \neq S \subseteq \neg T$ then $\neg(S \cup T) = \varnothing$.

4.3 The Definition of the Core

Suppose a set $T \subsetneq I$ of players has committed to $a_T \in A_T$. The core $\mathscr{K}(T, a_T)$ is defined recursively as follows.

1. If $\neg T = \{i\}$ for some $i \in I$, then $\mathscr{K}(T, a_T)$ is the set of all the solutions for $\{i\}$ that maximizes i's payoff among all the feasible solutions given (T, a_T).

2. If $|\neg T| > 1$, a solutions $\sigma_{\neg T}$ for $\neg T$ belongs to the core $\mathscr{K}(T, a_T)$ iff $\sigma_{\neg T}$ is feasible given (T, a_T) and there exists a coalitional response $(\mathscr{P}_{T,a_T}, \mathscr{R}_{T,a_T})$ such that

   a. $\sigma_{\neg T}$ is not blocked with respect to $(\mathscr{P}_{T,a_T}, \mathscr{R}_{T,a_T})$,

   b. if any $S \subsetneq \neg T$ requests commitment according to some $\alpha_S \in \Delta A_S$, then the preemptive response $\left(\mathscr{P}^i_{T,a_T}(S, \alpha_S)\right)_{i \in \neg(S \cup T)}$ constitutes a Nash equilibrium, i.e., for all $i \in \neg(S \cup T)$,

$$\begin{aligned}
&\frac{1}{|\neg T|} \sum_{a_i \in A_i} \left(\mathscr{P}^i_{T,a_T}(S, \alpha_S)(a_i)\right) u_i\left(\pi^\alpha\left(\mathscr{R}_{T,a_T}(\{i\}, a_i)\right), a_i, a_T\right) \\
&+ \sum_{j \in \neg(S \cup T \cup \{i\})} \left(\frac{1}{|\neg T|} - \frac{1}{|\neg T| - 1}\right) \sum_{a_j \in A_j} \left(\mathscr{P}^j_{T,a_T}(S, \alpha_S)(a_j)\right) u_i\left(\pi_i\left(\mathscr{R}_{T,a_T}(\{j\}, a_j)\right)\right) \\
&+ \left(\frac{|S|}{|\neg T|} - \frac{|S|}{|\neg T| - 1}\right) \sum_{a_S \in A_S} \alpha_S(a_S) u_i\left(\pi_i\left(\mathscr{R}_{T,a_T}(S, a_S)\right)\right) \geq 0 \qquad (5)
\end{aligned}$$

and

$$\mathscr{P}^i_{T,a_T}(S,\alpha_S) \quad \text{solves} \quad \max_{\alpha_i\in\Delta A_i} \sum_{a_i\in A_i} \alpha_i(a_i) u_i\left(\pi^\alpha\left(\mathscr{R}_{T,a_T}(\{i\},a_i)\right), a_i, a_T\right), \tag{6}$$

c. if any nonempty set $S \subsetneq \neg T$ has committed to any $a_S \in A_S$, then $\mathscr{R}_{T,a_T}(S,a_S)$ is in the core $\mathscr{K}(T\cup S,(a_T,a_S))$, which is well-defined by recursion and $S \neq \varnothing$.

Ineq. (5) says that player i cannot profit from not submitting a competing commitment request against the deviating coalition S. The first line of (5) corresponds to the event that player i wins the commitment device. Note that the payoff u_i there uses the notation defined in (3). The second line corresponds to the event that the commitment device is won by another outsider of S, whose probability is affected by player i's decision on whether to compete for commitment. The third line corresponds to the event that the deviating coalition S gets to commit, whose probability is again affected by i's decision. Eq. (6) says that, conditional on i's competing for commitment, the preemptive response $\mathscr{P}^i_{T,a_T}(S,\alpha_S)$ maximizes i's expected payoff among all the lotteries $\alpha_i \in \Delta A_i$. The equation does not contain the other components in i's expected payoff, because those terms do not vary with α_i.

4.4 The Pollution Example Revisited

Let us illustrate the above definition with the pollution example in §2. Suppose factory 1 has committed to Pollute. We shall establish the claim "the other two factories play Clean" with the notations introduced here. Given factory 1's commitment, we have $T = \{1\}$ and $a_T = a_1 =$ Pollute. The solution for factories 2 and 3 is the Dirac measure

$$\sigma_{\{2,3\}} = \mathbf{1}_{(\text{Clean},\text{Clean};-b-c,-b-c)}.$$

The claim is that $\sigma_{\{2,3\}}$ is a core-solution for $\neg T = \{2,3\}$ given $(T,a_T) = (\{1\},\text{Pollute})$.

This solution is feasible given (T,a_T), because

$$\begin{aligned} &\operatorname{supp}\sigma_{\{2,3\}} = (\text{Clean},\text{Clean}; -b-c, -b-c) =: (a_2,a_3;z_2,z_3) \\ &z_2+z_3 = -b-c-b-c = (-b)\times 2 + (-c)\times 2\times 1 \quad \text{by (2)}. \end{aligned}$$

The solution $\sigma_{\{2,3\}}$ is supported by the following coalitional response, where $i \neq j$ and $\{i,j\} = \{2,3\}$, $\alpha_j \neq \mathbf{1}_{\text{Clean}}$, and $a_j \in \{\text{Clean},\text{Pollute}\}$.

$$\mathscr{P}^i_{T,a_T}(j,\alpha_j) \quad := \quad \text{nil} \tag{7}$$

$$\mathscr{R}_{T,a_T}(j,a_j) \quad := \quad \mathbf{1}_{(\text{Pollute},-2c-c\mathbf{1}_{a_j=\text{Pollute}})}. \tag{8}$$

Eq. (7) says that factory i's preemptive response is to not compete for commitment against the deviating j, so j knows i will react with Pollute according to (8) if j commits unilaterally.

The first core-condition 2a. says that $\sigma_{\{2,3\}}$ cannot be blocked given $(\mathscr{P}_{T,a_T}, \mathscr{R}_{T,a_T})$. To verify that, consider the coalition $\{2,3\}$ $(= \neg T)$ first. For it to block $\sigma_{\{2,3\}}$, Ineq. (4) requires

$$u_i(\pi_i(\sigma'_{\{2,3\}})) > u_i(\pi_i(\sigma_{\{2,3\}})) = -b - c$$

for both $i \in \{2,3\}$, which is impossible by the fact that, with factory 1 committed to Pollute, the total payoff for the other two factories is maximized when they both play Clean, as already shown in §2. Next consider a singleton coalition $S := \{j\} \subseteq \{2,3\}$ so that $\neg(S \cup T) = \{i\}$. If j deviates, i.e., requests for the commitment device unilaterally, then by (7) player i will not compete for the device, so j gets to commit to some action. Then player i will react with Pollute by Eq. (8). Thus, the payoff for j to deviate is equal to

$$-2c - b\mathbf{1}_{a_j=\text{Clean}} - c\mathbf{1}_{a_j=\text{Pollute}} \leq -3c$$

since $b > c$ by (1). By contrast, if j does not deviate, then its payoff is equal to $-c - b$, which is greater than $-3c$ since $b < 2c$ by (1).

The second core-condition 2b. says that the preemptive response defined by (7) constitutes a Nash equilibrium given that factory 1 has committed to Pollute and another factory j has requested for commitment to an action that is not Clean for sure. In our notation here, $S = \{j\}$ and $\neg(S \cup T) = \neg\{j, 1\} = \{i\}$. Hence the condition becomes that it is a best response for i to not submit a competing request for commitment. To verify the condition, calculate factory i's expected payoff if it requests commitment to Pollute, which is equal to

$$\frac{1}{2} \underbrace{[\alpha_j(\text{Pollute})(-3c) + \alpha_j(\text{Clean})(-2c)]}_{\text{the deviating } j \text{ wins the commitment device}} + \frac{1}{2}(-3c), \tag{9}$$

which is based on the reactive response (8). By the same token, if factory i submits a competing request to commit to Clean, its expected payoff is equal to

$$\frac{1}{2}\left[\alpha_j(\text{Pollute})(-3c) + \alpha_j(\text{Clean})(-2c)\right] + \frac{1}{2}(-2c - b),$$

which is less than (9) because $b > c$ by (1). If i does not compete for commitment at all as (7) recommends, its expected payoff is equal to

$$\alpha_j(\text{Pollute})(-3c) + \alpha_j(\text{Clean})(-2c),$$

which is greater than or equal to (9) since $\alpha_j(\text{Pollute}) \le 1$. Hence condition 2b. is satisfied.

The third core-condition 2c. says that the reactive response $\mathbf{1}_{(\text{Pollute}, -2c - c\mathbf{1}_{a_j = \text{Pollute}})}$ defined by (8) belongs to the core $\mathscr{K}(\{1,j\}, a_{\{1,j\}})$. Since $\neg\{1,j\} = \{i\}$ is singleton, this core-condition amounts to that Pollute is a best response for factory i when the actions of the other factories have been fixed. That follows from $b > c$ by (1). Thus, all the three conditions for $\sigma_{\{2,3\}}$ to belong to the core $\mathscr{K}(\{1\}, \text{Pollute})$ are satisfied.

5 A Noncooperative Foundation for the Core

The institutional background which we informally describe in §4 to motivate the definition of the core is formalized into a noncooperative game here.

5.1 Principals and Contracts

We now add two constructs, principals and contracts. In addition to the set I of players, there is a finite set P of principals such that $I \cap P = \varnothing$. As detailed later, a principal is to coordinate the actions among some players and possibly get a consumption bundle in return. We may think of principals as abstract representations of brokers, attorneys, etc. We assume that every principal k is von Neumann Morgenstern rational and values good 1 and good 1 only in any consumption bundle so that its utility $u_k(z_{k1})$ is strictly increasing in the quantity z_{k1} of good 1 that k consumes. For reasons that will be clear later, we assume

$$|P| \ge |I| + 1. \tag{10}$$

A *contract* is a tuple $(S, \alpha_S, x_S, \tau_S \mid a_T)$ where:

- $S \subseteq \neg T$ is the coalition to be bound by the contract (if $S = \varnothing$ then the contract is nil)
- $\alpha_S \in \Delta A_S$ is the lottery that determines the action for S to commit to
- $T \subseteq \neg S$ has committed to a_T
- (x_S, τ_S) is a function

  $$(\operatorname{supp} \alpha_S) \times A_{\neg(S \cup T)} \to \left(\mathbb{R}^l\right)^S \times \mathbb{R}$$

  such that if the lottery α_S has selected an action a_S for the coalition S and if the set $\neg(S \cup T)$ of all the other players takes the action $a_{(\neg(S\cup T)}$ then $\tau_i(a_S, a_{\neg(S \cup T)})$ is the quantity of good 1 transferred from i to the principal and $x_i(a_S, a_{\neg(S\cup T)})$ is the consumption bundle for $i \in S$ after the transfer, subject to the feasibility condition

$$\left(x_i(a_S, a_{\neg(S\cup T)}) + \left(\tau_i(a_S, a_{\neg(S\cup T)}), 0, \ldots, 0\right)\right)_{i\in S} \in F_S\left(a_S, a_T, a_{\neg(S\cup T)}\right),$$

and the principal receives a *tax revenue* in the form of the following quantity of good 1:

$$\sum_{i\in S} \tau_i(a_S, a_{\neg(S\cup T)}).$$

The above notion of contracts implicitly assumes that a player can transfer payoffs (good 1) to a principal, although transfers need not be transferable directly across players.

Let $\mathscr{C}(T, a_T)$ denote the set of contracts given that T is the set of players who have committed to action a_T. Denote $\mathscr{S}(C)$ for the set of players covered by contract C, i.e.,

$$\mathscr{S}\left(S, \alpha_S, x_S, \tau_S \mid a_T\right) := S.$$

5.2 The Contracting Game

In this game, principals compete by offering contracts to the players, who now act in the role of agents. The game is defined by the following algorithm.

1. Initiate the state variables $T := \varnothing$, $a_T :=$ nil, and $\tilde{P} := \varnothing$.

2. A new *iteration* starts. Every principal in $P \setminus \tilde{P}$ independently chooses a contract from the set $\mathscr{C}(T, a_T)$. Let C^k denote the one chosen by principal k.

3. Partition the set $\{C^k : k \in P \setminus \tilde{P}\}$ into cells such that every cell consists of all the contracts identical to one another; in each cell, exactly one element in the set is randomly selected with uniform probability, and the others are eliminated.

4. If C^k is not eliminated, then it is offered to every member of $\mathscr{S}(C^k)$, and it is common knowledge among I.

5. Every player who is offered a contract independently makes a response to the contract so that a response is either Accept or Reject and a player can accept at most one contract in an iteration.

6. If an offered contract C^k is accepted by all the members of $\mathscr{S}(C^k)$, then principal k submits a request for the commitment device. With equal probabilities, exactly one of such k is randomly selected to be the *coalitional contender* for the commitment device. Denote the coalitional contender by k_*.

7. Observing C^{k_*}, every player $i \in \neg(\mathscr{S}(C^{k_*}) \cup T)$ independently chooses whether to become an *individual contender* for the commitment device and, if Yes, a lottery $\alpha_i \in \Delta A_i$ according to which i will pick an action to commit.

8. For any $m = 0, 1, 2, \ldots$, if players $i_1, \ldots, i_m$ are the individual contenders, then the usage of the commitment device is randomly assigned to exactly one contender. The probability for the coalitional contender k_* to win the usage is equal to

$$\frac{\left|\mathscr{S}(C^{k_*})\right|}{\left|\mathscr{S}(C^{k_*})\right| + m},$$

and the probability for an individual contender to win is equal to $1/\left(\left|\mathscr{S}(C^{k_*})\right| + m\right)$.

9. Denote S^* for the set of players who win the usage of the commitment device. They commit to an action profile $a_{S^*} \in A_{S^*}$ selected according to the lottery α_{S^*}. If $S^* = \mathscr{S}(C^{k_*})$ then C^{k_*} becomes a *winning contract*.

10. Update the state variables

$$\tilde{P} := \tilde{P} \cup \{k_*\}, \quad a_T := (a_T, a_{S^*}), \quad T := T \cup S^*. \tag{11}$$

11. If $T = I$, then all committed actions and existing winning contracts are carried out, any principal who belongs to $P \setminus \tilde{P}$ gets zero tax revenue, and the game ends.

12. If $T \neq I$, then go back to Step 2 with the variables updated in (11), thereby starting a new iteration.

At any node of the above game, a principal is said *available* iff it has not become a coalitional contender, i.e., not yet a member of the set $\tilde{P}$ in (11). As required in Step 2 of the game, only those principals who are still available are allowed to offer contracts.

Lemma 1 *At any node of the contracting game, if $T \subseteq I$ is the set of players who have committed to some actions, then there are at least $|I| - |T| + 1$ available principals.*

Proof In every iteration of the contracting game, either at least one player commits to an action or no player does so. If some player commits, the player does it through exactly one of two channels, either participating in the contract of a principal who turns out to be the coalitional contender in the current iteration, or being an individual contender without a principal. Thus, with T being the set of players who have committed, there are in the history of the game at most $|T|$ iterations where at least one player commits, and hence there are at most $|T|$ principals who have become coalitional contenders. Thus, there are still at least $|P| - |T|$ available principals, and $|P| - |T| \geq |I| + 1 - |T|$ by (10). ■

5.3 The Equilibrium Concept

For any history h of the contracting game, let $\mathcal{T}(h)$ denote the set of players who have already committed to some actions, with the profile of their actions being $a_{\mathcal{T}(h)}$.

A principal k's strategy is a function ω^k that maps any history h where k is still an available principal to an element of $\Delta\mathscr{C}(\mathcal{T}(h), a_{\mathcal{T}(h)})$, i.e., a lottery of contracts that k offers at the history h.

For a player $i \in I$, there are two kinds of strategic nodes, at neither of which i has committed. The first kind is when i is offered some contracts and i is to respond to the offers (Step 5 in the contracting game). Call them the *offered nodes* for player i. The second kind is when a coalitional contender for the commitment device has been formed and the coalition does not include i, so i is to decide whether to be an individual contender (Step 7 in the contracting game). Call them the *competitive nodes* for i. If a history is an offered node for i, let $\tilde{C}_i(h)$ denote the set of contracts offered to i (so that $i \in \mathscr{S}(C)$ for all $C \in \tilde{C}_i(h)$).

The strategy for a player $i \in I$ contains two functions, ρ^i and ψ^i. ρ^i maps any offered node h for player i to an element of $\Delta\left(\tilde{C}_i(h) \cup \{\text{nil}\}\right)$. I.e., $\rho^i(h)$ is a lottery that selects i's response to the contracts offered to i; if an element of $\tilde{C}_i(h)$ is selected then it is the contract that i accepts; if nil is selected then i rejects all the contracts. ψ^i maps any competitive node for player i to any element of $(\Delta A_i) \cup \text{nil}$. I.e., $\psi^i(h)$ tells whether i submits a competing request for commitment (if No then nil is selected) and, if Yes, is a lottery that selects which action i will commit to if i wins the commitment device.

Even when a history h is not a strategic node for principal k or player i, we shall write $\omega^k(h)$, $\rho^i(h)$, and $\psi^i(h)$, which mean inaction in that case. That is to simplify the notations.

Denote a strategy profile by $\left((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\right)$.

Our equilibrium concept is subgame perfect equilibrium subject to two refinement conditions. The first condition is that the equilibrium strategy profile be Markov, though a player is allowed to deviate to non-Markov strategies. Here a strategy is *Markov* means that the action prescribed by the strategy depends only upon the set of agents who have committed to some actions and their committed actions.

Formally, a strategy profile $((\psi^k)_{k\in B}, (\rho^i)_{i\in I})$ is said *Markov* if

$$\begin{aligned} & \big(\mathcal{T}(h), a_{\mathcal{T}(h)}\big) = \big(\mathcal{T}(h'), a_{\mathcal{T}(h')}\big) \\ \Rightarrow \quad & \big((\omega^k(h))_{k\in P}, (\rho^i(h), \psi^i(h))_{i\in I}\big) = \big((\omega^k(h'))_{k\in P}, (\rho^i(h'), \psi^i(h'))_{i\in I}\big)\,. \end{aligned} \tag{12}$$

Our second refinement condition is to guarantee that, when a principal makes a unilateral deviation of offering an off-path contract, the principal can expect the players covered by the contract to accept it as long as the contract, if carried out, makes them all better-off. Without such a condition, a contract may be rejected just because the players expect that it is rejected. We formalize this condition next.

Given any history h and any strategy profile $\big((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\big)$, if the contracting game is at node h and everyone will abide by $\big((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\big)$ from now on, the lottery on the consumption bundles that a player i may eventually get is uniquely determined. Denote the lottery by

$$\pi_i\left(h, \big((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\big)\right).$$

By the same token, if at the current history h, everyone will abide by $\big((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\big)$ except that a contract $C \in \mathscr{C}(\mathcal{T}(h), a_{\mathcal{T}(h)})$ becomes a winning contract against the prescription of $\big((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\big)$, then the lottery

$$\pi_i\left(h, \big((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\big) \mid C\right)$$

on the consumption bundles for player i is also uniquely determined.

A strategy profile $\big((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\big)$ is *open-minded* if for any history h and for any contract C such that $C \in \tilde{C}_i(h)$ for some $i \in I$ (i.e., C is offered to some player at the current history h),

$$\forall i \in \mathscr{S}(C)\left[\pi_i\left(h, \big((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\big) \mid C\right) \succ_i \pi_i\left(h, \big((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\big)\right)\right] \tag{13}$$

$$\Rightarrow \forall i \in \mathscr{S}(C)\left[\rho^i(h) = \mathbf{1}_C\right]. \tag{14}$$

Here (13) says that, compared to the outcome resulting from everyone's obedience to the strategy profile $\left((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\right)$, every player covered by the contract C would be strictly better-off if C is unanimously accepted and becomes a winning contract, and (14) says that all the players covered by the contract accept it.

The open-minded refinement condition is consistent with and weaker than an implicit assumption commonly used in mechanism design, which says that a principal assumes that the agents will be obedient and honest in *any* mechanism chosen by the principal as long as obedience and honesty in the mechanism are weakly preferred. In a multiple-principal model, Bernheim and Whinston [5] have applied this implicit assumption successfully. Similar to our open-minded condition, their assumption is that, whenever a principal deviates by offering an off-path contract, the agent will accept it as long as the agent weakly prefers the deviant contract to the status quo. Our open-minded condition weakens their assumption by requiring that the off-path contract be strictly preferred to the status quo.

We are ready to define the notion of equilibrium. An *equilibrium* of the contracting game is a strategy profile $\left((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\right)$ that is a subgame perfect equilibrium of the game and is Markov, satisfying (12), and open-minded, satisfying (13)–(14).

At any subgame starting from a history h, an equilibrium of the subgame is the projection of some equilibrium of the entire game onto this subgame. For any $T \subseteq I$ and $a_T \in A_T$, at any history h such that $\mathcal{T}(h) = T$ and $a_{\mathcal{T}(h)} = a_T$, the set of equilibria of any the subgame starting from h is uniquely determined by (T, a_T) due to the Markov condition. Thus, denote $\mathscr{E}(T, a_T)$ for the set of equilibria in any such a subgame.

5.4 Competitive Behaviors among the Principals

Let $e \in \mathscr{E}(T, a_T)$, i.e., an equilibrium in a subgame given that $T \subsetneq I$ is the set of players who have committed and $a_T \in A_T$ is their committed action. For any contract C, let $\bar{\tau}(C)$ denote the expected value of the tax revenue specified in contract C such that the expected value is calculated according to the equilibrium e.

For any contract $C \in \mathscr{C}(T, a_T)$ such that $\bar{\tau}(C) > 0$, an *undercut* of contract C, denoted $\overline{C}$, is a contract that is identical to C except that the tax revenue is reduced in expectation (with respect to e) so that (i) the reduction is equally distributed to the players in $\mathscr{S}(C)$ and (ii) the reduction is so small that $\bar{\tau}\left(\overline{C}\right) > 0$.

In the contracting game, a principal receives no tax revenue unless it offers a contract that is accepted, becomes the coalitional contender, and then wins the usage of the commitment device. There are two possible cases in which a principal may run into direct rivalry with other principals. The first is a *tie*, in which at least two principals offer an identical contract in the same iteration of the game. By rule 3 of the game, only one of these principals gets to make this offer to the players, and that one is chosen randomly with uniform probabilities. The second case is a *congestion*, where at least two contracts are accepted in the same iteration. By rule 6 of the game, only one of these contracts, chosen randomly with uniform probabilities, gets to be the coalitional contender in the current iteration.

We shall show that ties and congestion do not occur at equilibrium unless the contracts involved do not bring any tax revenue to the principals. Let us start with an observation that a principal may deviate without fearing penalty in the event where the current deviation does not result in winning a contract. This lemma follows directly from the Markov condition that we require for an equilibrium.

Lemma 2 *At any equilibrium $e \in \mathscr{E}(T, a_T)$, if at the current iteration a principal makes a deviation and if it does not become the coalitional contender in this iteration, the principal's continuation payoff is not affected by the deviation.*

Lemma 3 (no tie) *If $\succeq_i$ is strongly monotone on $\mathbb{R}^l$ for all $i \in I$, and if there is a tie at equilibrium $e \in \mathscr{E}(T, a_T)$ with a contract C such that $\bar{\tau}(C) > 0$, then any principal involved in the tie has a profitable deviation of offering an undercut of C instead of C itself.*

Proof Suppose a principal k involved in the tie deviates from e by offering an undercut $\overline{C}$ instead of the original contract C. Proposing a contract different from C, principal k gets to offer $\overline{C}$ to the players in $\mathscr{S}(C)$ for sure instead of splitting the probability of being the offerer with the other principals in the tie. With strongly monotone preferences, all the players in $\mathscr{S}(C)$ strictly prefer $\overline{C}$ to C. Thus, by the open-minded condition for equilibrium, (13)–(14), they all accept $\overline{C}$ and reject C. Since the reduction in tax revenues can be arbitrarily small, and since the principal's future payoff is not affected by the deviation if the principal remains available (Lemma 2), this is a profitable deviation. ■

Lemma 4 (no congestion) *If $\succeq_i$ is strongly monotone on $\mathbb{R}^l$ for all $i \in I$, and if there is a congestion at equilibrium $e \in \mathscr{E}(T, a_T)$ such that at least one of the accepted contracts specifies strictly positive tax revenues in expectation, then at least one principal involved in the tie has a profitable deviation.*

Proof Pick any two distinct principals k and j involved in the congestion such that $\bar{\tau}(C^k) \leq \bar{\tau}(C^j)$ and $\bar{\tau}(C^j) > 0$ (such j exists by hypothesis). Suppose principal k deviates from e by offering an undercut $\overline{C^j}$ of j's contract C^j instead of k's original contract C^k. Principal k can avoid ties by adjusting the tax revenue in $\overline{C^j}$ slightly so that it is different from any other contract being proposed. With strongly monotone preferences, all the players in $\mathscr{S}(C^j)$ strictly prefer $\overline{C^j}$ to C^j. Thus, by the open-minded condition for equilibrium, they all accept $\overline{C^j}$ and reject j's offer C^j, which they would have accepted had k not deviated. Eliminating one of its rivals, principal k increases its probability of being the coalitional winner. Since the reduction from $\bar{\tau}(C^j)$ to $\bar{\tau}\left(\overline{C^j}\right)$ can be arbitrarily small and since $\bar{\tau}(C^k) \leq \bar{\tau}(C^j)$, this is a profitable deviation for k by Lemma 2. ■

Proposition 1 (zero profit) *If $\succeq_i$ is strongly monotone on $\mathbb{R}^l$ for all $i \in I$, then at any equilibrium $e \in \mathscr{E}(T, a_T)$, the expected value of the tax revenue for every principal $k \in P$ is equal to zero.*

Proof Since a principal can offer the null contract (whose corresponding coalition is the empty set), its expected revenue at equilibrium e is nonnegative. We shall show that the expected revenue cannot be greater than zero.

Let Prob_e denote the probability measure of the events on the path of e assessed at the beginning of the subgame given (T, a_T).

First, we claim that on the path of e there is at least one principal who knows that its expected payoff is equal to zero based on Prob_e. The claim is true if there is at least one principal who knows at the outset that its probability of being a coalitional contender is zero. Thus, suppose that there is no such a principal. By Lemma 1,

$$\exists k \in P : 0 < \text{Prob}_e \{k \text{ becomes a coalitional contender on the path of } e\} < 1. \tag{15}$$

For principal k to have a strictly positive expected tax revenue, there must be a contract C^k such that, with E denoting the following event, $0 < \text{Prob}_e(E) < 1$:

E : principal k becomes the coalitional contender via offering C^k, with $\bar{\tau}(C^k) > 0$.

Then the probability for k to become the coalitional contender with C^k conditional on principal k's choosing C^k to offer, denoted

$$\text{Prob}_e\{\text{contender} \mid C^k\},$$

must be less than one, otherwise k's probability of choosing C^k must be less than one, but then k has a profitable deviation of increasing the probability of choosing C^k. (By Lemma 2, the principal does not fear penalty for the deviation.) Now that $\text{Prob}_e\{\text{contender} \mid C^k\} < 1$, if that is due a tie, then Lemma 3 applies; else it is due to congestion, then Lemma 4 applies. In either case, there is a profitable deviation from the equilibrium e. Hence the claim follows.

To complete the proof, suppose, to the contrary of the proposition, that there is a principal j whose expected payoff at e is greater than zero. Then j must offer a contract C^j that becomes a winning contract with strictly positive probability and $\bar{\tau}(C^j) > 0$. Then a principal k whose expected revenue is zero, who exists by the claim we have just proved, has a profitable deviation of offering an undercut $\overline{C^j}$ of j's contract C^j. With strongly monotone preferences, all the players in $\mathscr{S}(C^j)$ strictly prefer $\overline{C^j}$ to C^j. Thus, by the open-minded condition for equilibrium, they all accept k's offer $\overline{C^j}$ and reject j's C^j, which they would have accepted had k not deviated. Since the reduction from $\bar{\tau}(C^j)$ to $\bar{\tau}\left(\overline{C^j}\right)$ can be arbitrarily small, this is a profitable deviation for k by Lemma 2. ■

5.5 Equivalence between the Core and the Set of Equilibria

For any $T \subseteq I$ and $a_T \in A_T$, if $e \in \mathscr{E}(T, a_T)$, i.e., if e is an equilibrium of a subgame that starts from a history where T is the set of players who have committed to action a_T, then e determines a lottery $\alpha_{\neg T}$ of the actions for $\neg T$. This lottery is generated by the winning contracts and the preemptive responses that occur on the path of e. The lottery, coupled with these winning contracts, generates a contingent allocation $x_{\neg T}$ among the players in $\neg T$. Thus, e determines a solution for $\neg T$ in the language of §4.1. Denote $\tilde{\sigma}(e)$ for this solution. Thus, given (T, a_T), the set $\mathscr{E}(T, a_T)$ of equilibria and the core $\mathscr{K}(T, a_T)$ are comparable. Moreover, the set

$$\tilde{\sigma}\left(\mathscr{E}(T, a_T)\right) := \{\tilde{\sigma}(e) : e \in \mathscr{E}(T, a_T)\}$$

and the core $\mathscr{K}(T, a_T)$ are subsets of the same set.

Theorem 1 *For any $T \subsetneq I$ and any $a_T \in A_T$, $\mathscr{K}(T, a_T) \subseteq \tilde{\sigma}\left(\mathscr{E}(K, a_T)\right)$.*

Proof We shall use induction on the size of $\neg T$. The case of $|\neg T| = 1$ is trivial, as the equilibrium and core conditions in this case are the same, that the solution be a best response to a_T for the single player in $\neg T$.

Consider any case such that $|\neg T| > 1$. Pick any $\sigma_{\neg T} \in \mathscr{K}(T, a_T)$. The proof is complete if $\sigma_{\neg T} = \tilde{\sigma}(e)$ for some $e \in \mathscr{K}(T, a_T)$. By the definition of the core, $\sigma_{\neg T}$ is supported by a coalitional response $(\mathscr{P}_{T,a_T}, \mathscr{R}_{T,a_T})$. If $\varnothing \neq S \subsetneq \neg T$ and $a_S \in A_S$, the core condition 2c. implies that $\mathscr{R}_{T,a_T}(S, a_T) \in \mathscr{K}(T \cup S, (a_T, a_S))$, hence the induction hypothesis implies that

$$\exists \tilde{e}(S, a_S) \in \mathscr{E}(T \cup S, (a_T, a_S)) : \mathscr{R}_{T,a_T}(S, a_T) = \tilde{\sigma}(\tilde{e}(S, a_S)). \tag{16}$$

Define a contract

$$C := (\neg T, \alpha_{\neg T}, x_{\neg T}, \tau_{\neg T} \mid a_T)$$

such that $(\alpha_{\neg T}, x_{\neg T}) = \sigma_{\neg T}$ and $\tau_i := 0$ for all $i \in \neg T$. Hence C covers all the players who have not committed, and the core solution $\sigma_{\neg T}$ for $\neg T$ is carried out if C is carried out.

It follows that we are done if the following strategy profile is an equilibrium in $\mathscr{E}(T, a_T)$, as C is the winning contract on the path of the strategy profile.

1. Every principals who is still available offers to the coalition $\neg T$ the contract C.

2. If C is among those offered, every player in $\neg T$ accepts C and rejects any other contract.

3. If C is not among those offered, then—

   a. for any offered contract C', the outcome "C' is accepted by all players in $\mathscr{S}(C')$" is evaluated according to the expectation that every player $i \notin \mathscr{S}(C')$ takes the preemptive response $\mathscr{P}^i_{T,a_T}(\mathscr{S}(C'), \tilde{\alpha}(C'))$ if C' becomes the coalitional contender and that $\neg(\mathscr{S}(C') \cup T)$ takes the reactive response $\mathscr{R}_{T,a_T}(\mathscr{S}(C'), \tilde{\alpha}(C'))$,

   b. the players who are offered a contract respond according to a Nash equilibrium of the strategic-form game defined by the outcome-evaluation system described in provision 3a., then the continuation play follows provisions 4–6.

4. If a nonempty $S \subsetneq \neg T$ becomes a coalitional contender with a lottery α_S of actions, then every player $i \in \neg(S \cup T)$ takes $\mathscr{P}^i_{T,a_T}(S, \alpha_S)$ as the preemptive response.

5. If a nonempty $S \subsetneq \neg T$ is the set of players who commit to some action $a_T \in A_T$, then the play of the subgame thereof is $\tilde{e}(S, a_S)$ specified by (16).

6. If no player has committed, then the subgame repeats the procedure starting from provision 1, where every available principal offers C.

To verify the equilibrium condition for the above strategy profile, first recall from Lemma 1 that there are at least two available principals at the start of this subgame. Expecting the other principals to offer contract C, every principal knows that it cannot profit from offering a different contract, which is rejected by provision 2. Hence provision 1 is any available principal's best response.

In the case of provision 2, suppose a player i is offered both C and another contract $C' \neq C$. If $\mathscr{S}(C') \neq \{i\}$, then it is a best response for i to reject C' since i expects the players in $\mathscr{S}(C') \setminus \{i\}$ to do so. If $\mathscr{S}(C') = \{i\}$, then C' cannot make i better-off even if C' becomes the winning contract, otherwise the singleton coalition $\{i\}$ would block the solution $\sigma_{\neg T}$. Rejecting all the contracts is not a profitable deviation, because if all offered contracts are rejected then provision 6 implies that only the contract C will be offered in the subgame. Hence again it is a best response for i to accept C and reject any other C'.

In the case of provision 3, the outcome-evalution method 3a. is based on provisions 4–6. A Nash equilibrium specified by provision 3b. exist because the strategic form game is finite, as there are only finitely many available principals and hence finitely many contracts for each player to choose.

In the case of provision 4, the preemptive response $\mathscr{P}^i_{T,a_T}(S, \alpha_S)$ is player i's best response, due to condition 2b. in the definition of the core. Provision 5 constitutes an equilibrium in that subgame due to the induction hypothesis, which applies since $S \neq \varnothing$. Since provision 6 is self-referencing the whole strategy profile, it admits no profitable deviation, otherwise there would have been a profitable deviation from one of the previous provisions.

We have therefore verified that the strategy profile constitutes an equilibrium. Since it implements the core solution $\sigma_{\neg T}$, the proof is complete. ■

Theorem 2 *If $\succeq_i$ is strongly monotone and lower semicontinuous on $\mathbb{R}^l$ for all $i \in I$, then for any $T \subsetneq I$ and any $a_T \in A_T$, $\mathscr{K}(T, a_T) \supseteq \tilde{\sigma}\left(\mathscr{E}(K, a_T)\right)$.*

Proof We shall use induction on the size of $\neg T$. As in the proof of Theorem 1, the case of $|\neg T| = 1$ is trivial.

Consider any case such that $|\neg T| > 1$. Pick any equilibrium $e := \left((\omega^k)_{k\in P}, (\rho^i, \psi^i)_{i\in I}\right) \in \mathscr{E}(T, a_T)$. The proof is complete if $\tilde{\sigma}(e) \in \mathscr{K}(T, a_T)$.

To this end, construct a coalitional response $(\mathscr{P}_{T,a_T}, \mathscr{R}_{T,a_T})$. For any $S \subsetneq \neg T$ and any $a_S \in A_S$, let $h(T \cup S, a_T, a_S)$ denote any history of the contracting game where $T \cup S$ is the set of players who have committed and their committed action is (a_T, a_S). Denote $e|_{S,a_S}$ for the projection of the equilibrium e onto the subgame starting from history $h(T \cup S, a_T, a_S)$. (The Markov condition in our equilibrium concept implies that $e|_{S,a_S}$ is uniquely determined by e and $(T \cup S, a_T, a_S)$.) Clearly, $e|_{S,a_S} \in \mathscr{E}(T \cup S, (a_T, a_S))$. Thus, by the induction hypothesis,

$$\varnothing \neq S \subsetneq \neg T \Longrightarrow \mathscr{R}_{T,a_T}(S, a_S) := \tilde{\sigma}\left(e|_{S,a_S}\right) \in \mathscr{K}(T \cup S, (a_T, a_S)). \tag{17}$$

Also, for any $S \subsetneq \neg T$, let

$$\forall i \in \neg(T \cup S) : \mathscr{P}^i_{T,a_T}(S, a_S) := \psi^i(h(T \cup S, a_T, a_S)). \tag{18}$$

We are done if $\tilde{\sigma}(e)$ satisfies the core conditions 2a.–2c. with respect to $(\mathscr{P}_{T,a_T}, \mathscr{R}_{T,a_T})$. Condition 2c. follows directly from (17). Condition 2b. follows from (18), as the preemptive responses $\psi^i(h(T \cup S, a_T, a_S))$ best reply one another by the equilibrium property of e.

To verify condition 2a., suppose, to the contrary, that $\tilde{\sigma}(e)$ is blocked, with respect to $(\mathscr{P}_{T,a_T}, \mathscr{R}_{T,a_T})$, by a coalition $S \subseteq \neg T$ with a feasible solution σ'_S for S. Hence the blocking condition, Ineq. (4), is satisfied with $\tilde{\sigma}(e)$ taking the role of $\sigma_{\neg T}$ there. With lower semicontinuous preferences, the strict inequality (4) remains true after σ'_S is modified by reducing the good-1 consumption for every $i \in S$ by a sufficiently small amount $\tau_i(a_S, a_{\neg(T \cup S)})$ for all $a_S \in \operatorname{supp} \pi^\alpha(\tilde{\sigma}(e))$ and all $a_{\neg(T \cup S} \in A_{\neg(T \cup S)}$. Then for the contract

$$C := (S, \sigma'_S, \tau_S \mid a_T),$$

$\bar{\tau}(C) > 0$ with respect to the equilibrium e. By Lemma 1, there is a principal k who is still available. By Proposition 1, k's expected tax revenue is zero at e. Suppose k deviates by offering C to the players in S. For every player $i \in S$, Ineq. (4) implies that i strictly prefers C, conditional on the event that C becomes the coalitional contender, to the solution $\tilde{\sigma}(e)$ induced by e. Conditional on the complementary event where another, on-path contract becomes the coalitional contender, i is indifferent because i will follow the on-path preemptive response to that on-path coalitional contender anyway. Thus, both cases considered, player i strictly prefers C to $\tilde{\sigma}(e)$. Then the open-minded condition for equilibrium implies that everyone in S accepts C. Since $\bar{\tau}(C) > 0$, the deviation is profitable for k. This contradiction shows that the core condition 2a. holds, as asserted. ∎

6 The Core in the General Pollution Problem

Let us consider the pollution problem discussed in §2 except that the number of factories is for any general integer $n = 2, 3, 4, \ldots$. We characterize the core in the general case here.

Proposition 2 *In the pollution problem with n factories, the core is nonempty and, in any core solution, every factory takes the action Clean.*

Proof Since any Pareto dominated solution is blocked ($S = \neg T$ in the blocking condition (4)), the action of every factory is necessarily Clean in any core solution.

For any $T \subsetneq I$ with $|\neg T| \geq 2$ and $a_T \in A_T$ ($= \{\text{Pollute}, \text{Clean}\}^T$), we shall prove that "every factory take the action Clean and gets payoff $-b - c|\{i \in T : a_i = \text{Pollute}\}|$" is a solution in the core $\mathscr{K}(T, a_T)$ given that T is the set of factories who have committed and their committed action is a_T. The proof is by induction on $|\neg T|$. The case with $|\neg T| = 2$ has been demonstrated in §4.4. Let $|\neg T| \geq 3$. Given (T, a_T), $-c|\{i \in T : a_i = \text{Pollute}\}|$ is a constant component in the payoff for every player in $\neg T$. Thus, in the following we calculate a player i's payoff in terms of *net payoff* z_i, which is i's payoff subtracted by the constant $-c|\{i \in T : a_i = \text{Pollute}\}|$.

To support the "everyone cleans" solution, construct a coalitional response as follows. For any $S \subsetneq \neg T$ and $\alpha_S \in \Delta A_S$, let

$$\mathscr{P}^i_{T,a_T}(S, \alpha_S) := \begin{cases} \mathbf{1}_{\text{Pollute}} & \text{if } |\neg(T \cup S)| \geq 2 \\ \text{nil} & \text{if } \neg(T \cup S)| = 1 \end{cases} \tag{19}$$

$$\mathscr{R}_{T,a_T}(S, a_S) := \begin{cases} \mathbf{1}_{\forall i \in \neg(T \cup S): a_i = \text{Clean}, z_i = -b - c|\{j \in S: a_j = \text{Pollute}\}|} & \text{if } |\neg(T \cup S)| \geq 2 \\ \mathbf{1}_{(\text{Pollute}, -c|\{j \in S: a_j = \text{Pollute}\}|)} & \text{if } \neg(T \cup S)| = 1 \end{cases} \tag{20}$$

The core condition 2c. follows from (20) and the induction hypothesis. For the other two conditions, 2a. and 2c., the case $|\neg(T \cup S)| = 1$ has been demonstrated in §4.4, hence we suppose $|\neg(T \cup S)| \geq 2$ for the rest of the proof.

Let us verify the core condition 2b.. Let $m(\alpha_S)$ be the expected value of $|\{i \in S : a_i = \text{Pollute}\}|$ according to the lottery α_S. For every $i \in \neg(T \cup S)$, seeing that S is requesting commitment and expecting the other players in $\neg(T \cup S)$ to respond according to (19), i's expected net payoff from following (the upper branch of) (19) is equal to

$$z_i := \frac{1}{|\neg T|}(-c) + \frac{|S|}{|\neg T|}\left(m(\alpha_S)(-c) - b\right) + \frac{|\neg T| - 1 - |S|}{|\neg T|}(-c - b), \tag{21}$$

where the first term corresponds to the event where player i gets to commit, the second term the event where the coalition S gets to commit, and the third is when some other player gets to commit. In calculating each term we have used the upper branch of (20). If player i requests commitment to Clean instead of Pollute, then i's expected net payoff is the same as (21) except that the first term $\frac{1}{|\neg T|}(-c)$ is replaced by $\frac{1}{|\neg T|}(-b)$. Since $b > c$ by (1), this replacement lowers the payoff. The other alternative for player i is not to request commitment. Then i's expected net payoff is equal to

$$z_i' := \frac{|S|}{|\neg T| - 1}\left(m(\alpha_S)(-c) - b\right) + \frac{|\neg T| - 1 - |S|}{|\neg T| - 1}(-c - b), \tag{22}$$

where the first term corresponds to the event where the coalition S wins the commitment device and the second term is where someone in $\neg(T \cup S \cup \{i\})$ wins. By (21)–(22),

$$z_i - z_i' = \frac{1}{|\neg T|}\left(b + \frac{|S|\left(m(\alpha_S) - 1\right)}{|\neg T| - 1}c\right) \geq \frac{1}{|\neg T|}\left(b + \frac{|S|\,(-1)}{|\neg T| - 1}c\right) \geq \frac{1}{|\neg T|}(b - c) \overset{(1)}{>} 0,$$

where the second last inequality is due to $|(\neg T) \setminus S| = |\neg(T \cup S)| \geq 2$. Thus, it is a best response for i to follow the preemption (19), hence condition 2b. is satisfied.

To verify the core condition 2a., we calculate the total of the net payoffs for coalition S when it tries to block with α_S. Summing the left-hand side of (4) across members of S yields

$$\begin{aligned}
& \frac{|S|}{|\neg T|}\left(-c|S|m(\alpha_S) - b\left(|S| - m(\alpha_S)\right)\right) + \frac{|\neg T| - |S|}{|\neg T|}\left(-c|S| - b|S|\right) \\
= \; & -c\left(\frac{|S|}{|\neg T|}|S|m(\alpha_S) + \frac{|\neg T| - |S|}{|\neg T|}|S|\right) - b\left(\frac{|S|}{|\neg T|}\left(|S| - m(\alpha_S)\right) + \frac{|\neg T| - |S|}{|\neg T|}|S|\right) \\
= \; & \frac{|S|}{|\neg T|}\left[-c\left(|S|m(\alpha_S) + |\neg T| - |S|\right) - b\left(|S| - m(\alpha_S) + |\neg T| - |S|\right)\right] \\
\overset{(1)}{<} \; & \frac{|S|}{|\neg T|}\left[-\frac{b}{2}\left(|S|m(\alpha_S) + |\neg T| - |S|\right) - b\left(-m(\alpha_S) + |\neg T|\right)\right] \\
= \; & -\frac{|S|}{|\neg T|}b\left[\frac{1}{2}\left(|S|m(\alpha_S) + |\neg T| - |S|\right) - m(\alpha_S) + |\neg T|\right] \\
= \; & -b|S|\left[1 + \frac{|S|m(\alpha_S) + |\neg T| - |S| - 2m(\alpha_S)}{2|\neg T|}\right] \\
< \; & -b|S| \quad \text{since } |\neg T| \geq 3.
\end{aligned}$$

Since $-b|S|$ is the total net payoff for the entire coalition S if it abides by the "all clean" solution, it is impossible for (4) to hold for all $i \in S$. Hence S cannot block it. ■

To appreciate Proposition 2, we can look at the alternative approaches to achieve Pareto optimal cooperation among players with externality. One approach is mechanism design, which relies on a neutral trustworthy principal to design and enforce a Pareto optimal mechanism. Such a requirement begs the question "But who will guard the guardians?", which dates back to Plato and has yet to be answered. The approach taken in this paper, by contrast, is decentralized. In the noncooperative game that serves as the institutional foundation for our notion of the core, the principals are themselves profit-seekers. Pareto optimal outcome emerges out of their competition.

The other approach is repeated games. But Pareto optimal cooperation is only one of the many equilibria allowed by repeated games. Furthermore, cooperation via repeated games needs the players to be sufficiently patient, i.e., to be stuck with one another in the future with a sufficiently large probability. In a modern society where people are quite movable, repeated games may be fragile compared to the commitment devices such as the contracts considered in this paper.

References

[1] Robert J. Aumann. Acceptable points in general cooperative n-person games. In H. W. Kuhn and R. D. Luce, editors, *Contributions to the Theory of Games IV*, pages 287–324. Princeton: Princeton University Press, 1959.

[2] Robert J. Aumann. The core of a cooperative game without side payments. *Transactions of the American Mathematical Society*, 98(3):539–552, Mar. 1961.

[3] Robert J. Aumann and Michael Maschler. The bargaining set for cooperative games. In M. Dresher, L. S. Shapley, and A. W. Tucker, editors, *Advances in Game Theory*, pages 443–447. Princeton: Princeton University Press, 1964.

[4] B. Douglas Bernheim, Bezalel Peleg, and Michael Whinston. Coalition-proof Nash equilibria I: Concepts. *Journal of Economic Theory*, 42(1):1–12, June 1987.

[5] B. Douglas Bernheim and Michael Whinston. Common agency. *Econometrica*, 54(4):923–942, July 1986.

[6] Diego Moreno and John Wooders. Coalition-proof equilibrium. *Games and Economic Behavior*, 17:80–112, 1996.

[7] Roger B. Myerson. Incentive-dual methods in game theory: Virtual utility and dual reduction revisited. Mimeo, Department of Economics, University of Chicago, July 2002.

[8] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. The MIT Press: London, 1994.

[9] J. David Perez-Castrillo. Cooperative outcomes through noncooperative games. *Games and Economic Behavior*, 7:428–440, 1994.

[10] Motty Perry and Philip Reny. A noncooperative view of coalition formation and the core. *Econometrica*, 62(4):795–817, July 1994.

[11] Debraj Ray. *A Game-Theoretic Perspective on Coalition Formation*. University Press, Oxford, 2007.

[12] Debraj Ray and Rajiv Vohra. Coalitional power and public goods. *Journal of Political Economy*, 109(6):1355–1384, 2001.

[13] Roberto Serrano. A market to implement the core. *Journal of Economic Theory*, 67:285–294, 1995.

[14] Siyang Xiong and Charles Z. Zheng. Core equivalence theorem with production. *Journal of Economic Theory*, 137(1):246–270, November 2007.

Index