A Coase Theorem Based on a New Concept of the Core

Charles Zhoucheng Zheng

Working Paper No. 09003
March 2009
Revised on February 2010

IOWA STATE UNIVERSITY
Department of Economics
Ames, Iowa, 50011-1070

Iowa State University does not discriminate on the basis of race, color, age, religion, national origin, sexual orientation, gender identity, sex, marital status, disability, or status as a U.S. veteran. Inquiries can be directed to the Director of Equal Opportunity and Diversity, 3680 Beardshear Hall, (515) 294-7612.

A Coase Theorem Based on a New Concept of the Core[*]

Charles Z. Zheng[†]

May 14, 2010

Abstract

The core is reformulated to incorporate the externality typical in strategic form games. Any coalition of players may deviate by trying to commit to a profile of actions different from a status quo. The outsiders of the coalition may take a coordinated measure, incentive-feasibly for themselves, to preempt the coalition's commitment. If a coalition succeeds in committing to its action profile, the outsiders' reactions constitute a core solution among themselves. A core solution is robust against the deviations of coalitions which expect such preemptive and reactive responses from the other players. In an externality problem where pollution is the dominant action, the core is nonempty. In any two-player strategic form game, the core is also nonempty.

[*] I thank Bo Chen, David Frankel, Kyle Hydnman, Rich McLean, Tom Sjöström, Cheng Wang, and Siyang Xiong for discussions of a preliminary draft. I also thank Marcus Berliant, Debraj Ray and Roberto Serrano for literature advice, and the Southern Methodist University, Rutgers University, and the University of Maryland for seminar accommodations.

[†] Before July 2010: Department of Economics, Iowa State University, 260 Heady Hall, Ames, IA 50011, czheng@iastate.edu, http://www.econ.iastate.edu/sites/default/files/profile/cv/CharlesZhengCV.pdf. Starting from July 2010: Department of Economics, University of Western Ontario.

1 Introduction

Presented here is a new concept of the core that incorporates the strategic interactions between any coalitions. Applied to an externality problem where pollution is the dominant action, the new core is nonempty and equal to the set of Pareto optima. This result may be interpreted as a justification for the "Coase Theorem" [9] without assuming property rights or complete markets.

In his original paper, Coase was vague about the institutional setups. Researchers have proposed a few institutional setups in which the Coase Theorem might be true such as property rights, complete markets, and the core.[1] Among them, the core has the merit of not relying on external enforcement. Even if property rights have been defined, it remains to be an issue whether they can be enforced and who can be trusted to enforce them. And the main justification for the solution concept for complete markets, competitive equilibrium, is still based on the core à la the core equivalence theorem (e.g., Debreu and Scarf [10]).

The approach to the Coase Theorem through the core has been unsettling, however. Aivazian and Callen [1] have presented a three-player externality problem where the core is empty. Far worse than the possibility of empty core, the problem is What does the core mean at the presence of externality? Aivazian and Callen's notion of the core is based on a characteristic function that assigns a value to each coalition. Such a traditional notion of the core assumes that the payoff for a coalition is independent of the actions of its outsiders. This assumption is inadequate for externality problems.

Let us consider for example a typical externality problem. There are three players and each may play Pollute or Clean. Playing Clean costs only the player b dollars. Playing Pollute costs the player itself, as well as everyone else, c dollars. Assume that $c < b$, so that Clean is strictly dominated by Pollute. Also assume that $b < 2c$, so that the unique Pareto optimum is for everyone to play Clean. Now that pollution is the unique Nash equilibrium, the coalition-proof equilibrium refinements in noncooperative game theory[2] cannot support the Pareto optimum as an equilibrium. To use cooperative game theory, however, we immediately run into a problem: With the payoffs for any non-grand coalition affected by the actions of its outsiders, how would the complementary coalition respond to a deviating coalition?

[1] See Starrett [22] and Boyd and Conley [8] for two different treatments of the Coase Theorem through the approach of complete markets based on property rights.

[2] Aumann [3], Bernheim, Peleg and Whinston [7], and Moreno and Wooders [16], etc.

An obvious possibility is, extending the idea of Nash equilibrium for normal form games, to assume that the outsiders of a deviating coalition stick to the status quo regardless of whether this coalition deviates or not. This method does not work for our externality problem, as Pollute strictly dominates Clean.

In order to allow the outsiders of a deviating coalition to respond, I adopt the perspective that a coalitional deviation is an attempt to commit to some actions without the consent of those outside the coalition. If a coalition manages to make such a commitment, the response from the others is easy to formulate. Taking the committed actions as given, the outsiders of the coalition would play the "subgame" à la Stackelberg followers. As an equilibrium approach would use the equilibrium condition to predict the reactions of the Stackelberg followers, we shall use our notion of the core to predict the *reactions* to a deviating coalition that manages to commit. This recursive application of the core condition may be called *recursive principle*.

To see how the recursive principle may work, consider our externality problem with only two players and suppose that they can commit to an action either unilaterally or bilaterally as a coalition. If player 1 unilaterally commits to Pollute, then player 2's best response is uniquely Pollute. Thus, in deviating to Pollute, player 1 bears a c-dollar cost due to his own pollution action and another c-dollar cost due to player 2's pollution as her best response. Since $2c > b$, player 1 would rather bear the cost b dollars by playing Clean together with player 2. Same reasoning goes for player 2. Thus, the best- or core-response from the Stackelberg follower might deter a coalition from trying to be a deviant leader.

A question stands in the way, however: Who gets to commit first? To see why this question is unavoidable, recall the third player that we assumed away temporarily. Now suppose that this player 3 manages to commit to Pollute before the other two can make any commitment, then by our previous calculations the other two would stick to Clean. Neither of players 1 and 2 want to deviate to Pollute, nor would they want to deviate as a cooperative coalition, because the only Pareto optimum for players 1 and 2 is both playing Clean ($b < 2c$). Thus, the player who commits first gets a payoff $-c$ while the other two each get $-b - c$. Consequently, every player in our example has a strict incentive to grab the commitment opportunity from the other players.

To handle this question, I propose a notion of preemption. The assumption is that even when it has reached an agreement internally, a coalition is still uncertain whether it can commit to the agreement, as players outside the coalition can challenge it. For instance, some outsiders may try to make a commitment before the former coalition can commit thereby grabbing the first-mover advantage. Or some outsiders may petition the court to nullify the agreement of the coalition on the ground of its negative externality. If the coalition is challenged, the winner is chosen randomly between the contending parties according to a probability distribution that depends on the configuration of the contending parties. The winning probabilities may be determined purely by the sizes of the contending parties, or they may depend on other resources such as political clout and the means of violence.

Naturally the players outside a deviating coalition would like to preempt the deviating commitment. The question is How to predict their preemptive measures? As commitment has not yet been made, our Stackelberg-like core condition for the post-commitment reactions cannot be replicated here. The new construct for the pre-commitment phase is *preemptive mechanism*. The mechanism partitions the complementary coalition into subcoalitions and recommends some of them to try to commit to certain actions before the deviating coalition does. Furthermore, to each subcoalition the mechanism provides a contingency plan on how the participating members of the subcoalition should adjust their actions if some of its members deviate from the recommended preemptive measures. For a preemptive mechanism to be credible, we require an incentive feasibility condition analogous to the incentive condition in mechanism design. The only difference is that incentive feasibility is required not only for individuals but also for any subset of each subcoalition.

Let us illustrate the notion of preemption with the three-player externality problem. Suppose player 3 wishes to commit to Pollute unilaterally. He knows that if he attempts to do so, players 1 and 2 will independently challenge him by trying to be the first to commit to Pollute. Suppose that each may win with probability 1/3. Then player 3 incurs the cost c for sure since someone will commit to Pollute. With probability 1/3 he wins and so his total payoff is $-c$ as the other two players will react, according to the core of the subgame, with Clean. With probability 2/3, someone else commits to Pollute first and so player 3 will react with Clean, getting a payoff $-c-b$. Thus, player 3's expected payoff from deviation is equal to $-\frac{2}{3}b - c$. Since $c > \frac{1}{3}b$, this is less than his payoff $-b$ from abiding by the "everyone cleans" status quo. Thus, player 3 would not deviate if he expects the preemptive measure "everyone else will challenge me independently."

This preemptive measure is a credible threat to player 3, as it is incentive feasible for players 1 and 2. Pick player 1 for example. Expecting player 2 to participate in this preemptive measure, player 1 expects an expected payoff $-\frac{2}{3}b - c$ from participation. The calculation of this payoff is exactly the same as that for the deviating player 3. If player 1 does not participate, then only player 2 challenges player 3, so player 1's expected payoff is $-c-b$ for sure because one of players 2 and 3 will commit to Pollute first and the other player and player 1 will react with Clean according to the core of the subgame. Thus, player 1 strictly prefers participation in the preemptive measure against the deviating player 3.

With the recursive principle and the notion of preemptive mechanisms, the answer to the question "What would a coalition expect when it deviates?" now follows. Any deviation from a status quo will be met by a preemptive response coordinated by an incentive feasible preemptive mechanism. Any coalition who manages to commit will be followed by a core-like reaction. With such anticipation, a coalition can calculate the net gain from deviating from a status quo. Then the notion of the core is at hand.

This notion of the core is based on an environment that allows all the externalities captured by strategic form games. Spelled out in §2, the primitives in my model are a payoff matrix as in any strategic form game and a competitive protocol for coalitional commitment. The notion of preemptive mechanisms and the new concept of the core are defined in §3. The definition is illustrated in §4 with a three-player Battle of Sexes game. In the externality problem discussed above, the preemptive mechanism is simply for each player outside a deviating coalition to challenge the deviator independently, so coordination across the outsiders is not needed. By contrast, in the game analyzed in §4, coordination is necessary, which explains the complexity in our notion of preemptive mechanisms. The main result, Theorem 1, is in §5. It asserts that the our core is nonempty (and hence Pareto optimal) in the n-player generalization of the externality problem discussed above. Theorem 2 asserts that the core is nonempty for any two-player finite strategic form game.

The related works will be reviewed in §6. An index is at the end of the paper.

2 The Primitives

2.1 Players, Actions, and Payoffs

The set of players is I. The set of feasible actions for $i \in I$ is A_i. 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 $\neg S := I \setminus S$. Hence a_S is a profile of actions of the players in S.

Every player $i \in I$ has a von Neumann Morgenstern utility function

$$u_i : A_I \to \mathbb{R}, \tag{1}$$

meaning that $u_i(a_I)$ is the payoff for player i if the profile of actions across all players is a_I. Externality is incorporated, as $u_i(a_I)$ may vary with the component $a_{\neg S}$ in a_I.

For example, in the pollution example illustrated in the Introduction, the action set for each player i is $A_i = \{\text{Clean}, \text{Pollute}\}$. With $|T|$ denoting the size of any set T,

$$u_i(a_I) = -b\mathbf{1}_{a_i = \text{Clean}} - c\left|\{j \in I : a_j = \text{Pollute}\}\right|.$$

The above model includes the traditional characteristic-function-based model and transferable utility games as special cases. See Appendix A.

2.2 The Competition for Coalitional Commitment

Each nonempty set S of players is endowed with a *power* $\pi(S)$ if S acts as a coalition, where

$$\pi : 2^I \setminus \varnothing \to \mathbb{R}_{++}$$

is a function that associates a positive number to each nonempty subset S of I.

The players choose their actions through a multistage commitment process.

1. Any set of uncommitted players may reach a mutual agreement on what actions they shall take. If a set $S \subseteq I$ of players has done so, its agreement is in the form of an action profile $a_S \in A_S$. Any player $i \in S$ is said to be a member of coalition S.

2. A player cannot be a member of two distinct coalitions.

3. If a coalition S has reached an agreement a'_S within itself, the players outside S who have not committed choose whether to *challenge* S or not.

   a. If any set J of these outsiders choose to challenge S as a coalition, the agreements that involve any member of J are suspended and J tries to commit, before S does, to an action profile $a'_J \in A_J$, which need not be the same as the previous agreements. A player outside S can join only one of such challenging coalitions.

   b. If the coalitions challenging S are $J_1, \ldots, J_m$, then for any $E \in \{S, J_1, \ldots, J_m\}$, coalition E wins with probability

   $$\frac{\pi(E)}{\pi(S) + \sum_{k=1}^{m} \pi(J_k)}.$$

   The winning coalition E commits to its intended action profile a'_E, and the losing coalitions cannot commit to any action until the winner has committed (so a losing coalition may change its plan afterwards).

4. If a coalition S has reached an agreement within itself, if it is not challenged by anyone, and if none of its members challenge other coalitions, then S has the option to either commit to its agreed upon action profile or cancel its plan.

5. The committed players and their committed actions become common knowledge.

6. If all players have committed, the process ends; otherwise another iteration begins.

In the above protocol, the procedure of a challenge may be interpreted as a race for commitment such that the winner gets to commit before the loser and the loser, seeing the winner's commitment, may adjust its previously agreed upon action profile. A second interpretation is that there is a court and the challenging coalition asks the court to nullify the challenged coalition's agreed upon action profile on the ground of its externality. In a system with defined property rights, a challenge could be a challenging coalition's attempt to invade or pillage the challenged coalition's properties. Sometimes a challenge may have vacuous effect. For instance, in a pure exchange private ownership economy where no one is allowed to intervene the trading actions among others, a coalition J may "challenge" a disjoint coalition S by being the first to commit to a trade agreement within J, but that cannot prevent S from committing to its own trading agreement in the next iteration.

3 Defining the Core

The focus is to extend unilateral deviations to coalitional deviations. As a strategy profile in a dynamic game tells a unilateral deviator what to expect of the other players, a coalitional response defined in §3.2 tells a deviating coalition what to expect of the complementary coalition. The response contains two parts, a preemptive measure that the complementary coalition takes before the deviating coalition can commit to its intended action and a reaction to the party that manages to commit.

To make a reaction credible, we simply apply the recursive principle mentioned in the Introduction. Making a preemption credible is more complicated. As a deviation may be coalitional, a preemption against the deviation may also be coalitional, hence the outsiders of a deviating coalition should be allowed to act as a group. But then such a group should also need to know how to adjust its action if some members of the group deviate from the group action. Hence a preemption needs to specify the contingency plans in case of internal deviations. That is why we need to introduce a notion of preemptive mechanism in §3.1. In §3.4 we ensure the credibility of a preemption by imposing an incentive feasibility condition on the associated preemptive mechanism.

3.1 Preemptive Mechanisms

For any $T \subsetneq I$ and $a_T \in A_T$, let $\mathscr{G}(T, a_T)$ denote the subgame where T is the set of the players who have committed and a_T the committed action profile. Hence $\neg T$ is the set of players whose actions are yet to be determined.

For any $\varnothing \neq F \subsetneq E \subseteq \neg T$, $a_F \in \Delta A_F$, and $a_E \in \Delta A_E$, if there does not exist $a_{E \setminus F} \in A_{E \setminus F}$ such that $a_E = \big(a_F, a_{E\setminus F}\big)$, then we say a_F is *not aligned with* a_E and denote it by $a_F \nmid a_E$.

For any $S \subsetneq \neg T$, a *preemptive mechanism* against S in $\mathscr{G}(T, a_T)$ consists of—

- a partition of $(\neg T) \setminus S$, with partition cells $E_1, \ldots, E_m$ for some integer m, and
- for each $k = 1, \ldots, m$,
  - an $a_{E_k} \in A_E \cup \{\text{nil}\}$, and
  - for any $F \subsetneq E_k$ and any a'_F not aligned with a_{E_k}, a $\beta_k(F, a'_F) \in A_{E_k \setminus F} \cup \{\text{nil}\}$.

The profile $(a_{E_k})_{k=1}^m$ is called the on-path *preemptive response* to S.

If a coalition S deviates from some status quo of $\mathscr{G}(T, a_T)$, a preemptive mechanism

$$\left(\left(E_k, a_{E_k}, (\beta_k(F, a'_F))_{\varnothing \neq F \subsetneq E_k,\, a'_F \nmid a_{E_k}}\right)_{k=1}^m\right) \tag{2}$$

against S is interpreted as follows. The complementary coalition $(\neg T) \setminus S$ of S is partitioned into *subcoalitions* $E_1, \ldots, E_m$. Each subcoalition E_k independently decides whether to challenge the deviating coalition S or not. If $a_{E_k} = \text{nil}$, then E_k does not challenge S, otherwise a_{E_k} is an element of the set A_{E_k} of action profiles for E_k and E_k will commit to a_{E_k} if E_k wins the commitment opportunity at this challenge. Within each subcoalition E_k, if the members in the subset F play a preemptive response a'_F not aligned with a_{E_k}, then the other members, who constitute the set $E_k \setminus F$, switches to the response $\beta_k(F, a'_F)$.

In the special case where subcoalition E_k is a singleton for each k, the preemptive mechanism is completely decentralized and every player outside the deviating coalition S makes a preemptive response independently. In the special case where there is only one subcoalition, which is the entire complementary coalition $(\neg T)\setminus S$, the preemptive mechanism is completely coordinated and every subset $E_k \setminus F$ of $(\neg T) \setminus S$ has a contingent plan for its preemptive response in the event that all those in F do not abide by the preemptive response for the entire complementary coalition.

Given a profile $(a_{E_k})_{k=1}^m \in \prod_{k=1}^m (A_{E_k} \cup \{\text{nil}\})$ of preemptive responses, the set of subcoalitons challenging the deviating coalition S is

$$N\left((a_{E_k})_{k=1}^m\right) := \{E_k : a_{E_k} \neq \text{nil}; k = 1, \ldots, m\}, \tag{3}$$

so the probability with which S wins the challenge is

$$\frac{\pi(S)}{\pi(S) + \sum_{J \in N\left(a_{E_k})_{k=1}^m\right)} \pi(J)}$$

and likewise for a challenging subcoalition E_k.

Note that a preemptive mechanism does not require the existence of a principal or mediator to carry out any action, reward, or penalty. As in the repeated games theory, everything is carried out by players themselves, and the entire mechanism may be just an implicit expectation commonly held by the players.

3.2 Coalitional Responses

Suppose that $T \subsetneq I$ is the set of players who have committed and $a_T \in A_T$ the committed action profile. A *coalitional response* in the subgame $\mathscr{G}(T, a_T)$ is a pair $(\mathscr{P}, \mathscr{R})$ of two functions such that $\mathscr{P}$ associates to each (S, a_S) ($S \subsetneq \neg T$ and $a_S \in A_S$) a preemptive mechanism against S and $\mathscr{R}$ associates to each (S, a_S) ($S \subsetneq \neg T$ and $a_S \in A_S$) an element in $\Delta A_{(\neg T)\setminus S}$, with $\Delta A_{(\neg T)\setminus S}$ denoting the set of lotteries on $A_{(\neg T)\setminus S}$.

As explained in §3.1, $\mathscr{P}(S, a_S)$ is a preemptive mechanism against S when S deviates by trying to commit to an action profile a_S other than the status quo. If

$$\mathscr{P}(S, a_S) = \left(\left(E_k, a_{E_k}, (\beta_k(F, a'_F))_{\varnothing \neq F \subsetneq E_k,\, a'_F \nmid a_{E_k}}\right)_{k=1}^{m}\right)$$

then we write

$$\mathscr{P}_o(S, a_S) := (a_{E_k})_{k=1}^{m}\,, \tag{4}$$

which is the on-path preemptive response generated by the preemptive mechanism.

The part $\mathscr{R}(S, a_S)$ of a coaltional response is called *reaction* to the fact that the coalition S has committed to an action profile a_S. The coalition S need not be the one that deviated from the status quo. It could be a subcoalition that challenges a deviating coalition and wins at the challenge. A reaction to (S, a_S) may be implemented in one shot or in multiple stages where different coalitions in $(\neg T) \setminus S$ make commitments at different stages. The reaction may also be the outcome of a race for commitment between multiple coalitions. In that case, as the party that commits first may be randomly chosen, the reaction may be a lottery of action profiles.

3.3 Expected Payoffs given a Coalitional Response

Let $T \subsetneq I$ and $a_T \in A_T$. Pick any coalitional response $(\mathscr{P}, \mathscr{R})$ in the subgame $\mathscr{G}(T, a_T)$. Suppose a coalition $S \subsetneq \neg T$ deviates by trying to commit to some $a_S \in A_S$. If $(a_{E_k})_{k=1}^m$ is the profile of the preemptive responses against S, then the set of subcoalitions that choose to challenge S is the $N((a_{E_k})_{k=1}^m)$ in (3), and the expected payoff for any player $i \in I$ is

$$u_i\left(a_S, (a_{E_k})_{k=1}^m \mid \mathscr{R}, a_T\right) = \sum_{K \in S \cup N((a_{E_k})_{k=1}^m)} \frac{\pi(K) u_i\left(a_T, a_K, \mathscr{R}(K, a_K)\right)}{\sum_{J \in S \cup N((a_{E_k})_{k=1}^m)} \pi(J)}, \tag{5}$$

where

$$u_i\left(a_T, a_K, \mathscr{R}(K, a_K)\right) = \sum_{a_{(\neg T)\setminus K} \in A_{(\neg T)\setminus K}} \left(\mathscr{R}(K, a_K)(a_{(\neg T)\setminus K})\right) u_i(a_T, a_K, a_{(\neg T)\setminus K}).$$

Eq. (5) takes into account the reactive response $\mathscr{R}$ to any possible outcome of the challenge.

3.4 Incentive Feasible Preemptive Mechanisms

Let $T \subsetneq I$ and $a_T \in A_T$. Pick any coalitional response $(\mathscr{P}, \mathscr{R})$ in the subgame $\mathscr{G}(T, a_T)$. Suppose a coalition $S \subsetneq \neg T$ deviates by trying to commit to some $a'_S \in A_S$. A preemptive mechanism denoted by (2) is *incentive feasible* with respect to the reaction function $\mathscr{R}$ iff, for any $k = 1, \ldots, m$ (m being the number of subcoalitions of $(\neg T) \setminus S$ according to the mechanism),

a. the preemptive response a_{E_k} maximizes the total expected payoff

$$\sum_{i \in E_k} u_i \left(a'_S, \left(a'_{E_k}, \left(a_{E_j} \right)_{j \neq k} \right) \mid \mathscr{R}, a_T \right)$$

for subcoalition E_k among all a'_{E_k} in $A_k \cup \{\text{nil}\}$, and

b. for any nonempty $F \subsetneq E_k$, with $a_{E_k} = (a_F, a_{E_k \setminus F})$,

   i. a_F maximizes the total expected payoff

$$\sum_{i \in F} u_i \left(a'_S, \left(\left(a'_F, \beta_k(F, a'_F) \right), \left(a_{E_j} \right)_{j \neq k} \right) \mid \mathscr{R}, a_T \right)$$

   for the set F of players among all $a'_F \in A_F \cup \{\text{nil}\}$, and

   ii. for any $a'_F \neq a_{E_k}$, $\beta_k(F, a'_F)$ maximizes the total expected payoff

$$\sum_{i \in E_k \setminus F} u_i \left(a'_S, \left(\left(a'_F, a'_{E_k \setminus F} \right), \left(a_{E_j} \right)_{j \neq k} \right) \mid \mathscr{R}, a_T \right)$$

   for the set $E_k \setminus F$ of players among all $a'_{E_k \setminus F}$ in $A_{E_k \setminus F} \cup \{\text{nil}\}$.

I.e., every nonempty subset F of every subcoalition E_k best replies to the preemptive actions that the mechanism prescribes for the other subcoalitions, as well as to the contingency plan that the other members of the subcoalition E_k would adopt if F disobeys the mechanism. When the subset F contains multiple players, the word "best" here means a case of Pareto optimality within F, maximization of the sum of these players' expected payoffs.

3.5 The Definition of Blocking

Let $T \subsetneq I$ and $a_T \in A_T$. In the subgame $\mathscr{G}(T, a_T)$, a lottery $\alpha_{\neg T} \in \Delta A_{\neg T}$ of action profiles for $\neg T$ is *blocked* by a coalition $S \subseteq \neg T$ with respect to a coalitional response $(\mathscr{P}, \mathscr{R})$ iff there exists an action profile $a'_S \in A_S$ such that

$$u_i\left(a'_S, \mathscr{P}_o(S, a'_S) \mid \mathscr{R}, a_T\right) > \sum_{a_{\neg T} \in A_{\neg T}} \alpha_{\neg T}(a_{\neg T}) u_i\left(a_T, a_{\neg T}\right), \tag{6}$$

where the left hand-side is defined by (5) and the $\mathscr{P}_o(S, \alpha'_S)$ defined by (4). I.e., every member i of coalition S gets a higher expected payoff from the deviation of trying to commit to a'_S rather than abiding by the mixed action $\alpha_{\neg T}$, expecting the coalitional response $(\mathscr{P}, \mathscr{R})$.

When $\neg T = \{i\}$ for some player i, the above notion that a mixed action α_i for player i is blocked becomes equivalent to the notion that α_i is not a best response to $a_{\neg\{i\}}$ for player i.

3.6 The Definition of the Core

Suppose $T \subsetneq I$ is the set of players who have committed and $a_T \in A_T$ the committed action profile. The core $\mathscr{K}(T, a_T)$ for the set $\neg T$ of the players who have not committed 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 best responses for player i to the action profile a_T of T.

2. If $|\neg T| > 1$, an $\alpha_{\neg T} \in \Delta A_{\neg T}$ for $\neg T$ belongs to the core $\mathscr{K}(T, a_T)$ iff there exists a coalitional response $(\mathscr{P}, \mathscr{R})$ with the following properties:

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

   b. if any $S \subsetneq \neg T$ deviates by trying to commit to some $a'_S \in A_S$, then the preemptive mechanism $\mathscr{P}(S, a'_S)$ is incentive feasible with respect to $\mathscr{R}$,

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

Condition 2a. needs no justification. Condition 2c. is the recursive principle. It has been applied in some other models by Huang and Sjöström [13] and Piccione and Razin [18]. It is the natural extension of subgame perfection to our coalitional setting.

Condition 2b. is new. It is based on the idea that, as a blocking attempt α'_S is a coordinated action of the members of the blocking coalition S, the preemptive response to the blocking attempt should also allow for coordination among the players in the complementary coalition $(\neg T) \setminus S$. That leads to the notion of preemptive mechanism, which allows for coordination within $(\neg T) \setminus S$. If a preemptive mechanism recommends coordinated preemptive measures, for it to be credible, the mechanism must also specify a contingency plan if some players in $(\neg T) \setminus S$ do not participate in the recommended preemptive measures, and such contingency plan should be Pareto optimal for those who abide by the recommendation. That leads to the incentive feasibility condition for the preemptive mechanism.

Why do we impose an incentive feasibility condition on the preemptive response but not on blocking? The reason is analogous to a standard treatment in noncooperative game theory where we impose a best response condition on any equilibrium strategy profile but not on a deviation from the equilibrium. In our context, the preemptive response is part of an "equilibrium," while blocking is merely a deviation from the "equilibrium."

Why not extend the recursive principle to a preemption by requiring it to be in the core among the outsiders of the deviating coalition? That is because decisions on preemptive measures against a commitment attempt are in a different time frame than decisions on how to react to some action that has been committed to. If we insist on predicting the preemptive measures with a core-like notion, there seems to be only two alternatives. We may make up a different notion of the core to deal with the pre-commitment phase, thereby losing the symmetrically aesthetic appeal of the recursive principle. Alternatively we may embed a substructure in any pre-commitment phase such that within the substructure these outsiders play a preemption-reaction game while they as a whole are still looking for a preemptive response to the original deviating coalition. That may lead to intractable infinite regress.

4 A Three-Player Battle of Sexes Game

Let us illustrate the new notion of the core with the O'Neil Game (cited from Myerson [17]).

A_2 and A_3

	x_3		y_3	
A_1	x_2	y_2	x_2	y_2
x_1	$0, 0, 0$	$6, 5, 4$	$4, 6, 5$	$0, 0, 0$
y_1	$5, 4, 6$	$0, 0, 0$	$0, 0, 0$	$0, 0, 0$

Assume that in the power of every coalition $S \subseteq I$ is equal to the size of the coalition:

$$\pi(S) = |S|. \tag{7}$$

Why Not Use Coalition-Proof Equilibrium Concepts If we follow the coalition-proof equilibrium approach, we would not get a self-stable prediction of this game. Specifically, strong Nash equilibrium of this game does not exist. For any Nash equilibrium, say (x_1, y_2, x_3), which gives the payoff vector $(6, 5, 4)$, there are two players who strictly prefer to deviate from the equilibrium, assuming the other player abiding by the equilibrium. In the case of (x_1, y_2, x_3), players 2 and 3 would deviate to the action profile (x_2, y_3).[3]

Why the Core Should Allow for Lotteries of Actions Let us apply our notion of the core. Our first observation is that a core element must involve mixed strategies. If a core element is in pure strategies, then by the Pareto optimality of the core, the core element can only be (y_1, x_2, x_3) or (x_1, y_2, x_3) or (x_1, x_2, y_3). Pick without loss of generality (y_1, x_2, x_3), which gives player 2 a payoff 4. But then player 2 alone can block (y_1, x_2, x_3) by trying to commit to y_2 before the other players commit. If he manages to commit to y_2 first, the reaction from the other players is to play (x_1, x_3) according to the core condition 2c. and so player 2's payoff is 5. If he is defeated in the race for commitment, so it is given that another player say player 1 is playing y_1, then again condition 2c. implies that the reaction is (x_2, x_3) so that player 2's payoff cannot be lower than 4. Thus, player 2's expected payoff from the deviation is greater than 4.

The above observation shows that the core needs to allow for mixed strategies. That is why in our concept of the core an element of the core is a lottery of action profiles rather than a pure action profile.

[3] In fact, the outcome resulting from the coalitional deviation, (x_1, x_2, y_3), is itself a Nash equilibrium, as Myerson [17] points out.

A Solution in the Core In the rest of this section, we shall demonstrate that the following solution is in the core:

$$\frac{2}{3}(x_1, y_2, x_3) + \frac{1}{3}(x_1, x_2, y_3), \tag{8}$$

yielding a profile of expected payoffs across players 1, 2, and 3:

$$\frac{2}{3}(6,5,4) + \frac{1}{3}(4,6,5) = \left(\frac{16}{3}, \frac{16}{3}, \frac{13}{3}\right).$$

To construct the supporting coalitional response, denote $+$ for the addition operation modulo 3, so $3+1=1$ and $3+2=2$. Note that any two-player coalition can be denoted as $\{i, i+1\}$ for some $i \in \{1,2,3\}$.

The Reaction Function The reactive function $\mathscr{R}$ is defined trivially. If only player i has committed and his action is x_i, then the other two players commit to the action profile (x_{i+1}, y_{i+2}), i.e.,

$$\mathscr{R}(\{i\}, x_i) = (x_{i+1}, y_{i+2}), \tag{9}$$

which we display for future reference. If only player i has committed and his action is y_i, then the other two play (x_{i+1}, x_{i+2}) (recall from our notation of $+$ that player $i+2$ is player $i-1$). If only players i and $i+1$ have committed, then player $i+3$ plays whatever action such that the resulting payoff vector is not $(0,0,0)$. Clearly each of these reactions belongs to the core of the subgame after the early commitment has been made. Hence $\mathscr{R}$ satisfies the core condition 2c. in our definition of the core.

The Preemption Function against 2-Player Deviating Coalitions In this case, the incentive feasibility condition of a preemptive mechanism, core condition 2b., becomes the best response condition. Hence the preemptive mechanism is easily constructed as a best response to the deviation, given the $\mathscr{R}$ constructed above. With our notation of $+$ modulo 3, any two-player deviating coalition can be denoted by $\{i, i+1\}$. If $\{i, i+1\}$ wants to commit to (x_i, y_{i+1}), which would give player $i+2$ a payoff 4, the lowest among all Pareto optima, if $\{i, i+1\}$ gets to commit first. Thus, player $i+2$'s best response is to challenge $\{i, i+1\}$ and commits to y_{i+2} upon winning. (Trying to commit to x_{i+2} upon winning is suboptimal for player $i+2$ because according to (9) if he manages to commit to x_{i+2} the other two players will react with (x_i, y_{i+1}), giving him only a payoff 4.) If $\{i, i+1\}$ wants to commit to (y_i, x_{i+1}), however, player i does not challenge it and plays x_{i+2} after the coalition has committed, so player $i+2$ gets a payoff 6, the highest among all Pareto optima. If $\{i, i+1\}$ wants to commit to (x_i, x_{i+1}), then player $i+2$ does not challenge them and after they have committed will react with y_{i+2}; that gives him a payoff 5. If player $i+2$ challenges them with action y_{i+2}, his payoff is the same whether he wins or loses. Challenging the deviating coalition with x_{i+2} is suboptimal, as pointed out earlier in this paragraph. In sum,

$$\begin{aligned}
\mathscr{P}(\{i, i+1\}, (x_i, y_{i+1})) &= (\{i+2\}, y_{i+2}) && (10)\\
\mathscr{P}(\{i, i+1\}, (y_i, x_{i+1})) &= (\{i+2\}, a_{i+2} = \text{nil}) && (11)\\
\mathscr{P}(\{i, i+1\}, (x_i, x_{i+1})) &= (\{i+2\}, a_{i+2} = \text{nil}) && (12)\\
\mathscr{P}(\{i, i+1\}, (y_i, y_{i+1})) &= (\{i+2\}, y_{i+2})\,. && (13)
\end{aligned}$$

Why Coordination Is Necessary for Preemptive Measures When a deviating coalition is a singleton, however, construction of the preemptive mechanism is less straightforward. Suppose the outsiders of a deviating coalition can only make their preemptive responses independently, then the solution (8) would be blocked by player 3 alone with the plan of trying to commit to y_3 unilaterally. That is because in the normal form game where players 1 and 2 pick their preemptive actions independently, from player 2's viewpoint, both x_2 and y_2 are strictly dominated by nil (i.e., not challenging player 3). In the next matrix we list the calculations of the expected payoffs for player 2 in this two-player normal form game:

$$\begin{array}{cccc}
 & x_2 & y_2 & \text{nil} \\
x_1 & \frac{1}{3}(6+4+6) & \frac{1}{3}(6+5+6) & \frac{1}{2}(6+6) \\
y_1 & \frac{1}{3}(4+4+6) & \frac{1}{3}(4+5+6) & \frac{1}{2}(4+6) \\
\text{nil} & \frac{1}{2}(4+6) & \frac{1}{2}(5+6) & 6
\end{array}$$

Consider the (x_1, x_2)-case for example. In that case, players 1 and 2 challenge player 3 by trying to be the first to commit to their x-action. By (7), the three players each win with probability 1/3. If player 1 wins, he commits to x_1, then the reaction from players 2 and 3 is to play (x_2, y_3) according to (9), so player 2 gets a payoff 6. If player 2 wins, player 2 commits to x_2; again (9) implies that the reaction is (x_3, y_1) so player 2 gets a payoff 4. If player 3 wins and so commits to y_3, the reaction can only be (x_1, x_2) and so player 2 gets a payoff 6. The other cases are calculated in the similar fashion. Now that player 2 does not challenge player 3 when player 3 tries to commit to y_3 unilaterally, player 3's expected payoff is no less than $\frac{1}{2}(4+6)$, which is greater than the 13/3 that he would get at the solution (8).

Thus, to support (8) as a core solution, the preemptive measures needs to be coordinated across the players outside a deviating coalition. That is why we formulate the notion of preemptive mechanism, which allows for coordination among the non-deviating players.

The Preemption Function against Singleton Deviating Coalitions If players 1 and 2 can coordinate their preemptive measures, when players 1 and 2 form a challenging coalition and try to commit to (x_1, y_2) before player 3 commits. Hence with probability 2/3 players 1 and 2 win and the outcome of the game is (x_1, y_2, x_3), giving player 3 a payoff 4.[4] Thus, player 3's expected payoff from the deviation is equal to $\frac{2}{3}4 + \frac{1}{3}5 = \frac{13}{3}$, no greater than what he gets from the status quo.

Incentive Feasibility for the Challenging Coalition This preemptive measure maximizes the total expected payoff for players 1 and 2. That is because given that the payoff vector will be $(4, 6, 5)$ if player 3 gets to commit first (recall (9)), challenging player 3 with action (x_1, y_2) maximizes the probability for the payoff vector to be $(6, 5, 4)$, which gives the highest total payoff to players 1 and 2 among all possible payoff vectors. Thus, the preemptive measure is incentive feasible for players 1 and 2 as a single coalition, satisfying condition (a) of incentive feasibility defined in §3.4.

Incentive Feasibility for Individuals in the Challenging Coalition We still need to ensure that the preemptive measure is also incentive feasible for each of players 1 and 2 individually (condition (b) defined in §3.4). In the event that player 2 does not abide by the coordinated preemptive measure (x_1, y_2), calculate the expected payoff vector (u_1, u_2)

[4] Note that the coordinated preemptive measure (x_1, y_2) taken by players 1 and 2 as a single coalition is different from player 1 and player 2 independently try to commit to x_1 and y_2 as separate coalitions.

for players 1 and 2 for each possible preemptive action:

$$\begin{array}{rrl}
a_2 = \text{nil}: & a_1 = x_1: & \frac{1}{2}(4,6) + \frac{1}{2}(4,6) = (4,6) \\
& a_1 = y_1: & \frac{1}{2}(5,4) + \frac{1}{2}(4,6) = (4.5,5) \\
& a_1 = \text{nil}: & (4,6) \\
a_2 = x_2: & a_1 = x_1: & \frac{1}{3}(4,6) + \frac{1}{3}(5,4) + \frac{1}{3}(4,6) = (13/3, 16/3) \\
& a_1 = y_1: & \frac{1}{3}(5,4) + \frac{1}{3}(5,4) + \frac{1}{3}(4,6) = (14/3, 14/3) \\
& a_1 = \text{nil}: & \frac{1}{2}(5,4) + \frac{1}{2}(4,6) = (4.5,5)
\end{array}$$

Thus, for both cases the best reply for player 1 is to try to commit to y_1. That gives player 2 an expected payoff 5 (if $a_2 = \text{nil}$) or 14/3 (if $a_2 = x_2$), both lower than the payoff 16/3 had player 2 abides by the coordinated preemptive measure (x_1, y_2).

Analogously we calculate the expected payoff vectors for players 1 and 2 given player 1's disobedience to the preemptive measure:

$$\begin{array}{rrl}
a_1 = \text{nil}: & a_2 = x_2: & \frac{1}{2}(5,4) + \frac{1}{2}(4,6) = (4.5,5) \\
& a_2 = y_2: & \frac{1}{2}(6,5) + \frac{1}{2}(4,6) = (5,5.5) \\
& a_2 = \text{nil}: & (4,6) \\
a_1 = y_1: & a_2 = x_2: & \frac{1}{3}(5,4) + \frac{1}{3}(5,4) + \frac{1}{3}(4,6) = (14/3, 14/3) \\
& a_2 = y_2: & \frac{1}{3}(5,4) + \frac{1}{3}6,5) + \frac{1}{3}(4,6) = (5,5) \\
& a_2 = \text{nil}: & \frac{1}{2}(5,4) + \frac{1}{2}(4,6) = (4.5,5)
\end{array}$$

Thus, for both cases the best reply for player 2 is to not challenge player 3. That gives player 1 an expected payoff 4 or 4.5, both lower than the payoff 16/3 had player 1 abides by the preemptive measure. We have hence constructed an incentive feasible preemptive mechanism, in the notations of (2),

$$\mathscr{P}(\{3\}, y_3) = (\{1,2\}, (a_1, a_2) = (x_1, y_2), \beta_1(a_2 \neq y_2) = y_1, \beta_1(a_1 \neq x_1) = \text{nil})\,. \tag{14}$$

If the deviating coalition is a singleton $\{i\}$ such that $i \neq 3$ and if the action i tries to commit to is y_i, then the preemptive mechanism is the same as above with players $i+1$

and $i+2$ playing the role of players 1 and 2 in the above paragraph:

$$\mathscr{P}(\{i\}, y_i) = \left(\begin{array}{c} \{i+1, i+2\}, (a_{i+1}, a_{i+2}) = (x_{i+1}, y_{i+2}), \\ \beta_{i+1}(a_{i+2} = \text{nil}) = y_{i+1}, \beta_{i+2}(a_{i+1} \neq x_{i+1}) = \text{nil} \end{array} \right). \tag{15}$$

Giving player i an expected payoff $\frac{1}{3}5 + \frac{2}{3}4 = \frac{13}{3}$ instead of the 16/3 at the status quo, this preemptive mechanism makes such deviation unprofitable for player i. The preemptive mechanism is incentive feasible for players $i+1$ and $i+2$ by exactly the same reason why (14) is incentive feasible for players 1 and 2. (Although players 1 and $i+1$ may have different expected payoffs at the status quo (8), such payoffs are irrelevant to the calculations of their preemptive responses conditional on someone's deviation.)

Incentive Feasibility of Preemptive Measures That Need No Coordination If the deviating coalition is a singleton $\{i\}$ trying to commit to x_i, the other two players do not challenge him, anticipating the post-commitment reaction (x_{i+1}, y_{i+2}) by Eq. (9).

$$\mathscr{P}(\{i\}, x_i) = (\{i+1\}, \{i+2\}, a_{i+1} = \text{nil}, a_{i+2} = \text{nil}). \tag{16}$$

Since this preemptive mechanism partitions players $i+1$ and $i+2$ into two singleton subcoalitions, the mechanism does not need to specify the contingency plan when a proper subset of a subcoalition does not abide by the mechanism. With the two players taking preemptive measures independently, incentive feasibility of the mechanism becomes the Nash condition that their preemptive measures best reply each other. Expecting the post-commitment reaction (x_{i+1}, y_{i+2}), which will give player $i+1$ the highest possible payoff 6, player $i+1$ best replies by not challenging player i. With the same expectation, player $i+2$'s payoff from not challenging player i will be 5, while challenging player i with y_{i+2} will give player $i+2$ the same payoff 5 whether he wins or not. Challenging player i with x_{i+2} is suboptimal because if he wins the reaction will be the penalizing (x_i, y_{i+1}) by Eq. (9).

Verifying the Unblockability of the Core Solution (8) That amounts to verifying the core condition 2a. with respect to the coalitional response constructed above. First, the coalition $\{1, 2\}$ cannot profit from deviation. To gain the most from deviation, the coalition can only try to commit to (x_1, y_2). That gives the coalition the same expected payoff vector as in (8), due to player 3's preemptive response y_3 by Eq. (10) and the winning probabilities by (7).

Second, $\{2,3\}$ cannot block (8). To make player 3 better-off than in (8), the coalition needs to commit to (x_2, y_3) or (x_2, x_3) in the event that it wins. But committing to (x_2, x_3) would hurt player 2, as Eq. (12) says that player 1 will take nil preemptive measure and react with y_1 after the commitment. With players 2 and 3 trying to commit to (x_2, y_3), player 1 challenges the coalition with action y_1, according to Eq. (10). Hence player 2's expected payoff from the deviation is equal to

$$\frac{2}{3}6 + \frac{1}{3}4 = \frac{16}{3},$$

which is not higher than what he gets from (8). Similar reasoning holds for coalition $\{1,3\}$.

Third, a singleton coalition $\{i\}$ cannot block (8). If player i tries to commit to x_i, then by (16) the other two players do not challenge it and will react with (x_{i+1}, y_{i+2}) by (9) to give player i the lowest positive payoff, 4. If player i tries to commit to y_i, then we may assume without loss that $i = 3$, because the other players j can only get worse-off than (8) by playing y_j. Expecting the preemptive response (x_1, y_2) by (14), i's expected payoff is

$$\frac{1}{3}5 + \frac{2}{3}4 = \frac{13}{3},$$

which is not higher than his expected payoff in (8).

Thus, (8) is in the core.

A Contrast with Correlated Equilibrium Clearly, the core solution (8) is a correlated equilibrium of the original game. However, not all correlated equilibria of the original game belong to the core. For example, $\frac{1}{3}(y_1, x_2, x_3) + \frac{1}{3}(x_1, y_2, x_3) + \frac{1}{3}(x_1, x_2, y_3)$ is blocked. If $\{1,2\}$ deviates by trying to commit to (x_1, y_2), the uniquely best preemptive response for player 3 is to challenge $\{1,2\}$ with action y_3 to avoid the lowest payoff 4. Then the expected payoff vector for players 1 and 2 is

$$\frac{2}{3}(6,5) + \frac{1}{3}(4,6) = \left(\frac{16}{3}, \frac{16}{3}\right) > (5,5).$$

Hence they can block the correlated equilibrium $\frac{1}{3}(y_1, x_2, x_3) + \frac{1}{3}(x_1, y_2, x_3) + \frac{1}{3}(x_1, x_2, y_3)$.

5 A Coase Theorem in a Pollution Problem

Let us apply the new concept of the core to the n-player generalization of the pollution problem discussed in the Introduction. There are $n \geq 2$ players. Each player's action set is

$\{\text{Pollute}, \text{Clean}\}$. If there are exactly k players who play Pollute, then a player's payoff is equal to $-b - kc$ if he plays Clean and otherwise is $-kc$. Assume that

$$c < b. \tag{17}$$

Thus, Pollute strictly dominates Clean if we consider the situation as a strategic form game.

Despite the fact that pollution is the dominant action for each player, Theorem 1 asserts that the core is nonempty. The main reason why the positive result holds is that a coalition is torn between two countervailing forces. On one hand, to increase the chance of being the first to commit to Pollute, a coalition may want to include more members. On the other hand, having more members might make it suboptimal for the coalition to play Pollute. To see that, simply note that the total payoff for a coalition S to all play Pollute is equal to $-c|S|^2$ and the total payoff for S to all play Clean is equal to $-b|S|$, ceteris paribus. Thus, if a coalition is left alone, the Pareto optimum for the coalition is for all its members to play Pollute if its size is smaller than b/c and for all its members to play Clean if its size is larger than b/c. Therefore, there is an upper bound for the size of a coalition that wishes to pollute. With winning probabilities increasing in coalition sizes, such a coalition's expected payoff from playing Pollute may be outweighed by the payoff from not playing Pollute.

To establish the theorem, we assume the following for any coalition S and any collection $(E_k)_{k=1}^m$ of disjoint subcoalitions of $\neg S$ that challenge S:

$$|S| < \frac{b}{c} \leq \sum_{k=1}^{m} |E_k| \;\Rightarrow\; \frac{\pi(S)}{\pi(S) + \sum_{k=1}^{m} \pi(E_i)} \leq \frac{c}{b}. \tag{18}$$

Condition (18) is a way to formalize the idea in the previous paragraph that a small coalition does not enjoy a tremendously large probability of winning.

Theorem 1 *In the pollution problem with n players, if (18) holds then the core is nonempty and contains the set of Pareto optima that treat the players symmetrically.*

Proof For notation simplicity, consider only the case where no one has committed yet, i.e., $T = \varnothing$ so $I = \neg T$. In the general case where $T \neq \varnothing$, the action profile say a_T of T are already given, so the externality $-c|\{i \in T : a_i = \text{Pollute}\}|$ of a_T is a constant component in the payoff for every player in $\neg T$. Thus, in that general case we need only to replace I with $\neg T$ and calculate a player i's payoff in terms of i's payoff subtracted by the constant $-c|\{i \in T : a_i = \text{Pollute}\}|$.

Throughout the proof, S denotes a deviating coalition, and λ the fraction of polluters in S if it gets to commit to its deviating action profile. I.e., if a_S is the action profile S tries to commit to,

$$\lambda = \frac{|\{i \in S : a_i = \text{Pollute}\}|}{|S|}. \tag{19}$$

Lemma 1 *If $|I| < b/c$, everyone's playing Pollute, denoted $(a_i = \text{Pollute})_{\forall i}$, is Pareto optimal. If $|I| > b/c$, everyone's playing Clean, denoted $(a_i = \text{Clean})_{\forall i}$, is Pareto optimal. If $|I| = b/c$, the two constitute the set of Pareto optima that treat the players symmetrically, and any other Pareto optimum is blocked by an individual player who is supposed to play Clean (while someone else plays Pollute) at that optimum.*

Proof: Note that the total payoff for I is $-c|I|^2$ if everyone plays Pollute and $-b|I|$ if everyone plays Clean. Hence the case where $|I| < b/c$ or $|I| > b/c$ follows trivially. For any asymmetric Pareto optimum, where some players play Clean and the others play Pollute such that $0 < \lambda < 1$ for the λ defined in (19), with $S = I$, any individual who belongs to the first category would deviate by trying to commit to Pollute. In doing so, with the worst penalty for him being everyone else playing Pollute, the deviating player's payoff is at least $-c|I|$, which is higher than his expected payoff at the asymmetric Pareto optimum, $-b - \lambda|I|c = -|I|c - \lambda|I|c = -(1+\lambda)|I|c$. □

Lemma 2 *If $|I| \leq b/c$, $(a_i = \text{Pollute})_{\forall i}$, belongs to the core.*

Proof: With $|I| \leq b/c$, Lemma 1 implies that $(a_i = \text{Pollute})_{\forall i}$ is Pareto optimal. Let S be a deviating coalition. We may assume without loss that $S \subsetneq I$, so $|S| < |I| \leq b/c$. Suppose that S deviates with a polluters fraction $\lambda \in [0, 1]$ defined in (19). As the best case scenario for S is that none of the players in $\neg S$ play Pollute, the highest possible total payoff for S is equal to $-c\lambda|S||S| - b(1-\lambda)|S|$. As $|S| < b/c$, this total payoff is less than or equal to $-c|S|^2$, which is what S would have got without blocking. Thus, no coalition in I can block $(a_i = \text{Pollute})_{\forall i}$, as claimed. □

In the rest of the proof, we suppose $|I| \geq b/c$. Then "everyone playing Clean," denoted $(a_i = \text{Clean})_{\forall i}$, is Pareto optimal. We shall prove that it also belongs to the core with the following coalitional response:

$$\begin{aligned} [|\neg S| < b/c \text{ and } |I| - 1 < b/c] &\implies \mathscr{P}(S, \alpha_S) := ((\{i\}, a_i = \text{nil})_{i \in \neg S}) && (20) \\ |\neg S| < b/c &\implies \mathscr{R}(S, a_S) := (a_i = \text{Pollute})_{\forall i \in \neg S} && (21) \\ [|\neg S| \geq b/c \text{ or } |I| - 1 \geq b/c] &\implies \mathscr{P}(S, \alpha_S) := ((\{i\}, a_i = \text{Pollute})_{i \in \neg S}) && (22) \\ |\neg S| \geq b/c &\implies \mathscr{R}(S, a_S) := (a_i = \text{Clean})_{\forall i \in \neg S}. && (23) \end{aligned}$$

Note that the preemptive mechanism partitions $\neg S$ into singletons so that every player in $\neg S$ takes the preemptive measure independently. That has two implications. First, the mechanism does not need to specify a contingency action in the event that a proper subset of a partition cell does not abide by the preemptive mechanism. Second, the incentive feasibility condition for the preemptive mechanism (§3.4) is now equivalent to the condition that abiding by the mechanism constitutes a Nash equilibrium among the players in $\neg S$.

We prove that $(a_i = \text{Clean})_{\forall i}$ belongs to the core by induction on $|I|$.

Lemma 3 *If* $|I| = \min\{k = 1, 2, \ldots : k \geq b/c\}$, $(a_i = \text{Clean})_{\forall i}$ *belongs to the core.*

Proof: Pick any deviating coalition S. As $(a_i = \text{Clean})_{\forall i}$ is Pareto optimal (Lemma 1), we may assume without loss that $S \subsetneq I$, so $|S| \leq |I| - 1 < b/c$. Suppose that S deviates with a polluters fraction $\lambda \in [0, 1]$ defined in (19). Then by the coalitonal response (20)–(21) the coalition gets to commit and afterwards the other players all play Pollute, so the per-capita payoff for S generated by this deviation is equal to

$$-c\lambda|S| - b(1 - \lambda) - c|\neg S| < -c\lambda|S| - c|S|(1 - \lambda) - c|\neg S| = -c|I| \leq -b, \qquad (24)$$

where the first inequality uses the fact that $|S| < b/c$ and the last uses the fact $|I| \geq b/c$. As $-b$ is the per-capita payoff at $(a_i = \text{Clean})_{\forall i}$, coalition S cannot block it. Hence the core condition 2a. is satisfied. To verify core condition 2b., as pointed out previously, it suffices to show that the preemptive measure (20) constitutes a Nash equilibrium among players in $\neg S$ expecting (21). Hence pick any $i \in \neg S$. If i chooses inaction as recommended by the preemptive mechanism, his expected payoff is

$$-c\lambda|S| - c|\neg S| \geq -c|I|.$$

If i challenges S by trying to commit to some action first, then his payoff is different from the above only if he wins, and whatever action he commits to, the number of the other players is

$|I| - 1 < b/c$ and so they all react with Pollute by (21), so i's payoff is either $-c\,(|I|-1) - c$ or $-c\,(|I|-1) - b$. As $b > c$, neither payoff exceeds the one from abiding by the preemptive mechanism. Thus, condition 2b. is satisfied. The core condition 2c. follows directly from the fact that $(a_i = \text{Pollute})_{\forall i}$ belongs to the core when the number of uncommitted players is less than b/c, which has been proved previously. □

To complete the proof of the theorem, we consider the case $|I| > \min\{k = 1, 2, \ldots : k \geq b/c\}$.

The core condition 2c., that the reaction recommended by (21) and (23) belongs to the core in the subgame, follows from the induction hypothesis (for the case $|\neg S| \geq b/c$) and Lemma 2 (for the case $|\neg S| < b/c$).

Now we verify condition 2a., that $(a_i = \text{Clean})_{\forall i}$ cannot be blocked with respect to the coalitional response. Pick any coalition S. As $(a_i = \text{Clean})_{\forall i}$ is Pareto optimal, we may assume without loss that $S \neq I$.

Lemma 4 *If $|I| > \min\{k = 1, 2, \ldots : k \geq b/c\}$ and $|\neg S| < b/c$, $(a_i = \text{Clean})_{\forall i}$ cannot be blocked by S with respect to the coalitional response (20)–(23).*

Proof: With $|I| > \min\{k = 1, 2, \ldots : k \geq b/c\}$, (22) applies. Thus, if $|\neg S| < b/c$, everyone in $\neg S$ tries to be the first to commit to Pollute; if S wins this competition, everyone in $\neg S$ reacts with Pollute according to (21). Thus, there are only two possibilities:

a. S wins. Then the per-capita payoff for the deviating S follows the calculation in (24) and cannot exceed $-b$, which they could have got without deviation.

b. S loses, so an $i \in \neg S$ commits to Pollute, and $S \cup (\neg S) \setminus \{i\}$ reacts according to (21) and (23):

   i. if $|S \cup (\neg S) \setminus \{i\}| < b/c$, then the reaction from $S \cup (\neg S) \setminus \{i\}$ is "everyone plays Pollute" and so the per-capita payoff for S is $-c|I| < -b$.

   ii. otherwise, the reaction from $S \cup (\neg S) \setminus \{i\}$ is "everyone plays Clean" and so the per-capita payoff for S is $-b - c < -b$.

Thus, S cannot block $(a_i = \text{Clean})_{\forall i}$ when $|\neg S| < b/c$. □

Lemma 5 *Suppose S deviates with a polluters fraction λ defined in (19). If $|S| \leq b/c$ then it is optimal for S to set $\lambda = 1$; if $|S| \geq b/c$ then it is optimal for S to set $\lambda = 0$.*

Proof: Since the coalitional response (20)–(23) is independent of λ, the per-capita expected payoff for S is equal to a constant (with respect to λ) plus

$$-c\lambda|S| - b(1-\lambda) = \lambda(b - c|S|) - b.$$

Hence the lemma follows. □

Lemma 6 *If* $|I| > \min\{k = 1, 2, \ldots : k \geq b/c\}$ *and* $|\neg S| \geq b/c$*,* $(a_i = \text{Clean})_{\forall i}$ *cannot be blocked by* S *with respect to the coalitional response (20)–(23).*

Proof: By Lemma 5, we may assume without loss that $|S| < b/c$, otherwise the coalition S cannot get more than $-b$ per member, the per-capita payoff at the status quo. With $|S| < b/c$, everyone in S tries to commit to Pollute by Lemma 5; with the coalition response (22)–(23), the per-capita expected payoff for S is

$$-c|S|\frac{\pi(S)}{\pi(S) + \sum_{i\in\neg S}\pi(\{i\})} + (-c-b)\left(1 - \frac{\pi(S)}{\pi(S) + \sum_{i\in\neg S}\pi(\{i\})}\right).$$

The $-c-b$ in the second term reflects the fact that, if S loses, some individual say i gets to commit to Pollute (hence the $-c$) and all the other players, including S, would react with Clean, because $|\neg\{i\}| = |I| - 1 \geq b/c$ and so (23) applies. Thus, a member of S does not profit from the deviation if and only if

$$-c|S|\frac{\pi(S)}{\pi(S) + \sum_{i\in\neg S}\pi(\{i\})} + (-c-b)\left(1 - \frac{\pi(S)}{\pi(S) + \sum_{i\in\neg S}\pi(\{i\})}\right) \leq -b,$$

i.e.,

$$\frac{\pi(S)}{\pi(S) + \sum_{i\in\neg S}\pi(\{i\})}\left(\frac{b}{c} - |S| + 1\right) \leq 1. \tag{25}$$

Since $|S| < b/c \leq |\neg S|$ and everyone in $\neg S$ challenges S, (18) implies that $\frac{\pi(S)}{\pi(S)+\sum_{i\in\neg S}\pi(\{i\})} \leq c/b$. Thus, (25) follows from the fact that $|S| \geq 1$ and $b/c - |S| + 1 \geq 0$. Hence $(a_i = \text{Clean})_{\forall i}$ cannot be blocked by the coalition S. □

Finally, Let us verify the core condition 2b., i.e., incentive feasibility of the preemptive mechanism. As the preemptive mechanism (22) is completely decentralized, this condition is equivalent to the condition that (22) constitutes a Nash equilibrium. Pick any deviating coalition $S \subsetneq I$.

Lemma 7 *If* $|\neg S| < b/c$ *and* $|I| - 1 < b/c$*, then the preemptive mechanism prescribed by (20)–(23) is incentive feasible.*

Proof: With $|\neg S| < b/c$ and $|I| - 1 < b/c$, (20) recommends $\neg S$ to not challenge S. Pick any $i \in \neg S$. If player i does not challenge S, then S gets to commit and afterwards $\neg S$ will react with all playing Pollute, as $|\neg S| \leq |I| - 1 < b/c$ and so (21) applies. If player i challenges S, then his payoff can be different only if i defeats S, but after he has committed, the reaction from $I \setminus \{i\}$ is all playing Pollute, as $|I| - 1 < b/c$ and so (21) applies. In sum, if player i challenges S, this behavior can make a difference on his payoff only if he beats S, but then he is in the worst case scenario where everyone else will play Pollute. Thus, his best response is to not challenge S as recommended. □

Lemma 8 *If $|\neg S| \geq b/c$ or $|I| - 1 \geq b/c$, the preemptive mechanism prescribed by (20)–(23) is incentive feasible.*

Proof: Suppose S deviates with a polluters fraction λ defined in (19). For every $i \in \neg S$, seeing S deviating and expecting the other players in $\neg S$ to respond according to (22), i's expected net payoff from following (22) is equal to

$$\begin{aligned} z_i \quad := \quad & -c \cdot \frac{\pi(\{i\})}{\pi(S) + \sum_{j \in \neg S} \pi(\{j\})} - (\lambda|S|c + b) \frac{\pi(S)}{\pi(S) + \sum_{j \in \neg S} \pi(\{j\})} \\ & -(c+b)\left(1 - \frac{\pi(\{i\})}{\pi(S) + \sum_{j \in \neg S} \pi(\{j\})} - \frac{\pi(S)}{\pi(S) + \sum_{j \in \neg S} \pi(\{j\})}\right), \end{aligned} \tag{26}$$

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 the first and the third cases, only one player gets to commit first, and there are still at least b/c players who have not committed (since $|\neg S| \geq b/c$ or $|I| - 1 \geq b/c$ by hypothesis), so the reaction function (23) implies that these players will react with $(a_i = \text{Clean})_{\forall i}$. In the second case, the deviating coalition S gets to commit first, with $|\neg S| \geq b/c$, (23) implies that $\neg S$ will react with $(a_i = \text{Clean})_{\forall i}$.

If player i challenges S with the action Clean instead of Pollute, then i's expected net payoff is the same as (26) except that the $-c$ in the first term is replaced by $-b$. Since $b > c$ by (17), this replacement lowers the net payoff. The other alternative for player i is to not challenge S. Then i's expected net payoff is equal to

$$z_i' := -(\lambda|S|c + b) \cdot \frac{\pi(S)}{\pi(S) + \sum_{j \in (\neg S)\setminus\{i\}} \pi(\{j\})} - (c+b) \cdot \frac{\sum_{j \in (\neg S)\setminus\{i\}} \pi(\{j\})}{\pi(S) + \sum_{j \in (\neg S)\setminus\{i\}} \pi(\{j\})}, \tag{27}$$

where the first term corresponds to the event that the coalition S wins the commitment device and the second term is that someone in $\neg(S \cup \{i\})$ wins. We show that $z_i > z_i'$:

$$\begin{aligned}
z_i \overset{(26)}{=} & -\frac{\pi(\{i\})}{\pi(S) + \sum_{j \in \neg S} \pi(\{j\})} c \\
& - \left(1 - \frac{\pi(\{i\})}{\pi(S) + \sum_{j \in \neg S} \pi(\{j\})}\right) \\
& \times \left(\frac{\pi(S)}{\pi(S) + \sum_{j \in (\neg S) \setminus \{i\}} \pi(\{j\})} (\lambda |S| c + b) + \frac{\sum_{j \in (\neg S) \setminus \{i\}} \pi(\{j\})}{\pi(S) + \sum_{j \in (\neg S) \setminus \{i\}} \pi(\{j\})} (c + b)\right) \\
\overset{(27)}{=} & \frac{\pi(\{i\})}{\pi(S) + \sum_{j \in \neg S} \pi(\{j\})} (-c) + \left(1 - \frac{\pi(\{i\})}{\pi(S) + \sum_{j \in \neg S} \pi(\{j\})}\right) z_i'.
\end{aligned}$$

Thus, z_i is a strict convex combination between $-c$ and the lottery z_i'. Since each realized value of z_i' is less than $-c$ by (17), $z_i > z_i'$. Thus, it is a best response for i to follow the preemption (22). □

Thus, condition 2b. is satisfied. Hence $(a_i = \text{Clean})_{\forall i}$ is an element of the core when $|I| \geq b/c$. By Lemmas 1–2, $(a_i = \text{Pollute})_{\forall i}$ is Pareto optimal and a core element when $|I| \leq b/c$. Thus, the core contains the symmetric Pareto optima for any size of I. ■

To appreciate Theorem 1, note the decentralized feature of the approach. There is no a priori hierarchy among the players. Players are free to make agreements with one another. No central planner is needed to oversee the efficiency of their arrangements. The social contract where everyone commits to the Pareto optimum at the outset is achieved merely by the mutual threats of preemption, blocking, and reaction among the players themselves.

The Pareto optimum in such pollution problems may also be supported, without centralized device, as an equilibrium in repeated games. But that requires the players to be sufficiently patient. Interpreted to real-world setups, such patience typically corresponds to situations where players are stuck with one another for long such as in traditional agricultural societies. Such situations may become obsolete with the globalization of markets and societies. Then contractual commitments such as the kind implicitly assumed here may be natural substitutes for repeated games.

In the literature, there is of course the principal-agent approach, assuming that an external principal enforces the Pareto optimum for the players. The problem is that such approach to implement Pareto optima relies on the neutrality and incorruptibility of the principal. If we take the principal's incentive into account, treating it as a player, then a principal-agent solution amounts to a specific arrangement within a coalition consisting of the principal and the agents involved, which ultimately leads to a coalitional approach.

6 Bibliography Note

On the question how the complementary coalition responds to a deviating coalition, early literatures typically assumed that a deviating coalition expects the worst-case scenario of its outsiders.[5] Recent literature usually uses the noncooperative method of predicting the behaviors of coalitions by the Nash equilibria of the noncooperative game played between coalitions such that each coalition is treated as a "composite players." Contributions to this approach include Ichiishi [15], Zhao [24],[6] Ray and Vohra [19], and Hyndman and Ray [14]. This method does not work in our externality problem. The stumbling block is that, as long as the coalitions act independently of one another, Pollute strictly dominates Clean for singleton coalitions, so the Pareto optimum is ruined.

The recursive principle, which I use to formulate one of the conditions for the core, has been used in the coalitional games literature such as Ray and Vohra [19], Huang and Sjöström [13], and Piccione and Razin [18]. Huang and Sjöström called it r-theory.[7]

On the question who gets to commit at which time, the typical treatment in the literature is to assume an a priori sequential protocol governing the sequence in which players take turn to decide on which coalitions to join (e.g., Aumann and Myerson [6], Ray and Vohra [20], and Hafalir [12]). The problem is that the prediction is sensitive to the particular sequential protocol and not necessarily Pareto optimal. If applied to our externality problem, the prediction is suboptimal, since any first mover plays Pollute. My model does not need any exogenous sequential ranking on the players or coalitions.

This paper is not much related to literature on bargaining sets, initiated by Aumann and Maschler [5], with Piccione and Razin [18] a recent development. That literature requires

[5] That was assumed in the origin of our example, Shapley and Shubik [21]. Aumann's [4] β-effect, which requires that a blocking plan should benefit the coalition no matter what the complementary coalition does, also implies the worst-case assumption.

[6] Zhao replaces the "best response" condition in the equilibrium notion by the condition that a coalition's response to the other coalitions belongs to the core within the coalition.

[7] Hafalir's [12] r-core is different, as the reaction from a complementary coalition in his notion is not necessarily robust against deviations among the subcoalitions within the complement.

that a coalitional deviation should be credible. My approach, focusing on the opposite direction, requires that a coalitional deviation should take into account the outsiders' response and that the outsiders' response should be credible.

The traditional notion of the core has been applied to the general equilibrium theory to produce the elegant core equivalence theorem (Debreu and Scarf [10] and Anderson [2], etc.). Due to the traditional assumption that a coalition is self-sufficient, the theorem is restricted to exchange economies and production economies where production technologies are available to every coalition. Foley's [11] formulation of the core with public goods is subject to the same restriction. This restriction is removed recently by Xiong and Zheng [23]. There we incorporate a kind of externality between coalitions that share a firm.

I have found no precedent to the notion of coordinated preemptive measures, or preemptive mechanism. The notion of coalitional response also appears to be new.

A Appendix: Deriving the Characteristic Function

In the traditional model of the core, a coalition is self-sufficient and its "value" is given by an exogenous characteristic function. The notations set up in §2.1 include as a special case the traditional model, with the characteristic function derived from the primitives rather than assumed as a primitive.

For any nonempty $S \subseteq I$, call an $a_S \in A_S$ *independent* action profile for S iff

$$\forall a_{\neg S} \in A_{\neg S} \; \forall a'_{\neg S} \in A_{\neg S} \; \forall i \in S : u_i(a_S, a_{\neg S}) = u_i(a_S, a'_{\neg S}). \tag{28}$$

For example, any player i's any action a_i has a component $a_i^n \subseteq I$ such that $i \in a_i^n$, meaning that i proposes to form a self-sufficient club consisting of exactly the players in the set a_i^n; for any nonempty $S \subseteq I$, an action $a_S \in A_S$ is independent for S if $a_i^n = S$ for all $i \in S$, i.e., everyone in S agrees to form a self-sufficient club containing exactly the members of S. When (28) holds, denote for each $i \in S$ $v_i(S, a_S) := u_i(a_S, a_{\neg S})$ with $a_{\neg S}$ any element of $A_{\neg S}$.

By (28), $(a_S, a_{S'})$ is an independent action for $S \cup S'$ if a_S is an independent action for S and $a_{S'}$ an independent action for S'. I.e., the union of any two self-sufficient clubs can be regarded as a self-sufficient club.

For any nonempty $S \subseteq I$, let A_S^{ind} be the subset of A_S that contains all the independent actions for S. To specialize to the traditional model we make three assumptions. The first is that $A_S^{\text{ind}} \neq \varnothing$ for every nonempty $S \subseteq I$, i.e., every coalition can avoid externalities from its outsiders by turning itself into a self-sufficient club. Secondly,

$$\left[S \cap T = \varnothing \text{ and } a_S \in A_S^{\text{ind}} \text{ and } a_T \notin A_T^{\text{ind}}\right] \Longrightarrow (a_S, a_T) \notin A_{S \cup T}^{\text{ind}}. \tag{29}$$

I.e., the union of a self-sufficient club S and another set T of players who want to be included does not constitute a self-sufficient club, as S does not need, nor offers help to, outsiders. The third assumption is

$$\exists c \in \mathbb{R}\ \forall i \in I : \left[\exists S \subseteq I : i \in S \text{ and } a_S \in A_S^{\text{ind}}\right] \Longleftrightarrow u_i(a_S, a_{\neg S}) > c. \tag{30}$$

In other words, player i's payoff is above the constant c if and only if it belongs to a self-sufficient club. Coupled with the first assumption, which implies that $A_{\{i\}}^{\text{ind}} \neq \varnothing$ for any player i, (30) implies that any player i can secure a payoff above c.[8]

From the above assumptions, it follows that any coalition unanimously prefer to play only independent actions.

Utilities are *transferable* iff for any nonempty $S \subseteq I$ and any $a_S \in A_S^{\text{ind}}$, if $(z_i)_{i \in S} \in (z, \infty)^S$ and $\sum_{i \in S} z_i = \sum_{i \in S} v_i(S, a_S)$, then there exists an $a'_S \in A_S^{\text{ind}}$ such that $v_i(S, a'_S) = z_i$ for each $i \in S$. I.e., if a total payoff for a coalition is attainable, then any allocation $(z_i)_{i \in S}$ of the total payoff among the coalition members is attainable.[9]

Given transferable utilities, a characteristic function is derived, rather than assumed exogenously in the traditional model, as a mapping that associates to each coalition S the maximum total payoff $\sum_{i \in S} v_i(S, a_S)$ among all independent actions a_S for S.

B Appendix: Nonempty Core in Two-Player Games

Theorem 2 *For any two-player finite game, the core is nonempty.*

[8] Note that (30) does not imply that for any $i \in S$, if $a_S \notin A_S^{\text{ind}}$ then $u_i(a_S, a_{\neg S}) \leq c$. Consider for example an S being the disjoint union of S_1 and S_2, with action $a_S = (a_{S_1}, a_{S_2})$ such that $a_{S_1} \in A_{S_1}^{\text{ind}}$ and $a_{S_2} \notin A_{S_2}^{\text{ind}}$. Then $a_S \notin A_S^{\text{ind}}$ by (29). For any $i \in S_1$, $u_i(a_S, a_{\neg S}) > c$ for all $a_{\neg S}$, because $a_{S_1} \in A_{S_1}^{\text{ind}}$.

[9] Transferability of utilities requires that the action sets are sufficiently rich. For example, any action a_i of any player i contains two components, a_i^n and a_i^d. The component a_i^n is the self-sufficient club to which i wants to belong and a_i^d is the payoff demanded by i conditional on the formation of the club. An action $a_S \in A_S$ is independent for coalition S if and only if $a_i^n = S$ and $a_i^d = v_i(S, a_S)$ for all $i \in S$.

Proof For each player i, with the action set A_i finite, there exists an $a_i^* \in A_i$ such that for some $\gamma_{-i}(a_i^*) \in A_{-i} \cap \mathrm{BR}_{-i}(a_i^*)$,

$$u_i(a_i^*, \gamma_{-i}(a_i^*)) \geq u_i(a_i, a_{-i}) \tag{31}$$

for any $a_i \in A_i$ and any $a_{-i} \in \mathrm{BR}_{-i}(a_i) \cap A_{-i}$. Let

$$\alpha^* := \frac{\pi(\{1\})}{\pi(\{1\}) + \pi(\{2\})} \left(a_i^*, \gamma_{-i}(a_i^*)\right) + \frac{\pi(\{2\})}{\pi(\{1\}) + \pi(\{2\})} \left(\gamma_i(a_{-i}^*), a_{-i}^*\right).$$

If there is no $(a_1', a_2') \in A_1 \times A_2$ that Pareto dominates α^*, then α^* belongs to the core. The supporting coalitional response is:

i. if any player i deviates by trying to commit to any action a_i, player $-i$ challenges i by trying to commit to a_{-i}^*;

ii. if any player i has committed to any action a_i, player $-i$ plays any element in $\mathrm{BR}_{-i}(a_i)$ if $a_i \neq a_i^*$ and plays $\gamma_{-i}(a_i^*)$ if $a_i = a_i^*$.

Let us verify the core conditions for α^*. Core condition 2c. follows trivially from the reactive response specified by provision (ii). Core condition 2b. follows from the preemptive response specified by provision (i) and the fact that a_{-i}^* is the best action for player $-i$ to commit to due to Ineq. (31). To verify core condition 2a., first note α^* cannot be blocked by the grand coalition $\{1, 2\}$, as α^* is not Pareto dominated by any $(a_1', a_2') \in A_1 \times A_2$. Furthermore, α^* cannot be blocked by any individual player. Say player i deviates by trying to commit to some a_i. Then player $-i$ will also try to commit to a_{-i}^*. With probability $\frac{\pi(\{1\})}{\pi(\{1\})+\pi(\{2\})}$, player i wins and ends with the outcome (a_i, a_{-i}) for some $a_{-i} \in \mathrm{BR}_{-i}(a_i)$, which does not give player i a higher payoff than the outcome $(a_i^*, \gamma_{-i}(a_i^*))$ that i would have obtained with the same probability had it not deviated. With probability $1 - \frac{\pi(\{1\})}{\pi(\{1\})+\pi(\{2\})}$, player i loses, then we have the same outcome $(a_{-i}^*), a_{-i}^*)$ as in the case when player i does not deviate. Thus, i's unilateral deviation is not profitable. Hence core condition 2a. is satisfied.

If an $(a_1', a_2') \in A_1 \times A_2$ Pareto dominates α^*, then with $A_1 \times A_2$ finite, there exists an $(a_1'', a_2'') \in A_1 \times A_2$ that Pareto dominates α^* and is itself a Pareto optimum (as Pareto dominance is a transitive relation). Then (a_1'', a_2'') belongs to the core, supported by the same coalitional response as above. The proof is the same as the previous paragraph. ■

The construction in the above proof can be easily applied to the Nash demand game and the Battle of Sexes game and yield predictions different from existing solution concepts.

References

[1] Varouj A. Aivazian and Jefferey L. Callen. The Coase theorem and the empty core. *Journal of Law and Economics*, 24(1):175–181, April 1981. 1

[2] Robert Anderson. An elementary core equivalence theorem. *Econometrica*, 46(6):1483–1487, 1978. 6

[3] 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

[4] Robert J. Aumann. The core of a cooperative game without side payments. *Transactions of the American Mathematical Society*, 98(3):539–552, Mar. 1961. 5

[5] 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. 6

[6] Robert J. Aumann and Roger B. Myerson. Endogenous formation of links between players and of coalitions: An application of the Shapley value. In A. Roth, editor, *The Shapley Value: Essays in Honor of Lloyd Shapley*, pages 175–191. Cambridge University Press, Cambridge, UK, 1988. 6

[7] B. Douglas Bernheim, Bezalel Peleg, and Michael Whinston. Coalition-proof Nash equilibria I: Concepts. *Journal of Economic Theory*, 42(1):1–12, June 1987. 2

[8] John H. Boyd and John P. Conley. Fundamental nonconvexities in Arrovian markets and a Coasian solution to the problem of externalities. *Journal of Economic Theory*, 72:388–407, 1997. 1

[9] R. H. Coase. The problem of social cost. *Journal of Law and Economics*, 3:1–44, October 1960. 1

[10] Gerard Debreu and Herbert Scarf. A limit theorem on the core of an economy. *International Economic Review*, 4(3):235–246, September 1963. 1, 6

[11] Duncan K. Foley. Lindahl's solution and the core of an economy with public goods. *Econometrica*, 38(1):66–72, January 1970. 6

[12] Isa Hafalir. Efficiency in coalition games with externalities. *Games and Economic Behavior*, 61:242–258, 2007. 6, 7

[13] Chen-Ying Huang and Tomas Sjöström. Consistent solutions for cooperative games with externalities. *Games and Economic Behavior*, 43:196–213, 2003. 3.6, 6

[14] Kyle Hyndman and Debraj Ray. Coalition formation with binding agreements. *Review of Economic Studies*, 74:1125–1147, 2007. 6

[15] Tatsuro Ichiishi. A social coalitional equilibrium existence lemma. *Econometrica*, 49(2):369–377, March 1981. 6

[16] Diego Moreno and John Wooders. Coalition-proof equilibrium. *Games and Economic Behavior*, 17:80–112, 1996. 2

[17] Roger B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 1991. 4, 3

[18] Michele Piccione and Ronny Razin. Coalition formation under power relation. *Theoretical Economics*, 4:1–15, 2009. 3.6, 6

[19] Debraj Ray and Rajiv Vohra. Equilibrium binding agreements. *Journal of Economic Theory*, 73:30–78, 1997. 6

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

[21] Lloyd Shapley and Martin Shubik. On the core of an economic system with externalities. *American Economic Review*, 59(4):678–684, September 1969. 5

[22] D. Starrett. Fundamental nonconvexities in the theory of externalities. *Journal of Economic Theory*, 4:180–199, 1972. 1

[23] Siyang Xiong and Charles Z. Zheng. Core equivalence theorem with production. *Journal of Economic Theory*, 137(1):246–270, November 2007. 6

[24] Jingang Zhao. The hybrid solutions of an n-person game. *Games and Economic Behavior*, 4(1):145–160, 1992. 6

Index