

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

Charles Z. Zheng[†]

February 9, 2010

Abstract

The core is reformulated to incorporate the externality typical in strategic form games, where a player's payoff may depend on the profile of actions across all players. Any coalition of players may reach an agreement on who takes what action within the coalition. The outsiders of the coalition may coordinate a response to preempt the coalition's commitment to its agreement. If a coalition succeeds in committing to its action profile, the outsiders' reactions constitute a core solution among themselves. In an externality problem where pollution is the dominant action, the core is nonempty and consists of exactly the Pareto optima.

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

[†]Before July 2010: 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>. Starting from July 2010: Department of Economics, University of Western Ontario.

1 Introduction

Presented in this paper is a new concept of the core that incorporates the strategic interactions between any coalitions. Applied to an n -player externality problem where pollution is the dominant action, the new core is nonempty and consists of exactly the Pareto optima. This result may be interpreted as a justification for the “Coase Theorem” [6] without assuming property rights or complete markets.

In his original paper, Coase was reluctant to stick to a particular kind of institutional backgrounds. Researchers have proposed a few institutional backgrounds in which the Coase Theorem might be true such as property rights, complete markets, and the core. Among them, the core has the merit of not relying on any institution that requires external enforcement. Even if property rights have been well-defined, it remains as an issue whether they can be enforced. And the main justification for the solution concept for complete markets, competitive equilibrium, is still based on the core (e.g., Debreu and Scarf [8]).

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 is based on an assumption 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 pollution 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 (Aumann [2], Bernheim, Peleg and Whinston [5], and Moreno and Wooders [12], etc.) cannot support the Pareto optimum as an equilibrium. To apply the coalitional method, however, we immediately run into a problem. Now that the payoffs for any non-grand coalition depend on the actions of its outsiders, how would the complementary coalition respond to a deviating coalition?

In the example, as pollution is the unique Nash equilibrium, it does not help to assume

that the outsiders of a deviating coalition stick to the status quo. A typical treatment in the traditional coalitional theory is to assume the worst-case scenario for a blocking coalition.¹ In our example, that means everyone outside the coalition plays Pollute. The problem is that this threat is not credible. To see that, suppose that player 1 has somehow committed to Pollute. Then, with $b < 2c$, the only Pareto optimum for players 2 and 3 is both playing Clean. Furthermore, neither of them would deviate by somehow committing to Pollute. If player 2 commits to Pollute, then the actions of both players 1 and 2 have been given; consequently, player 3's best response is uniquely Pollute. Thus, in deviating to Pollute, player 2 bears a c -dollar cost due to its own pollution action and another c -dollar cost due to player 3's pollution as player 3's best response. Since $2c > b$, player 2 would rather bear the cost b dollars by playing Clean. Same reasoning goes for player 3. Thus, it is not credible for players 2 and 3 to both play Pollute in response to player 1's deviation.

The above calculation suggests an idea. If a coalition has committed to a deviation, the reaction from the outsiders should be a solution in the core within the outsiders. Given player 1's commitment to Pollute, the prediction that players 2 and 3 both play Clean is robust against any unilateral and coalitional deviations among players 2 and 3.

This idea of applying the core condition recursively has been hinted at by Ray and Vohra [16] and formalized by Huang and Sjöström [10] into a notion called r -core. One may call it the recursive principle. Based on some notion that a coalitional deviation involves commitment, the principle postulates that, once a coalition commits, the complementary coalition reacts with an action that belongs to the core within the complementary coalition.

A question stands in the way, however: Who gets to commit first? To see why this question is unavoidable, recall our pollution example. There we have shown that if a player manages to commit to Pollute before the other two can make any commitment, then the other two would stick to Clean. 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.

In the existing literature, the answer for this question is to assume that there is an a priori sequential protocol according to which players take turn to decide on their commit-

¹ That is assumed in the origin of our example, Shapley and Shubik [17]. Aumann's [3] 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.

ment. The problem is that the prediction is sensitive to the particular sequential protocol. For instance, in the protocol of Ray and Vohra, an early mover has an advantage over a late mover, and their equilibria are usually not Pareto optimal. By contrast, in a reversed protocol proposed by Hafalir [9], a late mover proposes a realignment of the coalitions formed by early movers and pockets the surplus if its proposed realignment is accepted, and the inefficiency prediction is overturned. Even if the sequential protocol were assumed to be randomly drawn by nature at the outset, a question remains whether the protocol may itself be susceptible to coalitional deviations. In the case of Ray and Vohra, late movers may want to skip to the front. In the case of Hafalir, early movers may want to hide until the end.

The answer proposed in this paper is that the commitment opportunity for a coalition is uncertain. When a coalition has reached an agreement within itself to commit to an action profile, the other players can challenge the coalition. For instance, the complementary coalition may try to make a commitment before the former coalition does thereby grabbing the first-mover advantage. Or it may ask the court to nullify the contract of the former coalition on the ground of its negative externality. If the coalition is challenged, the winner is chosen randomly between the contending parties. The probability of winning the commitment opportunity is an exogenous function of the configuration of the contending parties. It may be determined purely by the sizes of the contending parties, or it may depend on other resources such as political clout and the means of violence.

Introducing this competitive setup for commitment opportunities, this paper proposes an answer to the crucial question how the complementary coalition responds to a deviating coalition. Any set of players may deviate from a status quo coalitionally. In deviating, the coalition expects that, if it gets to commit to an action, the reactions from the outsiders constitute an element in the core within the outsiders. The coalition also expects that, when it tries to commit, the outsiders may coordinate a *preemptive response* to either challenge the coalition or stay put. To *block* the status quo, a coalition needs to profit from an alternative action profile across its members based on these expectations. A profile of actions belongs to the *core* if no coalition can block it in such a manner. Since the grand coalition has no outsider, it can always block Pareto inferior action profiles, so any element in this reformulated core is Pareto optimal.

This notion of the core incorporates whatever externalities that a strategic form game can capture. 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 new concept of the core is defined in §3 and contrasted with some benchmark solution concepts in §4. The traditional theory of the core corresponds to a special case in our model where coalitions can avoid externalities by forming self-sufficient clubs. Proposition 1 asserts that our notion of the core coincides with the traditional core in such externality-avoidable environments. Illustrated in §4.2 is a three-player coordination game where strong Nash equilibrium does not exist while our core is nonempty. The main result, Theorem 1, is in §5. It asserts that the Pareto optimum is the unique solution in our core in an n -player generalization of the above problem of pollution externality.

The externalities considered by the existing coalitional literature so far have been the possibility that the payoffs for a coalition depend on the coalition configuration among the outsiders. For such environments, a main approach in the literature is to analyze a non-cooperative game where the players' actions are their votes on the coalition configurations, proposed according to a protocol. A recent contribution is Hyndman and Ray [11]. There is also an axiomatic approach of de Clippel and Serrano [7] that extends the Shapley value to such externality environments. More related to this paper is the third approach, which is to formulate new concepts of the core, as in Huang and Sjöström [10] and Hafalir [9]. The closest one is Huang and Sjöström. They have developed a recursive principle, which is similar to condition 2c. in my notion of the core. The literature has not addressed the issue Who gets to commit first. The novelty of this paper is in its treatment for this issue.

Assuming full commitment abilities for any deviating coalition, this paper has not considered the possibility that a deviating coalition may be susceptible to the betrayal of some of its members who collude with some outsiders. This possibility would be my interpretation of the vast literature on objections and counterobjections, initiated by Aumann and Maschler [4], with Piccione and Razin [15] a recent development. The implicit institutional interpretation of those notions has been a centralized conference where coalitions debate various proposals until the grand coalition reaches an agreement (e.g., Myerson [13]). By contrast, the institutional interpretation of my notion is a somewhat decentralized setting where coalitions may reach agreements on their own and commit to their agreements if they are not challenged or if they win a challenge.

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 $\emptyset \neq S \subseteq I$, denote

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

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 , which we simply call an *action* of *coalition* S .

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

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

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 ,

$$u_i(a_I) = -b\mathbf{1}_{a_i = \text{Clean}} - c|\{j \in I : a_j = \text{Pollute}\}|. \tag{2}$$

This model includes the traditional characteristic-function-based model as a special case and allows for utility transfers across players. That will be established in §4.1.

2.2 The Competition for Coalitional Commitment

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 a mixed action $\alpha_S \in \Delta A_S$, with ΔK the set of lotteries on the set K . 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 within itself, the players outside S who have not committed choose whether to *challenge* S or not.

- a. If they choose to challenge S , they suspend their agreements and try to commit, before S does, to a profile of actions, which need not be the same as the profile of actions that they have agreed upon.
 - b. If the set of players challenging S is J , then with probability $\pi[S, J]$ the coalition S wins and otherwise the coalition J wins, with probability $1 - \pi[S, J]$.
 - c. The winning coalition commits to its agreed upon action profile, and the losing coalition cannot commit to any action until the winner has committed (so the 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.

The probability $\pi[S, J]$ in the above protocol signifies the relative power between coalitions S and J in the event that J challenges S . For instance, $\pi[S, J]$ may be equal to $|S|/(|S| + |J|)$. In general, however, π may take various forms.

3 Defining the Core

3.1 Coalitional Responses

This is a coalitional counterpart for the notion of strategies. Suppose that $T \subsetneq I$ is the set of players who have committed and $a_T \in A_T$ the committed action profile. For the rest of the economy, a *coalitional response* is a pair of two functions:

$$\mathcal{P}_{T,a_T} : (S, \alpha_S) \mapsto \mathcal{P}_{T,a_T}(S, \alpha_S)$$

maps any coalition $S \subsetneq -T$ and any $\alpha_S \in \Delta A_S$ to an element of $\Delta \left(\prod_{i \in (-T) \setminus S} (A_i \cup \{\text{nil}\}) \right)$, and

$$\mathcal{R}_{T,a_T} : (S, a_S) \mapsto \mathcal{R}_{T,a_T}(S, a_S)$$

maps any coalition $S \subsetneq -T$ and any action profile $a_S \in A_S$ to an element of $\Delta A_{(-T) \setminus S}$.

To interpret the above construct, suppose that there is an action profile to which the set $-T$ of all the players who have not committed is expected to agree to commit. A coalition $S \subsetneq -T$ may want to deviate from that expectation and reach an agreement within S to commit to an alternative action profile through a lottery α_S of action profiles. Seeing this agreement, the other players, constituting the set $(-T) \setminus S$, coordinate a *preemptive response* $\mathcal{P}_{T,a_T}(S, \alpha_S)$, which is a lottery, say $\alpha_{(-T) \setminus S}$, that selects a profile $a_{(-T) \setminus S} \in \prod_{i \in (-T) \setminus S} (A_i \cup \{\text{nil}\})$. If $a_i = \text{nil}$ according to this profile, player i does not participate in the coalition to challenge the deviating coalition S ; if $a_i \neq \text{nil}$, then player i participates in the challenging coalition and, in the event that the challenging coalition defeats S , player i commits to the action a_i .

Thus, if $a_{(-T) \setminus S}$ is a realization of the preemptive response, the challenging coalition is

$$N(a_{(-T) \setminus S}) := \{j \in \neg(T \cup S) : a_j \neq \text{nil}\}. \quad (3)$$

According to the protocol modeled in §2.2, with probability $\pi[S, N(a_{(-T) \setminus S})]$ the deviating coalition S wins and commits to an action profile selected by the lottery α_S , otherwise the challenging coalition wins and commits to $a_{N(a_{(-T) \setminus S})}$. Whichever event happens, the outcome is that some coalition S' (either S or $N(a_{(-T) \setminus S})$) commits to some action profile $a_{S'} \in A_{S'}$. Taking that as given, the other uncommitted players, constituting the set $((-T) \setminus S) \setminus S'$, will commit to an action profile according to a lottery $\mathcal{R}_{T,a_T}(S', a_{S'}) \in \Delta A_{-S'}$. This lottery we call *reactive response*. It may be implemented in one shot or in multiple stages where different coalitions in $((-T) \setminus S) \setminus S'$ make commitments at different stages.

3.2 Expected Payoffs given a Coalitional Response

Suppose that $T \subsetneq I$ is the set of players who have committed and $a_T \in A_T$ the committed action profile. Also suppose that a coalition $S \subsetneq -T$ has agreed within itself to commit via lottery $\alpha'_S \in \Delta A_S$ and the complementary coalition $(-T) \setminus S$ has coordinated a preemptive response $\alpha'_{(-T) \setminus S}$. Given any realization $a_{(-T) \setminus S}$ of $\alpha'_{(-T) \setminus S}$, the challenging coalition is the $N(a_{(-T) \setminus S})$ in (3), and the expected payoff for any player $i \in I$ is

$$u_i(a_T, \alpha'_S, a_{(-T) \setminus S} \mid \mathcal{R}_{T, a_T}) = \pi[S, N(a_{(-T) \setminus S})] \sum_{a_S \in A_S} \alpha'_S(a_S) u_i(a_T, a_S, \mathcal{R}_{T, a_T}(S, a_S)) \\ + (1 - \pi[S, N(a_{(-T) \setminus S})]) u_i(a_T, a_{N(a_{(-T) \setminus S})}, \mathcal{R}_{T, a_T}(\{j\}, a_{N(a_{(-T) \setminus S})})) \quad (4)$$

where $\pi[S, N(a_{(-T) \setminus S})]$ is the winning probability for coalition S defined in §2.2. Through the $\mathcal{R}_{T, a_T}(S, a_S)$ and $\mathcal{R}_{T, a_T}(\{j\}, a_{N(a_{(-T) \setminus S})})$ on the right-hand side, Eq. (4) takes into account the reactive response \mathcal{R} to either outcome of the commitment competition.

Integrating the above payoff across all possible realizations $a_{(-T) \setminus S}$ of the preemptive response $\alpha'_{(-T) \setminus S}$, we know that the expected payoff for any player $i \in I$ is

$$u_i(a_T, \alpha'_S, \alpha'_{(-T) \setminus S} \mid \mathcal{R}_{T, a_T}) \\ = \sum_{a_{(-T) \setminus S} \in \prod_{j \in (-T) \setminus S} (A_j \cup \{\text{nil}\})} \alpha'_{(-T) \setminus S}(a_{(-T) \setminus S}) u_i(a_T, \alpha'_S, a_{(-T) \setminus S} \mid \mathcal{R}_{T, a_T}). \quad (5)$$

3.3 The Definition of Blocking

If $T \subsetneq I$ is the set of players who have committed and $a_T \in A_T$ the committed action profile, a mixed action $\alpha_{-T} \in \Delta A_{-T}$ of $-T$ is *blocked* by a coalition $S \subseteq -T$ with respect to a coalitional response $(\mathcal{P}_{T, a_T}, \mathcal{R}_{T, a_T})$ iff there exists a mixed action $\alpha'_S \in \Delta A_S$ such that

$$u_i(a_T, \alpha'_S, \mathcal{P}_{T, a_T}(S, \alpha'_S) \mid \mathcal{R}_{T, a_T}) > \sum_{a_{-T} \in A_{-T}} \alpha_{-T}(a_{-T}) u_i(a_T, a_{-T}), \quad (6)$$

where the left hand-side is defined by (5). I.e., every member i of coalition S gets a higher expected payoff from the deviation of agreeing to commit via lottery α'_S rather than abiding by the mixed action α_{-T} , knowing the coalitional response $(\mathcal{P}_{T, a_T}, \mathcal{R}_{T, a_T})$.

When $-T = \{i\}$ 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 $a_{-\{i\}}$ for player i . That is because if $\emptyset \neq S \subseteq -T$ then $(-T) \setminus S = \emptyset$.

3.4 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 $\mathcal{K}(T, a_T)$ 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 $\mathcal{K}(T, a_T)$ 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 $\mathcal{K}(T, a_T)$ iff there exists a coalitional response $(\mathcal{P}_{T, a_T}, \mathcal{R}_{T, a_T})$ with the following properties:
 - a. $\alpha_{\neg T}$ is not blocked with respect to $(\mathcal{P}_{T, a_T}, \mathcal{R}_{T, a_T})$;
 - b. if any $S \subsetneq \neg T$ deviates to some $\alpha'_S \in \Delta A_S$, then the preemptive response $\mathcal{P}_{T, a_T}(S, \alpha'_S)$ is not $(\neg T) \setminus S$ -Pareto dominated by any preemptive response, given the fact that S will commit to α'_S unless it is defeated by a challenging coalition,
 - c. if any nonempty set $S \subsetneq \neg T$ has committed to any $a_S \in A_S$, then $\mathcal{R}_{T, a_T}(S, a_S)$ is in the core $\mathcal{K}(T \cup S, (a_T, a_S))$, which is well-defined by recursion.

Why do we define the core in this way? Let us focus on the nontrivial case $|\neg T| > 1$, where there are still multiple players who have not committed. Condition 2a. needs no justification. Condition 2c., which we may call the recursive principle, has been applied in some other models by Huang and Sjöström [10] and Piccione and Razin [15]. 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 is also coordinated among the players in the complementary coalition $(\neg T) \setminus S$. From the perspective that the preemptive response is based on the unanimous cooperation among the members of the $(\neg T) \setminus S$, it is natural to require the $(\neg T) \setminus S$ -Pareto optimality condition, i.e, condition 2b..

Why not impose stronger conditions on a preemptive response such as robustness to unilateral or coalitional deviations within the coalition $(\neg T) \setminus S$? Adding such conditions would skew the balance between a blocking coalition S and its complement $(\neg T) \setminus S$ by making it harder to preempt a blocking attempt. To restore the balance with the additional conditions, we would need to impose similar incentive compatibility and core-like conditions upon a blocking attempt. But then our solution concept may become too complicated.

4 Comparison with Some Benchmarks

To explain the new notion of the core, let us illustrate its relationship with the traditional notion of the core and contrast it with the coalition-proof equilibrium concepts.

4.1 The Traditional Core as a Special Case

The traditional coalitional model does not have explicit actions for the players. Rather, assumed as primitive is a *characteristic function* that maps any coalition to either a set of payoff allocations feasible for the coalition or simply a total payoff for the coalition to distribute among its members. There the implicit assumption is that the players of a coalition can somehow form a self-sufficient club if they desire so unanimously. We shall demonstrate that our model includes this traditional setup as a special case and in this special case the reformulated core coincides with the traditional core.

4.1.1 Independent Actions and Externality-Free Environments

If $\emptyset \neq S \subseteq I$, $a_S \in A_S$ is an *independent* action profile for S iff

$$\forall a_{-S} \in A_{-S} \forall a'_{-S} \in A_{-S} \forall i \in S : u_i(a_S, a_{-S}) = u_i(a_S, a'_{-S}). \quad (7)$$

An interpretation is that an independent action profile for coalition S implies that the players in S form a self-sufficient club. For example, any action a_i of any player i has a component $a_i^n \subseteq I$ such that $i \in a_i^n$, meaning that i proposes to form a network with and only with the set $a_i^n \setminus \{i\}$ of players; 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 network containing exactly the members of S .

Whenever (7) holds, denote for each $i \in S$

$$v_i(S, a_S) := u_i(a_S, a_{-S}) \quad (8)$$

with a_{-S} any element of A_{-S} . By (7), $(a_S, a_{S'})$ 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 within this

section. The first is that $A_S^{\text{ind}} \neq \emptyset$ 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,

$$[S \cap T = \emptyset \text{ and } a_S \in A_S^{\text{ind}} \text{ and } a_T \notin A_T^{\text{ind}}] \implies (a_S, a_T) \notin A_{S \cup T}^{\text{ind}}. \quad (9)$$

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

$$\exists c \in \mathbb{R} \forall i \in I : [\exists S \subseteq I : i \in S \text{ and } a_S \in A_S^{\text{ind}}] \iff u_i(a_S, a_{-S}) > c. \quad (10)$$

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 $A_{\{i\}}^{\text{ind}} \neq \emptyset$ for any player i , (10) implies that any player i can secure a payoff above c .²

In summary, the traditional coalitional model, which we call *externality-free* environment, is an environment where $A_S^{\text{ind}} \neq \emptyset$ for every nonempty $S \subseteq I$ and (9)–(10) hold.

4.1.2 The Traditional Core

Derived from the above primitives, the characteristic function is the mapping from S to the set of payoff vectors $(v_i(S, a_S))_{i \in S}$, with a_S ranging through A_S^{ind} . The traditional model would take this function as the primitive and assume away the set $A_S \setminus A_S^{\text{ind}}$ of actions that do not secure independence for coalition S .

As the traditional core is defined based on the characteristic function, it is formalized as follows. For any $T \subsetneq I$ and any $a_T \in A_T^{\text{ind}}$, a mixed action $\alpha_{-T} \in \Delta A_{-T}^{\text{ind}}$ belongs to the *traditional core* $\mathcal{H}_{\text{old}}(T, a_T)$ for $-T$ iff there does not exist a nonempty $S \subseteq -T$ such that, for some $a'_S \in A_S^{\text{ind}}$, $v_i(S, a'_S) > \sum_{a_{-T} \in A_{-T}^{\text{ind}}} \alpha_{-T}(a_{-T}) v_i(-T, a_{-T})$ for all $i \in S$.

4.1.3 Commitment without Interference

To establish the equivalence between the traditional core and our reformulated core, we make two additional assumptions within this section. The first is transferable utilities. Utilities are *perfectly transferable* iff for any nonempty $S \subseteq I$ and any $a_S \in A_S^{\text{ind}}$, if $(z_i)_{i \in S} \in (z, \infty)^S$

² Note that (10) does not imply that for any $i \in S$, if $a_S \notin A_S^{\text{ind}}$ then $u_i(a_S, a_{-S}) \leq c$. 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 (9). For any $i \in S_1$, $u_i(a_S, a_{-S}) > c$ for all a_{-S} , because $a_{S_1} \in A_{S_1}^{\text{ind}}$.

and $\sum_{i \in S} z_i = \sum_{i \in S} v_i(S, a_S)$, then there exists $a'_S \in A_S^{\text{ind}}$ 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.³

The second assumption is

$$\forall T \subsetneq I \quad \forall a_T \in A_T : \emptyset \neq \mathcal{K}(T, a_T) \subseteq \Delta A_{-T}^{\text{ind}}. \quad (11)$$

I.e., if any coalition T has committed to an action profile, any reaction from the outsiders according to our reformulated core is to form a self-sufficient club. Combined with (9) and the “ \Leftarrow ” direction of (10), this assumption implies that, for any coalition T that wants to commit by itself, some member in T would get hurt by a payoff below c unless the action to which T commits is to form a self-sufficient club. That rationalizes a coalition’s pessimistic expectation typically implicitly assumed in the traditional coalitional literature.

Proposition 1 *In any externality-free environment that satisfies Eq. (11) and perfect transferability of utilities, for any $T \subsetneq I$ and any $a_T \in A_T^{\text{ind}}$,*

a. $\mathcal{K}_{\text{old}}(T, a_T) = \mathcal{K}(T, a_T)$, and

b. if $\alpha_{-T} \in \mathcal{K}_{\text{old}}(T, a_T)$ then α_{-T} is supported as an element of $\mathcal{K}(T, a_T)$ by a coalitional response such that any deviating coalition gets to commit without being challenged.

Claim (b) of the proposition suggests the proof for the “ \subseteq ” direction of claim (a). Specifically, if α_{-T} is a solution in the traditional core $\mathcal{K}_{\text{old}}(T, a_T)$, it is supported as a solution in the reformulated core $\mathcal{K}(T, a_T)$ by the coalitional response: for all $S \subsetneq -T$, $\alpha'_S \in \Delta A_S$, $a_S \in A_S$ and $i \in (-T) \setminus S$, let

$$\mathcal{P}_{T, a_T}(S, \alpha'_S) := \mathbf{1}_{a_i = \text{nil} \forall i \in (-T) \setminus S} \quad (12)$$

$$\mathcal{R}_{T, a_T}(S, a_S) := \text{an arbitrary element of } \mathcal{K}(T \cup S, a_T, a_S), \quad (13)$$

where Eq. (13) is meaningful due to the assumption (11). According to this coalitional response, when a coalition S deviates by agreeing to make a commitment by itself, the

³ 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 $i \in S$.

outsiders of S do not challenge S . They simply let S make its commitment and then commit to a core solution among themselves.

To guarantee that no coalition has a profitable deviation with respect to this coalition, we use the “pessimistic assumptions” (9)–(11). Together, they implied that any deviating coalition would consider only independent actions and that its complementary would react with only independent actions. Then one can show that blocking a traditional core solution in our reformulated sense would mean blocking the solution in the traditional sense.

For the converse, the “ \supseteq ” direction of claim (a), the proof is to show that if a solution is blocked in the traditional sense then it is blocked in our reformulated sense. To block in the reformulated sense, a coalition faces the uncertainty that it need not win when it is challenged by the outsiders. (A deviating coalition may be challenged by outsiders, as the aforementioned “commitment without interference” result applies only to the “ \subseteq ” direction.) It is at this step that the assumption of transferable utilities is used. It allows a coalition to spread the risk among the coalition members by utility transfers in the event that they win despite being challenged.

4.2 Difference from Coalition-Proof Equilibria

To contrast the core with the coalition-proof equilibrium concepts, let us consider the O’Neil Game (cited from Myerson [14]), which may be called battle of three sexes:

		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 event of a challenge the winner is chosen via the majority vote, i.e.,

$$\pi[S, J] = \frac{|S|}{|S| + |J|}$$

for any disjoint subsets S and J of the set $\{1, 2, 3\}$ of players.

The game admits no strong Nash equilibrium. For any Nash equilibrium, say (x_1, y_2, x_3) , which gives the payoff vector $(6, 5, 4)$, there is a 2-player coalition that “blocks”—in the traditional sense—the equilibrium, assuming the other player abiding by the equilibrium. In

the case of (x_1, y_2, x_3) , the blocking coalition consists of players 2 and 3 taking the alternative action (x_2, y_3) .⁴

By contrast, the core according to our reformulation is nonempty. A core solution is

$$\frac{2}{3}(x_1, y_2, x_3) + \frac{1}{3}(x_1, x_2, y_3), \quad (14)$$

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 $3 + 1 = 1$. Note that any two of the three players can be denoted as i and $i + 1$ for some $i \in \{1, 2, 3\}$.

The reactive response \mathcal{R} can be defined trivially. For instance, 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}) ; if only players i and $i + 1$ have committed, then player $i + 3$ plays whatever action such that the resulting payoff vector is not $(0, 0, 0)$. Such an \mathcal{R} satisfies the recursive core condition 2c. in the definition of the core.

Next is the preemptive response \mathcal{P} . If the deviating coalition is $\{i, i + 1\}$, then the other player, $i + 2$, picks a preemptive response to maximize his expected payoff given the deviation. For instance, if $\{i, i + 1\}$ wants to commit to (x_i, y_{i+1}) , then player $i + 2$ challenges it and commits to y_{i+2} upon winning; whereas, if $\{i, i + 1\}$ wants to commit to (y_i, x_{i+1}) , player i does not challenge it and plays x_{i+2} after the coalition has committed. That ensures that the preemptive response is not $\{i + 2\}$ -Pareto dominated by any other preemptive response. Suppose the deviating coalition is a singleton $\{i\}$. If i wants to commit to x_i then the other two players do not challenge him and, after he has committed, they play (x_{i+1}, y_{i+2}) to penalize i . If i wants to commit to y_i , then the other two challenge i with

$$\mathcal{P}(\{i\}, y_i) := \mathbf{1}_{(a_{i+1}, a_{i+2}) = (x_{i+1}, y_{i+1})}. \quad (15)$$

One can check that this preemptive response is not $\{i + 1, i + 2\}$ -Pareto dominated by any other preemptive response.

Let us check that the coalitional response sketched above deters any coalitional deviation from (14). First, the coalition $\{1, 2\}$ cannot profit from deviation. To gain the most

⁴ In fact, the outcome resulting from the coalitional deviation, (x_1, x_2, y_3) , is itself a Nash equilibrium, as Myerson [14] points out.

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 (14), due to player 3's preemptive response $\mathbf{1}_{a_3=y_3}$ and the assumption that the coalition wins with probability $2/3$ in the event of being challenged.

Second, $\{2, 3\}$ cannot block (14). To make player 3 better-off than in (14), the coalition needs to commit to (x_2, y_3) in the event that it wins. Then player 1 challenges the coalition with action y_1 , according to the preemptive response sketched above. 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 (14). Similar reasoning holds for coalition $\{1, 3\}$.

Third, a singleton coalition, say $\{i\}$, cannot block (14). If player i tries to commit to x_i , then the other two players do not challenge it and will react with (x_{i+1}, y_{i+2}) to give player i the lowest positive payoff, 4. If player i tries to commit to y_i , then given the preemptive response (15), i 's expected payoff is

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

which is not higher than the payoff for anyone in (14). Thus, (14) is in the core.

Clearly, the core solution (14) can be viewed as a correlated equilibrium. However, not all correlated equilibria 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 by $\{i, i + 1\}$ for any $i \in \{1, 2, 3\}$. If $\{1, 2\}$ deviates by trying to commit to (x_1, y_2) , the only $\{3\}$ -Pareto undominated 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 the deviating coalition is

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

5 A Coase Theorem in a Pollution Problem

To show the power of the new notion of the core, let us apply it to the n -player generalization of the pollution problem discussed in the Introduction. There are $n \geq 2$ players. Each player's action is $\{\text{Pollute}, \text{Clean}\}$. Let k denote the number of players who play Pollute. If a player plays Clean, its payoff is equal to $-b - kc$; else the payoff is $-kc$. Assume that

$$c < b. \tag{16}$$

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 ensure a high probability of committing 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 coalition size affecting the winning probabilities, such a coalition's expected payoff from playing Pollute is outweighed by the payoff from not playing Pollute.

Theorem 1 *In the pollution problem with n players, assume:*

$$\forall S, J \subseteq I : \left[S \cap J = \emptyset, |S| < \frac{b}{c} \leq |J| \right] \Rightarrow \pi[S, J] \leq \frac{c}{b}. \quad (17)$$

Then the core is nonempty and consists of all the Pareto optima.

Let us sketch the proof of Theorem 1. Consider only the case where no one has committed yet, i.e., $T = \emptyset$ so $I = \neg T$. If $|I| < b/c$, then $-c|I|^2 < -b|I|$, so the Pareto optimum is uniquely that everyone plays Pollute. This is also in the core because for any deviating coalition S of I , $|S| \leq |I| < b/c$ and so S cannot profit from playing Clean at all.

Now consider the nontrivial case where $|I| > b/c$. Now the Pareto optimum is uniquely that everyone plays Clean. Thus, the core is nonempty if and only if “everyone plays Clean” belongs to the core. We construct a coalitional response to support this solution.

First, we construct the reactive response. By induction and the previous calculation on how the Pareto optimum depends on the coalition size, the only candidate for the reactive response is (where $\mathbf{1}_x$ denotes the singular probability measure concentrated at the point x)

$$\mathcal{R}(S, a_S) := \begin{cases} \mathbf{1}_{a_i = \text{Clean} \forall i \in \neg S} & \text{if } |\neg S| \geq b/c \\ \mathbf{1}_{a_i = \text{Pollute} \forall i \in \neg S} & \text{if } |\neg S| < b/c. \end{cases} \quad (18)$$

Second, let us construct the preemptive response. Without loss of generality, we may assume that $|S| < b/c$, otherwise the coalition S cannot profit from deviation from “everyone plays Clean.”

If $|\neg S| < b/c$, then $\neg S$ does not need to challenge the deviating coalition S at all, because once S has committed, everyone in $\neg S$ will react with Pollute according to the lower branch of (18), and so each member of the deviating coalition S gets $-c(|S| + |\neg S|) = -c|I|$ from the deviation, while he could have got the higher payoff $-b$ without deviation.

Thus, consider only the subcase where $|\neg S| \geq b/c$. Let us temporarily assume that the complementary coalition $\neg S$ is better-off challenging S than not challenging it at all. To deter coalitional deviations, the preemptive response needs to penalize a deviating coalition by trying to commit to Pollute with sufficiently large probability. However, the probability of committing to Pollute should be large only enough to deter coalitional deviations, otherwise the preemptive response would be Pareto inferior for the complementary coalition. Thus, the preemptive response may take the following form:

$$\mathcal{P}(S, a_S) := \begin{cases} \sum_{i \in \neg S} \frac{1}{|\neg S|} (\sigma_S \mathbf{1}_{a_i = \text{Pollute}} + (1 - \sigma_S) \mathbf{1}_{a_i = \text{Clean}}) & \text{if } |\neg S| \geq b/c \\ \mathbf{1}_{a_i = \text{nil}} \forall i \in \neg S & \text{if } |\neg S| < b/c, \end{cases} \quad (19)$$

where $\sigma_S \in (0, 1]$ is the probability to be determined. This preemptive response says: if S deviates and if the complementary coalition $\neg S$ is so small that the $\neg S$ -Pareto optimum is to all play Pollute, then $\neg S$ does not challenge S at all; if the size of $\neg S$ is bigger than the threshold, then each member of $\neg S$ joins the challenging coalition and has an equal probability of being the designated polluter; if $\neg S$ wins then the designated polluter commits to Pollute with probability σ_S and otherwise commits to Clean, and every other member of $\neg S$ commits to Clean.

We now calculate the minimum σ_S that deters S from deviation in the only nontrivial case where $|S| < b/c$ and $|\neg S| \geq b/c$. If S deviates, then with $|S| < b/c$, everyone in S tries to commit to Pollute, hence then the payoff for each of its member is

$$-c|S|\pi[S, \neg S] + (-c\sigma_S - b)(1 - \pi[S, \neg S]).$$

Thus, a member of S does not profit from the deviation if and only if

$$-c|S|\pi[S, \neg S] + (-c\sigma_S - b)(1 - \pi[S, \neg S]) \leq -b,$$

i.e.,

$$\sigma_S \geq \frac{\pi[S, \neg S]}{1 - \pi[S, \neg S]} \left(\frac{b}{c} - |S| \right). \quad (20)$$

By (17), the right-hand side of (20) is less than or equal to

$$\frac{1 - (c/b)|S|}{1 - c/b} \in (0, 1).$$

Thus, it is legitimate to define the probability

$$\sigma_S := \frac{\pi[S, \neg S]}{1 - \pi[S, \neg S]} \left(\frac{b}{c} - |S| \right). \quad (21)$$

Thus, we have constructed a coalitional response, defined by Eqs. (18), (19), and (21).

Finally, we remove the temporary assumption that the complementary coalition $\neg S$ is better-off challenging S than not challenging it at all. Consider a player $i \in \neg S$. If $\neg S$ does not challenge S at all, then i 's payoff is $-c|S| - b$. If the preemptive response defined above is followed, then i 's payoff is equal to either $-c|S| - b$ in the event that S wins or

$$\frac{1}{|\neg S|} \sigma_S (-c) + \left(1 - \frac{1}{|\neg S|} \sigma_S \right) (-b) > -c|S| - b$$

in the event that S loses. Thus, i prefers the preemptive response even without the temporary assumption. That completes the sketch of the proof for the theorem.

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.

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 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. [1](#)
- [3] Robert J. Aumann. The core of a cooperative game without side payments. *Transactions of the American Mathematical Society*, 98(3):539–552, Mar. 1961. [1](#)
- [4] 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. [1](#)
- [5] B. Douglas Bernheim, Bezalel Peleg, and Michael Whinston. Coalition-proof Nash equilibria I: Concepts. *Journal of Economic Theory*, 42(1):1–12, June 1987. [1](#)
- [6] R. H. Coase. The problem of social cost. *Journal of Law and Economics*, 3:1–44, October 1960. [1](#)
- [7] Geoffroy de Clippel and Roberto Serrano. Marginal contributions and externalities in the value. *Econometrica*, 76(6):1413–1436, November 2008. [1](#)
- [8] Gerard Debreu and Herbert Scarf. A limit theorem on the core of an economy. *International Economic Review*, 4(3):235–246, September 1963. [1](#)
- [9] Isa Hafalir. Efficiency in coalition games with externalities. *Games and Economic Behavior*, 61:242–258, 2007. [1](#)
- [10] Chen-Ying Huang and tomas Sjöström. Consistent solutions for cooperative games with externalities. *Games and Economic Behavior*, 43:196–213, 2003. [1](#), [3.4](#)
- [11] Kyle Hyndman and Debraj Ray. Coalition formation with binding agreements. *Review of Economic Studies*, 74:1125–1147, 2007. [1](#)
- [12] Diego Moreno and John Wooders. Coalition-proof equilibrium. *Games and Economic Behavior*, 17:80–112, 1996. [1](#)

- [13] Roger B. Myerson. Conference structures and fair allocation rules. *International Journal of Game Theory*, 9(3):169–182, 1980. [1](#)
- [14] Roger B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 1991. [4.2](#), [4](#)
- [15] Michele Piccione and Ronny Razin. Coalition formation under power relation. *Theoretical Economics*, 4:1–15, 2009. [1](#), [3.4](#)
- [16] Debraj Ray and Rajiv Vohra. Coalitional power and public goods. *Journal of Political Economy*, 109(6):1355–1384, 2001. [1](#)
- [17] Lloyd Shapley and Martin Shubik. On the core of an economic system with externalities. *American Economic Review*, 59(4):678–684, September 1969. [1](#)

Index

A_S , 6

A_S^{ind} , 11

A_i , 6

I , 6

$N(a_{(-T)\setminus S})$, 8

Δ , 6

α_{-T} , 9

$|T|$, $|\neg T|$, 6

$\mathcal{H}(T, a_T)$, 10

$\mathcal{H}_{\text{old}}(T, a_T)$, 12

\mathcal{P}_{T, a_T} , 8

\mathcal{R}_{T, a_T} , 8

$\neg S$, 6

a_S , 6

$u_i(a_T, \alpha'_S, a_{-(T \cup S)} \mid \mathcal{R}_{T, a_T})$, 9

$v_i(S, a_S)$, 11

action, 6

block, 4, 9

challenge, 6

characteristic function, 11, 12

coalition, 6

core, 4, 10

 traditional, 12

externality-free, 12

independent action, 11

Interdependency, 6

perfectly transferable utilities, 12

preemptive response, 4, 8

reactive response, 8

recursive principle, 3, 10

traditional core, 12