Spoilers, Blocking Coalitions, and the Core

Brennan C. Platt*†

August 6, 2008

Abstract

This paper surveys noncooperative implementations of the core which tell an intuitive story of coalition formation. Under the core solution concept, if a blocking coalition exists those agents abandon the current allocation without regard for the consequences to players outside the blocking coalition. Yet in certain circumstances, these players have an incentive to prevent formation of any blocking coalition; a game analyzed in Lagunoff (1994) is vulnerable to such circumstances. To obtain all core allocations and only core allocations, a mechanism must either restrict the actions of non-members of a proposed coalition, or ensure that non-members are unharmed by the departure of the coalition. These requirements illustrate the core's nonchalance towards agents not in blocking coalitions.

Journal of Economic Literature Classification Number: C71,C72.

Key Words: core, noncooperative implementation, coalitions

1 Introduction

Core allocations of an economy suggest a particular form of stability. If agents receive a core allocation, they have no compelling reason to seek an alternative allocation by breaking away from the rest of the economy (possibly in cooperation with a coalition of other agents), even when given full freedom to do so. Any feasible allocation proposed by the coalition would leave at least one member of the coalition weakly worse off. This state is loosely

*Department of Economics, Brigham Young University, 149 Faculty Office Building, Provo, UT 84602, brennan_platt@byu.edu, (801) 422-8904

†The author gives special thanks to Beth Allen and Andy McLennan for their direction, and gratefully acknowledges valuable discussions with Roger Lagunoff, Hugo Sonnenschein, John Ledyard, Myrna Wooders, Nuray Akin, as well as comments from participants at the 2004 Spring Midwest Economic Theory and the Society for Economic Design conferences and from anonymous referees. All errors remain my own. Partial funding came from National Science Foundation grants DMI-0070257 and DMI-0217974

described by Edgeworth (1881) as a "final settlement," from which no subset of agents is able to "recontract" to some other allocation.

The argument for stability is even more convincing when contrasted with non-core allocations. When an opportunity exists for a coalition to recontract — achieving a strictly preferred bundle for all agents involved — we naturally expect the agents to quickly take advantage of it. A non-core allocation seems unlikely to persist in an economy. Thus, the core seems to mark a dividing line between stable and unstable allocations, which is at least one of the reasons that economists are interested in this solution concept.

However, the core possesses some subtle features that cast doubt on this intuition. In a non-core allocation, we expect some blocking coalition to form, but we are not told what will occur to agents outside the coalition — the core is nonchalant towards these agents. Yet their fate could be bad enough that they will have an incentive to thwart the formation of the blocking coalition by luring away some of its members. If so, these agents (acting as spoilers) could cause a non-core allocation to persist indefinitely. In this paper, we present several such examples.

Of course, the core does not explicitly model the process by which coalitions form, or the means by which proposals to recontract are made, yet these are necessary details to determine how agents can react to various proposals. To model these strategic decisions by the agents, it is natural to turn to noncooperative game theory, where we can precisely specify the choices available to each agent, the order in which they interact, and the results of their choices. Considerable effort has been spent in developing mechanisms whose noncooperative outcomes coincide with the core (*i.e.* mechanisms that *fully implement* the core). There is no question as to whether the core can be implemented: viewed as a social choice correspondence, the core is Maskin-monotonic and satisfies no veto power — the sufficient conditions for Nash Implementation (see Maskin (1977), Maskin (1985), and Reny (1997)). Similar conditions ensure Subgame Perfect Implementation (see Moore and Repullo (1988) and Abreu and Sen (1990)).

Thus, the focus has been on developing "natural" mechanisms that depict the process of agents proposing a coalition and allocation, and, after approving it, breaking away from the rest of the economy. This body of research which provides noncooperative foundations for cooperative solution concepts is referred to as the Nash program. Some researchers have separated the Nash program from the broader theory of implementation because these mechanisms often require knowledge that a social planner is assumed not to have in implementation problems (such as the characteristic function or the agents' preferences). Serrano (1997) argues for an interpretation of the characteristic function which could reconcile this gap.[1] Without declaring this issue settled, I borrow the language of implementation theory throughout this paper's discussion of the Nash program. Section 4 provides a survey of this literature.

Among these natural implementations of the core, Lagunoff (1994) presents a mechanism with a concise and intuitive story. The economy begins at a status quo, which serves as a disagreement point. Each agent is sequentially given a single opportunity to propose a new coalition and allocation. If a particular coalition and allocation are proposed by all of its members, then the coalition forms; but any disagreement prevents formation. If any coalition forms, those who are left out of a coalition are pushed into a singleton coalition. If *no* coalition forms, all participants are left with the status quo. This story is particularly compelling because of its congruence with the recontracting described by Edgeworth, and allows us to investigate when the status quo is maintained (*i.e.* persists). This model is briefly reviewed in Section 2.

In this environment, the concept of fully implementing the core differs from that used elsewhere in the Nash program. The archetypical implementation would have a single non-cooperative game, with every equilibrium resulting in a core allocation and every core allocation obtainable as an equilibrium of that game. Instead, Lagunoff provides a *class* of games[2] meant to preserve the status quo if and only if it is a core allocation. The analogy

[1] Some work has approached noncooperative foundations fully within the framework of implementation theory, with appropriate limitations on the knowledge of the social planner. Examples include Kolpin (1989), Serrano (1995), and Bergin and Duggan (1999).

[2] The use of a class of games to reach the core is not entirely unique. Other mechanisms have relied on an exogenous probability distribution, only implementing the core when taking the union of all equilibrium outcomes across all probability distributions (Chatterjee *et al.*, 1993; Serrano, 1995; Yan, 2002, *e.g.*). That is, given a particular probability distribution, the mechanism may arrive at only some (not all) of the core

seems only a slight abuse, and facilitates comparison to other research on implementing the core; indeed, his work is cited in seven of the articles surveyed in Section 4. In the mechanism, the status quo is exogenous, and so there is no explanation as to how it arises.[3] In an economic environment, one could interpret this as an initial endowment bestowed by heredity or luck.

Lagunoff found that whenever the status quo allocation is in the core, it will be preserved as the outcome of some subgame perfect equilibrium of the game. However, it is also possible that the mechanism will preserve a non-core status quo in equilibrium, contrary to the claim in the main theorem. Indeed, in Section 3, several examples are presented where the mechanism is vulnerable to spoilers who can prevent the formation of blocking coalitions and thus maintain a non-core allocation as the equilibrium outcome. I also identify the error in the theorem's proof.

The counterexamples provided here are not disruptive for all implementations of the core. Section 4 examines the features of other core implementations that protect them from such spoilers. Contrasting these with Lagunoff's mechanism we find that to succeed at fully implementing the core, a mechanism must limit the ability of agents outside a proposed coalition from interfering with the proposal. This is because the core is nonchalant towards those not in a blocking coalition.

This feature can be replicated in a non-cooperative game by using some combination of the following methods: restricting information about previous or current proposals, or preventing non-members of a proposed coalition from acting until the members have either approved or rejected the proposal, or changing the payoffs of those who do not successfully enter a coalition. However, any mechanism in which agents can condition their strategies on the entire history and are able to interrupt proposals are potentially vulnerable to spoilers as in the counterexamples.

---

allocations.

[3] Serrano and Vohra (1997) provides a two stage mechanism, where the first stage is used to select a status quo while the second stage bears some resemblance to Lagunoff's mechanism. Section 4 discusses this work in further detail.

2 Overview of the Model

2.1 Environment

The economy has $m < \infty$ goods. There are n agents (so the set of agents is $N = \{1, \ldots, n\}$), each with a preference relation $\succcurlyeq_i$ defined on $\mathbb{R}^m$. These preferences are only required to be complete preorders.

An *allocation* z specifies a bundle of goods for each agent; *i.e.* z_i denotes the bundle received by agent i. The set of *feasible* allocations is given by $V \subset \mathbb{R}^{m \cdot n}$, which is assumed to be closed, compact, and convex.

For each coalition $C \subseteq N$, an *outcome characteristic function*[4], $v(C)$ specifies a non-empty, closed subset of $\mathbb{R}^{m \cdot |C|}$ indicating which bundles can be attained by members of that coalition. An allocation z is *coalition-feasible* for C if $z^C \in v(C)$, where z^C is the restriction of z to agents in C. It is assumed that $v(N) = V$, and that v satisfies *super-additivity*, which is that for all $C, D \subset N$ where $C \cap D = \emptyset$, $v(C) \times v(D) \subseteq v(C \cup D)$.

An economy E consists of a pair $(v, \{\succcurlyeq_i\}_{i \in N})$. Any economy that satisfies the restrictions above is said to belong to a set of economies, $\mathscr{E}$.

2.2 The Core

For any economy $E \in \mathscr{E}$, an allocation $x \in V$ is said to be *blocked by coalition* $C \subseteq N$ if there exists $z^C \in v(C)$ such that $z_i^C \succ_i x_i$ for all $i \in C$. The *core* of this economy consists of all feasible allocations that cannot be blocked by any coalition, and is denoted by $\mathscr{C}(E) \equiv \{x \in V : \forall C \subseteq N, C \text{ cannot block } x\}$.

[4]Note that this form excludes side payments, and is analogous to an non-transferable utility game.

2.3 Lagunoff's Mechanism

In accordance with the story, agents begin with an exogenous *status quo allocation*, x, and must decide whether to recontract by forming a coalition (understanding that they will receive the status quo otherwise). The particular method by which they may attempt to recontract is through a sequence of proposals[5], one from each agent. The proposals are made in a predetermined order, represented by a permutation $\phi : N \to N$, where agent i proposes $\phi(i)$th. The set Φ is composed of all possible permutations (*i.e.* the various orders in which proposals may be made).

Available actions are the same for all players at all nodes: each must choose both a coalition and an allocation for that coalition. The empty coalition is an acceptable proposal, and in effect will serve to ratify the status quo. Formally, the set of actions is $S^i = 2^N \times V$, and $S = \prod_{i \in N} S^i$.

For a coalition to form, each of its agents must propose the same coalition and the same coalition-feasible allocation. If no coalition forms, all agents are left with the status quo allocation. If any coalition forms, those agents excluded from all of the formed coalitions are placed into a singleton coalition. Such an agent will choose the best allocation available to him. For each player $i \in N$, define $\bar{z}^i$ such that for all $z^i \in v(i)$, $\bar{z}^i \succcurlyeq_i z^i$. (If there is more than one such $\bar{z}^i$, we can arbitrarily choose any of them).

Thus, for each profile of proposal $s \in S$, we can define the set of all coalitions that will form under s as $\mathscr{R}(s) = \{R \subseteq N : \forall i \in R, s^i = (R, z^R),\ z^R \in v(R)\}$. The *outcome function* $f : S \to V$ indicates what each agent receives under a particular proposal profile:

$$f(s) \equiv \begin{cases} z = \left(\left(z^R\right)_{R \in \mathscr{R}(s)},\ \left(\bar{z}^i\right)_{i \notin \bigcup_{R \in \mathscr{R}(s)} R} \right) & \text{if } \mathscr{R}(s) \neq \emptyset \\ x & \text{if } \mathscr{R}(s) = \emptyset \end{cases}$$

[5] Lagunoff refers to these as "votes," but the term "proposals" is more common in the literature.

2.4 The Noncooperative Game and Solution Concept

This mechanism constitutes a perfect-information, extensive form game. The *game* is defined by the elements $G \equiv (f, \phi, x, E)$ for any $E \in \mathscr{E}, x \in V,$ and $\phi \in \Phi$. Subgame perfection is used as the solution concept, in order to eliminate outcomes based on incredible threats. If proposals occur in the order of the agent's indices, $\phi(i) = i$, then the *history* for player i consists of all proposals up to i, *i.e.* $h^i = \left(s^1, \ldots, s^{i-1}\right)$. If ϕ indicates a different order, histories would be similar but would require more notation; this is omitted for simplicity. Denote H^i as the set of all possible histories. A *strategy* $\sigma^i : H^i \to S^i$ for player i maps from histories into proposals made by agent i.

A strategy profile $\sigma^* = \left(\sigma^{*1}, \ldots, \sigma^{*n}\right)$ is called a *subgame perfect equilibrium (SPE)* if for any i, h^i, and s^i, $f_i\left(h^i, \sigma^{*i}, \sigma^{*i+1}, \ldots, \sigma^{*n}\right) \succcurlyeq_i f_i\left(h^i, s^i, \sigma^{*i+1}, \ldots, \sigma^{*n}\right)$; *i.e.* the strategy satisfies sequential rationality. The resulting allocation, $f(\sigma^*)$, of a SPE σ^* is called a *subgame perfect outcome (SPO).*

2.5 Original Theorem

The result claimed by Lagunoff is that this mechanism *strategically separates* core allocations from non-core allocations. This is to say, if the status quo is a core allocation, it can be supported as an outcome of the mechanism in any proposal order. If the status quo is not in the core, it cannot be supported as an outcome for any proposal order.

Theorem 1 (Lagunoff, 1994)**.** *For each $E \in \mathscr{E}$ such that $\mathscr{C}(E) \neq \emptyset$, and each $x \in V$*

1. *if $x \in \mathscr{C}(E)$, then x is a SPO of $G(f, \phi, x, E)$ for all ϕ.*
2. *if x is a SPO of $G(f, \phi, x, E)$ for some ϕ, then $x \in \mathscr{C}(E)$.*

Note that the second claim is not merely the converse of the first; it is slightly stronger, and the two combined claims would imply that if x is a SPO for *some* proposal order, it is a SPO for *all* proposal orders. This stronger statement is needed to obtain strategic separation (described above in the contrapositive form); it would be unsatisfactory if non-core allocations are preserved for some (even if not all) proposal orders.

This result is what we might naturally have hoped for in this line of research; it provides a clear link between cooperative and non-cooperative concepts, using a relatively simple and intuitive mechanism. Unfortunately, the second claim does not hold for all admissible economies, as demonstrated in the following counterexamples.

3 Counterexamples

The counterexamples stem from two weaknesses in the mechanism, both of which are interesting in their own right. The first is surprisingly simple, and yet is the more serious of the two. Here, an agent (who is not in the blocking coalition) is afraid of being forced into a singleton coalition, and hence purposefully thwarts the formation of a blocking coalition.

The second weakness is more subtle. Here, an agent (again not in the blocking coalition) is indifferent between maintaining the status quo and allowing the blocking coalition to form, yet he is still able to thwart its formation by luring away members of the blocking coalition. The second counterexample illustrates this in it simplest form.

3.1 A Threatened Agent

Consider the following economy, E, consisting of one good and three agents with monotone preferences.[6]

[6] One could interpret this environment as three children playing catch with a baseball. A singleton gets no utility because he or she has no partner for playing. When paired with one friend, a child's possible utility is not affected by the utility of his friend — they are not in competition with each other. For whatever reasons, agents 2 and 3 are better players with each other (leading to higher possible utility) than either is with agent 1. When all three agents play together, there is some competition among the friends (as the throwing and catching need not be equally shared). Hence the cap on the sum of their utility.

$$v(\emptyset) = v(\{i\}) = \{0\} \qquad v(\{1,2\}) = [0,2] \times [0,2]$$
$$v(\{1,3\}) = [0,2] \times [0,2] \qquad v(\{2,3\}) = [0,3] \times [0,3]$$
$$V = v(\{1,2,3\}) = \{x \in [0,4] \times [0,4] \times [0,4] : x_1 + x_2 + x_3 \leq 6\}$$

Note that super-additivity holds, along with other assumptions required for $E \in \mathscr{E}$. Let $x = (1,4,1)$ be the status quo.[7] Note that the core is non-empty (*i.e.* $(0,3,3)$ is in the core); but $x \notin \mathscr{C}(E)$, since the coalition $\{1,3\}$ can block x using $(2,2) \in v(\{1,3\})$.

This admittedly stylized example is designed primarily for simplicity in expressing equilibrium strategies; the specific form of the characteristic function is not crucial.[8] Rather, what truly matters is (1) a blocking coalition exists that is not the grand coalition, (2) agent 2 would be harmed by any change from the status quo, and (3) agent 2 can offer more to agent 3 than the blocking coalition can. Because of these features, x can be supported as a SPO in the game where agents vote in the same order as their labels ($\phi(i) = i$), using the following SPE strategy (parenthetical clauses in the strategy mean "if applicable"):

Let $h^3 \equiv ((C_1, y), (C_2, z)), h^2 \equiv ((C_1, y))$ be used for notation of histories.

$$\sigma^{*3}\left(h^3\right) = \begin{cases} (C_2, z) & \text{if } C_2 = \{2,3\} \text{ and } z_3 \geq 1 \text{ (and } z_3 > y_3) \\ & \quad \text{or } C_1 = C_2 = \{1,2,3\}, y = z, \text{ and } z_3 \geq 1 \\ (C_1, y) & \text{if } C_1 = \{1,3\} \text{ and } y_3 \geq 1 \text{ (and } y_3 \geq z_3) \\ (\emptyset, x) & \text{otherwise} \end{cases}$$

$$\sigma^{*2}\left(h^2\right) = \begin{cases} (\{2,3\}, (3,3)) & \text{if } C_1 = \{1,3\} \text{ and } y_3 \geq 1, \text{ or } C_1 = \{1\} \\ (\emptyset, x) & \text{otherwise} \end{cases}$$

$$\sigma^{*1} = (\emptyset, x)$$

Since agent 3 proposes last, she can at most accept or reject the offers made by agent 1 or 2, if any — by the time her turn arrives, the proposals of 1 and 2 are fixed. Thus, she

[7] *e.g.* child 2 is dominating play time among the three.

[8] In fact, the same example can be written with a v that resembles transferrable utility (summing to 3, 3, and 4 for the respective coalitions) except for at the grand coalition. The equilibrium strategy would just be more complicated.

has up to four options: exactly match (C_1, y), exactly match (C_2, z), propose $(\{3\}, 0)$, or propose $(\emptyset, x)$. Her strategy directs her to choose the best response among these options. At nodes where she is indifferent between two options, if the chosen strategy were altered to some other best response, it would still be a subgame perfect equilibrium with the same outcome.

Agent 2 has a payoff of 4 in the status quo; yet if the blocking coalition were to form, he would be stuck with a singleton payoff of 0. Thus threatened, he has a strong incentive to disrupt this blocking coalition. He can do this by "tempting away" agent 3, offering her strictly more than she gets in both the status quo and the blocking coalition. However, agent 2 only obtains a payoff of 3 in this tempting coalition, so he only offers it when he must; *i.e.* when the blocking coalition would otherwise match. At nearly all nodes, the action prescribed under σ^{*2} is strictly preferred. The only exception is when $(\{1, 2, 3\}, x)$ is proposed, in which case agent 2 is indifferent; yet if the strategy were altered to have him ratify such a proposal, it would still be a SPE with the same equilibrium outcome.

As agent 1 selects his strategy, he realizes that proposing the blocking coalition would actually cause the tempting coalition to match, leaving him with a payoff of 0. The singleton proposal would also give him a payoff of 0, so both of these are strictly dominated by the status quo proposal. Any $\{1, 2\}$ proposal will be rejected by agent 2. Thus, agent 1 has no choice but to propose the status quo (or any other strategy that will not trigger a match). Again, he could propose $(\{1, 2, 3\}, x)$ instead, but even if agent 2 were to match it, it results in the same outcome.

Thus, agent 2 has skillfully defended the status quo by offering just the right temptation, thereby preventing the formation of the blocking coalition. Yet remarkably, his tempting coalition does not match on the equilibrium path either. Thus the non-core status quo will persist as the outcome of this game.

Indeed, one can quickly verify that x is the *unique* subgame perfect outcome in this voting order (even though it can be supported by a number of different subgame perfect equilibrium strategies). Notice also that indifference plays no role in this outcome; indeed, counterexamples of this flavor are generic, still arising even when payoffs are perturbed.

3.2 An Indifferent Spoiler

Consider the following economy, E', which is a slight modification of E.

$$v(\emptyset) = v(\{i\}) = \{0\} \qquad v(\{1,2\}) = [0,2] \times \{0\}$$
$$v(\{1,3\}) = [0,2] \times [0,2] \qquad v(\{2,3\}) = \{0\} \times [0,3]$$
$$V = v(\{1,2,3\}) = [0,4] \times \{0\} \times [0,4]$$

Note that $E' \in \mathscr{E}$. Here, agent 2 always get a payoff of 0; yet his presence in a coalition can be beneficial to the other agents. Let $x' = (1,0,1)$, which is not in the core of E'; $\{1,3\}$ can block the allocation. Yet x' can be supported and a SPO using the following strategy:

$$\sigma^{*3}\left(h^3\right) = \begin{cases} (C_2, z) & \text{if } C_2 = \{2,3\} \text{ and } z_3 > 1 \text{ (and } z_3 \geq y_3\text{)} \\ & \quad \text{or } C_1 = C_2 = \{1,2,3\}, y = z, \text{ and } z_3 > 1 \\ (C_1, y) & \text{if } C_1 = \{1,3\} \text{ and } y_3 > 1 \text{ (and } y_3 > z_3\text{)} \\ (\emptyset, x') & \text{otherwise} \end{cases}$$
$$\sigma^{*2}\left(h^2\right) = \begin{cases} (\{2,3\},(0,3)) & \text{if } C_1 = \{1,3\} \text{ and } y_3 > 1 \\ (\emptyset, x') & \text{otherwise} \end{cases}$$
$$\sigma^{*1} = \left(\emptyset, x'\right)$$

Agent 2 is indifferent among all his strategies since he will get a payoff of 0 regardless of the actions of others. But, by conditioning his strategies on the proposal of agent 1, he can prevent the blocking coalition from matching — whenever it becomes a possibility, he tempts away the third agent. He has no particular reason to do this, but nothing prevents him from it either. He is, in this sense, purely a spoiler.

Thus, agent 1 realizes that a {1,3} coalition will never match, given 2 and 3's strategies, and in fact, proposing such a coalition will lead to a {2,3} match, and leave agent 1 with his singleton coalition payoff. Since he strictly prefers the status quo payoff to the singleton payoff, he always votes for the status quo.

3.3 The Error in Claim 2 of the Theorem

The mechanism crucially relies upon the fact that proposals occur in sequential order and are public information. By observing prior proposals (and trusting in the credible choices of those who follow), an agent is able to coordinate with others to form mutually beneficial coalitions. Without these sequential proposals, we could only ensure that agents will obtain individually rational allocations, not necessarily core allocations (as was the case in the mechanism developed by Kalai *et al.* (1979)).

However, as the counterexamples show, this system of sequential proposals can be abused by a spoiler. Agents can condition their strategies on prior proposals, which permits a spoiler to offer a tempting coalition at certain nodes (such as when a blocking coalition would otherwise form) and withhold such an offer at other nodes.

The proof of the theorem's second statement errs because it fails to fully appreciate the sequential nature of the proposals. The proof proceeds by contradiction, assuming that some status quo x is supported as a SPO, and yet can be blocked by a coalition A. The blocking coalition did not form on the equilibrium path, since the status quo is the outcome. Lagunoff then correctly observes that, under this equilibrium strategy profile, if an agent were to deviate and propose the blocking coalition, it must cause some other coalition B to match which steals away some members of A.

Lagunoff then argues that this cannot happen, because if it did, B would have matched on the equilibrium path, and the agents $i \in B \cap A$ must still receive something strictly better than the status quo: $z_i^B \succcurlyeq_i z_i^A \succ_i x_i$. But herein lies the error: B need not match unconditionally.

It was apparently presumed that, if B is chosen by its members when A was proposed, B should also be chosen when A was not proposed. Yet this is not necessarily true, since we might have some agent $j \in B/A$ for whom $x_j \succcurlyeq_j z_j^B \succcurlyeq_j \bar{z}^j$. This is precisely the case with the spoilers in our counterexamples. Such an agent can credibly propose (z^B, B) conditional on A having been proposed by an earlier agent.

Hence, the proof does not analyze the case of an economy such as in the counterexamples.

3.4 Discussion of the Counterexamples

Knowing that these exceptions to the theorem exist, we might ask, how common are they? The second counterexample relied on the indifference of agent 2 to allow him to act as a spoiler; hence, one would rightly guess that such situations are not generic. However, in the first counterexample, our spoiler strictly preferred the status quo to both the tempting allocation and what he would have received as a singleton; hence, his behavior would be robust to small perturbations of payoffs.

One may also observe that both of the preceding economies only provide a counterexample under a couple of the proposal orders. The threatened agent could not credibly offer the tempting allocation if he proposed last (this is not a problem for the indifferent spoiler, however). In both counterexamples, agent 2 could not conditionally offer the tempting allocation if he were required to propose first — the spoiler would either offer the tempting coalition, which would subsequently match, or he would not, and the blocking coalition would form. Either way, x or x' are not longer supportable as a SPO in these proposal orders. This suggests that perhaps a counterexample can occur in only some (but not all) of the proposal orders.

In actuality, the problem is more general than that. I have also constructed a counterexample (presented in the appendix) in which it is possible to preserve the same non-core status quo in any proposal order.[9]

[9] This requires two spoilers to work together, but it utilizes similar techniques of luring away members of the blocking coalition whenever it is in danger of forming.

We might also ask, for what economies does the theorem hold; that is, what are the weakest assumptions we could add to rule out any spoilers? Below, we present conditions for a three-agent game (all one- and two-agent games abide by the theorem without additional assumptions).

Assumption (Spoiler-Proof)**.** *Let $N = 3$, and let $x \in V$. If x is not in the core, then either x can be blocked by a singleton coalition, or there exists a non-singleton blocking coalition and allocation (C, z^C) such that for every two-agent coalition and allocation (D, z^D) where $D \neq C, z^D \in v(D)$:*

- *either $z_i^C \succ_i z_i^D$ for all $i \in C \cap D$*
- *or $\bar{z}^j \succ_j z_j^D$ for some $j \in D$*
- *or $z_j^D \succ_j x_j$ for all $j \in D$*

This assumption restricts what agents outside of the blocking coalition are capable of offering, relative to the status quo and to the blocking allocation. If the first condition holds, none of the members of the blocking coalition can be tempted away. If the second condition holds — in particular, for a spoiler not in the blocking coalition — then the spoiler cannot credibly offer this proposal. The first two conditions can still permit a tempting allocation; however, the third condition ensures that if the blocking coalition does not form due to the temptation, the tempting coalition will form on the equilibrium path. (It would be accepted if proposed, so proposing it would be strictly better than using a strategy that leads to the status quo.)

The Spoiler-Proof assumption is sufficient to ensure that the second claim of the theorem holds for three-agent games. This can be proven by assuming x to be a SPO, then working backwards from the last agent to rule out the possibility of blocking coalitions (based on the fact that each agent did not propose any such coalitions).[10] Any analogous

[10]It is possible to weaken this assumption, though the conditions would become significantly more complex. Even so, many of the economies excluded by the Spoiler-Proof assumption will yield a counterexample, which can be constructed by using the previous ones as a guide.

Spoiler-Proof assumption for games with more than three agents is certain to be significantly more complex.

Even so, this should not be viewed as an entirely satisfactory remedy for the troubles created by the counterexamples. The assumption rules out certain economies in which the mechanism will not perform as intended; however, these conditions are not easy to verify. To do so would require complete knowledge of the preferences and the feasibility for all agents. Of course, this non-cooperative game implicitly assumes that all agents know the preferences and feasibility of the other agents. However, a regulator or an impartial observer of the game might not have such information, yet would like to conclude, based on the observed outcome, whether the economy's status quo belonged to the core. One of the attractive features of Lagunoff's paper is that with very weak assumptions (*i.e.* very little knowledge about agents preferences and feasibility), a regulator could have concluded that if a status quo is not overthrown (recontracted) as a result of the non-cooperative game, it must be in the core. But in light of the error in Claim 2, that conclusion can only be confidently made in economies where the Spoiler-Proof Assumption holds — which requires much more knowledge on the regulator's part.

4 Other Mechanisms Fully Implementing the Core

Lagunoff (1994) is only one of several implementations of the core. One might wonder if the other mechanisms are vulnerable to such spoilers. The answer is no. Indeed, by comparing why other mechanisms are spared while this one is foiled, we underscore an important lesson about the core: it is nonchalant towards those agents not in the blocking coalition.

This issue is briefly mentioned by Mas-Colell: "A weakness of the Core as a solution concept in economics and game theory is that it depends on the notion that when a coalition objects to a proposed allocation, *i.e.* engages on an improving move, it neglects to take into account the repercussions triggered by the move" (Mas-Colell, 1989, p. 129). Greenberg (1990, p. 66-69) provides additional discussion of the issue. He concludes that the core describes the result of a negotiation process where forming a coalition S commits those members to never negotiate with players outside of S, although members of S may renegotiate among themselves.

These comments strike at the heart of what gave our counterexamples their potency. When a coalition is proposed, the core suggests that it will be approved so long as all of its members are strictly improved. It does not consider how those outside of this coalition will be affected by the departure of these agents, nor the incentive that this gives non-members to prevent formation of the blocking coalition.

Lagunoff's mechanism provides these non-members an opportunity to *interrupt* the formation of a blocking coalition by tempting away some members. Yet the tempting proposal is only offered *conditionally*, hence neither the tempting nor the blocking coalitions form in equilibrium. The key features of the mechanism which enable the spoilers in this way are:

1. The *disagreement point* — what agents receive if no coalition successfully forms.
2. The *outcome of non-members* — what is given to those who are not part of the blocking coalition, should it form.
3. The *strategic involvement of non-members* — while a coalition is under consideration, how much opportunity non-members are given to interrupt.
4. The *memory of spoilers* — whether agents are able to make offers conditional on all of the previous history.

The particular treatment of these four issues will determine whether a mechanism is vulnerable to spoilers.

To begin, we should note that Lagunoff (1994) is unique in the first point. His mechanism sets a status quo as the disagreement point, with the intent to see if the mechanism preserves the status quo if and only if it is a core allocation. This seems appropriate in addressing the issues addressed by Edgeworth (1881), which asks whether agents will overturn some precedent by recontracting. The rest of the literature, in contrast, does not start the game with any precedent; rather, agents "start from scratch" and must decide with whom they wish to cooperate. The intent is to see if the mechanism can induce agents to choose any core allocation and only core allocations.

As we survey other implementations of the core, it may not be possible to precisely map the environment of one paper into the other. Some of these papers assume that underlying cooperative game to have transferable utility; others use non-transferable utility. Lagunoff's environment is akin to non-transferable utility, with v indicating feasible bundles of goods rather than feasible utilities.[11] However, the counterexamples are not merely an artifact of using a particular specification of the cooperative game; indeed, we have created TU counterexamples. Rather, the mechanism does not fully correspond to the incentives in the core solution concept. Thus, the pertinent question is, would other mechanisms run into similar difficulties with spoilers in one way or another?

4.1 Two stage games

Pérez-Castrillo (1994), Serrano (1995), Serrano and Vohra (1997), and Bergin and Duggan (1999) proceed in two stages: in the first, a status quo is selected, and in the second, agents have an opportunity to form a coalition or simply accept the status quo. At first glance, these seem to simply add a pre-game to Lagunoff's mechanism; yet the structure of the second stage has critical differences that spare them from spoilers.

The mechanism in Serrano and Vohra (1997) allows the least amount of strategic involvement from non-members. In the first stage of the game, agents simultaneously propose a status quo and an order for agents in stage two. The disagreement point for stage one is to force agents into singleton coalitions. In stage two, the first agent proposes a coalition and allocation. Then, in order, each of the members of the coalition approve or reject the proposal. Any rejections leads all agents to receive the status quo; if approved, the members

[11] As long as the preferences have utility representation, his environment could be translated into a typical NTU game.

receive the chosen allocation, while others are placed in singleton coalitions.

This mechanism always arrives at core allocations. A spoiler has no opportunity to conditionally offer a tempting coalition, because if the blocking coalition is proposed at the beginning of stage 2, the non-member will not play at any later nodes. For such an agent, her only opportunity to influence the outcome is in the first stage, playing an unconditional strategy. She might affect which core allocation is reached, but she cannot exploit the mechanism to reach a non-core allocation.

The initial stage of the Bergin and Duggan (1999, Section 6) mechanism[12] again features a simultaneous proposal from all agents, but with two dimensions. Everyone must choose a status quo allocation for the grand coalition, but they also have an option to propose a separate coalition and accompanying allocation. If they exercise that option, they also submit a schedule stating a (distinct) time for each coalition member to accept or reject the proposal. If any coalition proposals were made in stage 1, then only the proposal with the earliest scheduled ratifier is evaluated.

If all members of the proposed coalition accept, they receive that payoff and the non-members simultaneously choose a replacement status quo, feasible for their coalition. If any member rejects the proposal or if no proposals were made, a default rule is triggered, and the same rule applies to non-members if a coalition matched. The rule is: if two or more agents (or non-members, when a coalition has matched) disagreed in their choice of status quo, an exogenous status quo is imposed; otherwise, the (almost unanimously) chosen status quo is imposed. As in the Serrano and Vohra mechanism, at most one coalition proposal will be evaluated, and after a proposal begins to be evaluated, there are no nodes where non-members make choices.

In Serrano (1995), one of the agents is randomly[13] selected as a "broker" who selects

[12]Their mechanism was built to accommodate cooperative games where one coalition may impose externalities on another. While that goes beyond the scope of the other mechanisms analyzed here, the mechanism still fully implements the core for cooperative games where a coalition's characteristic function is independent of choices made by non-members.

[13]To obtain a particular core allocation may require a particular probability distribution over the agents; so here, implementation means that the core is equal to the union of all subgame perfect outcomes over all probability distributions.

an allocation x that is feasible for the grand coalition. x_i can be thought of as a market price for buying the participation of agent i in a coalition. In the second stage, each of the remaining agents respond in an exogenous sequential order, choosing one of two options: they can accept, obtaining a payoff of x_i, or they can decline and instead propose to "buy" a coalition B_i at the same prices listed in x.

The members of B_i do not directly respond to this proposal. Rather, at the end of the game, proposals are evaluated in the same order (except for the broker, who ends up with a coalition of all leftover agents). The proposal B_i will match if all of its members accepted x_i and are not part of a previous match. Agents who declined and failed to match are left in a singleton coalition. Therefore, anyone who accepted will obtain x_i, regardless of which coalition they entered. Anyone who proposed a coalition that matched will receive the value of that coalition (plus himself), minus the allocation promised in x to the other coalition members.

Note that in the mechanisms of Lagunoff, Serrano and Vohra, and Bergin and Duggan, the status quo allocation is only received if coalitions do not form. In this brokered market of Serrano, the allocation selected in the first stage is a guarantee to the agents. They will receive that payoff whether as part of a proposed coalition or the broker's coalition of leftovers, and will only decline if this leads to a higher payoff. Thus, even if x is not in the core, agents outside of the blocking coalition are not threatened by its formation. Serrano also defuses the possibility of indifferent spoilers by imposing an assumption that whenever an agent is indifferent between accepting or declining, he will accept. Under this assumption, any core allocation proposed by the broker will be accepted by all, and any non-core allocation will be declined by someone. Given this, Serrano proves that from the broker's perspective, the proposal of any non-core allocation is always strictly dominated.

Pérez-Castrillo (1994) has a similar feel, except that the noncooperative game is played by principals whose role is to set up coalitions among the agents (only agents are members of the underlying cooperative game). In the initial stage, each principal e simultaneously announces an allocation w^e (which need not be feasible) for all of the agents. In the second stage, each principal has an opportunity to form a coalition, proceeding in an exogenous order. A principal *must* include any agent i for which his offer w_i^e was strictly larger than all others; he *cannot* include any agent for which his offer was strictly less, or who was selected by an earlier principal. For all other agents (those who are unmatched and for which his offer is tied as the largest), the principal has the liberty to include them or not. The last principal receives all unclaimed agents. The principal's payoff is the value of his assembled coalition minus the allocation offered by the principal to his coalition's members.

The subgame perfect equilibrium of this game has all principals offering the same allocation and receiving a payoff of zero. Thus, in equilibrium, the offered allocation is in fact feasible (though, off the equilibrium path, it is possible for the principals to have a negative payoff, which corresponds to an infeasible allocation). If the characteristic function is super-additive, the set of equilibrium allocations equals the core.

4.2 Sequential proposal games

Okada (1992), Chatterjee *et al.* (1993), Moldovanu and Winter (1994), Evans (1997), Yan (2002), and Horniaček (2007) share much in common. These mechanisms allow for up to an infinite number of discrete stages, each one providing agents with an additional opportunity to agree on a coalition and allocation. In all of these mechanisms, a payoff is only obtained once a coalition has been formed.[14]

Each stage begins by selecting a proposer[15] from among the agents, who announces a coalition and allocation. In sequence,[16] each member of that coalition must approve or reject it. Quite significantly, no one outside the proposed coalition has an opportunity to move within the stage.

[14] Only in Chatterjee *et al.* and Yan are the the eventual payoffs discounted by the number of stages used to reach the outcome; they obtain full implementation as the discount factor approaches one. The others have no discounting.

[15] This may be assigned by an exogenously-fixed rotation (Chatterjee *et al.*; Horniaček), by competition through an all-pay bid (Evans), or by random selection.

[16] Usually this follows an exogenous order, though Moldovanu and Winter have the proposer or the preceding approver select who gets the next opportunity to approve.

If all of the coalition members approve, the coalition forms and receives its agreed upon payoff.[17] If any member of the coalition objects, a new stage begins with the objector as the proposer.[18] The bargaining may continue infinitely, in which case all unmatched agents receive their singleton payoff.[19]

Each of these mechanisms impose some form of *stationarity* in equilibrium strategies. This means that at each node where an agent faces the same proposal and with the same players still in the game, he should choose the same action. To be clear, the agent can still consider unilateral deviations from stationary strategies while selecting a best response at a given node; but he will take as given the stationary strategies of other agents and *even himself* at later nodes. This equilibrium refinement greatly reduces opportunities for spoilers by diminishing the memory of agents.

Suppose a blocking coalition is under consideration. A spoiler would need to tempt away one of its members, which can only be done based on a credible promise to propose something advantageous next round. A spoiler would want to offer this tempting allocation only when the blocking coalition has been proposed *and* rejected in a previous round. However, stationarity requires that the same offer must be made in any similar situation; the tempted agent must assume that the spoiler will not distinguish between a node where the blocking coalition has been rejected and any other nodes in which no coalition was formed.

Thus, any tempting must be done within the same stage that the blocking coalition is proposed. However, as previously noted, this is not possible since non-members of the proposed coalition are not permitted to act during the stage. The combination of stationarity between stages and non-member exclusion within stage completely diffuses any opportunity for spoilers. Thus all non-core proposals will be rejected because a blocking coalition can credibly form in a future stage.

[17] In Yan and Moldovanu and Winter, the game then concludes for everyone, forcing non-members in singleton coalitions. The other mechanisms permit the non-members to continue the same game among themselves in the next stage. Evans subtracts the agent's bids from his payoff.

[18] Or a new random selection or competition occurs in the cases of Yan or Evans, respectively.

[19] In Moldovanu and Winter the punishment is more extreme, setting the infinite delay payoff to $-\infty$. Evans assigns a singleton payoff minus the agent's bids.

4.3 Continuous time proposal game

The mechanism in Perry and Reny (1994) (also summarized in Reny, 1997) is modeled in continuous time. Rather than proposing in a specified order, agents can choose to make a proposal or be quiet at any time. Only one active proposal can be considered at a time; any agent can replace the active proposal by proposing something else. If all of the members of a proposed coalition match the proposal before it is replaced, it becomes a binding proposal. There may be more than one binding proposal, although each agent can only be in one binding proposal. Also, a binding proposal may be revised either to expand the coalition or change payoffs, but all members of the coalition would need to propose the revised proposal, and binding coalition members cannot switch to a different coalition. A member of a binding proposal can choose to exit at anytime, which forces the entire coalition to exit the game and receive the agreed-upon payoff. Those who never exit the game are forced into their singleton coalition on the infinite horizon.

At first glance, this mechanism provides significant opportunity for strategic involvement by non-members. Anyone can interrupt the acceptance of an active proposal by replacing it just a little bit earlier than it is accepted. Of course, the members of the proposed coalition can also accept sooner; so the interruption can only be an equilibrium strategy if the new proposal is weakly better for a member of the replaced coalition, *i.e.* if someone can be tempted.

However, this does not mean that a spoiler can thwart the formation of blocking coalitions and cause the mechanism to yield a non-core outcome. These hazards are entirely defused by stationarity. If a tempting proposal would match after the blocking coalition is interrupted, we are assured that the same tempting proposal would also match if proposed *before* the blocking coalition was proposed. The agent lured out of the blocking coalition will have an incentive to making this earlier proposal, since he strictly prefers the tempting proposal to the non-core outcome (See Perry and Reny, 1994, p. 808). Interestingly enough, this is precisely the reasoning that failed in Lagunoff's second claim. Here it succeeds because stationarity restricts the information on which agents may condition their strategies.

4.4 Comments

Although these mechanisms that successfully implement the core employ different methods to accomplish it, their common thread is neutralizing the impact of non-members on possible blocking coalitions. In Pérez-Castrillo (1994) and Serrano (1995), this is done by guaranteeing the same outcome for non-members regardless of which coalitions form. In Serrano and Vohra (1997) and Bergin and Duggan (1999), non-members are denied the opportunity to act once a coalition has been proposed. This also occurs within each stage of the sequential proposal games, along with restrictions on how an agent is able to condition his or her strategy on prior histories.[20] The necessity of such conditions illustrates how the core, as a solution concept, is nonchalant towards agents that do not belong to the blocking coalition. The core ignores how they will be impacted if the blocking coalition secedes from the grand coalition.

Two strands of the literature warrant mention before concluding. First, there is a considerable body of work relating walrasian equilibria of competitive economies with the core. Anderson (1992) provides a thorough survey of this work. In many instances, competitive markets are an effective mechanism to implement the core, though this typically relies on large numbers of agents, as opposed to the bargaining framework commonly used in the Nash program.

Second, some work in the Nash program has been directed towards situations in which one coalition may impose externalities on another. Feasibility for a coalition is no longer defined only by its membership but by the entire partition of the set of players and possibly their joint actions. To define the core in this environment, we must explicitly make an assumption about how the complementary set of players react to the departure of the blocking coalition. For instance, the α-core assumes that if a coalition leaves and selects a particular allocation, the remaining non-members will choose an allocation that causes the greatest harm to the departing coalition. The β-core is somewhat more optimistic,

[20] Okada and Winter (2002) provide an axiomatization of sequential proposal games that implement the core. It is noteworthy that they found subgame consistency (a generalization of stationarity) to be a necessary element for a game to succeed in implementation.

assuming that the non-members must commit to their punitive action first, and then the departing coalition can choose a best response to it. For examples of non-cooperative games which study these environments, see Bloch (1996), Yi (1997), Bergin and Duggan (1999), and Chander (2007).

At first glance, the α- or β-core seem related to the spoilers in the counterexamples presented here, since members of a (potential) coalition might be affected by the response of non-members. The difference is that these adaptations of the core are concerned with how non-members respond *after* the coalition has agreed to leave, while the present survey has considered how potential non-members might respond *while* the coalition is forming.

A more closely related cooperative solution concept to consider for these issues is the bargaining set (Aumann and Maschler, 1964; Serrano and Vohra, 2002) or the consistent bargaining set (Dutta *et al.*, 1989). These require that if any agent raises an objection (*i.e.* by indicating that he could form a blocking coalition), there should exist a counter-objection (*i.e.* an alternative coalition that excludes the original objector and provides at least as good an outcome to all its members as the original allocation and as the blocking coalition, if applicable). This is to say the threat of this blocking coalition can be ignored if one of the non-members can maintain his same allocation by assembling another coalition. The consistent bargaining set extends this analysis to chains of objections / counter-objections.

Both concepts explicitly consider how non-members might try to lure away members of a (would-be) blocking coalition. The concept of counter-objections have some resemblance to the spoiler-proof assumptions provided in Section 3. However, the counter-objections are broader than what we would consider (making the Bargaining Set larger than the set of SPO for Lagunoff's mechanism). Our counterexamples only occur when the objecting coalition has a non-empty intersection with the counter-objecting coalition.

5 Conclusion

The intuitive story of the core suggests that non-core allocations should not persist over time. When agents are given an opportunity to recontract among themselves, those who can strictly improve on their allocation should join together and do so. The model presented by Lagunoff seems to effectively capture this story; yet we have seen several instances in which blocking coalitions will not form, in spite of a non-core status quo.

There are several directions for addressing these counterexamples. The easiest option is to use the same mechanism with stronger assumptions on the economies considered, such as with a Spoiler-Proof assumption, to rule out the counterexamples. Unfortunately, the needed assumptions are quite restrictive and hard to verify without full knowledge of preferences and coalition feasibility.

Another approach is to adopt a mechanism which restrict the actions of non-members of a proposed coalition, as in successful implementations in the literature. Of course, this alters the story, either restricting how strategies may condition on previous proposals or preventing them from offering alternative proposals while one is being considered.

A final alternative is to use the same mechanism with an equilibrium refinement to eliminate the counterexamples. Indeed, with indifferent spoilers, this is an easy fix. Small perturbations of payoffs will disrupt the indifference; the spoiler will either offer the tempting proposal unconditionally, or he will withhold the offer unconditionally and the blocking coalition will form.

However, counterexamples caused by a threatened spoiler are not so easily eliminated. If he strictly prefers the status quo over any other feasible outcome, he will have a strong incentive to disrupt a blocking coalition from forming, even when payoffs are perturbed by small amounts. Other refinements, such as elimination of weakly dominated strategies, will also be insufficient due to the strict preference of the spoiler for the status quo.

The only approach the author has found to eliminate this incentive (while still allowing non-member agents to interrupt the formation of a coalition) is to change the outcome function such that all unmatched agents receive their status quo allocation rather than being forced into a singleton coalition. Under this new outcome function, an otherwise "threatened agent" has no reason to undermine a blocking coalition, since he will receive the same outcome regardless of its formation. However, this adaptation has the unpleasant effect that some action profiles lead to an outcome that is not feasible (or alternatively, feasibility must be restricted so that the status quo can be obtained as a singleton). This criticism is certainly valid, particularly since it pulls us away from the intuitive story that we hoped to maintain.[21] While this does not provide a satisfactory implementation of the core, it does sharply illustrates the nonchalance of the core towards those not in the blocking coalition: it treats them as if they will be left with their status quo outcome.

Indeed, this same lesson can be learned from each of these methods of correcting the theorem: *the core does not consider how agents outside a coalition will be affected by the coalition's departure*. To successfully implement the core, either the mechanism must deny these outsiders any opportunity to interrupt a proposed coalition, or one must assume that the outsiders will not be adversely affected. Otherwise, it is plausible that non-core allocations can be sustained over time based on the credible threats of a spoiler.

This is particularly troublesome when considering applications of coalition formation, such as in legislative caucuses. The process of making, considering, and accepting proposed coalitions is likely to be informal and fluid; it hardly seems appropriate to assume strong limitations on the ability to interject alternate proposals at any time, or require that strategies to be stationary. Consequently, a legislature could maintain a non-core allocation indefinitely due to the efforts of a spoiler who lures away members of the would-be blocking coalition. If so, we should not consider the persistence of the status quo as evidence that it is in the core.

[21] In an earlier version of this paper, the author presents a mechanism which combines this altered outcome function and a particular perturbation of payoffs. This altered mechanism successfully achieves the result intended by Lagunoff; status quo allocations will be preserved if and only if they are in the core. However, because of the lack of feasibility of some outcomes, this cannot be considered a full implementation of the core, at least not in the standard sense.

Appendix: An Order-independent Counterexample

The following economy demonstrates that a non-core allocation x can be supported as a SPO regardless of the order in which proposals occur.

Consider the economy, E, consisting of one good and four agents. Again, we assume monotone preferences.

$v(\emptyset) = v(\{i\}) = \{0\}$	$v(\{1,i\}) = \{0\} \times [0,1]$ for $i = 2,3,4$
$v(\{1,i,j\}) = \{0\} \times [0,1] \times [0,1]$ for $i,j = 2,3,4$	$v(\{i,j\}) = \{0\} \times \{0\}$ for $i,j = 2,3,4$
$V = v(\{1,2,3,4\}) = \{0\} \times [0,1] \times [0,1] \times [0,1]$	$v(\{2,3,4\}) = [0,1] \times [0,1] \times [0,1]$

Note that super-additivity holds, along with other assumptions required for $E \in \mathscr{E}$. Let $x = (0,0,0,0)$. $x \notin \mathscr{C}(E)$, because the coalition $\{2,3,4\}$ can block this allocation with $(1,1,1) \in v(\{2,3,4\})$. But x can be supported as a SPO of $G(f, \phi, x, E)$ for any ϕ.

Agents 2, 3, and 4 are interchangeable in this game; thus, any strategy we construct as a SPE under one proposal order can also be used when the order of these three agents are swaped. Of course, if we change agent 1's position in the proposal order, we will have to use a different strategy. Thus, we only need to examine four (of 24 possible) proposal orders.

In proposal orders where agent 1 does not propose first, x can be supported using a strategy similar to those used in the indifferent spoiler counterexamples. That is, if the first agent in the proposing order suggests the blocking coalition, agent 1 can propose a coalition with either of the remaining two members of $\{2,3,4\}$ which makes that member indifferent between the two coalitions. Hence, this tempted agent can credibly match 1's proposal, and the first proposer is dissuaded by this subgame perfect strategy from proposing a blocking coalition. All other coalitions of importance require the participation of agent 1, which he can credibly refuse to give.

The more difficult case is when agent 1 moves first. Here, agent 1 must set an unconditional strategy that somehow dissuades the blocking coalition from forming. Without loss of generality, take the case where agents vote in the same order as their labels ($\phi(i) = i$). Let $h^4 \equiv ((C_1, w), (C_2, y), (C_3, z))$, $h^3 \equiv ((C_1, w), (C_2, y))$, $h^2 \equiv ((C_1, w))$ be used for notation of histories.

$$
\sigma^{*4}\left(h^4\right) = \begin{cases}
(\{1,2,3,4\}, w) & \text{if } C_1 = C_2 = C_3 = \{1,2,3,4\}, z = y = w, \text{ and } w_4 > 0 \\
(\{1,3,4\}, w) & \text{if } C_1 = C_3 = \{1,3,4\}, z = w, \text{ and } w_4 > 0 \\
(\{1,3,4\}, w) & \text{if } C_1 = C_3 = \{1,3,4\}, z = w, w_4 = 0, \text{ and } C_2 = \{2,3,4\} \\
(\{2,3,4\}, y) & \text{if } C_2 = C_3 = \{2,3,4\}, y = z, \text{ and } y_4 > 0 \\
 & \quad (\text{and } y_4 > w_4 \text{ if } C_1 = \{1,4\}) \\
(\{1,2,4\}, w) & \text{if } C_1 = C_2 = \{1,2,4\}, w = y, \text{ and } w_4 > 0 \\
(\{1,4\}, w) & \text{if } C_1 = \{1,4\} \text{ and } w_4 > 0 \\
 & \quad (\text{and } w_4 \geq y_4 \text{ if } C_2 = C_3 = \{2,3,4\} \text{ and } y = z) \\
(\emptyset, x) & \text{otherwise}
\end{cases}
$$

$$
\sigma^{*3}\left(h^3\right) = \begin{cases}
(\{1,2,3,4\}, w) & \text{if } C_1 = C_2 = \{1,2,3,4\}, y = w, w_3 > 0, \text{ and } w_4 > 0 \\
(\{1,3,4\}, w) & \text{if } C_1 = \{1,3,4\}, w_3 > 0 \text{ and } w_4 > 0 \\
 & \quad (\text{and } w_3 \geq y_3 \text{ if } C_2 = \{2,3,4\}) \\
(\{1,3,4\}, w) & \text{if } C_1 = \{1,3,4\}, C_2 = \{2,3,4\}, w_3 \geq y_3 \text{ and } w_4 = 0 \\
(\{2,3,4\}, y) & \text{if } C_2 = \{2,3,4\} \text{ and } y_3, y_4 > 0 \\
 & \quad (\text{and } y_3 > w_3 \text{ if } C_1 = \{1,3\} \text{ or } C_1 = \{1,3,4\}) \\
(\{1,2,3\}, w) & \text{if } C_1 = C_2 = \{1,2,3\}, y = w, \text{ and } w_3 > 0 \\
(\{1,3\}, w) & \text{if } C_1 = \{1,3\} \text{ and } w_3 > 0 \\
 & \quad (\text{and } w_3 \geq y_3 \text{ if } C_2 = \{2,3,4\}) \\
(\emptyset, x) & \text{otherwise}
\end{cases}
$$

$$
\sigma^{*2}\left(h^2\right) = \begin{cases}
(\emptyset, x) & \text{if } C_1 = \{1,3,4\}, w_3 = 1 \text{ and } w_4 = 0 \\
(\{2,3,4\}, (1,1,1)) & \text{otherwise}
\end{cases}
$$

$$
\sigma^{*1} = (\{1,3,4\}, (0,1,0))
$$

These strategies were constructed by backward induction, with agent 4 (then 3, then 2) accepting proposals as required by subgame perfection. Wherever agents were indifferent between a proposal and the status quo, the strategy selects the status quo; for indifference between the tempting coalition and the blocking coalition, the strategy selects the tempting coalition. The only exception in the third line of agent 4's strategy: he accepts the tempting coalition with an allocation of 0 (even though it is no better than the status quo), but only when the blocking coalition has been rejected by 3 in favor of the tempting coalition.

Agent 3 is indifferent between the blocking coalition and the tempting coalition. So she can credibly choose the tempting coalition — when it is offered *and* will match. But if the blocking coalition is not proposed, the tempting coalition will not match, due to agent 4's strategy. In that situation, she can stick with the status quo.

Agent 2 could unconditionally propose the blocking coalition. It will be at least as good as any coalition proposed by 1 that includes 2, and if 1's proposal excludes 2, then the blocking coalition is the only possible way to get a payoff greater than 0. Hence, even if the proposal is unmatched, it is still a weak best response. Yet in one case, agent 2 decides to propose the status quo — when 1 has proposed the tempting coalition. In that case, proposing the blocking coalition would only cause the tempting coalition to form, leaving agent 2 unmatched and no better off than under the status quo.

In this way, agent 1 has unconditionally offered the perfect temptation for spoiling: the offer induces later agents to reject a blocking coalition, yet the offer itself is not accepted on the equilibrium path. Here, our spoiler is essentially pitting the agents against each other to prevent their cooperation.

While indifference may appear to be a prominent feature of this counterexample, it is mainly used to simplify the description of the history-dependent equilibrium strategies. The place where it plays a crucial role is that one member of the tempting coalition must be indifferent between the status quo and the tempting allocation. Furthermore, that agent must propose *after* another member of the blocking coalition. Because of the symmetry of 2, 3, and 4, this can be accomplished in any proposal order.

References

Abreu, D., and Sen, A. (1990) "Subgame Perfect Implementation: A Necessary and Almost Sufficient Condition," *J. Econ. Theory* **50**, 285-299.

Anderson, R. (1992) "The Core in Perfectly Competitive Economies," in *Handbook of Game Theory with Economic Applications* (R. Aumann and S. Hart), pp. 413-457, Amsterdam: North-Holland.

Aumann, R., and Maschler, M. (1964) "The Bargaining Set for Cooperative Games," in *Advances in Game Theory* (M. Dresher, L.S. Shapley, and A.W. Tucker), pp. 443-476. Princeton: Princeton University Press.

Bergin, J., and Duggan, J. (1999) "An Implementation-Theoretic Approach to Non-cooperative Foundations," *J. Econ. Theory* **86**, 50-76.

Bloch, F. (1996) "Sequential Formation of Coalitions in Games with Externalities and Fixed Payoff Division," *Games and Economic Behavior* **14**, 90-123.

Chander, P. (2007) "The gamma-core and coalition formation," *International Journal of Game Theory* **35**, 539556 .

Chatterjee, K., Dutta, B., Ray, D., and Sengupta, K. (1993) "Theory of Coalitional Bargaining," *Review of Economic Studies* **60**, 463-477.

Dutta, B., Ray, D., Sengupta, K., and Vohra, R. (1989) "A Consistent Bargaining Set," *J. of Econ. Theory* **49**, 93-112.

Edgeworth, F. Y. (1881) *Mathematical Psychics* London: Kegan Paul.

Evans, R. (1997) "Coalitional Bargaining with Competition to Make Offers," *Games and Economic Behavior* **19**, 211-220.

Greenberg, J. (1990) *The Theory of Social Situations: An Alternative Game-Theoretic Approach* Cambridge: Cambridge University Press.

Horniaček, M. (2007) "Negotiation, Preferences over Agreements, and the Core," *International Journal of Game Theory*, forthcoming.

Kalai, E., Postlewaite, E., and Roberts, J. (1979) "A Group Incentive Compatible Mechanism Yielding Core Allocations," *J. Econ. Theory* **20**, 13-22.

Kolpin, V. (1989) "Core Implementation via Dynamic Game Forms," *Social Choice and Welfare* **6**, 205-225.

Lagunoff, R. (1994) "A Simple Noncooperative Core Story," *Games and Economic Behavior* **7**, 54-61.

Mas-Colell, A. (1989) "An Equivalence Theorem for a Bargaining Set," *J. of Math. Econ.* **18**, 129-139.

Maskin, E. (1977) "Nash Equilibrium and Welfare Optimality," Mimeo, Massachusetts Institute of Technology.

Maskin, E. (1985) "The Theory of Implementation in Nash Equilibrium: A Survey," in *Social Goals and Organization: Essays in Memory of Elisha Pazner* (L. Hurwicz, D. Schmeidler, and H. Sonnenschein), pp. 173-204, Cambridge: Cambridge University Press.

Moldovanu, B., and Winter, E. "Core Implementation and Increasing Returns to Scale for Cooperation," *J. Math Econ.* **23**, 533-548.

Moore, J. and Repullo, R. "Subgame Perfect Implementation," *Econometrica* **56**, 1191-1220.

Okada, A. (1992) "Noncooperative Bargaining and the Core of an N-Person Characteristic Function Game," *Control and Cybernetics* **21**, 231-250.

Okada, A., and Winter, E. (2002) "A Non-Cooperative Axiomization of the Core," *Theory and Decision* **53**, 1-28.

Pérez-Castrillo, D. (1994) "Cooperative Outcomes through Noncooperative Games," *Games and Economic Behavior* **7**, 428-440.

Perry, M., and Reny, P. (1994) "A Noncooperative View of Coalition Formation and the Core," *Econometrica* **62**, 795-817.

Ray, D., and Vohra, R. (1999) "A Theory of Endogenous Coalition Structures," *Games and Economic Behavior* **26**, 286-336.

Reny, P. (1997) "Two Lectures on Implementation Under Complete Information: General Results and the Core," in *Cooperation: Game-Theoretic Approaches* (S. Hart, and A. Mas-Colell), pp.91-113. Berlin: Springer-Verlag.

Serrano, R. (1995) "A Market to Implement the Core," *J. Econ. Theory* **67**, 285-294.

Serrano, R. (1997) "A comment on the Nash program and the theory of implementation," *Economic Letters* **55**, 203-208.

Serrano, R., and Vohra, R. (1997) "Non-cooperative Implementation of the Core," *Social Choice and Welfare* **14**, 513-525.

Serrano, R., and Vohra, R. (2002) "Bargaining and Bargaining Sets," *Games and Economic Behavior* **39**, 292-308.

Yan, H. (2002) "Non-cooperative Selection of the Core," *Int'l J. of Game Theory* **31**, 527-540.

Yi, S. (1997) "Stable coalition structures with externalities," *Games and Economic Behavior* **20**, 201237.