Theoretical Economics 12 (2017), 1191–1227 1555-7561/20171191

Rational expectations and farsighted stability

Bhaskar Dutta
Department of Economics, University of Warwick and Department of Economics, Ashoka University

Rajiv Vohra
Department of Economics, Brown University

In the study of farsighted coalitional behavior, a central role is played by the von Neumann–Morgenstern (1944) stable set and its modification that incorporates farsightedness. Such a modification was first proposed by Harsanyi (1974) and was recently reformulated by Ray and Vohra (2015). The farsighted stable set is based on a notion of indirect dominance in which an outcome can be dominated by a chain of coalitional "moves" in which each coalition that is involved in the sequence *eventually* stands to gain. However, it does not require that each coalition make a *maximal* move, i.e., one that is not Pareto dominated (for the members of the coalition in question) by another. Consequently, when there are multiple continuation paths, the farsighted stable set can yield unreasonable predictions. We restrict coalitions to hold common, history independent expectations that incorporate maximality regarding the continuation path. This leads to two related solution concepts: the rational expectations farsighted stable set and the strong rational expectations farsighted stable set. We apply these concepts to simple games and to pillage games to illustrate the consequences of imposing rational expectations for farsighted stability.

Keywords. Stable sets, farsightedness, consistency, maximality, rational expectations, simple games, pillage games.

JEL classification. C71, D72, D74.

1. Introduction

Theories of coalitional stability are based on the notion of domination or objections by coalitions. A coalition is said to have an objection to the status quo if it can change the outcome to one in which all its members gain. Perhaps the most widely used solution concept in this literature is the *core*: the set of outcomes to which there is no objection. The original formulation of the theory by von Neumann and Morgenstern (1944), however, was concerned with a somewhat more sophisticated equilibrium concept, which they referred to simply as the "solution," but that has since become known

Bhaskar Dutta: b.dutta@warwick.ac.uk
Rajiv Vohra: Rajiv_Vohra@brown.edu
We are extremely grateful to Debraj Ray for many fruitful discussions on this subject. Thanks are also due Jean Jacques Herings, Mert Kimya, Dilip Mookherjee, and an anonymous referee.

DOI: 10.3982/TE2454

as the von Neumann–Morgenstern (vNM) stable set. A *vNM stable set* consists of outcomes that satisfy two properties: (i) *internal stability* in the sense that no stable outcome dominates any other stable outcome; (ii) *external stability* in the sense that every outcome not in the stable set is dominated by some stable outcome. There is a large literature on the stable set even though it has been notoriously difficult to work with.[1]

Both the core and the stable set are based on *myopic*, or one-shot, deviations by coalitions. If a change made by a coalition can be followed by other coalitional moves, then clearly we should require coalitions to be farsighted in their behavior. Each coalition should ask itself, "if we take action a, how will other coalitions react?" This is a direction of research that has attracted renewed interest; see, for example, Harsanyi (1974), Aumann and Myerson (1988), Chwe (1994), Bloch (1996), Ray and Vohra (1997, 1999), Xue (1998), Diamantoudi and Xue (2003), Konishi and Ray (2003), Mauleon and Vannetelbosch (2004), Herings et al. (2004, 2009), Ray (2007), Mauleon et al. (2011), Ray and Vohra (2014, 2015), Chander (2015), and Kimya (2015). It is by no means obvious how the classical theory should be modified to account for farsightedness, which partly explains the diversity of approaches taken in this literature. Ray and Vohra (2014) distinguish between two principal approaches: (a) the *blocking approach*, which follows traditional cooperative game theory in abstracting away from the details of the negotiation process and relying on a coalitional game to specify what each coalition is able to accomplish on its own, and (b) the *bargaining approach*, which is based on noncooperative coalition bargaining and relies on specifying details such as a protocol that describes the order of moves.

This paper studies farsightedness in the tradition of the blocking approach, with the underlying model described through a coalitional (or characteristic function) game. In particular, there is no prespecified set of "terminal" states, or a protocol specifying the order in which players or coalitions are allowed to move. Consequently, we cannot capture farsightedness through subgame perfection or some other solution concept based on backward induction. We nevertheless want to capture the idea that coalitional decision making is based not on the immediate effect of an initial "move," but on the "final outcome."[2] This immediately raises the question of how to determine what the final outcome is in a sequence of coalitional moves (in a model without the formal apparatus of an extensive form game). Suppose coalition S^1 replaces x with y, and then S^2 replaces y with z. If z is the final outcome and payoffs are realized only once, farsightedness would require S^1 to compare the utility of z to that of x and ignore its payoff at y. But this argument only works *if* z is known to be the final outcome. What is considered to be a final outcome must, of course, also be *stable*. Thus, testing the stability of a particular outcome against a sequence of moves requires us to know which of the other outcomes are stable. This is precisely the kind of circularity that the stable set is very adept at handling, making it a fruitful vehicle for incorporating farsightedness within the blocking approach.

[1] See Lucas (1992) for a survey.

[2] In a real-time model such as in Konishi and Ray (2003), what matters is the entire stream of (discounted) payoffs along a sequence of moves.

The idea of modifying the stable set by allowing for sequences of coalitional moves, with each coalition focused on the final outcome, goes back to Harsanyi (1974). Unfortunately, the Harsanyi stable set, including its more recent reformulation by Ray and Vohra (2015), does not fully capture the idea of optimal behavior embodied in backward induction. One difficulty is that coalitions involved in a farsighted objection are not required to make the *most* profitable moves (in a Pareto sense) that may be available to them. This is the issue of *maximality*. For example, coalition S^1 might move from x to y, anticipating that S^2 will then move to z. If z is a final outcome, and all players in S^1 prefer z to x and all those in S^2 prefer z to y, then this would be considered a legitimate farsighted objection to x. But what if the move by S^2 to z is not *maximal*, in the sense that at y it has another available move, say to z', which is also a final outcome, that is even better than z for all its members? It seems reasonable to require that if z and z' are the only moves available to S^2, then both S^2 and S^1 should focus on z' rather than z as the final outcome. But the farsighted stable set does not insist on this.

Another feature of farsighted objections is that they permit coalitions to hold different beliefs about the continuation path of coalitional moves. For instance, it is possible that x is not in the farsighted stable set because S^1 replaces it with y, anticipating a second, and final, move to z while at the same time x' is not stable because S^2 replaces it with y, expecting the next, and final, move to be to z' (not z). We refer to this as the issue of *consistent* beliefs. An alternative interpretation of this phenomenon is that rather than being inconsistent, beliefs are history dependent. Thus, agents may commonly believe that y will transition to z or z' depending on whether y was preceded by x or x'. In this paper we restrict beliefs to be consistent, or history independent, in the sense that the belief about the continuation from a given state does not depend on how the current state was reached. This corresponds to the Markovian assumption commonly made in dynamic models such as Konishi and Ray (2003). In Sections 2 and 3, we provide several examples to explain these issues in more depth.

The main aim of this paper is to incorporate maximality and consistency (or history independence) of beliefs in the notion of farsighted stability while maintaining the parsimony of the blocking approach, i.e., without the introduction of a protocol or other details of the negotiation process.[3] To accomplish this, the only new concept that needs to be added to the traditional framework of coalitional games is that of an *expectation function*, a tool we borrow from Jordan (2006). This describes the transition from one state to another, as well as the coalition that is supposed to effect the move. The expectation function represents the commonly held beliefs of all agents about the sequence of coalitional moves, if any, from every state. The use of a single expectation function immediately incorporates consistency or history independence. Of course, the coalition expected to move from one state to another must have the power as well as the incentive to do so. As we see, the explicit specification of an expectation function makes it possible to impose appropriate restrictions on it that serve to incorporate maximality. We consider two versions of maximality, one being stronger than the other.

[3]It would also be interesting to incorporate maximality while allowing for history dependence, but that is beyond the scope of the present paper.

A state is said to be *stationary* if the expectation function prescribes no further move once this state is reached. The set of such stationary states then has an inherent element of stability. This leads us to define two related solution concepts: the *rational expectations farsighted stable set* (REFS) and the *strong rational expectations farsighted stable set* (SREFS). These are the set of stationary points of an expectation function that satisfies one or the other notion of maximality as well as farsighted versions of internal and external stability. A key point of the paper is to show that although there are some cases in which a farsighted stable set, or even a vNM stable set, is a REFS or a SREFS, in general imposing rational expectations can be consequential for farsighted stability.

At some intuitive level, notions of farsightedness and maximality attempt to bring into coalitional games considerations that are similar to backward induction in noncooperative games. This will become clear in Section 2, where we describe the framework and present a series of examples showing that the failure to impose maximality and consistency can lead to counterintuitive predictions. Because these examples have "terminal nodes," allowing for backward induction, the connection between subgame perfection and maximality is obvious.[4] The difficulty, of course, is that coalitional games do not typically have the structure of an extensive form that allows for recursion,[5] and our main conceptual task is to formalize these ideas within the parsimony of the blocking approach (without introducing a protocol or other details about the negotiation process). We do so by incorporating an expectation function into the framework of coalitional games.

Section 3 contains formal definitions of an expectation function as well as the different versions of maximality and our solution concepts: REFS and SREFS. We show that one special but interesting case in which both these solution concepts coincide with the farsighted stable set is when the latter consists of states with a single payoff. However, this equivalence does not hold more generally. The following two sections apply these concepts to a couple of important economic models and illustrate how the imposition of rational expectations can result in predictions that are very different from those of existing versions of stable sets with farsighted players.

Section 4 provides an application to *simple games*, which have proved to be very useful in studying voting behavior and possess a rich literature on stable sets. We show that in this class of games, the farsighted stable sets identified in Ray and Vohra (2015) *do not* meet the consistency test or, if they do, it is only by allowing beliefs to be history dependent. Moreover, the restriction to rational expectations, as in REFS or SREFS, leads to sharply different predictions regarding farsighted stability. For this class of games, the contrast between a farsighted stable set and a REFS (or a SREFS) seems not to arise from the maximality issue; it hinges entirely on the consistency issue. We also establish the existence of a SREFS in a large class of simple games.

[4]More generally, the connection comes out perhaps most clearly in Kimya's (2015) concept of *equilibrium coalitional behavior* (ECB), which is defined for a model that has the advantage of being directly applicable to extensive form games.

[5]*Coalition-proof Nash equilibrium* in Bernheim et al. (1987) and *equilibrium binding agreements* in Ray and Vohra (1997) are able to make use of recursion by restricting attention to *internal blocking*, where each coalition in a sequence of objections is a subset of the previous one.

Section 5 studies *pillage games*. These are models of economies where property rights do not exist so that the more "powerful" can capture the assets or wealth of the less powerful. Jordan (2006) and Acemoglu et al. (2008) have studied farsighted cooperative behavior in these models.[6] The application of our concepts to pillage games is complementary to the exercise in the previous section. Here it is the maximality issue, rather than consistency, that turns out to make a crucial difference. In fact, in this model, the distinction between our two notions of maximality is also important; there exist REFSs that are not SREFSs. This is also the case for the Acemoglu et al. (2008) model of political power. Moreover, the SREFS provides a new characterization of their equilibrium notion, while the notions of REFSs or farsighted stable sets are not strong enough to yield precisely their solution.

2. Maximality and consistency

We consider a general setting, described by an *abstract game*, $(N, X, E, u_i(\cdot))$, where N is the set of players and X is the set of outcomes or states. Let $\mathcal{N}$ denote the set of all subsets of N. The effectivity correspondence, $E : X \times X \mapsto \mathcal{N}$, specifies the coalitions that have the ability to replace a state with another state: for $x, y \in X$, $E(x, y)$ is the (possibly empty) set of coalitions that can replace x with y. Finally, $u_i(x)$ is the utility of player i at state x.

The set of outcomes as well as the effectivity correspondence depend on the structure of the model being studied. For instance, in a *coalition (or characteristic function) game*, (N, V), there is a set of feasible utilities, $V(S)$, for every coalition S.[7] In this case, a state generally refers to a coalition structure and a corresponding payoff allocation that is feasible and efficient for each of the coalitions in the coalition structure. Historically, however, following von Neumann and Morgenstern (1944), much of the literature has treated the set of states to be the set of *imputations*—the Pareto efficient utility profiles in $V(N)$—and implicitly assumed that $S \in E(x, y)$ if and only if $y_S \in V(S)$. We explain below why this turns out to be unsatisfactory for studying farsightedness.

State y *dominates* x if there is $S \in E(x, y)$ such that $u_S(y) \gg u_S(x)$. In this case, we also say that (S, y) is an *objection* to x.

The *core* is the set of all states to which there is no objection.

A set $K \subseteq X$ is a *vNM stable set* if it satisfies the following statements:

(i) There do not exist $x, y \in K$ such that y dominates x: internal stability.

(ii) For every $x \notin K$, there exists $y \in K$ such that y dominates x: external stability.

For an abstract game, we define farsighted dominance as follows.

State y *farsightedly dominates* x (under E) if there is a sequence $y^0, (y^1, S^1), \ldots, (y^m, S^m)$, with $y^0 = x$ and $y^m = y$, such that for all $k = 1, \ldots, m$,

$$S^k \in E\big(y^{k-1}, y^k\big)$$

[6] Piccione and Rubinstein (2007) study the analogue of an exchange economy in the "jungle," which is similar to a world without property rights.

[7] A transferable utility (TU) coalitional game is denoted (N, v), where $V(S) = \{u \in R^S \mid \sum_i u_i \leq v(S)\}$ for all S.

and

$$u(y)_{S^k} \gg u\big(y^{k-1}\big)_{S^k}.$$

A set $F \subseteq X$ is a *farsighted stable set* if the following statements hold:

(i) There do not exist $x, y \in F$ such that y farsightedly dominates x: farsighted internal stability.

(ii) For every $x \notin F$, there exists $y \in F$ such that y farsightedly dominates x: farsighted external stability.

It is important to emphasize that the notion of effectivity is especially delicate in the context of farsightedness. Harsanyi (1974), in defining farsighted dominance for a characteristic function game, maintained the von Neumann–Morgenstern assumption that $S \in E(x, y)$ if and only if $y_S \in V(S)$.[8] This way to specify effectivity gives coalition S complete freedom in choosing y_{-S}, the payoffs to outsiders (provided y is an imputation and $y_S \in V(S)$). This obviously leads to the question of why or how coalition S can dictate the payoffs of coalition $N - S$. However, this questionable assumption has not received much attention until recently because it plays no role for myopic solutions such as the core and the stable set. But in the case of farsighted dominance, this is not only conceptually questionable but can significantly alter the nature of the farsighted stable set, as shown by Ray and Vohra (2015). They demonstrate that imposing reasonable restrictions on the effectivity correspondence results in a farsighted stable set that is very different from, and arguably more plausible than, that of Harsanyi (1974). We therefore need to be attentive to this issue when we consider specific models in Sections 4 and 5. Until then, to highlight the main concerns of this paper, we work in the generality of an abstract game, without any explicit restrictions on the effectivity correspondence.

The farsighted stable set is based on an optimistic view of the coalitions involved in a farsighted objection. A state is dominated if there exists *some* path that leads to a better outcome. Chwe (1994) proposed a farsighted solution concept based on conservative behavior, which is good at identifying states that cannot possibly be considered stable. A set $K \subseteq X$ is *consistent* if

$$K = \big\{x \in X \mid \text{for all } y \text{ and } S \text{ with } S \in E(x, y), \text{ there exists } z \in K \text{ such that } z = y \text{ or}$$
$$z \text{ farsightedly dominates } y \text{ and } u_i(z) \leq u_i(x) \text{ for some } i \in S\big\}.$$

Thus, any potential move from a point in a consistent set is deterred by *some* farsighted objection that ends in the set. Chwe shows that there exists one such set that contains all other consistent sets and he defines this to be the *largest consistent set* (LCS).

In general, both of these solution concepts are unsatisfactory because optimistic or pessimistic expectations are both ad hoc.[9]. Ideally, a solution concept should be based

[8] In fact, Harsanyi was following the standard practice of making this part of the dominance condition rather than presenting it through an effectivity correspondence. So it would be more precise to say that this is implicitly what Harsanyi assumed.

[9] For a comprehensive study of stable sets based on optimistic or pessimistic behavior see Greenberg (1990)

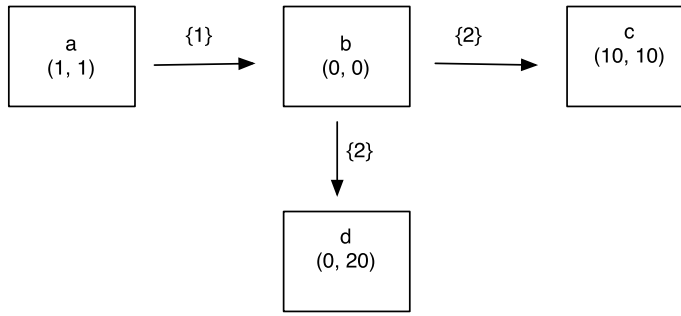

FIGURE 1. The maximality problem for a farsighted stable set.

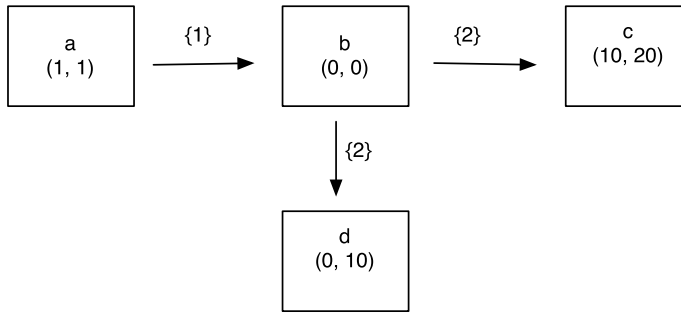

FIGURE 2. The maximality problem for a LCS.

on *optimal* behavior (which may of course turn out to be optimistic or pessimistic in particular examples). The following examples, based on similar ones in Xue (1998), Herings et al. (2004), and Ray and Vohra (2014), illustrate this problem vividly.

EXAMPLE 1. The game is depicted in Figure 1. Player 1 is effective in moving from state a to b, while player 2 can replace state b with either c or d, which are both "terminal" states. The numbers below each state denote the utilities to the players.

Both c and d belong to the farsighted stable set since they are terminal states. Since there is a farsighted objection from a to c, the former is not in the farsighted stable set. However, this is based on the expectation that player 2 will choose to replace b with c rather than d even though he/she prefers d to c. If player 2 is expected to move, rationally, to d, then a should be judged to be stable, contrary to the prediction of the farsighted stable set. Note that a belongs to the LCS because of the possibility that the final outcome is d. So in this example the LCS makes a more reasonable prediction than the farsighted stable set. ◊

EXAMPLE 2. This is a modification of Example 1 as shown in Figure 2.

Now the optimal move for player 2 is to choose c rather than d. The LCS and farsighted stable set remain unchanged. But now it is the LCS that provides the wrong answer because player 1 should not fear that player 2 will (irrationally) choose d instead of c. In this example, the farsighted stable set makes a more reasonable prediction. ◊

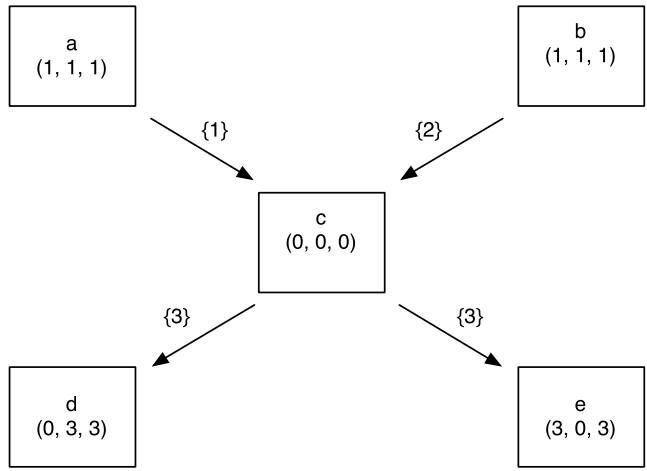

FIGURE 3. History dependence or inconsistency.

As the previous two examples show, both the LCS and the farsighted stable set suffer from the problem that they do not require coalitions (in these examples, player 2) to make moves that are *maximal* among all profitable moves. (A formal definition of maximality in our framework appears in the next section.)

Another problem that afflicts both the LCS and the farsighted stable set is that they may be based on expectations that are inconsistent in the sense that coalitions move based on different expectations about the continuation to follow. For the LCS this was pointed out by Konishi and Ray (2003). Our next example illustrates this problem for both the farsighted stable set and the LCS.

EXAMPLE 3. This is a three-player game with five states as shown in Figure 3.

In this example, the farsighted stable set is $\{d, e\}$ while the LCS is $\{a, b, d, e\}$. State a is not in the farsighted stable set because of a farsighted objection to e, and b is not in it because of a farsighted objection to d. However, from c player 3 can only move to either d or e. Thus, the exclusion of both a and b from the farsighted stable set is based on inconsistent expectations of the move from c. Alternatively, the inclusion of both a and b in the LCS is also based on inconsistent expectations. The "right" answer in this example should be that $\{a, d, e\}$ and $\{b, d, e\}$ are two "stable sets": the former if the expectation is that player 3 will move from c to d and the latter if the expectation is that 3 will move from c to e. Another interpretation of this phenomenon is that expectations under the farsighted stable set or the LCS allow for history dependence: the continuation from c depends on the history. In this sense, what we are asking for can be seen as the joint requirement of consistency *and* history independence. Formally, in a dynamic model this corresponds to the Markovian assumption, which is commonly made in the literature, as in Konishi and Ray (2003). ◊

To define optimal behavior, one needs to rely on players having (rational) expectations about the continuation path following any coalition move. In a dynamic setting, such as in Konishi and Ray (2003) or Ray and Vohra (2014), these expectations are

specified by a dynamic process of coalitional moves. An *equilibrium process of coalition formation* (EPCF) is a Markovian process in which coalitions take actions that are *maximally profitable* in terms of a value function.[10] One difference in these models is that Ray and Vohra (2014), unlike Konishi and Ray (2003), specify a protocol to explicitly determine the order in which coalitions are called upon to move at each stage. In spirit, though, in both cases the approach for incorporating consistency and rational expectations is similar to ours, even though we seek to accomplish this more directly within the static, blocking approach. Static models most closely related to our approach are Xue (1998) and Kimya (2015).

Xue (1998) argued that to resolve the maximality issue, we should consider a stable set defined over *paths* of coalition actions rather than on outcomes. In many cases, such as Examples 1 and 2, this can resolve the problem. However, it may push the choice between optimism and pessimism to another level. When a path is tested against a deviation by a coalition, the deviation can itself lead to multiple stable paths and so in evaluating these multiple paths the pessimism/optimism choice resurfaces. This leads Xue to define the *optimistic stable standard of behavior* and the *conservative stable standard of behavior*. In Example 3, the predictions of these two concepts match the farsighted stable set and the LCS, respectively. We are able to avoid this by considering stability in terms of a given expectation that describes transitions from *every* outcome. In our framework, once a coalition makes a change, there is no further ambiguity about the continuation path. In this respect, our approach is similar to Konishi and Ray (2003), Ray and Vohra (2014), and Kimya (2015), even though these papers propose solution concepts that are not defined in terms of stable sets. In these papers, an equilibrium path need not involve all coalitions doing *strictly* better, whereas in our framework, a sequence of coalition moves will be a farsighted objection, involving strict improvements; see Kimya (2015) for further discussion.

We should acknowledge that one reason all the examples in this section are so simple is because they concern abstract games with terminal nodes. The skeptical reader might wonder whether issues of maximality or consistency are consequential in more general models of economic interest. Our analysis of simple games and pillage games in Sections 4 and 5 demonstrates that these issues are indeed of more general importance. But we must first turn to the task of formally incorporating notions of maximality and consistency in a definition of farsightedness in coalitional games.

3. Farsightedness with rational expectations

Jordan (2006) formulates the idea that farsighted stability can be expressed in terms of commonly held consistent expectations regarding the final outcome from any state. He defines an expectation as a function $\phi : X \to X$ such that for every $x \in X$, $\phi(\phi(x)) = \phi(x)$. A stationary state of ϕ is x such that $\phi(x) = x$. Given a farsighted stable set, Z, it

[10] In Example 3, therefore, the prediction would be that the stable outcomes are either $\{a, d, e\}$ or $\{b, d, e\}$.

is straightforward to construct an expectation ϕ that is consistent with farsighted dominance and yields Z as the collection of all stationary outcomes. If $x \in Z$, let $\phi(x) = x$. If $x \notin Z$, let $\phi(x) = y$ for some $y \in Z$ that farsightedly dominates x.[11]

So as to deal with the issues discussed in Section 2, we modify Jordan's approach by interpreting an expectation to describe the transition from one state to another, not necessarily the final outcome from a state. In addition, we also find it important to keep track of the coalition that is expected to make the transition. With this in mind, we define an expectation as a function $F : X \to X \times \mathcal{N}$. For a state $x \in X$, denote $F(x) = (f(x), S(x))$, where $f(x)$ is the state that is expected to follow x and $S(x) \in E(x, f(x))$ is the coalition expected to implement this change. If $f(x) = x$, then $S(x) = \varnothing$, signifying the fact that no coalition is expected to change x. A stationary point of F is a state x such that $f(x) = x$. Given an expectation $F(\cdot) = (f(\cdot), S(\cdot))$, let f^k denote the k-fold composition of f. In particular, $f^2(x) = f(f(x))$. With a slight abuse of notation, let $F^k(x) = F(f^{k-1}(x))$.

An expectation is said to be *absorbing* if for every $x \in X$, there exists k such that $f^k(x)$ is stationary. In this case, let $f^*(x) = f^k(x)$, where $f^k(x)$ is stationary.

We seek to describe a set of stable outcomes $Z \subseteq X$ that is "justified" by an expectation in the sense that Z is the set of stationary points of an expectation F that embodies farsighted rationality.

An absorbing expectation F is said to be a *rational expectation* if it has the following properties:

(I) If x is stationary, then from x, no coalition is effective in making a profitable move (consistent with F), i.e., there does not exist $T \in E(x, y)$ such that $u_T(f^*(y)) \gg u_T(x)$.

(E) If x is a nonstationary state, then $F(x)$ must prescribe a path that is profitable for all the coalitions that are expected to implement it, i.e., $(x, F(x), F^2(x), \ldots, F^k(x))$ is a farsighted objection where $f^k(x) = f^*(x)$.

(M) If x is a nonstationary state, then $F(x)$ must prescribe an optimally profitable path for coalition $S(x)$ in the sense that there does not exist y such that $S(x) \in E(x, y)$ and $u_{S(x)}(f^*(y)) \gg u_{S(x)}(f^*(x))$.

The set of stationary points, $\Sigma(F)$, of a rational expectation F is said to be a *rational expectations farsighted stable set* (REFS).

Conditions (I) and (E) are related to but not the same as farsighted internal and external stability (conditions (i) and (ii) in the definition of a farsighted stable set), and the differences can be significant enough to generate very different results, as we will see. Since $\Sigma(F)$ is a set of stationary states, condition (I) clearly implies that $\Sigma(F)$ satisfies *myopic* internal stability in the traditional sense. It is weaker than farsighted internal stability since it requires internal stability only with respect to those farsighted objections that are *consistent* with the common expectation F.

[11] This bears some similarity to Harsanyi's (1974) attempt to relate the stationary set of an equilibrium in a noncooperative game to a version of a (farsighted) stable set.

Condition (E) states that to every $x \notin \Sigma(F)$, there is a farsighted objection (terminating in $\Sigma(F)$) consistent with the common expectation F. This is stronger than farsighted external stability since it requires consistency with F.

Condition (M) is the *maximality* condition—a translation of the corresponding condition of Konishi and Ray (2003) and Ray and Vohra (2014) into our framework. It requires that if $S(x)$ is expected to move from x to $f(x)$, then the final outcome resulting from this is not Pareto dominated for $S(x)$ by some other move this coalition could have made. For instance, it would require that in Example 1, $f(b) = d$, and in Example 2, $f(b) = c$. Condition (M) is clearly a minimal requirement of optimality. It is also a sufficient expression of optimality if one takes the view that at a nonstationary state x, $S(x)$ is the coalition that has the floor, which gives it the *sole* option to select a transition from x.

However, one could entertain models in which, under certain conditions, some other coalition may also have the right to intervene and change course. For instance, it may be possible for some subset of $S(x)$ to thwart $f(x)$ and move elsewhere. This motivates the following notion of *strong maximality*:

(M′) If x is a nonstationary state, then $F(x)$ must prescribe an optimally profitable path in the sense that no coalition has the power to change course and gain, i.e., there does not exist $T \in E(x, y)$ such that $T \cap S(x) \neq \varnothing$ and $u_T(f^*(y)) \gg u_T(f^*(x))$.

Condition (M′) strengthens (M) by allowing for the possibility that a coalition T that includes some players from $S(x)$ is allowed to change the transition. This is based on the idea that a move by $S(x)$ requires the unanimous consent of all its members, which means that another coalition may seize the initiative if it can enlist the support of at least one player in $S(x)$.

A expectation F satisfying (I), (E), and (M′) is a *strong rational expectation*. The set of stationary points of a strong rational expectation F is said to be a *strong rational expectations farsighted stable set* (SREFS).

Note that strong maximality continues to assume that a coalition disjoint from $S(x)$ cannot interfere in the expected move. An even stronger notion of maximality drops this assumption as well. While we do not pursue this here, see footnotes 12 and 16 below. We expect that these differences in maximality notions would get reflected in the details of the negotiation process embedded in any underlying noncooperative game that might explain a given notion of a farsighted stable set. An important direction for future work is to carry forward the general Nash program to the study of stable sets by studying noncooperative games that lead to one kind of stable set or another. Interestingly, Harsanyi (1974) can be viewed as one such exercise, showing that a (noncooperative) "equilibrium-point interpretation of stable sets" compels us to replace the dominance relation of von Neumann and Morgenstern with farsighted dominance.

Every SREFS is clearly a REFS. However, the converse is not true, as illustrated in the next couple of examples.

EXAMPLE 4. For the two-player game depicted in Figure 4, consider F such that $f(b) = d$ and $S(x) = \{1, 2\}$. Since $S(x)$ gains by moving from b to d and there is only one move

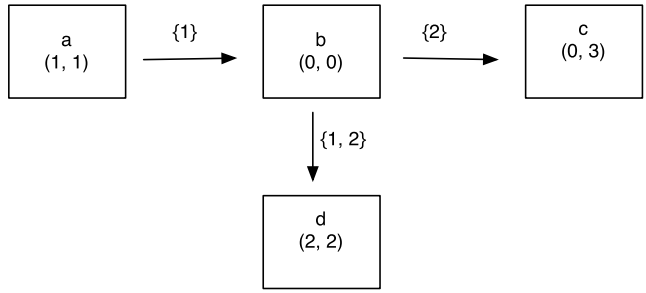

FIGURE 4. A REFS need not be a SREFS.

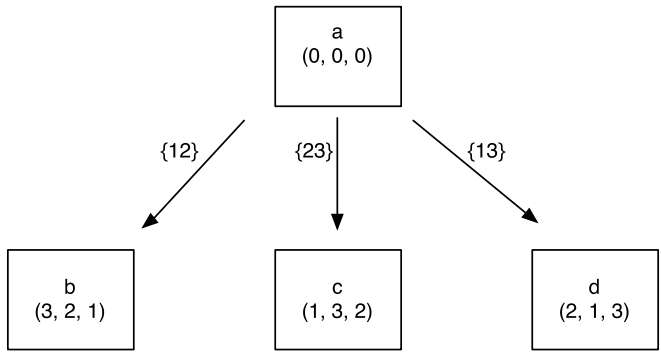

FIGURE 5. A REFS may exist but not a SREFS.

available to this coalition, this is clearly a maximal move. If player 1 anticipates that coalition $\{1, 2\}$ will form and move to d, then he/she will move from a to b. So the REFS corresponding to F is $\{c, d\}$. However, the move by $\{1, 2\}$ from b to d is not *strongly* maximal because $\{2\}$ could do even better by turning down this move and moving to c instead. In our language, F does not satisfy strong maximality, and so $\{c, d\}$ is not SREFS. A strongly maximal rational expectation, say F', must have the property that $f'(b) = c$ with $S(b) = \{2\}$. This dissuades $\{1\}$ from moving from a to b, ensuring that $\{a, c, d\}$ is a SREFS and, of course, also a REFS. Alternatively, $\{c, d\}$ is a REFS but not SREFS. In Section 5, we see a more general version of this phenomenon. ◊

Our next example shows that condition (M′) of the SREFS may be too demanding for existence, even though a REFS may exist.

EXAMPLE 5. In the game shown in Figure 5, there are three possible rational expectation functions prescribing moves from a to one of b, c, or d. All of them satisfy maximality. Of course, for each of these expectation functions, all three terminal states are stationary. So in each case, $\{b, c, d\}$ is a REFS. However, none of these expectation functions satisfies strong maximality. For instance, consider F such that $f(a) = b$ and $S(a) = \{1, 2\}$. This does not satisfy strong maximality since player 2 can join with 3 to move to c, a terminal state. The same argument applies to the other two rational expectation functions.

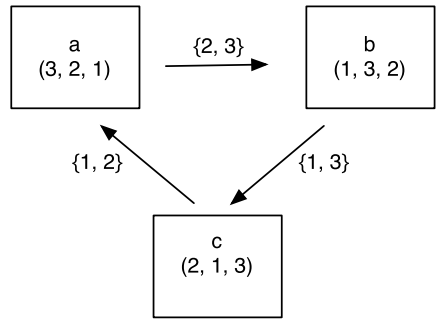

FIGURE 6. Non-existence.

So there is no expectation function for which the set $\{b, c, d\}$ is a SREFS. Alternatively, $\{a, b, c, d\}$ cannot be a SREFS because it would violate (I). So there is no SREFS in this example.[12] ◊

In general, even the existence of a REFS is not guaranteed. One example in which this is the case is the three-player nontransferable utility (NTU) "roommate game" depicted in Figure 6.

EXAMPLE 6. As shown in Figure 6, from every state there is one two-player coalition that gains by moving to another state.

It is easy to see that this game possesses no vNM stable set, no farsighted stable set, and no REFS. ◊

Fortunately, the failure of existence seen in Examples 5 and 6 does not extend to the applications we consider in Sections 4 and 5. There we are able to establish the existence of a SREFS under some mild conditions. Apart from showing existence, the main theme of these sections is that a REFS or SREFS can be very different from a farsighted stable set (Section 4) and a REFS can be very different from SREFS (Section 5). However, there is one interesting case in which a SREFS (or REFS) coincides with a farsighted stable set.

A set of states Z is a *single-payoff* set if $u(x) = u(y)$ for all $x, y \in Z$.

THEOREM 1. *If Z is a single-payoff REFS, it is a SREFS and a farsighted stable set. Conversely, if Z is a single-payoff farsighted stable set, then it is a SREFS.*

All proofs are provided in the Appendix.

[12] An even stronger notion of maximality, which we do not pursue, is the one adopted by Xue (1998). It allows the expected path to be altered by *any* coalition, even one that is disjoint from the coalition that is expected to move. For instance, modify Example 5 so that player 1 is effective in moving from a to b, player 2 from a to c, and player 3 from a to d. Now, REFS and SREFS are the same: they consist of $\{b, c, d\}$. But the extremely strong maximality condition that allows any coalition to change course and prevent someone else from making a move would result in nonexistence in this example. More general models in which it can lead to nonexistence are pillage games; see footnote 16 below.

REMARK 1. Consider a characteristic function game in which the interior of the core is nonempty. Then, by Theorem 2 of Ray and Vohra (2015), every state with a payoff in the interior of the core is a single-payoff farsighted stable set. By Theorem 1, all such games possess a SREFS.

4. SIMPLE GAMES

In this section, we study the class of monotonic, proper simple games (see von Neumann and Morgenstern 1944). These are TU games that have the following properties for each coalition S:

(i) Either $v(S) = 1$ or $v(S) = 0$.

(ii) If $v(S) = 1$, then $v(T) = 1$ for all $S \subseteq T$.

(iii) Moreover, if $v(S) = 1$, then $v(N - S) = 0$.

The set of efficient payoff allocations, or imputations, in any such game is the nonnegative N-dimensional unit simplex, $\triangle$. A coalition S such that $v(S) = 1$ is called a *winning* coalition. Let $\mathcal{W}$ denote the set of all winning coalitions. For simplicity, assume that no $i \in N$ is a *dummy player*; that is, each i belongs to at least one *minimal* winning coalition.[13]

A player i is called a *veto player* if i is a member of every winning coalition; that is, $i \in W$ for every $W \in \mathcal{W}$. The collection of all veto players, also known as the *collegium*, is denoted $C = \bigcap_{S \in \mathcal{W}} S$. A *collegial game* is one in which $C \neq \varnothing$. The collegium (and the corresponding game) will be called *oligarchic* if C is itself a winning coalition. Note that in the absence of dummy players, an oligarchic game is one in which $C = N$; it is a pure bargaining game, in which the grand coalition is the only winning coalition.

In a simple game, a state x specifies a coalition structure, denoted $\pi(x)$, and an associated payoff, $u(x)$, such that $\sum_{i \in W(x)} u_i(x) = 1$, where $W(x)$ is the winning coalition (if any) in $\pi(x)$. We use X^0 to denote the set of states where no winning coalition forms and so $u_i = 0$ for all i. States in X^0 are called *zero states*.

We show that under certain conditions, all such games possess a SREFS. Moreover, the consistency of expectations underlying a SREFS makes the structure of these sets very different from the farsighted stable sets identified in Ray and Vohra (2015)

We make the following assumption regarding the effectivity correspondence.

ASSUMPTION 1. *The effectivity correspondence satisfies the following restrictions:*

(a) For every $x \in X$, $S \subseteq N$, and $u \in R^S_+$ with $\sum_{i \in S} u_i = v(S)$, there is $y \in X$ such that $S \in E(x, y)$ and $u(y)_S = u$.

(b) If $S \in E(x, y)$, then $S \in \pi(y)$ and $T - S \in \pi(y)$ for every $T \in \pi(x)$.

(c) For all $x, y \in X$ and $T \subset N$, if $(W(x) - T) \in \mathcal{W}$, then $T \in E(x, y)$ only if $u_i(y) \geq u_i(x)$ for all $i \in W(x) - T$.

[13]The proof of our main result in this section can be easily modified to accommodate the presence of dummy players.

Condition (a) states that every coalition can form and divide its worth in any way among its members. Condition (b) states that when a coalition S forms, it does not affect any coalition that is disjoint from it, and if it includes some members of a coalition, then the residual remains intact. This is a natural way to describe the immediate change in the coalition structure resulting from the formation of a coalition. Condition (c) requires that if, with the formation of T, the residual in $W(x)$ remains winning, then the players in $W(x) - T$ cannot lose.[14] It includes the condition that if $W(x) \cap T = \varnothing$, then $u(x) = u(y)$. Conditions (b) and (c) should be interpreted as natural restrictions that prevent a coalition from reorganizing the payoffs or coalition structure of those outside it. We mentioned earlier the pitfalls of not imposing such restrictions in the context of farsighted solution concepts.

In an oligarchic game, any strictly positive payoff is in the interior of the core. By Remark 1, any state with a strictly positive payoff, along with coalition N, is a SREFS. In the remainder of this section, therefore, we concentrate on non-oligarchic games.

We discuss in detail an example that illustrates one of the main themes of this section—namely the difference between REFS or SREFS and the farsighted stable set in simple games.

4.1 *An example*

Consider the following three-person example.

Example 7 (A three-player, TU game, (N, v), with one veto player). We have $N = \{1, 2, 3\}$, $v(\{1, 2\}) = v(\{1, 3\}) = v(N) = 1$, and $v(S) = 0$ for all other S. ◊

Ray and Vohra (2015) show that under Assumption 1, every farsighted stable set in this game assigns a fixed payoff to the veto player, strictly between 0 and 1, while the remaining surplus can be divided in any way among players 2 and 3. More precisely, for every $a \in (0, 1)$, there is a farsighted stable set Z_a with the set of payoffs $\{u \in R^3_+ \mid u_1 = a, u_2 + u_3 = 1 - a\}$; see Figure 7, where the vertices of the simplex denote states at which the entire surplus is an allocation to one of the three players.

However, no set of the form Z_a can be a REFS because the external stability of Z_a (in the sense of a farsighted stable set) relies on inconsistent expectations. To see this, consider why the allocation u is not in Z_a. There is first an objection by $\{2, 3\}$, resulting in the coalition structure $\{\{1\}, \{2, 3\}\}$ and 0 payoff to all players. Call this state x^0. This is followed by a move by N to a point in A. Additionally, u' is not stable because there is first a move by $\{2, 3\}$ to x^0, followed by a move by N to a point in B. In the first case, x^0 is expected to be replaced by a point in A, while in the second case it is expected to be replaced by a point in B. This is precisely the kind of "inconsistent" expectation that must be ruled out in a REFS or SREFS. In other words, a farsighted stable set in this example cannot be a REFS.

[14] Of course, this also implies that $W(x) - T \in \pi(y)$. Condition (b) goes beyond this because it also applies to residuals that are not winning.

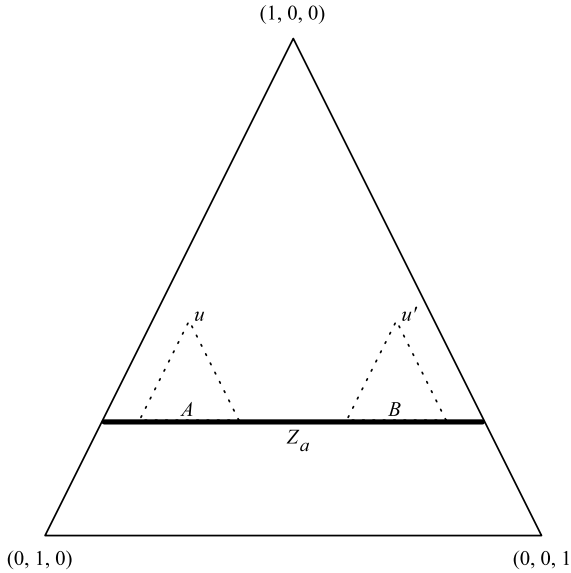

FIGURE 7. A farsighted stable set, Z_a, in Example 7.

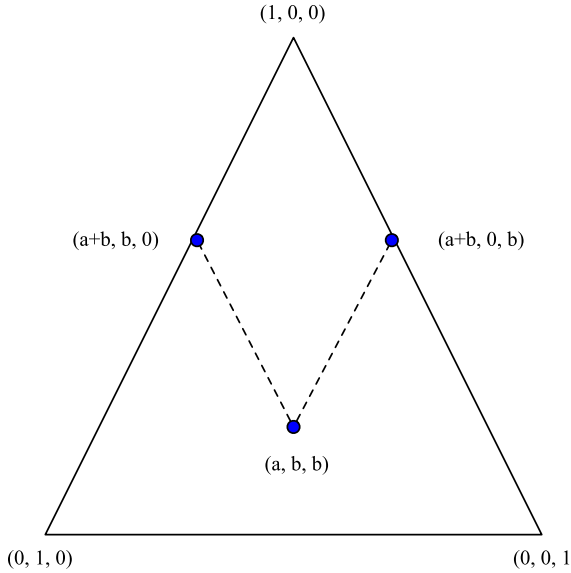

FIGURE 8. SREFS in Example 7.

We now show that in this example there is a SREFS, Z, consisting of a finite set of states $Z = \{(u^1, \pi^1), (u^2, \pi^2), (u^3, \pi^3)\}$, where for some $a \in (0, 1)$ and $b \equiv (1 - a)/2$,

$$u^1 = (a, b, b), \qquad \pi^1 = \{N\},$$
$$u^2 = (a + b, b, 0), \qquad \pi^2 = \big\{\{1, 2\}, \{3\}\big\},$$
$$u^3 = (a + b, 0, b), \qquad \pi^3 = \big\{\{1, 3\}, \{2\}\big\}.$$

The three imputations corresponding to Z are shown in Figure 8. To sustain Z as a SREFS, we construct an expectation described by the following rules. In what follows, we write $x = (u, \pi)$ and $x^i = (u^i, \pi^i)$, $i = 1, 2, 3$.

(i) For each $x \in Z$, $f(x) = x$.

(ii) If x is such that $u = (0, 0, 0)$, then $S(x) = N$ and $f(x) = x^1$.

(iii) For each $x \notin Z$ such that $u_1 \geq a + b$, $S(x) = \{2, 3\}$ and $f(x) = ((0, 0, 0), \pi)$.

(iv) For each $x \notin Z$ such that $u_1 < a + b$ and $u_2 < b$, $S(x) = \{1, 2\}$ and $f(x) = x^2$.

(v) If x is not covered by (i)–(iv) above, and $u_1 < a + b$ and $u_3 < b$, then $S(x) = \{1, 3\}$ and $f(x) = x^3$.

(vi) Finally, if x is not covered by (i)–(v) above and $u_1 < a$, then $S(x) = \{1\}$ and $f(x) = ((0, 0, 0), \pi)$.

This describes F for all $x \in X$.

Clearly, F satisfies (I) and (E). To check (M′), note that in all states in Z players 2 and 3 get either b or 0. This implies strong maximality in cases (ii) and (iii). The deviations in (iv) and (v) are strongly maximal since 1 gets $a + b$, her highest possible payoff in Z. In case (vi), (M′) is satisfied because players 2 and 3 have no reason to move, which means that player 1 does not possess a farsighted objection that could end up at x^2 or x^3.

This completes the demonstration that Z is a SREFS.

4.2 *The main theorem*

In more general non-oligarchic games, Ray and Vohra (2015) show that it is possible to construct a farsighted stable set in which veto players, and perhaps some others, receive a fixed payoff while the remainder of the surplus is shared in any arbitrary way among the remaining players. Such sets, known as discriminatory stable sets, also play an important role in vNM stable set theory, though with the rather important difference that in vNM stable sets it is *non-veto players* who received a fixed payoff. In contrast, SREFSs do not seem to have the structure of discriminatory stable sets. Instead, in most cases SREFSs yield finite payoff sets. We have of course already observed this in Example 7, but this is also the more general conclusion that emerges from the proof of our next result.

Although the notion of a farsighted stable set does not impose maximality, in non-oligarchic simple games this property does seem to hold. In this model, therefore, the difference between farsighted stable sets and SREFSs seems to stem from consistency and history independence.

For most simple games we have been able to constructively prove the existence of a SREFS. There is, however, one particular case for which existence has proven to be elusive. This is the case in which there is a three-player minimal winning coalition with precisely one veto player. Our existence result applies whenever there are two or more veto players or whenever the size of a minimal winning coalition exceeds three. We do need to make the following assumption.

ASSUMPTION 2. *There does not exist a three-player minimal winning coalition with precisely one veto player.*

Subject to this assumption we are able to construct a SREFS for all collegial games.

THEOREM 2. *A SREFS exists in every non-oligarchic collegial game satisfying Assumptions 1 and 2.*

Can Assumption 2 can be dispensed with or will this case yield an example in which a SREFS does not exist? As of now this question remains open.

Of course, a large class of simple games does not have any veto player, the simplest example being the majority game in which any majority of players constitutes a winning coalition. von Neumann and Morgenstern (1944) identified a class of constant-sum games that have a vNM stable set known as a *main simple solution*. Suppose there is $a \in \Re_+^N$ such that $\sum_{i\in S} a_i = 1$ for every minimal winning coalition S. For each minimal winning coalition S, define u^S to be the imputation such that $u_i^S = a_i$ for all $i \in S$ and $u_i^S = 0$ otherwise. If the game is a constant-sum game, then the set of all such imputations is a vNM stable set, known as the main simple solution. For instance, the imputation $(0.5, 0.5, 0)$ and its permutations constitute a main simple solution in the three-person majority game. It can be shown that the set of states corresponding to a main simple solution is a SREFS.

Suppose U is a main simple solution with associated vector $a \in \Re_+^N$. Let $Z(U) = \{x \in X \mid u(x) \in U\}$. We claim that $Z(U)$ is a SREFS. Since U is a vNM stable set, for every $x \notin Z(U)$, there is $S \subseteq N$ and $y \in Z(U)$ such that $S \in E(x, y)$ and $u_S(y) \gg u_S(x)$. For every $x \notin Z(U)$, pick any (S, y) with this property and set $F(x) = (y, S)$. If there are several such (S, y), pick one arbitrarily. For every $x \in Z(U)$, let $f(x) = x$. Clearly, F is an expectation that satisfies (E). Suppose it does not satisfy (I). Then there is $x \in Z(U)$ and $T \in E(x, y)$ such that $u_T(f^*(y)) \gg u_T(x)$. Let S be the minimal winning coalition such that $u(x) = u^S$. Since $u_i(x) = a_i$ for all $i \in S$, no $i \in S$ can get a higher payoff at any other state in $Z(U)$, which implies that $S \cap T = \varnothing$. Of course, S must be contained in the winning coalition at x. By Assumption 1(c), $u_S(y) = u_S(x) = a_S$, i.e., $y \in Z(U)$ and therefore $f^*(y) = y$. But then $u_T(f^*(y)) \gg u_T(x)$ means that $u_T(y) \gg u_T(x)$, with $y \in Z(U)$, which contradicts the myopic internal stability of U. Thus, F satisfies (I). It clearly satisfies strong maximality, (M′), because if any S gains by moving from x to $y \in Z(U)$, then $u_i(y) = a_i$ for all $i \in S$ and there is no other $y' \in Z(U)$ such that $u_i(y') > u_i(y)$ for any $i \in S$.

5. PILLAGE GAMES

In this section, we show that in pillage games, in contrast to simple games, maximality and strong maximality rather than consistency play a crucial role in identifying farsighted stability. In Jordan's (2006) model of "wealth is power," we find that there is an important distinction between REFS and SREFS, both of which can be different from the vNM stable set or the farsighted stable set. In the Acemoglu et al. (2008) model of

political power, we see another interesting illustration of the difference between REFS and SREFS. Acemoglu et al. (2008) propose the *unique ruling coalition* (URC) as a solution concept and characterize it both through axioms as well as the subgame perfect equilibrium of a noncooperative model of coalition formation. In this model, REFSs coincide with farsighted stable sets and include the URC, but they generally include other outcomes as well. We show that SREFS provides just the refinement of REFS that yields precisely the URC. Hence SREFS provides a new characterization of the URC.

In a pillage game, a coalition can appropriate the resources of any other coalition that has less power. Given a set of players N, the set of wealth allocations is $\triangle$, the unit simplex in R^N. We consider the class of pillage games in which wealth is power: the power of coalition S is simply its aggregate wealth. Given wealth allocations w and w', let $L(w,w') = \{i \in N \mid w'_i < w_i\}$ denote the set of players who lose in moving from w to w'. We define the effectivity correspondence in this model as[15]

$$S \in E(w,w') \quad \text{if and only if} \quad \sum_{i\in S} w_i > \sum_{i\in L(w,w')} w_i \text{ and } w_i = w'_i \text{for all } i \notin S \cup L(w,w'). \quad (1)$$

This expresses the notion that a coalition can pillage another only if its power is strictly greater than that of the victims. Moreover, only the winners' and losers' wealth payoffs can be affected through the act of pillaging. That is, if j is neither among those who have been pillaged nor part of the coalition that changes w to w', then $w_j = w'_j$. This last condition rules out a pillaging coalition sharing its spoils with others. While this condition is of no consequence for myopic notions of stability, it becomes important in the context of farsighted stability. As we remarked earlier, it makes no sense to allow a deviating coalition to affect the distribution of the payoff of outsiders. This is illustrated in Example 8 below, where a gift can turn out to be hazardous to the recipients—a Trojan horse. We therefore assume throughout this section that the effectivity correspondence is defined by (1).

Given the effectivity correspondence, notions of the core, myopic and farsighted stable set, REFS, and SREFS remain unchanged.

By way of background, it is useful to begin with Jordan's analysis of the (myopic) vNM stable set.

A number $a \in [0,1]$ is said to be *dyadic* if $a = 0$ or $a = 2^{-k}$ for some nonnegative integer k. For every positive integer k, let

$$D_k = \{w \in \triangle \mid \ w_i \text{ is dyadic for every } i \text{ and if } w_i > 0, \text{ then } w_i \geq 2^{-k}\}.$$

The set of all dyadic allocations is $D = \bigcup_k D_k$. The set of all allocations in which one player captures the entire surplus, D_0, is the set of *tyrannical allocations*. Of course, all such allocations are in the core. It is easy to see that the only other allocations in the core are ones in which two players share the surplus equally. In other words, the core is D_1.

Jordan (2006) provides the following characterization of the stable set.

[15]Although Jordan (2006) does not explicitly define an effectivity correspondence, our formulation is consistent with his.

THEOREM 3 (Jordan). *The unique vNM stable set is D.*

Jordan (2006) illustrates the issue of farsightedness by considering the three-player example in this model, where D consists of the allocations $(1,0,0)$, $(0.5,0.5,0)$, $(0.5,0.25,0.25)$, and all their permutations. From the allocation $(0.5,0.25,0.25)$, player 1, by pillaging 2, can achieve the allocation $(0.75,0,0.25)$. While the latter is not in the stable set, it allows player 1 to then pillage 3 and achieve the tyrannical allocation $(1,0,0)$, which is stable. In other words, $(1,0,0)$ is a farsighted objection to $(0.5,0.25,0.25)$. Note that if player 3 anticipates the second step in this move, she should not remain neutral when player 1 pillages 2. Jordan (2006) formalizes this idea by explicitly introducing expectations. He shows that if otherwise neutral players act in accordance with the expected (final) outcome, then the stable set, D, is indeed stable in a farsighted sense.

However, Jordan's analysis does not really conform to a framework in which the effectivity correspondence specifies which coalition(s) is (are) effective in changing a current state w^{k-1} to w^k, independently of where w^k ends up. For instance, in the three-player example, whether player 1 is effective in changing the allocation $(0.5,0.25,0.25)$ to $(0.75,0,0.25)$ cannot depend on any further changes that may be expected to take place. What is the farsighted stable set if we adopt the effectivity correspondence specified in (1)? As our next result shows, it turns out to be identical to D_1, the core.

THEOREM 4. *Suppose the effectivity correspondence is defined as in (1). Then the unique farsighted stable set is D_1, the core.*

We now turn to a consideration of SREFSs and REFSs in this model. Dyadic allocations in which players with positive wealth share equally play an important role in this analysis. For every nonnegative integer k, let $B_k = \{x \in \triangle \mid x_i = 0 \text{ or } x_i = 2^{-k}, \forall i\}$ and $B = \bigcup_k B_k$. Note that B_0 is the set of tyrannical allocations and $B_0 \cup B_1 = D_1$.

Our next example illustrates a crucial difference between REFSs and SREFSs in this model.

EXAMPLE 8. The pillage game with four players. The unique SREFS is B but there are two REFSs, B as well as $B_0 \cup B_1$. ◊

The core, or $D_1 = B_0 \cup B_1$, in this example is the set of all permutations of allocations of the form $(1,0,0,0)$ and $(0.5,0.5,0,0)$. Since B_2 consists of the equal-division allocation, $\bar{w} = (0.25,0.25,0.25,0.25)$, then $B = D_1 \cup \{\bar{w}\}$. It is easy to see, that there can be no further pillaging from any allocation in D_1. So every REFS, and therefore every SREFS, contains D_1. In fact, D_1 is a REFS and so is B, but only the latter is a SREFS. Theorems 5 and 6 provide general existence results for REFS and SREFS, respectively. But the four-player case is useful for illustrating the difference between REFS and SREFS.

For the four-player game, consider an expectation function F such that $\Sigma(F)$ is a SREFS. As we already observed, $D_1 \subseteq \Sigma(F)$. We now argue that this inclusion is strict by showing that $\bar{w} \in \Sigma(F)$. Observe that if w is such that only two players have positive worth, unless they are equal, F must specify that the more powerful player pillages the

weaker one to end up at a tyrannical allocation. Now consider w such that three players have positive wealth and one has zero wealth. Without loss of generality, suppose $w_1 \geq w_2 \geq w_3 > 0$ and $w_4 = 0$. We now describe why $w \notin \Sigma(F)$. There are two distinct cases.

Case 1. Suppose $w = (1/3, 1/3, 1/3, 0)$. Any two of the first three players can pillage the remaining player with positive wealth to provide 0.5 to each member of the coalition, which is stable. Since the only possible act of pillage at w must come from a two-player coalition, any such move satisfies strong maximality. Thus $f(w)$ must be of the form $f_i(w) = f_j(w) = 0.5$ for $i, j \in \{1, 2, 3\}$.

Case 2. Suppose $w_1 \geq w_2 > w_3 > w_4 = 0$. If player 1 pillages 3, he becomes more powerful than 2 and we know that F must then predict the tyrannical allocation $(1, 0, 0, 0)$. Since player 1 ends up with the entire surplus, the move by 1 to pillage 3 is strongly maximal. This establishes that $w \notin \Sigma(F)$ and that $S(w) = \{1\}$ is one possibility for the coalition that is expected to move at w. If $w_1 = w_2$ or if $w_1 = 0.5$ and $w_2 > w_3$, then $S(w) = \{2\}$ would be another possibility. But—and this is crucial for our arguments to follow—a strongly maximal F *rules out the possibility that* $S(w) = \{1, 2\}$. While $\{1, 2\}$ can pillage player 3, maximality demands that there be no further change. Otherwise one of the two players in $\{1, 2\}$ would not have made a profitable move in the first place. Thus $S(w) = \{1, 2\}$ must mean that $f(w) = (0.5, 0.5, 0, 0)$. While this would be maximal for $\{1, 2\}$, it would violate *strong* maximality since player 1 would have done even better by pillaging player 3 on her own to move to $(w_1 + w_3, w_2, 0, 0)$ and then to $(1, 0, 0, 0)$. (We saw a similar phenomenon in Example 4). Thus, strong maximality implies that either $S(w) = \{1\}$ or $S(w) = \{2\}$.[16] For concreteness, consider $w = (0.375, 0.375, 0.25, 0)$. It is a maximal move for $\{1, 2\}$ to pillage player 3 and move to $(0.5, 0.5, 0, 0)$. But this is not strongly maximal. For F to satisfy strong maximality, one of the stronger players must pillage player 3 on her own and end up with the entire surplus.

It is now easy to see that $\bar{w} \in \Sigma(F)$. The only possible acts of pillage at $\bar{w}$ are for a three-player coalition to pillage the fourth player or for a two-player coalition to pillage one of the remaining players. The former action cannot be consistent with F because we have already shown that if only three players have positive wealth, there must be a further act of pillage, resulting in one of the three players in the original move eventually losing. The latter action is also ruled out because, as we have shown in the previous paragraph, strong maximality leads to a tyrannical outcome; one of the two players eventually loses. This completes the proof that $\bar{w} \in \Sigma(F)$ and $\Sigma(F) \neq D_1$.

To complete the proof that $B = \Sigma(F)$, consider any allocation $w \neq \bar{w}$ such that all four players have positive wealth. Without loss of generality, let $w_1 \geq w_2 \geq w_3 \geq w_4 > 0$ with $w_1 > w_4$. It follows that $w_1 + w_2 > 0.5$. It is sufficient to show that w cannot be stationary under F. If $w_1 < 0.5$, then $w_2 < 0.5$ as well, and $\{1, 2\}$ can pillage players 3 and 4 to arrive at the stable allocation $(0.5, 0.5, 0, 0)$. By internal stability, this implies that $w \notin \Sigma(F)$. If $w_1 \geq 0.5$, since $w_1 > w_4 > 0$, player 1 can pillage player 4 to move to $(w_1 + w_4, w_2, w_3, 0)$, where $w_1 + w_4 > 0.5$ and $w_1 + w_4 > w_2$. Now it is easy to see that $f^*(w_1 + w_4, w_2, w_3, 0) =$

[16]Note that if $w_1 = w_2$, either case is possible, but it would then be impossible for F to satisfy the even stronger maximality condition described in footnote 12. If $S(x) = \{1\}$, then $\{2\}$ would like to disrupt this move and become the coalition that pillages 3 and vice versa if $S(x) = \{2\}$.

$(1,0,0,0)$. This means that player 1's move to pillage 4 is strongly maximal. By internal stability, this must again imply that $w \notin \Sigma(F)$.

Matters are quite different as far as REFSs are concerned. It is possible to construct a rational expectations function F such that $\Sigma(F) = D_1$ is a REFS. The construction of F (see the proof of Theorem 5 for details) has the property that from every unequal allocation, the most powerful pillages a least powerful player, and this leads sequentially to a tyrannical allocation. The only remaining issue is to define F in such a way that $\bar{w} = (0.25, 0.25, 0.25, 0.25)$ is rendered unstable. This can be done by defining F as follows:

(i) We have $f(\bar{w}) = (0.375, 0.375, 0.25, 0) \equiv w'$ with $S(\bar{w}) = \{1,2\}$. That is, players 1 and 2 jointly pillage player 4 and divide $\bar{w}_4$ equally.

(ii) We have $f(w') = f^*(w') = (0.5, 0.5, 0, 0)$ with $S(w') = \{1,2\}$. Here, players 1 and 2 pillage player 3 and share w'_3 equally.

Note that at $\bar{w}$ or w', if $\{1,2\}$ did not divide their gains equally, at the next stage the more powerful of the two will pillage the remaining player(s) and obtain her tyrannical allocation. Thus, the less powerful of the two would not have joined in the act of pillage. This means that $f(\bar{w})$ and $f(w')$ describe a maximal move by $\{1,2\}$ in the sense of condition (M). It is now easy to see that D_1 is a REFS. But as we have already seen, it cannot be a SREFS.

We now turn to existence for the general case of an arbitrary number of players. For any positive integer n, let $k(n)$ be the largest integer such that $2^k \leq n$.

THEOREM 5. *For any positive integer $k^* \leq k(n)$, $\bigcup_0^{k^*} B_k$ is a REFS.*

Recall that $B = \bigcup_0^{k(n)} B_k$ and is therefore the largest REFS identified by Theorem 5. And there are many others, including the unique farsighted stable set $B_0 \cup B_1 = D_1$. In this model, therefore, unlike simple games, the farsighted stable set can be justified on the basis of consistent and rational expectations. It does not, however, meet the strong maximality test, as shown by Example 8. In other words, Theorem 5 cannot be strengthened to assert that $\bigcup_0^{k^*} B_k$ is a SREFS for all $k > 0$. However, as our next result shows (formalizing the message from Example 8), one of the REFS identified in Theorem 5 is a SREFS.

THEOREM 6. *Suppose the effectivity correspondence is as in (1). Then B is a SREFS.*

We close this section with a discussion of Acemoglu et al. (2008). They study a model of political coalition formation in which the power of each player is exogenously given. For each $i \in N$, $\gamma_i > 0$ denotes i's political power. The power of coalition S is $\gamma_S = \sum_{i \in S} \gamma_i$. Coalition $S \subseteq T$ is *winning in* T if $\gamma_S > \alpha \gamma_T$, where $\alpha \in [0.5, 1)$. Denote by $\mathcal{W}(T)$ the set of subsets of T that are winning in T. If such a coalition exercises its power, it captures the entire surplus and becomes *the ruling coalition*. The other players are eliminated and play no further role. However, the ruling coalition may itself be subject to a new round of power grab from within.

The distribution of wealth is determined through an exogenous rule that depends only on the identity of the ruling coalition. Assume that for every player it is better to be in a ruling coalition than not. Moreover, it is better to be in a ruling coalition with lower aggregate power. As Acemoglu et al. (2008) point out, a particular example of such a rule, which we adopt for the sake of simplicity, is

$$w_i(S) = \begin{cases} \gamma_i/\gamma_S & \text{if } i \in S, \\ 0 & \text{otherwise.} \end{cases}$$

A state can now be defined as the ruling coalition: at state S, the ruling coalition is S and the wealth distribution is $w(S)$. The set of states is therefore $\mathcal{N}$. Winning coalitions are the ones effective in changing a state:

$$S \in E(T, S) \quad \text{if and only if} \quad S \text{ is winning in } T.$$

This means that if a change occurs, the new ruling coalition is smaller. If such a change is expected to lead to a further change, then a winning coalition will choose *not* to exercise its power. This is a result of the fact that any further change must leave some member(s) of the original winning coalition with a payoff of 0. In other words, if there is a farsighted objection $S, S^1, \ldots, S^m$ leading from S to S^m, it must be the case that $m = 1$; farsighted dominance is equivalent to dominance.[17] Harsanyi (1974) refers to a farsighted dominance relation with this property as *trivial* and points out that if this property holds for every farsighted dominance of one state over another, then the vNM stable set is equivalent to the farsighted stable set.

Another feature of this model that makes it very tractable is that objections can only come from subsets of the ruling coalition (internal blocking). This makes it possible to construct a stable set recursively. Of particular interest in these models are the stable sets that can be reached from N, inlcuding possibly N itself. We illustrate this through the following example.

Example 9 (Four-player example of the Acemoglu et al. 2008 model). Suppose $N = \{1, 2, 3, 4\}$, $\gamma = (2, 4, 6, 8)$, and $\alpha = 0.5$. ◊

A vNM stable set can be constructed as follows. Any ruling coalition consisting of one individual clearly belongs to the stable set (it is in the core). Any ruling coalition consisting of two players is not in the stable set because the more powerful player will eliminate the weaker one; there is an objection leading to a stable state. Next, consider the three-player coalitions. The coalition $\{1, 2, 4\}$ is not stable because player 4 has enough power to eliminate the other two. Let the collection of the other three-player coalitions be denoted $\mathcal{S} = \{\{1, 3, 4\}, \{2, 3, 4\}, \{1, 2, 3\}\}$. It is easy to see that no coalition in $\mathcal{S}$ is threatened by a single powerful player. In each instance, two of the players have enough power to eliminate the third, but the resulting outcome is not stable, as we have

[17] Recall that in Jordan's pillage game, a farsighted objection could last several steps, although at each step it would be the same coalition making the change. This difference stems from the fact that in the Acemoglu et al. model, only the winning coalition survives to the next stage; there are no neutral players.

just noted. This means that all coalitions in $\mathcal{S}$ are stable and N is not; it will be replaced by one of these three-player coalitions. Thus, the stable set consists of singletons and the collection $\mathcal{S}$. In fact, this is also a farsighted stable set because farsighted dominance is equivalent to myopic dominance in this model. To verify this directly in this example, it is only necessary to establish the farsighted internal stability for states in $\mathcal{S}$. While two players could eliminate a third, this cannot result in a farsighted objection ending in a stable state because the weaker of the two will get eliminated at the next stage. We conclude that N is not stable and will be replaced by one of the coalitions in $\mathcal{S}$.[18]

Recall that farsighted internal stability is stronger than condition (I) of the REFS and myopic external stability is stronger than condition (E) of the REFS. In fact, the equivalence of farsighted dominance and myopic dominance also yields equivalence with the REFS. To see this, consider a set of states Z that is a stable set as well as a farsighted stable set. For every state $S \in Z$, let $F(S) = S$. (Because $X = \mathcal{N}$, we abuse notation slightly to consider F as a function from $\mathcal{N} \to \mathcal{N}$.) For $S \notin Z$, define $F(S)$ to be a subcoalition of S that dominates it (myopically) and belongs to Z. If there are several such coalitions, pick one arbitrarily. By construction, F satisfies (E). It satisfies (I) because Z is a farsighted stable set. Finally, it satisfies (M) because for every nonstationary state, S, it prescribes a move by coalition $F(S)$ that ends with $F(S)$. Since this is the *only* profitable move available to $F(S)$, it is trivially maximal.

Acemoglu et al. (2008) provide an axiomatic characterization of a solution to this model, which they refer to as the unique ruling coalition (URC), and also show that it coincides with the subgame perfect equilibria of a noncooperative model of coalition formation. The URC is a refinement of the REFS or the stable set. This difference turns out to hinge on the difference between REFS and SREFS. In fact, in this model SREFS refines REFS precisely to the URC.[19] This can be illustrated through Example 9. As explained above, in constructing a rational expectation F, we have the freedom to choose $F(N)$ to be any one of the three coalitions in $\mathcal{S}$. In particular, we could define $F(N) = \{2, 3, 4\}$. But players 3 and 4 could do better by forming $\{1, 3, 4\}$; recall that the payoff to a player is higher in a coalition with lower aggregate power. In other words, F does not satisfy (M′). In fact, strong maximality in this model, not just in Example 9, reduces to the condition that if S is not a stationary state, then $F(S)$ has the *lowest* aggregate power among all stable coalitions that are winning in S:

$$\text{If } S \notin \Sigma(F), \quad \text{then } F(S) \in \operatorname*{arg\,min}_{T \in \Sigma(F) \cap \mathcal{W}^*(S)} \gamma_T, \quad \text{where } \mathcal{W}^*(S) \text{ denotes } \mathcal{W}(S) - S.$$

If γ is generic in the sense that $\gamma_S \neq \gamma_T$ for any S, T, $S \neq T$, then clearly $F(S)$ is unique for every S. The unique strong rational expectation can be computed recursively as follows. Of course, $F(S) = S$ if $|S| = 1$. Suppose $F(S)$ has been defined for all S such that

[18] The singletons are also stable, but none of those states is reachable from N.

[19] Ray and Vohra (2014) show that their notion of an EPCF yields the same predictions as an REFS, but an appropriately chosen protocol is needed to sharpen the equilibria to coincide with the URC. See also Kimya (2015).

$|S| < k$. Then for S with $|S| = k$,

$$F(S) = \begin{cases} \underset{T \in \Sigma(F) \cap \mathcal{W}^*(S)}{\arg\min} \ \gamma_T & \text{if } \Sigma(F) \cap \mathcal{W}^*(S) \neq \varnothing, \\ S & \text{otherwise.} \end{cases}$$

The term $F(S)$ is the same as $\phi(S)$, the Acemoglu et al. (2008) notion of the URC for player set S. Thus we have yet another characterization of the URC: it is the unique SREFS. In this model, the farsighted stable set or REFS generally yields a set of outcomes that strictly includes the URC. Strong maximality refines it precisely to the URC.

6. Concluding remarks

This paper provides a framework for the analysis of cooperative solution concepts that combines farsightedness with the principles underlying the von Neumann–Morgenstern stable sets. We go beyond the recent work of Ray and Vohra (2015) by restricting coalitional moves to be *maximal*. We do so under the assumption that all coalitions hold *common* or *consistent* beliefs about the continuation path of future coalitional moves. Another interpretation of consistency is *history independence*: future moves from any state are independent of previous moves culminating in the current state. An interesting avenue for future research is to incorporate maximality without assuming history independence.

We model consistency through the use of an *expectation function*, which describes the transition from one state to another as well as the coalition that makes the move. Analogues of internal and external stability that define a stable set are easily incorporated in this framework. These and two versions of maximality then give rise to two different solution concepts: the rational expectations farsighted stable set (REFS) and the strong rational expectations farsighted stable set (SREFS).

In a series of examples, we show that these solution concepts lead to predictions that are intuitively more appealing than the Chwe largest consistent set or the Ray–Vohra farsighted stable set. However, Theorem 1 identifies one situation where SREFS, REFS, and the farsighted stable set coincide. This is the case in which there is a unique payoff at all states in the farsighted stable set. As shown in Ray and Vohra (2015), any payoff in the interior of the core of a characteristic function game can be identified as such a farsighted stable set. According to Theorem 1 it is also a SREFS.

We apply our solution concepts to two broad classes of games. The first is the class of proper simple games. We prove constructively that under a mild assumption, all games with veto players have a nonempty SREFS. The SREFS turns out to be quite different from the farsighted stable set identified by Ray and Vohra (2015), the main driver for the difference being consistency. Our second application is for the class of pillage games, first studied by Jordan (2006) to analyze cooperative situations without well defined property rights over resources. We show that both a REFS and a SREFS exists for all such games. Again, these solution concepts turn out to make predictions that are quite different from that made by the farsighted stable set. In this case, the main driving

force for the difference is maximality. In this model there is also an important difference between REFSs and SREFSs.

Finally, we show that the SREFS provides another characterization of the ultimate ruling coalition, a solution concept advanced by Acemoglu et al. (2008) in their analysis of political coalition formation.

Appendix

Proof of Theorem 1. Suppose Z is a single-payoff REFS. Since all stationary states have the same payoff, Z is clearly also a SREFS and satisfies farsighted internal stability in the definition of a farsighted stable set. Moreover, condition (E) implies farsighted external stability in the definition of a farsighted stable set. Thus Z is a farsighted stable set.

To prove the second part of the theorem, consider a single-payoff farsighted stable set X^0. Define X^1 to be the set of states from which there is a farsighted objection to some state in X^0 in a single step. More precisely,

$$X^1 = \{x \in X - X^0 \mid \exists x^0 \in X^0, S \in E(x, x^0) \text{ with } u_S(x^0) \gg u_S(x)\}.$$

Since X^0 is a farsighted stable set, from every $x \in X - X^0$, there is a farsighted objection leading to some $x^0 \in X^0$: $x, (x^1, S^1), \ldots, (x', S'), (x^m, S^m)$. Clearly $x' \in X^1$, which establishes the nonemptiness of X^1. We now recursively define subsets of X from which there are farsighted objections leading to X^0 in a minimal number of steps. All of these sets are disjoint and cover X. The construction is as follows.

Suppose X^j is defined for all $j = 1, \ldots, k$. Define X^{k+1} to be the set of all other states from which there is a farsighted objection leading to X^0 such that the first step is a state in X^k:

$$X^{k+1} = \left\{ x \in X - \bigcup_{j=0}^{k} X^j \;\middle|\; \text{there is a farsighted objection } x, (x^1, S^1), \ldots, (x^m, S^m), \right.$$
$$\left. \text{with } x^1 \in X^k \text{ and } x^m \in X^0 \right\}.$$

Note that if $X^{k+1} = \varnothing$, then $\bigcup_{j=0}^{k} X^k = X$. To complete the proof, we construct a function $F : X \to X \times \mathcal{N}$, where $F(x) = (f(x), S(x))$ such that $f(x^0) = x^0$ for every $x^0 \in X^0$ and, for every $x \in X^{k+1}$, $f(x) \in X^k$. We know that from $x \in X^{k+1}$ there is a farsighted objection leading to some state in X^0 that proceeds by first moving to a point in X^k. We choose $f(x)$ as one such point along with a unique coalition that initiates such a farsighted objection. The function F is constructed recursively. For $x \in X^1$, define $f(x) = x^0 \in X^0$ and $S(x)$ to be a coalition that has a one step objection from x to x^0. If there are multiple such coalitions, choose one arbitrarily. This describes a unique transition from X^1 to X^0. Having defined $F : X^j \mapsto X^{j-1} \times \mathcal{N}$ for all $j = 1, \ldots, k$, if $X^{k+1} \neq \varnothing$,

for $x \in X^{k+1}$ let S^1 be a coalition that has a farsighted objection from x to $x^m \in X^0$, denoted $x, (x^1, S^1), \ldots, (x^m, S^m)$, such that $x^1 \in X^k$. Let $F(x) = (x^1, S^1)$. Note that there may be multiple such farsighted objections. In that case, pick $F(x)$ to be the first element of any such sequence. Proceeding in this way, we have constructed a function $F : X \to X \times \mathcal{N}$ with X^0 as its set of stationary points. It remains to be shown that F is a strong rational expectation.

Since the stationary points of F have the same payoff vector, it trivially satisfies condition (I) in the definition of a rational expectation.

To prove condition (E), consider $x \in X - X^0$. Of course, there is some $k \geq 0$ such that $x \in X^{k+1}$. From the construction of F, we know that there exists a farsighted objection from x to $x^m \in X^0$, say, $x, (x^1, S^1), (x^2, S^2), \ldots, (x^m, S^m)$, such that $F(x) = (x^1, S^1)$. (There is no presumption that $(x^2, S^2) = F^2(x)$ or that $m = k$.) Let u^0 denote the (common) payoff corresponding to each of the (single-payoff) states in X^0. Obviously, $u_{S^1}(x^m) = u^0_{S^1} \gg u_{S^1}(x)$. Since $f^k(x) \in X^0$, $u(f^k(x)) = u^0$, which implies that S^1 gains in moving along the path $(x, F(x), F^2(x), \ldots, F^k(x))$. By the same reasoning, S^2 also gains by moving from $f(x)$ along the path $(F^2(x), \ldots, F^k(x))$, and so on for all S^j, $j = 1, \ldots, S^k$. Thus, $(x, F(x), F^2(x), \ldots, F^k(x))$ is a farsighted objection from x to $f^k(x) \in X^0$.

To see that condition (M′) is satisfied, note that no player, and therefore no coalition, can gain by deviating from the path prescribed by F because any deviation leads to the same payoff vector, u^0. This establishes condition (M′) and completes the proof that F is a strong rational expectation with $\Sigma(F) = Z$. □

The proof of Theorem 2 makes use of the following lemma.

LEMMA 1. *There exists a positive number $d < 1/|C|$, a vector $b \in R^{N-C}_+$, and a nonempty collection of coalitions $\mathcal{J}$ in $N - C$ such that the following statements hold:*

(i) For every $J \in \mathcal{J}$, $C \cup J$ is a minimal winning coalition, $b_J \gg 0$ and $\sum_{j \in J} b_j = \epsilon \equiv 1 - d|C|$.

(ii) There does not exist a winning coalition $C \cup T'$ such that $\sum_{j \in T'} b_j < \epsilon$.

(iii) We have $1 > \sum_{j \in N-C} b_j > \epsilon$.

PROOF. Let $\mathcal{J}' = \{J \subset N - C \mid C \cup J \text{ is a minimal winning coalition}\}$. Without loss of generality, assume that the coalitions in $\mathcal{J}' = \{J^1, \ldots, J^K\}$ are ranked in nondecreasing order of cardinality, so $|J^k| \leq |J^{k+1}|$ for all $k = 1, \ldots, K-1$.

Choose $\epsilon \in (0, 1)$ such that $\epsilon < |J^1|/|N - C|$ and let $d \equiv (1 - \epsilon)/|C|$. We now construct an algorithm that yields b and $\mathcal{J} \subseteq \mathcal{J}'$ satisfying the desired properties.

Let J^f, $f > 1$, be the first coalition in $\mathcal{J}'$ that has a nonempty intersection with some J^k, $k < f$. If no such f exists, then set $f = K + 1$ and $J^{K+1} = \varnothing$.

Step 1. For all $k < f$, let $b_i = \epsilon/|J^k|$ for all $i \in J^k$. Note that for all $k, k' < f$, $J^k \cap J^{k'} = \varnothing$, so this construction is well defined. Clearly, b and J^k satisfy condition (i) of the lemma for all $k < f$. Moreover, since the coalitions in $\mathcal{J}'$ are in nondecreasing order of cardinality,

$$\text{for any } k < k' < f, i \in J^k, \text{ and } j \in J^{k'}, \quad \text{we have } \bar{b} = \frac{\epsilon}{|J^1|} \geq b_i \geq b_j > 0. \tag{2}$$

Define $A^1 = \bigcup_{k=1}^{f-1} J^k$. Then, for every $i \in A^1$, b_i as defined above is the "terminal" value. For $i \notin A^1$, we construct b_i iteratively.

For every $k \geq f$, we recursively define nonnegative numbers t_i^1 for all $i \in J^k - A^1$ as follows. Suppose t_j^1 have been defined for all $j \in G^k \equiv \bigcup_{j=1}^{k-1} J^j - A^1$. If $J^k - (G^k \cup A^1) \neq \varnothing$, for every $i \in J^k - (G^k \cup A^1)$ let

$$t_i^1 = \max\left[0, \frac{\epsilon - \sum_{j \in J^k \cap A^1} b_j - \sum_{j \in J^k \cap G^k} t_j^1}{|J^k - (G^k \cup A^1)|}\right].$$

Since there are no dummy players, every $i \in N - C$ belongs to at least one coalition in $\mathcal{J}'$, and has therefore been assigned a nonnegative number b_i or t_i^1. It follows from (2) that for all $i \notin A^1$, $t_i^1 \leq \bar{b}$.

Let $\mathcal{J}^1 = \{J \in \mathcal{J}' | \sum_{i \in J \cap A^1} b_i + \sum_{i \in J - A^1} t_i^1 < \epsilon\}$.

If $\mathcal{J}^1 = \varnothing$, terminate the algorithm, set $b_i = t_i^1$ for all $i \notin A^1$, and go to Step 3.

Suppose $\mathcal{J}^1 \neq \varnothing$. Since $J \in \mathcal{J}^1$ implies that $|J| \geq |J^k|$ for all $k < f$ and $J \in \mathcal{J}^1$, it follows from (2) that for every $J \in \mathcal{J}^1$, $J - A^1 \neq \varnothing$. Let J^{k_1} be a coalition in $\mathcal{J}^1$ that maximizes $(\epsilon - \sum_{j \in J^k \cap A^1} b_j)/|J^k - A^1|$, ties being broken arbitrarily. For each $i \in J^{k_1} - A^1$, set

$$b_i = \frac{\epsilon - \sum_{j \in J^{k_1} \cap A^1} b_j}{|J^{k_1} - A^1|}.$$

Since $J^{k_1} - A^1 \neq \varnothing$, $A^2 \equiv A^1 \cup J^{k_1}$ is a strict superset of A^1.

Step 2. Since some of the components of the t vector have increased, the remaining components may not be feasible. So now repeat Step 1 with A^2 replacing A^1.

For $k \geq f$, $k \neq k_1$, we define t_i^2 recursively as follows. Suppose t_i^2 have been defined for all $j \in G'^k \equiv \bigcup_{j=1}^{k-1} J^j - A^2$. If $J^k - (G'^k - A^2) \neq \varnothing$, for every $i \in J^k - (G'^k \cup A^2)$ let

$$t_i^2 = \max\left[0, \frac{\epsilon - \sum_{j \in J^k \cap A^2} b_j - \sum_{j \in J^k \cap G'^k} t_j^2}{|J^k - A^2|}\right].$$

As in Step 1, it follows from (2) that $t_i^2 \leq \bar{b}$ for all $i \notin A^2$. Let $\mathcal{J}^2 = \{J \in \mathcal{J} | \sum_{i \in J \cap A^2} b_i + \sum_{i \in (J - A^2)} t_i^2 < \epsilon\}$. By construction, $J^{k_1} \notin \mathcal{J}^2$. If $\mathcal{J}^2$ is empty, terminate the algorithm with $b_i = t_i^2$ for all $i \in N - (C \cup A^2)$ and move to Step 3. Otherwise, choose the coalition J^{k_2} that maximizes $(\epsilon - \sum_{j \in J^k \cap A^2} b_j)/|J^k - A^2|$ in this set. For each $i \in J^{k_2} - A^2$, set

$$b_i = \frac{\epsilon - \sum_{j \in J^{k_2} \cap A^2} b_j}{|J^{k_2} - A^2|}.$$

Claim 1. *The term $J^{k_2} - A^2$ is nonempty.*

Suppose not. Since $J^{k_2} \in \mathcal{J}^2$ and, by hypothesis, $J^{k_2} \subseteq A^2$, then $\sum_{i\in J^{k_2}\cap A^2} b_i = \sum_{i\in J^{k_2}\cap A^1} b_i + \sum_{i\in S} t_i^1 < \epsilon$. Recall that for $i \in S \subseteq J^{k_1} - A^1$, $t_i^1 = (\epsilon - \sum_{i\in A^1\cap J^{k_1}} b_i)/|J^{k_1} - A^1|$. But this means that

$$\frac{\epsilon - \sum_{i\in J^{k_2}\cap A^1} b_i}{|S|} > \frac{\epsilon - \sum_{i\in J^{k_1}\cap A^1} b_i}{\left|J^{k_1} - A^1\right|},$$

which contradicts the choice of J^{k_1}. Hence the claim is true and $A^3 \equiv A^2 \cup J^{k_2}$ is a strict superset of A^2.

Since the sets A^k are strictly increasing over stages, the algorithm terminates.

Step 3. At this stage we have constructed b such that $b_i \leq \bar{b}$ for all $i \in N - C$, and for all $J \in \mathcal{J}'$, $d|C| + \sum_{i\in J} b_j \geq 1$. Clearly then condition (ii) of the lemma holds. Define $\mathcal{J} = \{J \in \mathcal{J}' \mid b_J \gg 0 \text{ and } \sum_{i\in J} b_i = \epsilon\}$, which is nonempty because $J^k \in \mathcal{J}$ for all $k < f$. Of course, $\mathcal{J}$ satisfies condition (i).

Since $b_i \leq \bar{b}$ for all $i \in N - C$, then $\sum_{i\in N-C} b_i \leq |N - C|\bar{b} = (|N - C|\epsilon)/|J^1| < 1$. To establish condition (iii) of the lemma, it remains to show that $\sum_{j\in N-C} b_j > \epsilon$. Recall that for every $J \in \mathcal{J}$, $\sum_{i\in J} b_i = \epsilon$, which implies that this condition clearly holds if $f > 2$. Suppose $f = 2$, i.e., $J^1 \cap J^2 \neq \varnothing$. Since J^1 and J^2 are minimal winning coalitions, neither is a subset of the other. By construction, $b_i > 0$ for all $i \in J^1 \cup J^2$. This together with the fact that $\sum_{i\in J^1} b_1 = \epsilon$ implies that $\sum_{i\in N-C} b_i \geq \sum_{i\in J^1\cup J^2} b_i > \epsilon$. □

Proof of Theorem 2. We construct a SREFS for two distinct cases. The first is the case, as in Example 7, where minimal winning coalitions consist of C and any one of the non-veto players. For the second case, in which there is at least one minimal winning coalition with two or more non-veto players, our construction relies on Lemma 1 and Assumption 2.

Case 1: $C \cup \{j\} \in \mathcal{W}$ for all $j \notin C$.

Let $a > 0$ and $b > 0$ be such that $|C|a + |N - C|b = 1$. Define $\hat{x}$ so that $\pi(\hat{x}) = N$ and

$$u_i(\hat{x}) = \begin{cases} a & \text{if } i \in C, \\ b & \text{if } i \notin C. \end{cases}$$

For every $j \notin C$, let X^j be the set of all states x in which the winning coalition contains $C \cup \{j\}$ and the payoff vector has the property that $u_i(x) \geq a$ for all $i \in C$, $u_j(x) = b$, and $u_k(x) = 0$ for all $k \notin C \cup \{j\}$.

We claim that $Z = \bigcup_{j\notin C} X^j \cup \{\hat{x}\}$ is a SREFS: it is the set of stationary points of a strong rational expectation F satisfying the following properties:[20]

P1.1. For $x \in Z$, $f(x) = x$.

P1.2. For $x \in X^0$, $S(x) = N$ and $f(x) = \hat{x}$.

[20]In what follows it is understood that $\pi(f(x))$ is the immediate change in $\pi(x)$ resulting from the formation of $S(x)$, as formalized in Assumption 1(b).

The remainder of the rules for F relate to $x \in X - Z - X^0$.

P1.3. For x such that $u_j(x) < b$ for all $j \in N - C$, $S(x) = N - C$ and $f(x) \in X^0$.

P1.4. For x such that $u_j(x) \geq b$ for all $j \in N - C$, since $x \notin Z$, there must be a veto player i for whom $u_i(x) < a$. Let $\{i'\}$ be the lowest indexed player of this kind. In this case we have $S(x) = \{i'\}$ and $f(x) \in X^0$.

P1.5. For x such that $u_j(x) \geq b$ for some non-veto player but not all, let j' be the lowest indexed non-veto player such that $u_{j'}(x) < b$. There are now two distinct cases for describing F:

(a) Suppose $s(x) \equiv 1 - b - \sum_{i \in C} \max\{a, u_i(x)\} > 0$. Then $S(x) = C \cup \{j'\}$, $u_i(f(x)) = \max\{a, u_i(x)\} + s/|C|$, and $u_{j'}(f(x)) = b$. Note that $f(x) \in X^{j'}$.

(b) Suppose $s(x) \leq 0$. In this case, unlike the previous one, it is not possible to construct an objection that leads to Z in one step. However, it must be the case that there is a veto player i for whom $u_i(x) < a$. Otherwise, $s(x) = 0$, which contradicts the supposition that $x \notin Z$. In this case, as in P1.4, $S(x) = \{i'\}$ and $f(x) \in X^0$.

For $x \notin Z$, $f^*(x) = \hat{x}$ in all cases except P1.5(a). It is easy to see that in every instance all the players in $S(x)$ prefer $f^*(x)$ to x, which means that F satisfies (E).

Since all non-veto players weakly prefer $\hat{x}$ to any other state in Z and since a winning coalition must include at least one such player, it follows that (M′) is satisfied in case P1.2. The same reasoning applies to case P1.3. In case P1.4, player i' cannot construct another farsighted objection as none of the non-veto players are interested in moving, which implies that this move is strongly maximal. In case P1.5(a), there is no state in Z that all the veto players prefer to $f(x)$ and so (M′) holds. Finally, in case P1.5(b), it is clear that player i' cannot construct an objection leading to any other state in Z. Thus, in all cases (M′) is satisfied.

To see that F satisfies (I), consider a possible farsighted move from $\hat{x}$ that ends at some $x \in X^j$. Since all non-veto players weakly prefer $\hat{x}$ to x, none of them can be part of the first move from $\hat{x}$. But then the first move must lead to a state in X^0, which only results in returning to $\hat{x}$, making it impossible for the initiating coalition to gain. Next, consider a move from $x \in X^j$ that ends up at $\hat{x}$. Since all players in $C \cup \{j\}$ weakly prefer x to $\hat{x}$, the first move must come from non-veto players other than j. But then Assumption 1(c) on the effectivity correspondence implies that the payoff remains $u(x)$ through the entire sequence of moves, contradicting the supposition that the state eventually becomes $\hat{x}$. Finally, consider the possibility of a farsighted move that leads from $x^j \in X^j$ to some $x^k \in X^k$, $k \neq j$. Again, by Assumption 1(c), the first coalition in such a move cannot be from a coalition that is disjoint from $C \cup \{j\}$. Moreover, such a coalition cannot include j, since j prefers x^j to x^k. It cannot include C since it is impossible for all $i \in C$ to gain in moving from x^j to x^k. The only remaining possibility is that it includes some strict subset of C. But that leads to a state in X^0, from which the final outcome is $\hat{x}$, not x^k.

This completes the the proof of Case 1.

Case 2: There exists a minimal winning coalition $C \cup J$ such that $J \subset N - C$ and $|J| \geq 2$.

Let d, b, and $\mathcal{J}$ be as in Lemma 1. Define $a \equiv (1 - \sum_{i \in N-C} b_i)/|C|$. By condition (iii) of Lemma 1, $a \in (0, d)$.

Let $S^* = \bigcup_{J\in\mathcal{J}} J$ and let $N^* = C \cup S^*$. Since $C \cup J$ is a winning coalition for every $J \in \mathcal{J}$, clearly N^* is a winning coalition.

Define $\hat{X}$ to consist of all states x such that $W(x) \supseteq N^*$ and

$$u_i(x) = \begin{cases} a & \text{if } i \in C, \\ b_i & \text{if } i \notin C. \end{cases}$$

Corresponding to each $J^k \in \mathcal{J}$, define X^k as the set of states in which the winning coalition contains $C \cup J^k$ and the payoff vector corresponding to $x^k \in X^k$ is

$$u_i(x^k) = \begin{cases} d & \text{if } i \in C, \\ b_i & \text{if } i \in J^k, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\bar{X}$ be the set of all (nonzero) states x such that

$$u_i(x) = \begin{cases} d & \text{if } i \in C, \\ b_i & \text{if } i \in W(x) - C, \\ 0 & \text{otherwise.} \end{cases}$$

Of course, corresponding to every $J^k \in \mathcal{J}$, $X^k \subseteq \bar{X}$. However, $\bar{X}$ may also include a state x with $W(x) = C \cup K$ and $K \notin \mathcal{J}$ because b_K is not *strictly* positive.

We claim that $Z = \bar{X} \cup \hat{X}$ is a SREFS.

To prove this, we construct a rational expectations function F with the following properties: (a) $f(x) = x$ for all $x \in Z$, (b) for $x \in X - Z - X^0$, $f(x) \in X^0$, and (c) for $x \in X^0$, $f(x) \in Z$, depending on the nature of $\pi(x)$.

Let $T = \{i \in N - C \mid C \cup \{i\} \in \mathcal{W}\}$. Note that if $i \in T$, then $\{i\} \in A^1$ as constructed in the proof of Lemma 1. Moreover, given that we are considering Case 2, A^1 also includes at least one coalition J such that $|J| \geq 2$. This means that T is a strict subset of S^*. It is also easy to see from the proof of Lemma 1 that $\mathcal{J}$ includes at least two distinct coalitions, which implies that $|S^*| \geq 3$.

To describe the transition from zero states, we partition X^0 into *three* disjoint sets: X_1^0 is the set of all zero states in which the coalition structure contains precisely *one* two-player coalition consisting of one player from C and the other from S^*; X_2^0 is the set of zero states containing precisely *two* two-player coalitions, $\{i, j\}$ and $\{i', k\}$ such that $i, i' \in C$, $j \in T \subset S^*$ and $k \in S^* - T$; the set of all other zero states is denoted X_3^0.

For each $k \in S^*$, pick a unique $J(k) \in \mathcal{J}$ that contains k. The existence of such $J(k)$ follows from condition (i) of Lemma 1. With some abuse of notation, let X^k refer to the set of states in which the winning coalition is $C \cup J(k)$, and the payoff is d for all $i \in C$ and b_j for all $j \in J(k)$.

We now provide a complete description of F.

P2.1. For $x \in Z$, $f(x) = x$.

P2.2. For $x \in X_3^0$, $S(x) = N^*$ and $f(x) \in \hat{X}$.

P2.3. For $x \in X_1^0$, let (i,k) with $i \in C$ and $k \in S^*$ be the unique two-player coalition of this form in $\pi(x)$. Let $S(x) = C \cup J(k)$ and $f(x) \in X^k$.

P2.4. For $x \in X_2^0$, let (i,k) with $i \in C$ and $k \in S^* - T$ be the unique pair of this form in $\pi(x)$. Let $S(x) = C \cup J(k)$ and $f(x) \in X^k$.

The remainder of the rules for F relate to $x \in X - Z - X^0$.

P2.5. There is $k \in S^*$ such that $u_k(x) < b_k$ and $i \in C$ such that $u_i(x) < d$.

There are three subcases to consider:

(a) Either $|C| \neq 2$ or $W(x) = C \cup J$, where $|J| \geq 3$.

Let H be the two-player coalition consisting of the lowest indexed player $i \in C$ with $u_i(x) < d$ and the lowest indexed player $k \in S^*$ such that $u_k(x) < b_k$. Define $S(x) = H$. If H is a winning coalition, let $f(x) \in X^k$. Otherwise, $x' = f(x) \in X^0$. Now, since $|C| \neq 2$ or $|J| \geq 3$, invoking Assumption 1(b), H is the unique coalition in $\pi(x')$ consisting of one player from C and another from S^*. Thus, $x' \in X_1^0$ and $f(x') \in X^k$ with $u_i(f(x')) = d > u_i(x)$ and $u_k(f(x')) = b_k > u_k(x)$. Thus, H has a farsighted objection to x that leads to a state in X^k.

(b) We have $|C| = 2$ and $W(x) = C \cup J$, where $|J| = 2$.

Since $|S^*| \geq 3$ and $|J| = 2$, there exists $k \in S^* - J$. Since $k \notin W(x)$, $u_k(x) = 0$. Without loss of generality, let k be the lowest indexed player in $S^* - J$. Let $S(x) = H = \{i,k\}$. Clearly, for $x' = f(x)$, H is the unique coalition in $\pi(x')$ with one player from C and another from S^*. Thus, $x' \in X_1^0$ and, as in the previous paragraph, $f^*(x) \in X^k$.

(c) We have $|C| = 2$ and $W(x) = C \cup \{j\}$ for some $j \in T$.

Recall that T is a strict subset of S^*, i.e., there exists $k \in S^* - T$. Of course $k \neq j$ and $u_k(x) = 0 < b_k$. Let i be the lowest indexed player in C such that $u_i(x) < d$ and let $S(x) = H = \{i,k\}$. Note that $x' = f(x) \in X_2^0$, with H as the unique two-player coalition in $\pi(x')$ with one player from C and another from $S^* - T$. This means that $f(x') \in X^k$.

P2.6. Suppose $x \notin Z$ is such that there is $k \in S^*$ with $u_k(x) < b_k$ and $u_i(x) \geq d$ for all $i \in C$. We now define $S(x)$ to ensure that $f(x) \in X_3^0$.

Since $u_i(x) \geq d$ for all $i \in C$, we must have $R \equiv \{j \in W(x) - C \mid u_j(x) < b_j\} \neq \varnothing$. Otherwise, by (ii) of Lemma 1, $u_i(x) = d$ for all $i \in C$ and $u_j(x) = b_j$ for all $j \in W(x) - C$, which means that $x \in \bar{X}$, a contradiction to the hypothesis that $x \notin Z$. Let R' be a minimal subset of R such that $W(x) - R'$ is not winning; i.e., $W(x) - R'$ is not winning but the addition of any single player, j, from R' would make $W(x) - (R' - \{j\})$ a winning coalition. By Assumption 2, this winning coalition cannot be of the form $\{i,j,k\}$, where i is a veto player and j and k are non-veto players. This must mean that $W(x) - R'$ cannot consist of precisely one veto player and one non-veto player. Moreover, since $W(x) - R' \notin \mathcal{W}$, with $S(x) = R'$, $f(x) \in X_3^0$.

P2.7. If x is such that $u_j(x) \geq b_j$ for all $j \in S^*$, since $x \notin Z$, there must be a veto player i for whom $u_i(x) < a$. Let $\{i'\}$ be the lowest indexed player of this kind and let $S(x) = \{i'\}$. Since $|S^*| \geq 3$, by Assumption 1(b), $f(x) \in X_3^0$.

This completes the description of F. Note that in each case $x \notin Z$, the expectation leads in at most two steps to a state in Z. It is also easy to see that all players in $S(x)$ strictly gain in moving from x to $f^*(x)$. Thus F satisfies (E).

To show that F satisfies (M′), it is useful to note that a non-veto player $j \in N - C$ receives either b_j or 0 in Z while a veto player i receives either a or $d > a$. In cases P2.3,

P2.4, and P2.5, the move from x is initiated by a coalition of the form $\{i, k\}$ with $i \in C$ and $k \notin C$ leading either directly or in two steps to a state in X^k. Since neither i nor k prefers any other state in Z, strong maximality holds in all these cases. In case P2.2, $S(x)$ includes all non-veto players and every such j receives b_j at $f^*(x) = f(x)$. Thus, none of them can do better as part of some other objecting coalition, implying that (M′) is satisfied. The same argument also applies to case P2.6. In case P2.5, non-veto players have no reason to join any deviating coalition, and so any objection coalition must be a subset of C. Since $|S^*| \geq 3$, Assumption 1(b) implies that non-veto players can only move to X_3^0 (and then to $\hat{X}$). Thus, the move by $\{i'\}$ is strongly maximal.

Finally, we show that F satisfies (I).

Take any $\bar{x} \in \bar{X}$ and $\hat{x} \in \hat{X}$. Then all $i \in W(\bar{x})$ weakly prefer $\bar{x}$ to $\hat{x}$. So if there is a farsighted objection from $\bar{x}$ to $\hat{x}$, and K is the first coalition to move, then $K \subseteq N - W(\bar{x})$. From Assumption 1(c), it follows that if $K \in E(\bar{x}, x)$, then $u_i(x) \geq u_i(\bar{x})$ for all $i \in W(\bar{x})$. Repeated application of this argument rules out any farsighted objection. Notice that an identical argument ensures that there cannot be a farsighted objection from a state in $\bar{X}$ to another state in $\bar{X}$. Obviously there cannot be a farsighted objection from a state in $\hat{X}$ to another state in $\hat{X}$ since the payoff vector for all states in $\hat{X}$ is unique.

Finally, consider the possibility of a farsighted objection from $\hat{x}$ to $\bar{x}$, where $\hat{x} \in \hat{X}$ and $\bar{x} \in \bar{X}$. All members of $N - C$ weakly prefer $\hat{x}$ to $\bar{x}$. So the first deviation must come from some subset of C. Since $|S^*| \geq 3$, Assumption 1(b) implies that this leads to a state x in X_3^0, and $f(x) = \hat{x}$. This establishes that Z satisfies (I) and completes the proof that Z is a SREFS. □

Proof of Theorem 4. Suppose Z is a farsighted stable set. It is obvious that no players have the power to beneficially change a tyrannical allocation since one player has already captured the entire surplus. It must therefore belong to every farsighted stable set (as well as to every REFS). It is easy to see that allocations where two players get 0.5 are also stable in this sense. Thus, $D_1 \subseteq Z$.[21]

To complete the proof, we now show that for every $w \notin D_1$, there is a farsighted objection that terminates in D_1. There are two cases:

(i) A state $w \notin D_1$ is such that $w_i = w_j$ for all i, j such that $w_i > 0$, $w_j > 0$. This means that there are k players who receive $1/k$, where $k \geq 3$. Suppose two such players, say i and j, pillage a third and share the spoils equally. This increases the power of i and j. If there are any other players remaining with $1/k$, then i and j pillage one such player in the next step. This process continues until we arrive at an allocation in D_1 where i and j get 0.5 each. This is clearly a farsighted objection.

(ii) There are i and j such that $w_i > w_j > 0$. Let i' be a player such that $w_{i'} \geq w_i$ for all i. Of course, i' can pillage a player with lower wealth. This results in i' becoming more powerful, and she can now pillage any other player j, with $w_j > 0$, if there is any. Through this process of sequential pillaging, i' can achieve the tyrannical allocation in which she has the entire wealth. This describes a farsighted objection, leading from w to a tyrannical allocation in D_1. □

[21]The fact that the core is a subset of the farsighted core is a feature of pillage games. In general, it is possible that the core is disjoint from every farsighted stable set or REFS; recall Example 7.

PROOF OF THEOREM 5. We construct an expectation F as follows:

(i) Suppose w is such that $w_i > w_j > 0$ for some i, j. Let i' be the lowest indexed player such that $w_{i'} \geq w_i$ for all i and let j' be the lowest indexed player such that $w_{j'} < w_{i'}$. Then the expectation is that i' pillages j': $f(w) = w'$, where $w'_{i'} = w_{i'} + w_{j'}$, $w'_{j'} = 0$, and $w_k = w'_k$ for all $k \neq i', j'$, and $S(w) = \{i'\}$. Note that $f^*(w)$ is the tyrannical allocation where i' gets the entire wealth.

(ii) Suppose w is such that all players with positive wealth have the same wealth but this is not 2^{-k} for any integer k. In other words, m players get $1/m$ but $m \neq 2^k$ for any integer k. Let $\hat{k}$ be the largest k such that $2^k < m$. Then $f(w) \in B_{\hat{k}}$ and $S(w)$ is the coalition consisting of the lowest indexed $2^{\hat{k}}$ players getting $1/m$ at w. Note that $S(w)$ has the power to make this move since the total wealth of this coalition at w is $2^{\hat{k}}/m = 2^{\hat{k}+1}/2m > \frac{1}{2}$.

(iii) For $w \in B$, $f(w) = w$.

We have constructed F such that $\Sigma(F) = B$. It remains to be shown that F satisfies conditions (I), (E), and (M′).

Suppose F does not satisfy (I). Then there exists $w \in B_k$ and $S \in E(w, w')$ such that $f^*(w') = w'' \in B_{k'}$ and $w''_S \gg w_S$. The last inequality implies that $k' < k$.

First, suppose that w' is such that $w'_i > w'_j$ for some pair i, j. Then w'' is a tyrannical allocation, so that $|S| = 1$. But then $S \notin E(w, w')$ since $w_i = w_j$ if $w_i, w_j > 0$.

So w' must satisfy $w'_i = w'_j$ if $w'_i, w'_j > 0$. Also, since E satisfies the equation (1), $S = \{i | w'_i > w_i\}$. Putting these together, we must have $w' = w''$; that is, $w' \in B_{k'}$.

Since $w'_i > 0$ implies that $w'_i = 2^{-k'}$ and $w_i > 0$ implies that $w_i = 2^{-k}$, this means that $w'_i \geq 2w_i$ for $i \in S$—those with positive wealth at w' must have at least twice as much as they did at w. Since the added wealth must have been pillaged, those who were pillaged must have had at least as much wealth at w as the pillagers. So

$$\sum_{\{i \in S\}} w_i \leq \sum_{\{i : w'_i = 0\}} w_i.$$

This implies that $S \notin E(w, w')$.

To see that (E) is satisfied, consider $w \notin B$. If w is covered by case (i), the only coalition that moves at each step is the singleton consisting of the lowest indexed player with the highest wealth at w, and, at each step, this coalition does better by eventually attaining the tyrannical allocation. Thus (E) holds for w in case (i). For w covered by case (ii), $S(w)$ moves in one step to a stationary allocation, which is an improvement since it involves equal sharing in a smaller coalition, and condition (E) is therefore satisfied.

We now turn to condition (M′). For w in case (i), maximality is trivially satisfied since the singleton that moves ultimately achieves the tyrannical allocation. Suppose w is covered by case (ii) and (M′) is not satisfied. This means that there is a coalition T with $T \cap S(w) \neq \varnothing$ that does better than $f(w)$. Since all stationary allocations satisfy equal division among players with positive wealth, $|T| < |S(w)|$. Recall that $|S(w)| = 2^{\hat{k}}$, which implies that if $|T| = 2^{k'}$ for some integer $k' < \hat{k}$, then T does not have the power to change w. Thus, $T \neq 2^{-k}$ for some positive integer k, in which case the final outcome according to F results in a smaller coalition and some player in T gets 0. This contradicts the supposition that T can do better than $f(w)$. □

PROOF OF THEOREM 6. Choose any positive integer $k^* \leq k(n)$ and define $B(k^*) \equiv \bigcup_0^{k^*} B_k$.

For any w, define $H(w) = \{i \in N : w_i \geq w_j \ \forall j \in N\}$ and let $\bar{H}(w)$ be the subset of $H(w)$ consisting of the 2^k-lowest indexed players in $H(w)$, where k is the largest integer not exceeding k^* with $2^k \leq |H(w)|$.

Given k^*, F is defined as follows.

(i) If $w \in B(k^*)$, then $f(w) = w$.

(ii) If $w \notin B(k^*)$ and $|H(w)| \geq 2^{k^*}$, let j' be the lowest indexed agent not in $\bar{H}(w)$ with $w_{j'} > 0$. Such j' must exist since $w \notin B(k^*)$. Then $S(w) = \bar{H}(w)$ and $f(w) = w'$, where

$$w'_i = \begin{cases} w_i + \dfrac{w_{j'}}{|\bar{H}(w)|} & \text{if } i \in \bar{H}(w), \\ w_i & \text{if } i \notin \bar{H}(w) \cup \{j'\}, \\ 0 & \text{if } i = j'. \end{cases}$$

(iii) If $w \notin B(k^*)$, $|H(w)| < 2^{k^*}$, and there is a pair i, j with $w_i > w_j > 0$, then the lowest indexed agent $i^* \in H(w)$ pillages the lowest indexed agent $j^* \notin H(w)$. So $S(w) = \{i^*\}$ and $f(w) = w'$, where

$$w'_i = \begin{cases} w_i + w_{j^*} & \text{if } i = i^*, \\ w_i & \text{if } i \neq i^*, j^*, \\ 0 & \text{if } i = j^*. \end{cases}$$

(iv) If $w \notin B(k^*)$, $|H(w)| < 2^{k^*}$ and if $H(w) = \{i | w_i > 0\}$, then $S(w) = \bar{H}(w)$ and $f(w) = w'$, where

$$w'_i = \begin{cases} \dfrac{1}{|\bar{H}(w)|} & \text{if } i \in \bar{H}(w), \\ 0 & \text{if } i \notin \bar{H}(w). \end{cases}$$

We note that in (iv) above, $|\bar{H}(w)| > |(H(w) - \bar{H}(w))|$ and so $\bar{H}(w)$ can pillage the rest, and hence f is well defined.

The proof that F satisfies (I) is virtually identical to that in Theorem 5, and we only give a very short sketch of the proof. Again, suppose $w \in B_k$ with $k \leq k^*$ and some S has a farsighted objection ending in $B_{k'}$, where $k' < k$. Then, $|S| \geq 2$ since no singleton has the power to pillage anyone at w. But then the first move from w must be to some w' that is an equal allocation; any unequal allocation terminates in a tyrannical allocation. Just as before, w' itself must be a stationary allocation, and then S cannot have the power to pillage the remaining players.

We now check (M). In cases (ii) and (iv), each $i \in \bar{H}(w)$ ends up getting $1/|\bar{H}(w)|$. There cannot be a better deviation. In case (iii), the agent initiating the deviation ends getting 1. Hence, (M) is satisfied.

Finally, it is easy to check that (E) is satisfied. In each of cases (ii)–(iv), $S(w)$ has a farsighted objection culminating in some allocation in $B(k^*)$. □

References

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin (2008), "Coalition formation in non-democracies." *Review of Economic Studies*, 75, 987–1009. [1195, 1208, 1209, 1212, 1213, 1214, 1215, 1216]

Aumann, Robert J. and Roger B. Myerson (1988), "Endogenous formation of links between players and of coalitions, an application of the shapley value." In *The Shapley Value: Essays in Honor of Lloyd Shapley* (Alvin Roth, ed.), 175–191, Cambridge University Press, Cambridge. [1192]

Bernheim, B. Douglas, Bezalel Peleg, and Michael D. Whinston (1987), "Coalition-proof Nash equilibria. I. Concepts." *Journal of Economic Theory*, 42, 1–12. [1194]

Bloch, Francis (1996), "Sequential formation of coalitions in games with externalities and fixed payoff division." *Games and Economic Behavior*, 14, 90–123. [1192]

Chander, Parkash (2015), "An infinitely farsighted stable set." Unpublished, Jindal Global University. [1192]

Chwe, Michael Suk-Young (1994), "Farsighted coalitional stability." *Journal of Economic Theory*, 63, 299–325. [1192, 1196]

Diamantoudi, Effrosyni and Licun Xue (2003), "Farsighted stability in hedonic games." *Social Choice and Welfare*, 21, 39–61. [1192]

Greenberg, Joseph (1990), *The Theory of Social Situations: An Alternative Game-Theoretic Approach*. Cambridge University Press, Cambridge, Massachusetts. [1196]

Harsanyi, John C. (1974), "An equilibrium-point interpretation of stable sets and a proposed alternative definition." *Management Science*, 20, 1472–1495. [1191, 1192, 1193, 1196, 1200, 1201, 1213]

Herings, P. Jean-Jacques, Ana Mauleon, and Vincent J. Vannetelbosch (2004), "Rationalizability for social environments." *Games and Economic Behavior*, 49, 135–156. [1192, 1197]

Herings, P. Jean-Jacques, Ana Mauleon, and Vincent Vannetelbosch (2009), "Farsightedly stable networks." *Games and Economic Behavior*, 67, 526–541. [1192]

Jordan, James S. (2006), "Pillage and property." *Journal of Economic Theory*, 131, 26–44. [1193, 1195, 1199, 1208, 1209, 1210, 1215]

Kimya, Mert (2015), "Equilibrium coalitional behavior." Unpublished, Brown University. [1192, 1194, 1199, 1214]

Konishi, Hideo and Debraj Ray (2003), "Coalition formation as a dynamic process." *Journal of Economic Theory*, 110, 1–41. [1192, 1193, 1198, 1199, 1201]

Lucas, William (1992), "von Neumann–Morgenstern stable sets." In *Handbook of Game Theory*, volume 1 (Robert Aumann and Sergiu Hart, eds.), 543–590, North Holland: Elsevier. [1192]

Mauleon, Ana and Vincent J. Vannetelbosch (2004), "Farsightedness and cautiousness in coalition formation games with positive spillovers." *Theory and Decision*, 56, 291–324. [1192]

Mauleon, Ana, Vincent J. Vannetelbosch, and Wouter Vergote (2011), "von Neumann–Morgenstern farsighted stable sets in two-sided matching." *Theoretical Economics*, 6, 499–521. [1192]

Piccione, Michele and Ariel Rubinstein (2007), "Equilibrium in the jungle." *The Economic Journal*, 117, 883–896. [1195]

Ray, Debraj (2007), *A Game Theoretic Perspective on Coalition Formation*. Oxford University Press, Oxford. [1192]

Ray, Debraj and Rajiv Vohra (1997), "Equilibrium binding agreements." *Journal of Economic Theory*, 73, 30–78. [1192, 1194]

Ray, Debraj and Rajiv Vohra (1999), "A theory of endogenous coalition structures." *Games and Economic Behavior*, 26, 286–336. [1192]

Ray, Debraj and Rajiv Vohra (2014), "Coalition formation." In *Handbook of Game Theory*, Vol. 4 (H. Peyton Young and Shmuel Zamir, eds.), 239–326, North Holland: Elsevier. [1192, 1197, 1198, 1199, 1201, 1214]

Ray, Debraj and Rajiv Vohra (2015), "The farsighted stable set." *Econometrica*, 83, 977–1011. [1191, 1192, 1193, 1194, 1196, 1204, 1205, 1207, 1215]

von Neumann, John and Oscar Morgenstern (1944), *Theory of Games and Economic Behavior*. Princeton University Press, Princeton. [1191, 1195, 1204, 1208]

Xue, Licun (1998), "Coalitional stability under perfect foresight." *Economic Theory*, 11, 603–627. [1192, 1197, 1199, 1203]

Co-editor Dilip Mookherjee handled this manuscript.

Manuscript received 29 February, 2016; final version accepted 10 October, 2016; available online 14 October, 2016.