Rational Expectations and Farsighted Stability

Bhaskar Dutta

Warwick University and Ashoka University

Rajiv Vohra

Brown University

March 2016

The (classical) blocking approach to coalitional stability is based on the notion of domination or objections by coalitions.

An outcome y dominates x if there is a coalition S that can replace x with y and gain by doing so: (S, y) is an objection to x.

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

The von Neumann-Morgenstern (1944) stable set is a set Z such that:

1. If $x \in Z$, it is not dominated by any $y \in Z$ (internal stability),

2. If $x \notin Z$, it is dominated by some $y \in Z$ (external stability).

Note the circularity: we can't say that a particular outcome is stable, except in relation to a set of stable outcomes.

Farsightedness

The core and stable set assume one-shot coalitional moves.

If coalitional deviations/moves can be followed by other moves what matters to a coalition is not the immediate effect of a move but where things will end up: the 'final outcome' – farsightedness.

How do we define what is 'final'?

$x \rightarrow_{S^1} x^1$, $x^1 \rightarrow_{S^2} x^2 \rightarrow_{S^3} x^3$.

If x^3 is the final outcome, farsightedness would mean that S^1 compare the utility of x^3 to that of x (and ignore its payoff at x^1 and x^2).

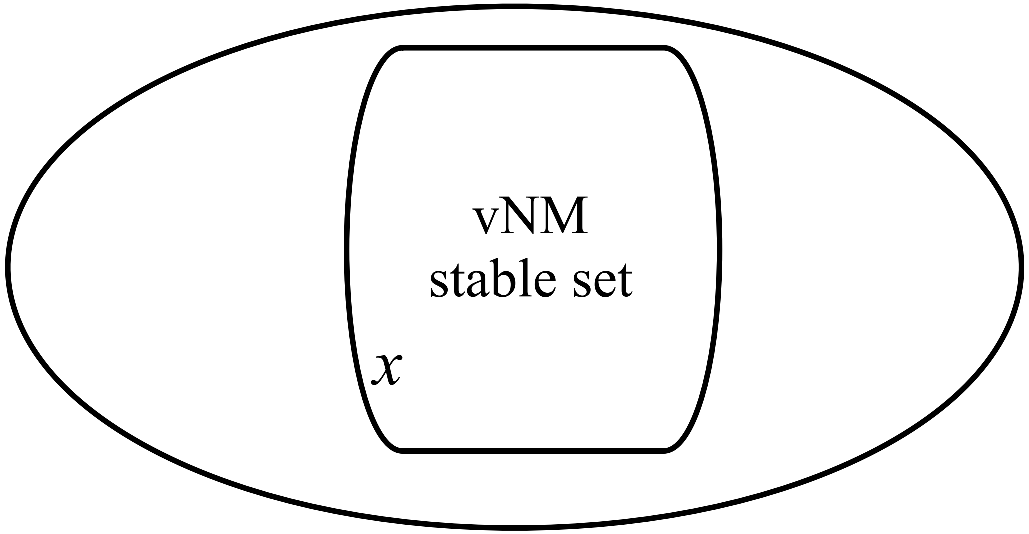
But this argument only works if x^3 is the 'final outcome'.

What is considered to be a final outcome must, of course, be stable.

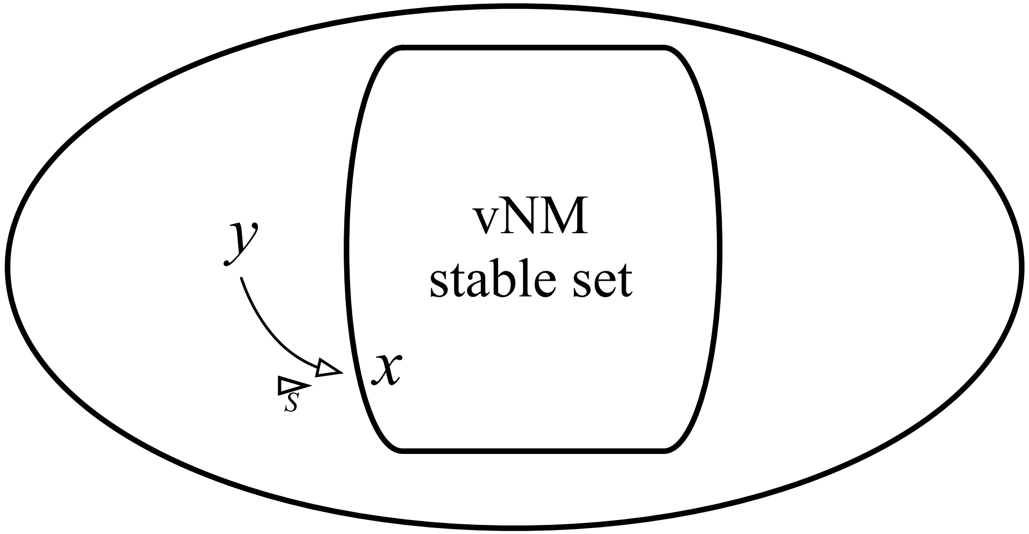
Thus, testing the stability of a particular outcome against a sequence of moves requires us to know which of the other outcomes are stable.

How should we handle this circularity?

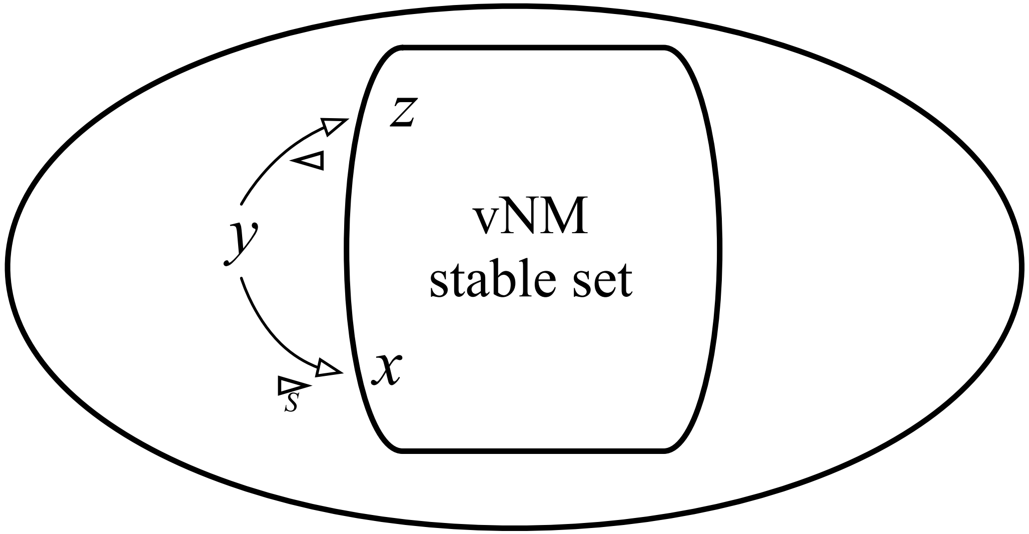
Harsanyi's Critique of vNM (myopic) stable set



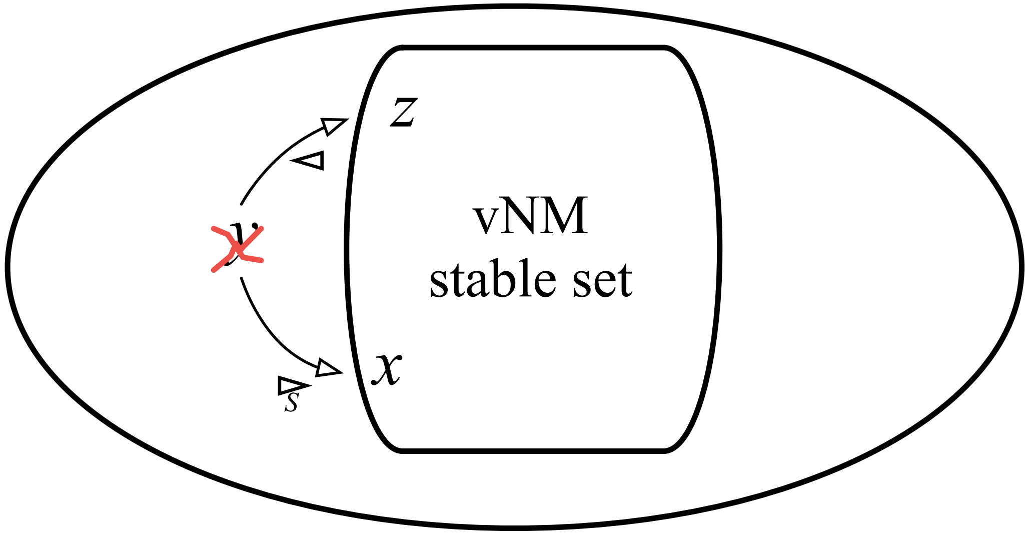
Harsanyi's Critique of vNM (myopic) stable set



Harsanyi's Critique of vNM (myopic) stable set



Harsanyi's Critique of vNM (myopic) stable set



Harsanyi: so what? IF $u(z)_S \gg u(x)_S$, S should still block x and replace it with y!

z farsightedly dominates x and z is supposed to be stable, so why should we judge x to be stable?

An abstract game $(N, X, E, u_i(.))$:

set of players N; outcomes or states X; utility of i at x is $u_i(x)$.

Effectivity correspondence E: $S \in E(x, y)$ is a coalition that can replace x with y. There will be natural restrictions on E depending on the context.

In a characteristic function game X will refer to the payoffs and associated coalition structure.

In the traditional theory states are taken to be efficient payoffs and effectivity is implicitly defined by saying that $S \in E(u, u')$ iff $u'_S \in V(S)$. This is fine for myopic concepts but strikingly wrong for farsightedness; Ray and Vohra (2015).

y dominates x if there is $S \in E(x, y)$ such that $u_S(y) \gg u_S(x)$.

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 $k = 1, \ldots m$:

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

and

$$u_{S^k}(y) \gg u_{S^k}(y^{k-1}).$$

A set of states $F \subseteq X$ is a farsighted stable set if

(1) If $x \in F$, $\nexists y \in F$ that farsightedly dominates x

(2) If $x \notin F$, $\exists y \in F$ that farsightedly dominates x.

There are two conceptual difficulties that remain with the farsighted stable set, as reformulated in Ray and Vohra (2015):

Maximality

Consistency and History Independence

And this paper attempts to rectify these.

Our first application, Simple Games, highlights consistency.

The second one, Pillage Games, highlights maximality.

We begin by illustrating these issues through simple examples of abstract games.

Maximality

The notion of farsighted dominance is based on an optimistic view.

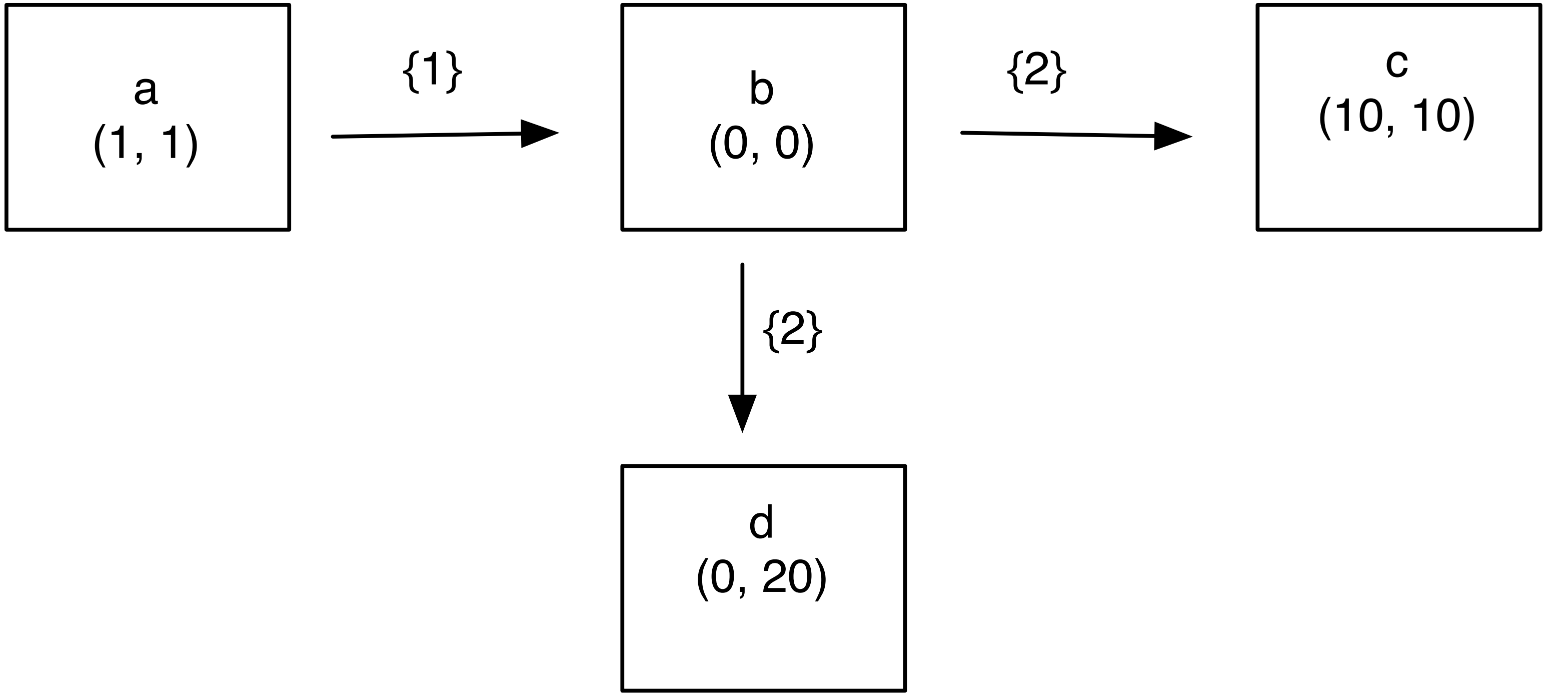
Suppose there is a farsighted objection: $x \to_{S^1} x^1$, $x^1 \to_{S^2} x^2$.

It's possible that at x^1 coalition S^2 could also have gained by moving to $\hat{x}^2$. Shouldn't S^1 worry about this possibility? To not worry is optimistic.

Chwe (1994) considers conservative behavior: largest consistent set.

But optimistic or pessimistic expectations are both ad hoc.

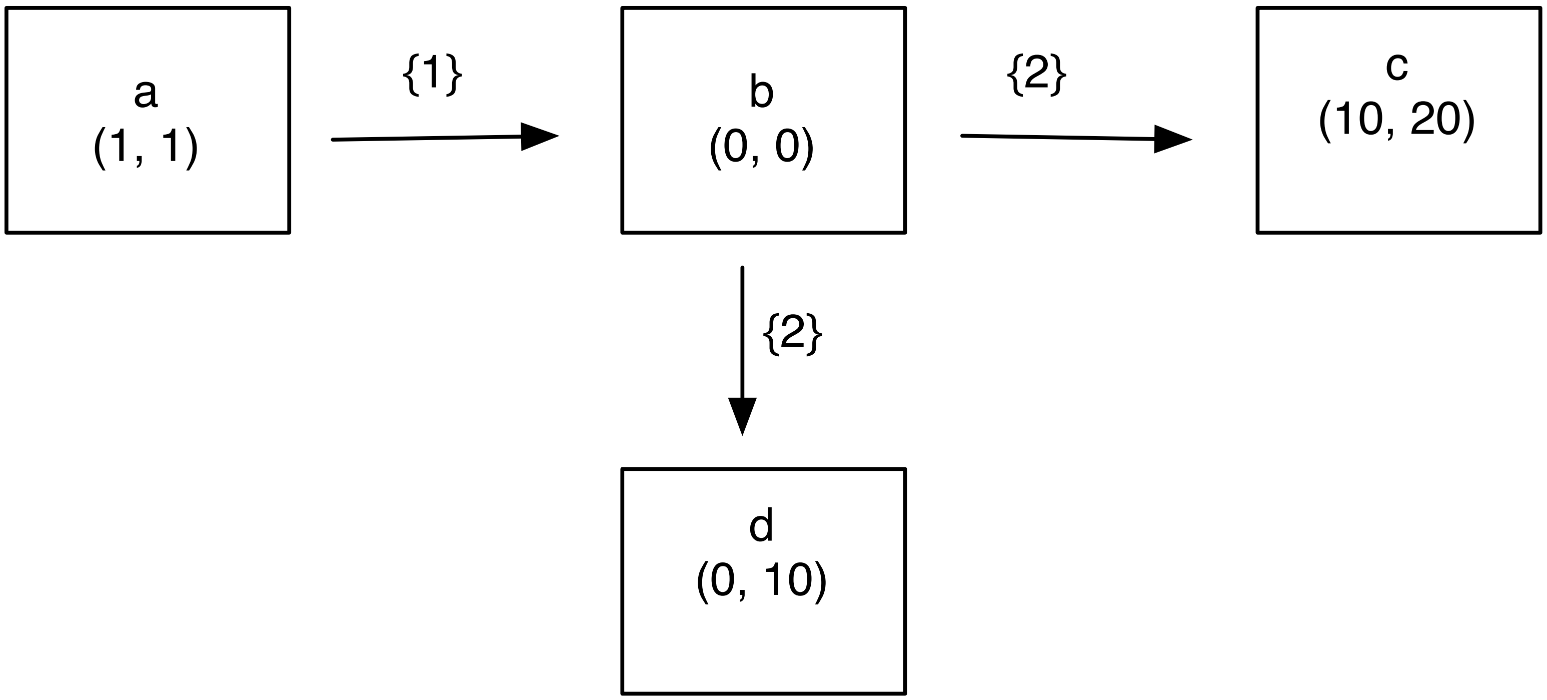
A solution concept should be based on optimal behavior.



Both c and d belong to the farsighted stable set.

Instability of a is based on the expectation that player 2 will choose c instead of d even though 2 prefers d to c.

a belongs to the LCS because of the possibility that the final outcome is d.

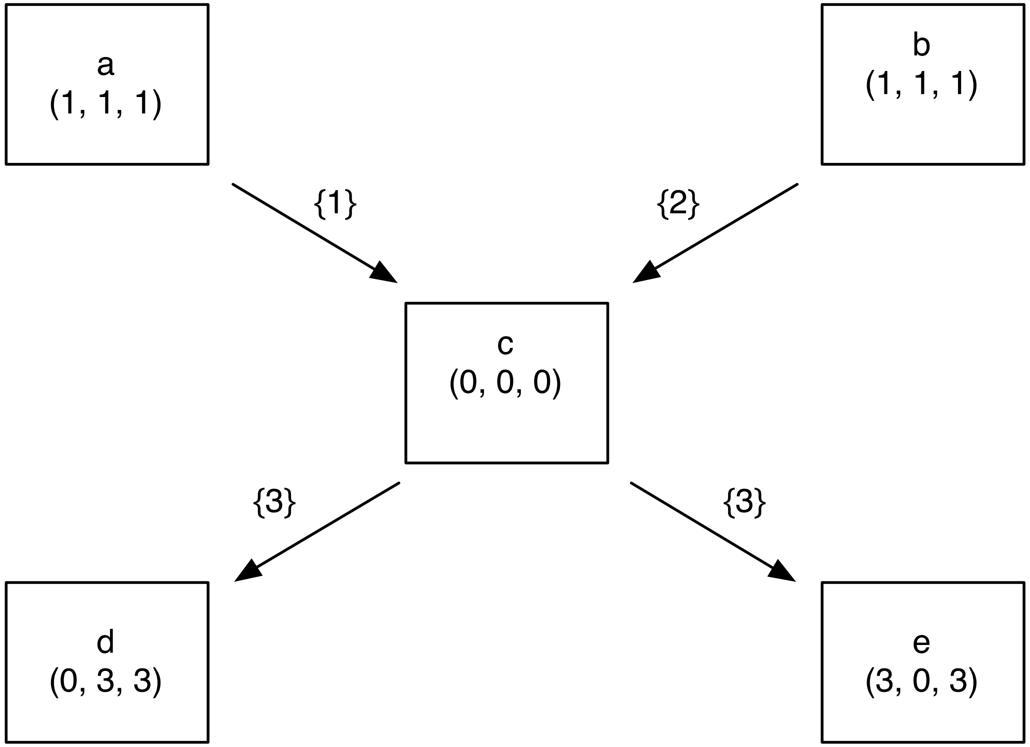


The LCS and farsighted stable set remain unchanged.

But now it is the LCS which provides the wrong answer: player 1 should not fear that 2 will (irrationally) choose d instead of c.

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.

Consistency and History Independence



The farsighted stable set is $\{d, e\}$ while the LCS is $\{a, b, d, e\}$.

The 'right' answer: $\{a, d, e\}$ and $\{b, d, e\}$ are two 'stable sets' depending on whether 3 is expected to move from c to d or to e.

To define optimal behavior one will need to rely on players having (rational) expectations about the continuation path following any coalition move.

In a dynamic setting, e.g., an EPCF as in Konishi and Ray (2003), these expectations are specified by a (Markovian) process of coalition formation and the condition that coalitions take actions that are maximally profitable.

We incorporate the idea of consistent and rational expectations in the traditional (static) framework of an abstract game: a new solution concept, related to the LCS and the farsighted stable but distinct.

Define an expectation as a function $F : X \to X \times \mathcal{N}$.

$F(x) = (f(x), S(x))$ specifies the state that is expected to follow x as well as the coalition expected to implement this change.

A stationary point of F is a state x such that $f(x) = x$.

f^k is the k-fold composition of f, e.g., $f^2(x) = f(f(x))$. $F^2(x) = F(f(x))$.

Having defined f^j for all integers $j < k$, $F^k(x) = F(f^{k-1}(x))$.

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.

A rational expectation F:

(i) If x is stationary, then no coalition is effective in making a profitable move in accordance with F: there does not exist $S \in E(x, y)$ such that $u_S(f^*(y)) \gg u_S(f^*(x))$.

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

(iii) If x is nonstationary, then $F(x)$ must prescribe an optimally profitable path for coalition $S(x)$: 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).

Condition (i) implies that $\Sigma(F)$ satisfies farsighted internal stability provided we restrict attention to farsighted objections consistent with F.

Condition (ii) implies farsighted external stability of $\Sigma(F)$. It is stronger than external farsighted stability because it states that to every $x \notin \Sigma(F)$ there is a farsighted objection, terminating in $\Sigma(F)$, consistent with the common expectation F.

Condition (iii) implies maximality of profitable moves as defined in Ray and Vohra (2014).

Maximality is the proper 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 sole priority in selecting the transition from x.

But one could entertain models in which, some other coalition may also have the right to intervene and change course. This motivates a stronger notion of maximality:

(iii') 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 $S \in E(x, y)$ such that $S \cap S(x) \neq \emptyset$ and $u_S(f^*(y)) \gg u_S(f^*(x))$.

A coalition disjoint from $S(x)$ cannot interfere in the expected move. But, based on the idea that a move by $S(x)$ requires the unanimous consent of all its members, another coalition S may take the initiative if it can enlist the support of at least one player in $S(x)$; $S \cap S(x) \neq \emptyset$

A expectation F satisfying (i), (ii) and (iii') 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).

Every SREFS is a REFS.

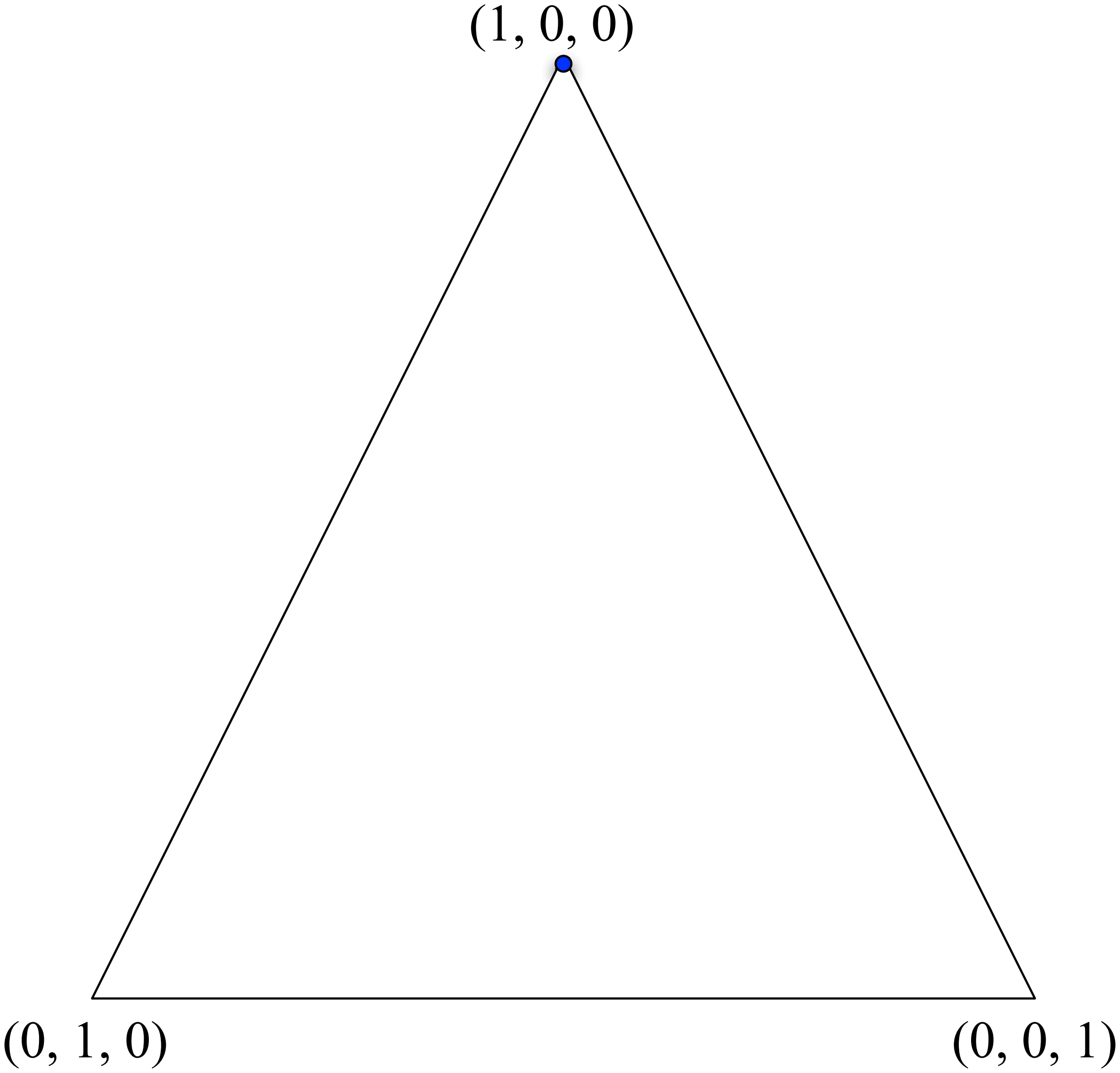
There is one interesting case in which an SREFS (or REFS) coincides with a farsighted stable set.

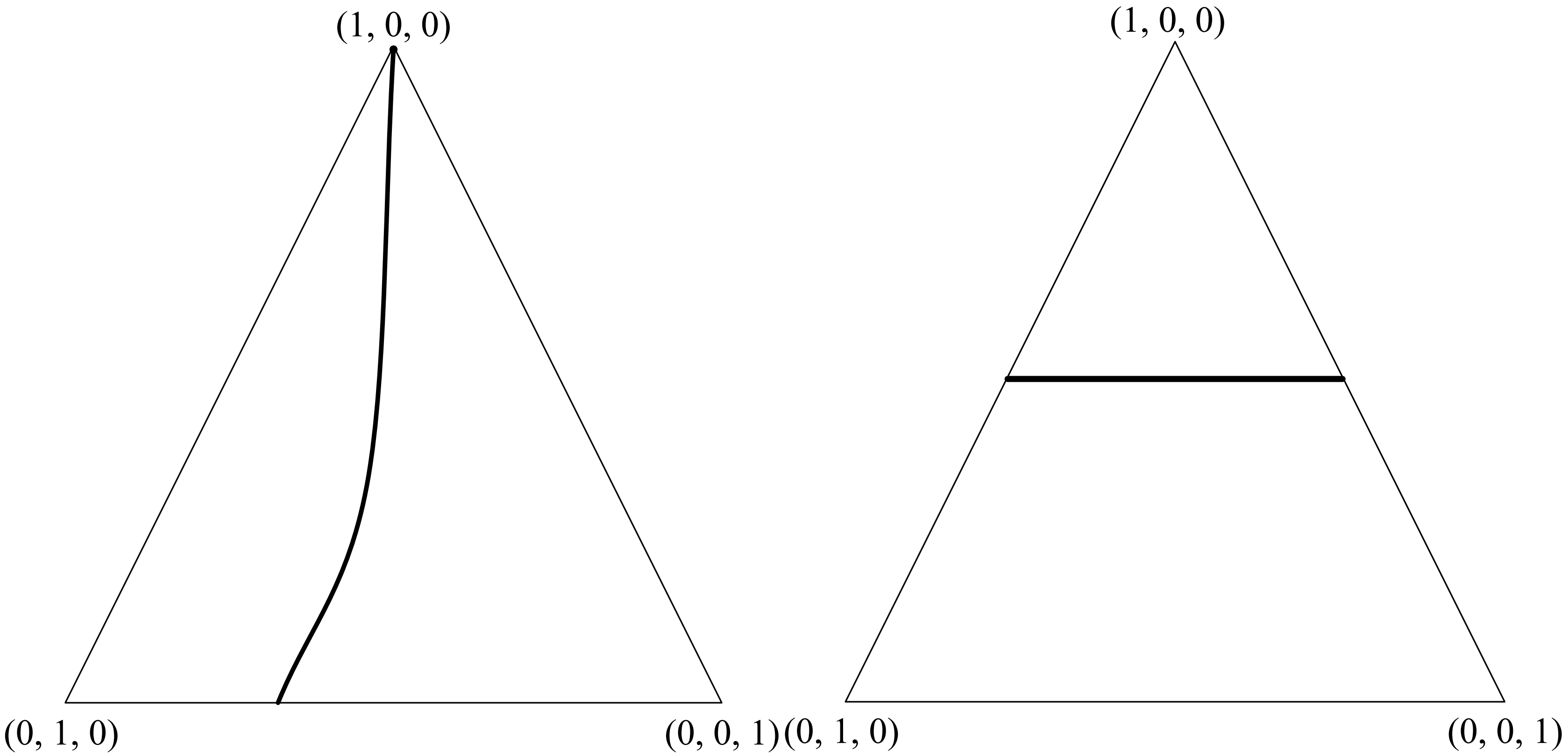
Theorem **1** *If Z is a single-payoff REFS it is a SREFS and a farsighted stable set. Conversely, if Z is a single-payoff farsighted stable, then it is a SREFS.*

Under appropriate restrictions on the effectivity correspondence, Ray and Vohra (2015) provide a sufficient condition for a payoff allocation in a characteristic function game to be a single-payoff farsighted stable set. This condition is satisfied by all allocations in the interior of the core.

In general, REFS or SREFS can be different from farsighted stable sets.

EXAMPLE 4. (Three-player veto game). $N = \{1, 2, 3\}$, $v(\{1, 2\}) = v(\{1, 3\}) = v(N) = 1$ and $v(S) = 0$ for all other S.





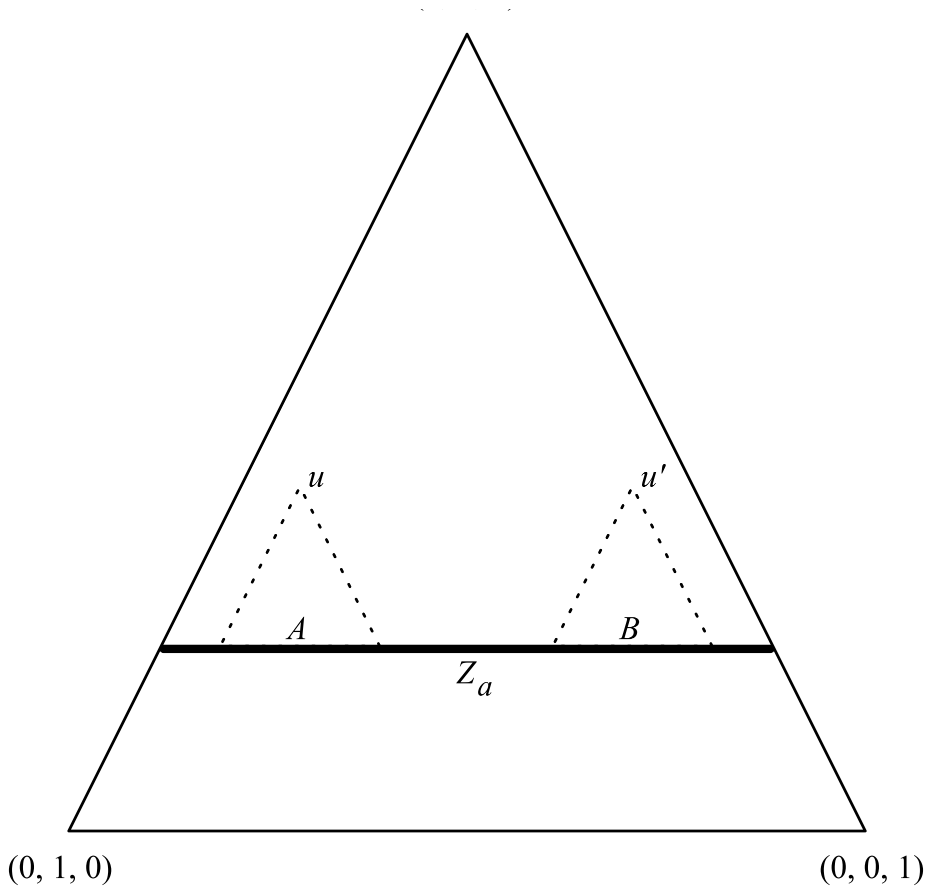
(a) vNM set

(b) Farsighted Stable Set

Ray and Vohra (2015) show that for every $a \in (0, 1)$, there is a farsighted stable set Z_a with the set of payoffs: $\{u \in R^3_+ \mid u_1 = a, u_2 + u_3 = 1 - a\}$. In fact every farsighted stable is of this form.

In this example maximality is not an issue with the farsighted stable set.

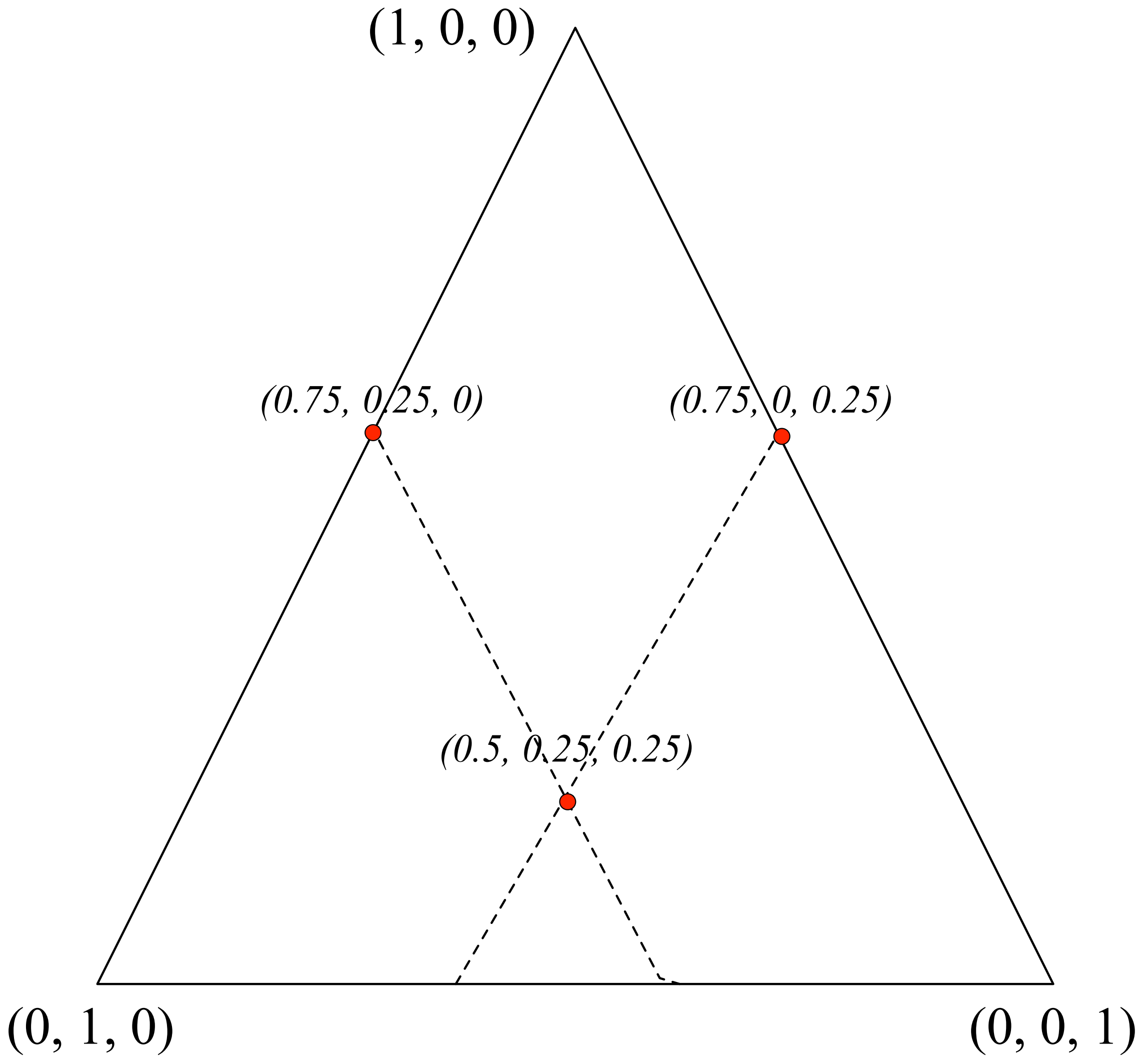
However, no set of the form Z_a can be a REFS because the external stability of Z_a (in the sense of a farsighted stable set) relies on inconsistent expectations (assuming history independence).



External stability of Z_a:

$u \to_{\{2,3\}} 0 \to_N A$ but $u' \to_{\{2,3\}} 0 \to_N B$.

SREFS



Simple Games

A simple game is a superadditive TU game s.t. for every S

$v(S) = 1$ or $v(S) = 0$, and if $v(S) = 1$, then $v(N - S) = 0$.

S is a winning coalition if $v(S) = 1$ and losing if $v(S) = 0$. Set of winning coalitions: $\mathcal{W}$.

A player is a veto player if she belongs to every winning coalition.

The collection of all veto players, the collegium: $S^* = \cap_{S \in \mathcal{W}} S$.

A collegial game is one in which $S^* \neq \emptyset$.

Assume $S^* \neq N$ (non-oligarchic game).

A state x specifies a coalition structure $\pi(x)$ and payoff, $u(x)$, such that $\sum_{i \in W(x)} u_i(x) = 1$, where $W(x)$ is the winning coalition.

Assumption 1 *The effectivity correspondence satisfies the following restrictions:*

(a) every coalition can form and divide its worth in any way among its members;

(b) When 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;

(c) If S includes members of $W(x)$ and the residual in $W(x)$ remains winning, then the players in $W(x) - S$ cannot lose.

Ray and Vohra (2015) 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: "discriminatory stable sets".

SREFS do not seem to have this structure (Example 4). Instead, they are finite payoff sets.

General Existence of SREFS?

One case has remained resistant to our efforts, so

Assumption 2 *There does not exist a winning coalition with one veto player and two non-veto players.*

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

Can Assumption 2 can be dropped or will this case yield an example of non existence? As of now this is an open question.

In our next application things are very different.

The farsighted stable set is a REFS but there are many others. And SREFS is a different refinement of REFS.

Pillage Games (Jordan, 2006)

Wealth-is-power

Set of players N; set of wealth allocations $\triangle$, the simplex in R^N.

For $w \in \triangle$ the power of coalition S is $\pi(S,w) = w_S \equiv \sum_{i \in S} w_i$.

Given w and w' let $L(w,w') = \{i \in N \mid w'_i < w_i\}$.

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

Those whose wealth is unchanged remain neutral.

$x \in [0,1]$ is dyadic if $x = 0$ or $x = 2^{-k}$ for some integer $k \geq 0$.

For every integer $k > 0$ let

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

The set of all dyadic allocations is $D = \cup_k D_k$.

It is easy to see that D_1 is the core.

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

What about farsightedness?

In the three-player case, D consists of the allocations $(1,0,0)$, $(0.5,0.5,0)$, $(0.5,0.25,0.25)$ and all their permutations.

But $(0.5,0.25,0.25) \rightarrow_1 (0.75,0,0.25) \rightarrow_1 (1,0,0)$.

So $(1,0,0)$ farsightedly dominates $(0.5,0.25,0.25)$.

Clearly, if player 3 anticipates the second step in this move, she should not remain neutral when player 1 pillages 2.

Jordan (2006) introduces expectations and shows that if otherwise neutral players act in accordance with the expected (final) outcome, then the stable set is indeed farsighted.

But this does not conform to a framework in which the effectivity correspondence specifies which coalition(s) are effective in changing w^{k-1} to w^k, independently of where w^k will end up.

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 when $S \in E(w, w')$ if and only if $w_S > w_{L(w,w')}$ and $w_i = w'_i$ for all $i \notin S \cup L(w, w')$?

It turns out to be identical to the core.

Theorem **4** *The unique farsighted stable set is D_1.*

What about REFS and SREFS?

For every integer $k \geq 0$, consider dyadic allocations in which all those with positive wealth have the same wealth:
$B_k = \{x \in \triangle \mid x_i = 0$ or $x_i = 2^{-k}, \forall i\}$ and $B = \cup_k B_k$.

Note that B_0 is the set of tyrannical allocations and $B_1 = D_1$.

Theorem **5** *B is a SREFS.*

Since B is a SREFS it is also a REFS.

But there are several other REFS, including the unique farsighted stable set: $B_0 \cup B_1 = D_1$.

So the farsighted stable set can be justified on the basis of consistent and rational expectations.

But it does not meet the strong maximality test.

Given n, let $k(n)$ be the largest integer such that $2^{-k(n)} \geq 1/n$.

Theorem 6 *For any $1 \leq k^* \leq k(n)$, $\cup_0^{k^*} B_k$ is a REFS.*

What is the reason for this difference between SREFS and REFS?

Take the 4-player example. Here the core, or $D_1 = B_0 \cup B_1$ consists of all permutations of $(1, 0, 0, 0)$ and $(0.5, 0.5, 0, 0)$. This is a REFS but not a SREFS.

SREFS also includes $B_2 = \{\bar{w}\} = (0.25, 0.25, 0.25, 0.25)$.

Suppose F is a rational expectation and $Z = \Sigma(F)$ is the associated REFS. It must contain D_1. An allocation with three having positive wealth can't be stable.

The question is whether or not it contains $\bar{w}$.

Consider $w' = (0.375, 0.375, 0.25, 0)$. There are three possible changes.

(a) $S(w') = \{1, 2\}$, i.e., players 1 and 2 pillage 3. In this case it is easy to see that $f(w') = f^*(w') = (0.5, 0.5, 0, 0)$;

(b) $S(w') = \{1\}$, $f(w') = (0.625, 0.375, 0, 0)$ and $f^*(w') = (1, 0, 0, 0)$;

(c) $S(w') = \{2\}$ and $f^*(w') = (0, 1, 0, 0)$.

Is it possible that $S(\bar{w})$ consists of two players i and j, who pillage a third and move to a w' (or a permutation thereof)?

Yes, if the next step, according to F we have (a). In this case $\bar{w} \notin Z$.

No, if F conforms to cases (b) or (c). In this case $\bar{w} \in Z$.

A rational expectation can be of either kind; both $B_0 \cup B_1$ and B can be shown to be REFS.

But there is an important difference.

F satisfying (a) cannot be a *strong* rational expectation.

Case (a) is not possible with a strong rational expectation because one of the players could have done better by pillaging player 3. In other words, B is a SREFS while $B_0 \cup B_1$ is not.

The unique farsighted stable set, B_1, is a REFS, but there are many other REFS.

SREFS is different because it relies on strong maximality.

Acemoglu *et al.* (2008): a model of political coalition formation in which $\gamma_i > 0$ denotes i's political power and $\gamma_S = \sum_{i \in S} \gamma_i$ is coalition S's power.

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.

Assume that if S is the ruling coalition, then

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

A state can be defined as the ruling coalition.

Winning coalitions are the ones effective in changing a state:

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

Ruling coalitions can only become smaller (internal blocking).

So a coalition will form if and only if there is no further change; a farsighted objection must be a myopic objection and the vNM stable set is equivalent to the farsighted stable set.

Of particular interest in these models is the stability of N, or the stable state(s) starting from N.

Example: $N = \{1, 2, 3, 4\}$, $\gamma = (2, 4, 6, 8)$ and $\alpha = 0.5$.

What is the stable set?

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.

The coalition $\{1, 2, 4\}$ is not stable because player 4 has enough power to eliminate the other two.

Let $\mathcal{S} = \{\{1, 3, 4\}, \{2, 3, 4\}, \{1, 2, 3\}\}$. All these are stable so N is not.

Any coalition in $\mathcal{S}$ is a REFS.

$N = \{1, 2, 3, 4\}$, $\gamma = (2, 4, 6, 8)$ and $\alpha = 0.5$.

$\mathcal{S} = \{\{1, 3, 4\}, \{2, 3, 4\}, \{1, 2, 3\}\}$.

Only $\{1, 2, 3\}$, the one with the least aggregate power, is a SREFS. This a result of the fact that every player prefers to be in a coalition with lower aggregate power.

$F(N) = \{2, 3, 4\}$ is not strongly maximal because players 3 and 4 could do better by forming $\{1, 3, 4\}$. Similarly for $\{1, 3, 4\}$.

SREFS turns out to be precisely the solution proposed by Acemoglu *et al.* (2008).