Maximality in the Farsighted Stable Set[1]

Debraj Ray
New York University and University of Warwick

and

Rajiv Vohra
Brown University

December 2017

Abstract. The stable set of von Neumann and Morgenstern imposes credibility on coalitional deviations. Their credibility notion can be extended to cover farsighted coalitional deviations, as proposed by Harsanyi (1974), and more recently reformulated by Ray and Vohra (2015). However, the resulting farsighted stable set suffers from a conceptual drawback: while coalitional deviations improve on existing outcomes, coalitions might do *even better* by moving elsewhere. Or other coalitions might intervene to impose their favored moves. We show that every farsighted stable set satisfying some reasonable, and easily verifiable, properties is unaffected by the imposition of this stringent maximality requirement. These properties are satisfied by many, but not all, known farsighted stable sets.

KEYWORDS: stable sets, farsightedness, maximality, history dependent expectations.

JEL CLASSIFICATION: C71, D72, D74

1. Introduction

The core is a classical solution concept: it identifies payoff profiles that no group, or coalition, can dominate with an allocation that is feasible for the coalition in question. But the core does not ask if the new allocation itself is threatened or "blocked" by *other* coalitions. In this conceptual sense the solution is too strong, possibly excluding allocations that would not be credibly dominated. The problem is that the definition of credibility is often circular — an allocation is not credible if it is not challenged by a credible allocation — and concepts such as the bargaining set (Aumann and Maschler 1964), which only try to build in an additional "round" of domination, are just not up to the task. But the vNM stable set (von Neumann and Morgenstern, 1944) can indeed be seen as such a theory: it cuts through that circularity. Say that a payoff profile is dominated by another profile if some coalition prefers the latter profile and can unilaterally implement the piece of the new profile that pertains to it. A set of feasible payoff profiles Z is *stable* if it satisfies two properties:

Internal Stability. If $u \in Z$, it is not dominated by $u' \in Z$.

External Stability. If $u \notin Z$, then there exists $u' \in Z$ which dominates u.

Notice how internal and external stability work in tandem to get around the circularity implicit in the definition of credibility. The set Z is to be viewed as a "standard of behavior" (Greenberg,

[1]Ray acknowledges funding from the National Science Foundation under grant SES-1629370.

1990). Once accepted, no allocation in the standard can be overturned by another allocation also satisfying the standard. Moreover, allocations within the standard jointly dominate all non-standard allocations. This perspective drives home the idea that the relevant solution concept is not *a* payoff profile, but a *set* of payoff profiles which work in unison. It is a beautiful definition.

Yet, and temporarily setting beauty aside, there are at least three problems with the concept:

1. *Harsanyi critique*. Suppose that u' dominates $u \in Z$, and that u' is in turn dominated by $u'' \in Z$, as required by vnM stability. Then it is true that u' isn't "credible," but so what? What if the coalition that proposes u' only does so to induce u'' in the first place, where it *is* better off? Harsanyi went on to propose a "farsighted version" of vNM stability, one that permits a coalition to anticipate a chain reaction of payoff profiles, and asking for a payoff improvement at the terminal node of this chain.

2. *Ray-Vohra critique*. Ray and Vohra (2015) point to a seemingly innocuous device adopted by von Neumann and Morgenstern. Dominance is defined over entire profiles of payoffs. As described above, profile u' dominates u when some coalition is better off under u' and can implement *its piece* of u' unilaterally. But what about the rest of u', which involves allocations of payoffs to others who have nothing to do with the coalition in question? Who allocates these payoffs, and what incentive to do they have to comply with the stipulated amounts? To this, von Neumann and Morgenstern would answer that it does not matter: payoffs to outsiders are irrelevant, and only a device for tracking all profiles in a common space. However, once the solution is modified along the lines of Harsanyi, the critique *does* matter: the payoffs accruing to others will fundamentally affect the chain reaction that follows. Their determination cannot be finessed.

3. *Maximality problem*. Domination requires every coalition participating in the chain reaction of proposals and counter-proposals be better off (relative to their starting points) once the process has come to a standstill. But it does not require coalitions to choose their *best* moves (Ray and Vohra 2014, Dutta and Vohra 2017), and it rules out possibly unwelcome interventions by other coalitions. This is of concern not just at any stage but along the entire farsighted blocking chain. That chain is supported by the anticipation that later coalitions participating in the chain will also be "better off" doing so. But now "better off" isn't good enough: what if they gain *even* more by doing something else, and that something else isn't good for the original deviator? What if another coalition intervenes? Faced with such potential complexities, the entire chain of proposals becomes suspect.

This third problem forms the subject of our paper.

To fix ideas, consider the following simple example (Example 5.8, Ray and Vohra 2014). There are two players, 1 and 2, and four states, a, b, c, d. The payoff profiles by state are $u(a) = (1, 1)$, $u(b) = (0, 0)$, $u(c) = (10, 10)$ and $u(d) = (0, 20)$. Suppose that state a can only changed by player 1, and that she can only move to b. From b, only player 2 can move, and she can move either to c or d, both of which are terminal states (no further move is possible from **c** or d). It is easy to see that the unique farsighted stable set is $\{c, d\}$. In particular, a is not in the farsighted stable set because it is dominated by a move by player 1 to b followed by a move by player 2 to

c; player 1 gains by replacing a with c and player 2 gains by replacing b with c. But this clearly violates maximality since player 2's optimal move from b is to d rather than to c.

This example suggests that something like subgame perfection, common in noncooperative game theory, needs to be added on to farsighted stability. But it should be noted that cooperative game theory attempts to model free-form negotiations. There is no protocol (deterministic or stochastic) that sets the "rules of the game," assigning a particular player or coalition to move at each node. Noncooperative game theory imposes such protocols, but the apparent gain in precision is in part illusory, for it is well known that the answers can be notoriously sensitive to the choice of the extensive form. In contrast, the theory of blocking is more open-ended: *any* coalition can move at any stage. Specifically, this implies that the problem of maximality is not just restricted to the coalition that actually moves, but it also applies to other coalitions that could *potentially* move. So, while maximality is certainly related to sequential rationality or subgame perfection, it goes beyond that. Indeed, different definitions of maximality are possible depending on which coalitions are "allowed" to move at any state. In this paper we take on board the strictest possible version: one that asks for immunity to all deviations, not just by the coalition that moves "in equilibrium," but also by all other coalitions. To distinguish this concept from weaker notions of maximality we refer to it as *absolute maximality*.

None of this is particularly germane to the example above; all the maximality concepts yield the same answer. But in a negotiation setting, there aren't states as in the example with a highly restricted, tree-like structure describing the moves that are possible. States are combinations of coalition structures and proposed payoff allocations, and while it is true that not all coalitions are capable of precipitating one state from another, it is possible to travel from any state to any other. Indeed, our main result shows that in the context of negotiations, the example is an outlier: *every farsighted stable set satisfying reasonable and easily verifiable properties is unaffected by the imposition of absolute maximality*.

These properties are described as A and B in Section 3.1, and the main result is stated as Theorem 1. The theorem is useful because the identification of farsighted stable sets, or even stable sets, is not always an easy task. Having to check if they satisfy maximality adds an additional layer of complexity. It would be extremely desirable if such a check could be sidestepped, and Properties A and B allow us to do just that. For instance, we show that in simple games satisfying a mild restriction, a farsighted stable set exhibiting Properties A and B — and consequently an absolutely maximal; farsighted stable set — always exists.

That said, Properties A and B need not always be satisfied. Two examples are constructed in Section 3.4. In one, Property A is satisfied for a farsighted stable set but not B; in the other, Property B is satisfied for the set but not A. In either example, the farsighted stable set fails to be absolutely maximal, demonstrating that Theorem 1 is tight.

2. Maximal Farsighted Stability

2.1. **Coalitional Games.** A coalitional game, or a characteristic function game, is described by a finite set N of players and a mapping V that assigns to each coalition S (a nonempty subset of N) a closed set of feasible payoff vectors $V(S) \subseteq \mathbb{R}^S$. Normalize the game so that singletons obtain zero, and assume that all coalitions can get nonnegative but bounded payoffs. So $V(S)$ is some nonempty compact subset of $\mathbb{R}^S_+$. A transferable utility (TU) game is one in which each coalition S has a *worth* $v(S)$ and $V(S) = \{u \in \mathbb{R}^S \mid \sum_{i \in S} u_i \leq v(S)\}$.

2.2. **States and Effectivity.** A *state* is a coalition structure π and a payoff profile u feasible for that structure. A typical state x is therefore a pair (u, π) (or $\{u(x), \pi(x)\}$ when we need to be explicit), where $u_S \in V(S)$ for each $S \in \pi$. An *effectivity correspondence* $E(x, y)$ specifies for each pair of states x and y the collection of coalitions that have the power to change x to y. Ray and Vohra (2015) argue that effectivity correspondences must satisfy natural restrictions for the resulting solution concepts to make sense. Specifically, we assume throughout:

(E.1) If $S \in E(x, y)$, $T \in \pi(x)$ and $T \cap S = \emptyset$, then $T \in \pi(y)$ and $u(x)_T = u(y)_T$.

(E.2) For every state x, coalition $S \subseteq N$, partition μ of S and payoff vector $v \in \mathbb{R}^{|S|}$ with $v_W \in V(W)$ for each $W \in \mu$, there is $y \in X$ such that $S \in E(x, y)$, $\mu \subseteq \pi(y)$ and $u_T(y) = v$.

Condition E.1 grants coalitional sovereignty to the untouched coalitions: the formation of S cannot influence the membership of coalitions that are entirely unrelated to S in the original coalition structure, nor can it influence the going payoffs to such coalitions. Condition E.2 grants some degree of sovereignty to the moving coalition. It says that if S wants to move from a going state, it can do so by reorganizing itself (breaking up into smaller pieces if it so wishes, captured by the sub-structure μ), provided that the resulting payoff to it, v, is feasible ($v_W \in V(W)$ for every $W \in \mu$). What happens "elsewhere," however, is not under its control (see, for instance, the sovereignty restriction E.1), which is why E.2 only asserts the existence of *some* new state y satisfying the sovereignty conditions.

2.3. **Farsighted Stability.** A *chain* is a finite collection of states $\{y^0, y^1, \ldots, y^m\}$ and a corresponding collection of coalitions, $\{S^1, \ldots, S^m\}$, such that for every $k \geq 1$, we have $y^{k-1} \neq y^k$, and S^k is effective in moving the state from y^{k-1} to y^k: $S^k \in E(y^{k-1}, y^k)$. A state y *farsightedly dominates* x if there is a chain with $y^0 = x$ and $y^m = y$ such that for all $k = 1, \ldots m$, $u(y)_{S^k} \gg u(y^{k-1})_{S^k}$. The associated chain will be called a *blocking chain*.

A set of states $F \subseteq X$ is a *farsighted stable set* if it satisfies two conditions:

(i) *Internal Farsighted Stability*. No state in F is farsightedly dominated by another state in F;

(ii) *External Farsighted Stability*. A state not in F is farsightedly dominated by some state in F.

Observe that farsightedness does not impose any optimization requirement on coalitional moves, except for requiring that coalitions must be eventually better off participating in the blocking chain rather than not participating at all. Below, we impose stringent maximality requirements.

2.4. **Absolutely Maximal Farsighted Stable Sets.** To incorporate the notion of maximality in a farsighted stable set, we will "embed" that set into an ambient history-dependent *negotiation process*. To this end, define a *history* h to be a finite sequence of states (where any change of state must be feasible), along with the coalitions that generate any state transitions. If there is no move, the empty coalition is recorded. An *initial history* is just a single state.[2] Let $x(h)$ be the last state in a history. A *negotiation process* is a map σ from histories to the new outcome. For each h, $\sigma(h) = \{y(h), S(h)\}$, where $y(h)$ is the state that follows $x(h)$ and $S(h) \in E(x(h), y(h))$ is the coalition implementing any change. If $x(h) = y(h)$, then $S(h) = \emptyset$. In this way, given any history h, σ induces a continuation chain.

A state x is *absorbing* under the process σ if at any history h with $x(h) = x$, $y(h) = x(h) = x$. That is, once at x the continuation chain displays x forever. Say that σ is an *absorbing process* if its continuation chain must terminate in an absorbing state starting from any history.[3] For every absorbing process σ and history h, let $x^\sigma(h)$ denote the absorbing state reached from h. Say that an absorbing process σ is *coalitionally acceptable* if for each history h, if $S(h)$ is nonempty, then $u_S(x^\sigma) \geq u_S(x(h))$. Finally, call an absorbing process σ *absolutely maximal* if at no history h does there exist a coalition T and a state y with $T \in E(x(h), y)$, such that $u_T(x^\sigma(h, y, T)) \gg u_T(x^\sigma(h))$. We discuss these concepts in detail below.

A farsighted stable set F is *absolutely maximal* if it can be embedded in some absorbing, coalitionally acceptable, and absolutely maximal process σ; that is,

(i) F is the set of all absorbing states of σ.

(ii) At any initial history $h = \{x\}$ where $x \notin F$, or $h = (x, (S, y))$ with $x \in F$, $S \in E(x, y)$ and $y \notin F$, the continuation chain from h is a blocking chain terminating in F.

2.5. **Discussion.** Condition (i) asks that the set F be the ultimate repository of all end-states of σ starting from *any* history. That is, we seek not just absorption, but absorption back into F. Condition (ii) seeks consistency with the "blocking chain" approach that was originally used to describe F. That is, starting from some state not in F, or following some replacement of a state in F by another outside it, the process prescribes a blocking chain leading back into F, just as envisaged in the traditional definition of stability.

But, of course, σ does more: it prescribes a continuation chain for *all* histories, not just the ones described in condition (ii) above. It is imperative to do this, because we need a setting where maximality can be discussed. That requires us to consider deviations from ongoing chains, deviations from deviations, and so on; σ handles all these. The requirement that σ be absolutely maximal is a central part of the embedding requirement for F.

We reiterate that this condition applies to *every* coalition, not just the coalition stipulated to move at the state in question. It is therefore stronger than the *maximality* condition of Dutta and Vohra (2017) which is imposed only on the coalition about to move.[4] Indeed, it is stronger than their *strong maximality* condition, imposed on all coalitions T that have a nonempty intersection with $S(h)$. It is arguably the strongest form of maximality that once could insist on.[5]

[2]For instance, players might all begin the negotiation process as standalone singletons, or it may be that some going arrangement or state is already in place.

[3]That is, there exists k such that $y^{(t)}(h) = x$ for all $t \geq k$, where $y^{(t)}$ is defined recursively in the obvious way.

[4]This is also true of the maximality conditions imposed in Konishi and Ray (2003) and Ray and Vohra (2014).

[5]This is not a formality. We provide an example (Example 1 in Section 3) of a farsighted stable set that satisfies strong maximality but not absolute maximality.

Maximality arises naturally from considerations of coalitional rationality. If an absorbing σ were not maximal in this sense, it would be disrupted after some history by a coalition that would stand to gain from that disruption. It could not describe an "equilibrium" negotiation process. But absorption *per se* does not follow from rationality: nothing dictates that a process *must* be absorbing: it could, for instance, cycle forever. By imposing absorption, we are imposing the presumption that a negotiation process must ultimately terminate.[6]

In similar vein, coalitional acceptability is not a necessary concomitant of rationality, though in some cases in could be.[7] It is more a philosophical statement that an existing state has status-quo-like properties, and any coalition that resolves to actively move from it must be made at least weakly better off. Now, every initial history, or one preceded by a stable state, is followed by a blocking chain for the particular farsighted stable set we seek to embed (and so leads to absorption into that set); this is Condition (ii). Because a blocking chain satisfies coalitional acceptability, the latter imposes no additional restriction for such histories. Elsewhere, our process is indeed constrained by absolute maximality, but without a restriction such as coalitional acceptability, this has little bite. For instance, pick any state x, however unpalatable for some or all of the players, and then define σ so that from any initial state it specifies a path back to x. Then no one-shot deviation by any coalition can be improving, and we would be able to embed $\{x\}$ within σ with all requirements met save for coalitional acceptability. That is absurd, because some coalition could be better off by refusing to go. The imposition of coalitional acceptability after every history rules out such absurd prescriptions. (In passing, adding this restriction makes our theorem harder to prove). More formally, because payoffs are received only at the "final stage," the one-shot deviation principle is not enough to capture all deviations by a coalition. It can be shown that in our setting, coalitional acceptability plus the one-shot deviation principle does indeed handle all deviations.

To understand how coalitional acceptability might arise, suppose that at some state, just one coalition is always called up to move, and is repeatedly asked to move to an unpalatable outcome. With only one-shot deviations allowed, as we have done, there is no way out of the impasse. But a slight perturbation of the model will fix that. Suppose that each coalition receives payoffs in real time, as in Konishi and Ray (2003) and Ray and Vohra (2014), and discounts future payoffs. Then coalitional acceptability would be *implied* by the no-single-deviation condition, and it would remain in place as the discount factor converges to one, which is the model analyzed here.

That said, it is important to point out that coalitional acceptability is not always implied by a discounted real-time model of negotiations. For instance, if a coalition refuses to move, then another coalition might be called upon to do so; and indeed it is possible to lock two coalitions into a coordination failure so that coalitional acceptability applies to neither of them. So, while (in some situations) coalitional acceptability can be derived from a model of rational play, this isn't inevitable. To come full circle: it is a property placed on the status quo and the final outcome, and we view as a desirable characteristic of the negotiation process.

---

[6] It implies a bit more than that, as it also asks for an absorbing state to be absorbing following *every* history leading to it.

[7] For instance, if there is a unique coalition that is asked to move at any state — as would be the case in a Markov negotiation process — then acceptability could be derived from coalitional rationality.

Indeed, we could strengthen coalitional acceptability even more: we could ask that after *every* history ending in a state not in F, a blocking chain must be used, thereby imposing farsighted dominance not just "on path," but following every conceivable history. That places even more restrictions on the process, because — as already noted — farsighted dominance implies coalitional acceptability. This extension is discussed in the Online Appendix.

A final remark on the concept: we seek conditions under which a farsighted stable set might satisfy maximality (absolute maximality, in our case). Alternatively, one might pursue a research program to understand the absorbing states of a negotiation process, without seeking to embed any existing solution concept such as the farsighted stable set. Such a program is introduced in Konishi and Ray (2003) and Dutta and Vohra (2017), along with the additional restriction that the negotiation process is Markovian.[8] History-dependent versions of these sets are studied in Hyndman and Ray (2007), Ray and Vohra (2014) and Dutta and Vartiainen (2017). These solutions need not be farsighted stable sets in their own right.[9]

3. The Maximality of Farsighted Stable sets in Coalitional Games

Our main theorem is that any farsighted stable set that satisfies two properties is absolutely maximal. In general, directly constructing an absorbing, coalitionally acceptable and absolutely maximal process in which to embed our set and thereby check for absolute maximality is not an easy task. Our result is useful precisely because that task is replaced by the verification of two simple properties.

3.1. Two Properties.

Consider the following conditions which can be directly checked for any farsighted stable set.

A. *Suppose there are two states a and b in F such that $u_j(b) > u_j(a)$ for some j. Then there exists a state $z \in F$ such that $u_j(z) \leq u_j(a)$, and $u_i(z) \geq u_i(b)$ for all $i \neq j$.*

B. *There is no a, b in F and coalition T with $u_T(b) \in V(T)$, $T \in \pi(b)$ and $u_T(b) \gg u_T(a)$.*

Property A states that if player j gets a strictly higher payoff at $b \in F$ compared to a, then it is possible to find another state at which j's payoff is capped at $u_j(a)$ without punishing the other players.

Property B states that given a state in F, there is no other state in F, with payoff higher than a for some coalition in that state *and* feasible for that coalition. This property bears a close resemblance to internal stability. In fact, in the classical literature, stemming from von Neumann

[8] See their concepts of *rational expectations farsighted stable sets*, which impose maximality, and *strong rational expectations farsighted stable sets*, which impose strong maximality.

[9] For instance, the solution concepts in Dutta and Vohra (2017) satisfy myopic internal stability but farsighted internal stability *only* for objection chains consistent with the given process. In fact, the the solutions they construct for simple games are not farsighted stable sets. The same is true of the solutions constructed by Dutta and Vartiainen (2017), allowing for history dependence, and using a weaker notion of absorption that we have defined above. Indeed, their solutions may not even satisfy myopic internal stability; they show that in a strictly superadditive game, the set of all strictly positive feasible (not necessarily efficient) payoffs is stable in their sense.

and Morgenstern (1944) and including Harsanyi (1974), it *is* internal stability.[10] It is only because of our insistence on the coalitional sovereignty conditions (E.1) and (E.2) that Property B could go beyond internal stability.[11]

Readers interested in more discussion of these Properties, and in their necessity in establishing the main theorem, are referred to Section 3.3.

3.2. **Main Theorem.** We can now state:

THEOREM **1.** *If a farsighted stable set satisfies Properties A and B, then it is absolutely maximal.*

Proof. The following lemma establishes that whenever there is a blocking chain from x to y, there exists what might be called a *canonical blocking chain* from x to y, in which each individual moves at most twice, once to form a singleton and again at the very last step, when "consolidation" occurs to generate the final state y.

LEMMA **1.** *Suppose that y farsightedly dominates x using the chain $\{\tilde{y}^0, \tilde{y}^1, \ldots, \tilde{y}^{\tilde{m}-1}, \tilde{y}^{\tilde{m}}\}$, $\{\tilde{S}^1, \ldots, \tilde{S}^{\tilde{m}}\}$, where $\tilde{y}^0 = x$ and $\tilde{y}^{\tilde{m}} = y$. We claim that there exists another blocking chain $\{y^0, y^1, \ldots, y^{m-1}, y^m\}$, $\{S^1, \ldots, S^m\}$, such that*

(i) $y^0 = x$ *and* $y^m = y$; *and*

(ii) S^i *and* S^j *are disjoint for all* i *and* j *between* 1 *and* $m - 1$.

Proof. To construct the new blocking chain, with $y^0 = x$, let $S^1 \equiv \tilde{S}^1$ and recursively, $S^k = \tilde{S}^k - \cup_{t<k}\tilde{S}^t$ for all $k = 2, \ldots, \tilde{m} - 1$. When coalition S^k moves at stage $k < \tilde{m}$, it does so by breaking into singletons. This means that for any $k = 1, \ldots, \tilde{m} - 1$, the corresponding coalition structure, π^k, is such that all players in $\cup_{t<k}\tilde{S}^t$ are in singletons, and (by Condition E.1) all other players belong to the same coalition as in $\tilde{y}^k$. At the last step, let $S^{\tilde{m}} = \cup_{k=1}^{\tilde{m}}\tilde{S}^k$. That is, we collect all the coalitions have already moved, along with all other individuals (if any) in $\tilde{S}^{\tilde{m}}$. Since $S^{\tilde{m}}$ is the set of *all* players who were involved in moving from x to y, it is clearly effective in moving to y. Have it do so, creating the final coalition structure, $\pi^{\tilde{m}} = \pi(y)$.

Denote by u^k the associated payoffs in the newly constructed chain and by $\tilde{u}^k = u(\tilde{y}^k)$ the payoffs generated by the original blocking coalition. Of course, $u^0 = \tilde{u}^0 = u(x)$, $u^{\tilde{m}} = \tilde{u}^{\tilde{m}} = u(y)$. Given the coalition structures $\pi^k, k = 1, \ldots, \tilde{m}$ in the new chain, it follows from Conditions E.1 and E.2 that

(1) for $k = 1, \ldots \tilde{m} - 1$, $\quad u_i^k = 0$ if $i \in \cup_{t\leq k}\tilde{S}^t$, and $u_i^k = \tilde{u}_i^k$ otherwise.

[10] Recall that in that literature, it is assumed that a coalition is effective in moving to any state as long the payoff restricted to this coalition is feasible for it; there is no restriction on the distribution of payoffs to outsiders. In that model, Property B is equivalent to internal (myopic) stability and is automatically satisfied by every farsighted (or myopic) stable set.

[11] If there are a, b, and T such that $u_T(b) \in V(T)$ and $u_T(b) \gg u_T(a)$, then by (E.2), T can move to some state, say b', where $u_T(b') = u_T(b)$. But it is possible that $T \notin E(a, b)$ because the coalition structure and or the payoffs of players outside T might differ across b and b'. Of course this possibility does not arise for a farsighted stable set in which every state has the grand coalition as the associated coalition structure; in that case Property B is equivalent to myopic internal stability.

Finally, let the associated states be $y^k = (u^k, \pi^k)$ for all $k = 1, \ldots, \tilde{m} - 1$, and $y^{\tilde{m}} = y$.

It is possible that for some stages $k < m$, S^k as defined is empty and the succeeding state y^{k+1} is identical to y^k.[12] In that case, remove the step at all such k. We are left with a chain of m steps, where $m \leq \tilde{m}$, and this is the chain to which the lemma refers. By construction, (i) and (ii) are satisfied.

We only need to check that the new chain is a blocking chain. That is, for every $k \geq 1$ and every $i \in S^k$, $u_i(y) > u_i^{k-1}$. But this is true because $u_i(y) > \tilde{u}_i^{k-1}$ since the original chain is a blocking chain, and by (1), $\tilde{u}_i^{k-1} \geq u_i^{k-1}$. ■

Consider any farsighted stable set F. If $x \in F$, define $\Psi(x) = x$. For each state $x \notin F$, fix any blocking chain, $\mathbf{c}(x)$, with y as the next state and $\Psi(x) \in F$ as the terminal state.

The next lemma uses Properties A and B and the existence of a canonical blocking chain to construct a particular chain that will be later used to deter deviations from some on-path process.

LEMMA **2.** *Consider a farsighted stable set F satisfying Properties A and B. Suppose T moves from state $x \notin F$ to state y, $\Psi(x) = a$ and $\Psi(y) = b$. Then there is a state $z \in F$ and a coalitionally acceptable chain from y to z such that $u_j(z) \leq u_j(a)$ for some $j \in T$.*

Proof. Fix states x, y, a, b and a coalition T as in the statement of the lemma. Because any nonempty blocking chain is acceptable, there is nothing to prove if $u_j(b) \leq u_j(a)$ for some $j \in T$; simply take $z = b$. So in what follows, assume that $u_T(b) \gg u_T(a)$. In that case, by Property B, no subset of T belongs to the coalition structure at b. This implies that $y \neq b$, so $y \notin F$. Lemma 1 then tells us that there is a canonical blocking chain from y to b. Fix one such canonical blocking chain, $\mathbf{c} = \{y, y^1, \ldots, y^{m-1}, y^m\}$, $\{S^1, \ldots, S^m\}$, where $(y^m, S^m) = b$. Since no subset of T belongs to the partition at b, every player in T is involved in some coalitional move in this blocking chain.

We now consider two cases:

Case 1. Some subset W of T moves *only* in the final step from y^{m-1} to b, and so is part of the coalition S^m. Pick any $j \in W$. Modify the original blocking chain by adding an extra step at y^{m-1} in which $W - j$ breaks up into singletons and moves from y^{m-1} to y'. Note that $\pi(y')$ has all the same coalitions as $\pi(y^{m-1})$, except that W appears as singletons, and $u_i(y') = u_i(y^{m-1})$ if $i \notin W$, while $u_i(y') = 0$ if $i \in W$. With y' as an added step between y^{m-1} and y^m we have a new chain $\mathbf{c}' = \{y, y^1, \ldots, y^{m-1}, y', y^m\}$, $\{S^1, \ldots, S^{m-1}, W - j, S^m\}$, with $(y^m, S^m) = b$.[13] Clearly, this new chain is also a blocking chain. The critical feature of this new blocking chain is that at state y' player j has yet to move and $u_j(y') = 0$.

Property A assures us of the existence of $z \in F$ such that $u_j(z) \leq u_j(a)$ and $u_i(z) \geq u_i(b)$ for all $i \neq j$. Modify the blocking chain $\mathbf{c}'$ by replacing the terminal state with z to construct the chain $\bar{\mathbf{c}} = \{y, y^1, \ldots, y^{m-1}, y', z\}$, $\{S^1, \ldots, S^{m-1}, W - j, N\}$. Since $u_i(z) \geq u_i(b)$ for

[12] This will happen when a new coalition belongs to the union of previous coalitions in the chain.

[13] If $W = \{j\}$ this step is redundant: $y' = y^{m-1}$. However, it is still the case that at y' player j has yet to move and $u_j(y') = 0$.

all $j \neq i$, this chain clearly satisfies the acceptability conditions for all players in $N - j$. Since player j only moves at the last step, from y' to z, and $u_j(y') = 0$, the acceptability condition also holds for player j. Thus, $\bar{\mathbf{c}}$ is a coalitionally acceptable chain from y to z such $u_j(z) \leq u_j(a)$.[14]

Case 2. $T \subseteq \cup_{t<m} S^t$; i.e., every member of T has made some move by the time the state y^{m-1} is reached. Let $k < m$ be the maximal index such that some member of T belongs to S^k, and let j be any such member of T. Since all players in $T - S^k$ have already moved, a 'move' by S^k must mean that S^k is not a singleton, i.e., $S^k - j \neq \emptyset$. Now interpret the move to y^k as one made by $S^k - j$.[15] Keep the rest of the process unchanged until y^{m-1}. With this interpretation we have a blocking chain in which there is a player $j \in T$ who at state y^{m-1} has yet to make a move. In other words, there is a subset of T that moves *only* in the final step from y^{m-1} to y^m. But then we are back in Case 1. ■

Recall that for a given farsighted stable set we have chosen for every $x \notin F$ *some* blocking chain $\mathbf{c}(x)$ with next state y and terminal state $\Psi(x)$. We will now embed this in an absorbing process σ satisfying absolute maximality and coalitional acceptability. Recall that for any history h, $x(h)$ denotes the current state. Let $\ell(h)$ denote the state immediately preceding $x(h)$, in case there is one.

For any history h with current state $x(h) \in F$, the process will prescribe no change, i.e., if $x(h) \in F$, $\sigma(h) = (x(h), \emptyset)$.

Turning to histories in which the current state is not in F, we begin with a single-state history, or a one-step history, $h = \{x\}$ (where $x \notin F$). Take the already chosen blocking chain $\mathbf{c}(x)$ that leads from x to a state $\Psi(x)$. This gives us a mapping $\{\mathbf{c}(h), y(h), S(h), \Psi(x)\}$ which assigns to h some minimal blocking chain $\mathbf{c}(h)$ with next state $y(h)$, the first coalition in that blocking chain $S(h)$ and terminal state $\Psi(x) \in F$. The negotiation process σ will specify $\sigma(h) = (y(h), S(h))$ and $x^\sigma(h) = \Psi(x)$.

Next, consider a history h such that $x(h) \notin F$ and is immediately preceded by a state $\ell(h) \in F$. In this case, we proceed according to the chain already defined (in the previous paragraph) for the single-state history $h' = \{x(h)\}$, i.e., $\{c(h), y(h), S(h), x^\sigma(h)\} = \{c(h'), y(h'), S(h'), \Psi(x(h)\}$.

It remains to define the process for histories of the form h where $x(h) \notin F$ and $\ell(h) \notin F$. Note that we have already the defined the process for single-state histories. Recursively, suppose that we have defined our process $\{c(h), y(h), S(h), x^\sigma(h)\}$ for histories with K steps or less, where $K \geq 1$. Now consider a history h with $K + 1$ steps. Let h_K denote the first K steps. There are now three possibilities:

[14]Any player other player j who was an active mover in the original chain $\mathbf{c}$ (i.e. any player in $S^m - j$) will also gain *strictly* in the chain $\bar{\mathbf{c}}$. There are precisely two reasons why $\bar{\mathbf{c}}$ may not be a blocking chain: (i) some players in $N - S^m$, who were not involved in making a move in $\mathbf{c}$, may be assigned to a different coalition in $\pi(z)$ and may not gain (strictly) in following the path to z, (ii) it's possible that $u_j(y') = u_j(z) = 0$ so j does not experience a strict improvement.

[15]Formally, replace S^k by $S^k - j$.

(i) If $\ell(h) = x(h_K) = x(h_{K+1})$, ignore the last "move" and continue with the chain $\mathbf{c}(h_K)$ at h, which defines $\{c(h_{K+1}), y(h_{K+1}), S(h_{K+1}), x^\sigma(h_{K+1})\} = \{c(h_K), y(h_K), S(h_K), x^\sigma(h_K)\}$ from h.

(ii) If $x(h) = y(h_K)$, then we ignore whether the intended coalition $S(h_K)$ or some other coalition generated the state $x(h)$, and simply continue, as in (i), with the chain $\mathbf{c}(h_K)$ at h, which then defines the objects $\{c(h), y(h), S(h), x^\sigma(h)\}$ from h.

(iii) If $x(h) \neq y(h_K)$ (and both $x(h)$ and $x(h_K)$ are not in F), let T be the associated coalition in the last step of the history h, to be interpreted as the coalition that moved the state from $x(h_K)$ to $x(h)$. Let a equal the "intended" terminal state from h_K; i.e., $a = x^\sigma(h_K)$, and let y equal $x(h)$. By Lemma 2, there is a state $z \in F$ and an acceptable chain $\mathbf{c}'$ from y to z such that $u_j(z) \leq u_j(a)$ for some $j \in T$. Fix any such chain $\mathbf{c}'$ and assign it to the history h, defining $\{c(h), y(h), S(h), x^\sigma(h)\}$ accordingly at h. This last step ensures that given any history h, no coalition can profitably deviate from the path prescribed by $\mathbf{c}(h)$.

Proceeding recursively in this way, we define $\{c(h), y(h), S(h), x^\sigma(h)\}$ for every history h. Let σ be the induced process of state formation from the second and third elements of this map. Clearly, this process embeds F and is coalitionally acceptable. In fact, except for case (iii), for every history h with $x(h) \notin F$, σ describes a blocking chain terminating in F. For a history h with $x(h) \in F$, absolute maximality follows from farsighted internal stability of F. If $x(h) \notin F$, absolute maximality follows from the last step of the previous paragraph. ■

3.3. **Remarks on Properties A and B.** Properties A and B are trivially satisfied by every farsighted stable set with a unique payoff profile. Theorem 2 of Ray and Vohra (2015) shows that such sets always exist in games with interior core outcomes.

Apart from such games, we know a lot about farsighted stable sets in particular classes of games. (The general structure of farsightedness is yet to be fully understood.) Specifically, Ray and Vohra (2015) provide a full analysis of *simple games*, which are TU games in which each coalition S is either "winning" ($v(S) = 1$) or "losing" ($v(S) = 0$), and if a coalition is winning, then its complement is losing. Despite the simple-sounding nomenclature, simple games describe a rich class of situations: parliaments, bargaining institutions, and committees have been studied with this device.[16] Simple games may or may not possess nonempty cores.

Every farsighted stable set in a simple game satisfies Property B, an assertion that is a fairly straightforward consequence of farsighted internal stability. To see this, suppose there is a farsighted stable set F for which Property B does not hold. Then there are states a and b in F and a coalition T such that $u_T(b) \gg u_T(a)$ and $\sum_{i \in T} u_i(b) \leq v(T)$. This must mean that T is a winning coalition, which means that the complement of T is losing. But then, at state a, T can precipitate the zero state (by breaking up into singletons) and then counting on the winning coalition for state b to move to b, making T better off. In other words, b farsightedly dominates a, which contradicts farsighted internal stability of F. So Property B must hold.

[16]Simple games have been extensively analyzed in the context of the vNM stable set (see, for example, Lucas 1992), are used in theories of bargaining with majority voting (Baron and Ferejohn, 1989) and have played a significant role in the analysis of political institutions; see, for example, Austen-Smith and Banks (1999) and Winter (1996).

To understand whether Property A holds, it is useful to distinguish between two sub-classes of simple games. A veto player is a singleton coalition with a losing complement (she can single-handedly precipitate the zero state). If the set of all veto players is winning, say that the game is oligarchic. Oligarchic games have singleton farsighted stable sets (Ray and Vohra 2015, Theorem 3), which trivially satisfy Property A. Otherwise, the game is oligarchic, and now there are no singleton farsighted stable sets.

And yet, Property A is satisfied by a class of sets that played a central role in von Neumann and Morgenstern's analysis of stable sets. These are what they termed *discriminatory sets*. For a set of players $K \subseteq N$, a discriminatory set in payoff space is a set

$$D(K, \mathbf{a}) = \{u \in \mathbb{R}^N_+ \mid u_i = \mathbf{a}_i \text{ for } i \in K\}.$$

Those in K, the "discriminated players," each get a fixed amount, while the remaining surplus is divided arbitrarily among the "bargaining players."

It is easy to see that every discriminatory set satisfies Property A. Moreover, Ray and Vohra (2015, Theorem 5) prove the existence of a discriminatory farsighted stable set in every non-oligarchic simple game that has a non-elitist veto coalition, i.e., a minimal veto coalition with no indispensable members.[17] To summarize, we have established:

COROLLARY **1.** *In every simple game that has a non-elitist veto coalition, there exists an absolutely maximal farsighted stable set.*

We should mention that we do not know if the qualification in the Corollary is necessary.

3.4. **The Necessity of Properties A and B.** Thus, there is a sizable class of games that have farsighted stable sets satisfying Properties A and B. On the other hand, these properties are nontrivial restrictions. In this section, we ask: (i) are these properties always satisfied by any farsighted stable set, and (ii) even if they are not, might Theorem 1 still be valid without them? The answer to each of these questions is no. That said, we had to work hard to obtain these counterexamples.

Example 1: A farsighted stable set that fails Property A, satisfies Property B, and is not absolutely maximal.

Consider a four-player simple game with minimal winning coalitions is $\{\{1, 2, 3\}$, $\{1, 4\}$, $\{2, 4\}$ and $\{3, 4\}\}$.[18] Let $m = (1/3, 1/3, 1/3, 2/3)$. For every minimal winning coalition S let u^S be the utility profile where $u^S_i = m_i$ for $i \in S$ and $u^S_i = 0$ for $i \notin S$. The collection of all such utility profiles — $(1/3, 1/3, 1/3, 0)$, $(1/3, 0, 0, 2/3)$, $(0, 1/3, 0, 2/3)$ and $(0, 0, 1/3, 2/3)$ — is a vNM stable set known as a *main simple solution*. Ray and Vohra (2015) show that the set of all states corresponding to such utility profiles is a farsighted stable set. Dutta and Vohra (2017) show that all such farsighted stable sets also satisfy strong maximality. However, as we will now show, the main simple solution in this example does not satisfy absolute maximality.

[17] That is, every member of the coalition can be removed and replaced by any individual not in it, and the coalition would remain veto.

[18] This game can also be represented as a weighted majority game where the players' weights are $(1, 1, 2, 3)$ and a coalition is winning if and only if its aggregate weight is more than 3.5.

Let F denote the farsighted stable set corresponding to the main simple solution (the four payoffs described above, along with the respective winning coalitions). Although F satisfies Property B, it does not satisfy Property A. To see this, consider the states $a, b \in F$ where $u(a) = (1/3, 0, 0, 2/3)$ and $u(b) = (1/3, 1/3, 1/3, 0)$. There is no $z \in F$ with $u_3(z) = 0$, $u_1(z) \geq 1/3, u_2(z) \geq 1/3$. So we cannot appeal to Theorem 1 to show that F is absolutely maximal. Indeed, F is not absolutely maximal.

Apart from (E.1) and (E.2) we impose another condition on the effectivity correspondence that prevents a coalition from influencing the payoff distribution among the players who may be left as residuals. More precisely, we assume that if a winning coalition becomes smaller (because of the departure of some its members) and remains a winning coalition, the resulting (non-zero) surplus is shared equally among the players that remain.[19]

Suppose there is an absorbing process σ that embeds F and satisfies coalitional acceptability as well as absolute maximality. Consider the state x where $u(x) = (0, 0, 0.36, 0.64)$ with winning coalition $W(x) = N$. Since $x \notin F$, there is $x' \in F$ that farsightedly dominates it and σ leads from the history $h = \{x\}$ to x'. Ray and Vohra (2015, Lemma 1) tells us that there are just two possibilities. Either x' myopically dominates x, or $W^+ = \{i \in N \mid u_i(x') > u_i(x)\}$ and $W(x) - W^+$ are both losing coalitions.[20]

The present example is one of a constant-sum game: a coalition is losing if and only if its complement is winning. Since $W(x) = N$, this means that W^+ and $W(x) - W^+$ cannot both be losing. Thus the second case is ruled out and x' myopically dominates x. The only two states in F that myopically dominate x are $((1/3, 0, 0, 2/3), \{1, 4\})$ and $((0, 1/3, 0, 2/3), \{2, 4\})$. In either case, $u_3(x') = 0$.

To show that σ violates absolute maximality we argue that player 3 has a profitable deviation from x. Suppose player 3 leaves the grand coalition at x resulting in state y. Note that the residual coalition, $\{1, 2, 4\}$ is winning. Given that the residual players share equally in the surplus released by 3's departure, $u(y) = (0.12, 0.12, 0, 0.76)$. Since $y \notin F$, σ must prescribe a continuation that is coalitionally acceptable. Using the same kind of argument as in the previous paragraph, it can be shown that $x^\sigma(y) = ((1/3, 1/3, 1/3, 0), \{1, 2, 3\})$.[21] Player 3 can therefore gain by interfering in this way with any process that attempts to proceed from x to x'. In other words, F does not satisfy absolute maximality.

Interestingly, there is a non-elitist veto coalition in Example 1, namely $\{1, 2, 3\}$. As a minimal winning coalition of a constant-sum game it is of course a minimal veto coalition. It is non-elitist because it remains a veto coalition if we replace any of its members with player 4. By Ray and Vohra (2015, Theorem 5), there exists a discriminatory farsighted stable set, for example $D(\{4\}, 0.1)$, in which player 4 receives 0.1 and players 1, 2 and 3 receive any arbitrary division of 0.9. From the discussion in Section 3.3, this set is absolutely maximal. Example 1 therefore

[19] See Ray and Vohra (2015) for a more general version of this "monotonicity condition".

[20] In case (ii), W^+ can precipitate the zero state by leaving $W(x)$, followed by a move by $W(x')$ to x'.

[21] For example, if $x^\sigma(y) = ((1/3, 0, 0, 2/3), \{1, 4\})$ coalitional rationality implies that in the first step player 1 must leave $W(y)$, resulting in the state $y' = ((0, 0.18, 0, 0.82), \{2, 4\})$. But from y' it is not possible, by coalitional rationality, to end up at $((1/3, 0, 0, 2/3), \{1, 4\})$.

illustrates how absolute maximality refines the set of farsighted stable sets: some may satisfy absolute maximality while others may not.

Example 2: A farsighted stable set that satisfies Property A, fails Property B, and is not absolutely maximal.

Consider a six-player game in which each coalition S has only one efficient payoff $\nu(S)$. (Such a game is referred to as a hedonic game). A few words of explanation will help. Players 1 and 5 are symmetric, so are players 2 and 4. These are the four players whose positive payoffs vary across different states. Player 3 gets a constant payoff whenever her payoff is positive. Player 6 always gets a zero payoff. Both players create synergies with other players. Player 3 benefits from those synergies when her own payoff turns positive while player 6 is completely indifferent throughout.

Formally, the coalitional game is described as follows:

$$\begin{aligned}
&\nu(\{1,2\}) = \nu(\{4,5\}) = (3,3), \quad \nu(\{1,3\}) = \nu(\{3,5\}) = (2,2),\\
&\nu(\{2,3,4\}) = (4,2,4), \quad \nu(\{1,3,5\}) = (1,2,1)\\
&\nu(\{1,3,4,5,6\}) = (3,2,4,3,0), \quad \nu(\{1,2,3,5,6\}) = (3,4,2,3,0)\\
&\nu(\{2,3,4,5,6\}) = (4,2,4,3,0), \quad \nu(\{1,2,3,4,6\}) = (3,4,2,4,0),\\
&\nu(S) = 0 \text{ for all other } S.
\end{aligned}$$

There are as many states as there are coalition structures. However, many of them have the same payoff profile and differ only in the way in which some zero-payoff players are partitioned. To describe the collection of states that have the same payoff we need some more notation. For every coalition S, let π_S denote a subpartition of S and let $\Pi^0(S) = \{\pi_S \mid \nu(T) = 0 \text{ for all } T \in \pi_S\}$ be the collection of subpartitions that result in every player in S getting 0.[22]

Consider the set of states $F = X^1 \cup X^2 \cup X^3 \cup \{x^4, x^5, x^6, x^7\}$ where

		Payoffs to Players					
States	Structures	1	2	3	4	5	6
X^1	$\{1,2\},\{3,5\},\Pi^0(\{4,6\})$	3	3	2	0	2	0
X^2	$\{1,3\},\{4,5\},\Pi^0(\{2,6\})$	2	0	2	3	3	0
X^3	$\{2,3,4\},\Pi^0(\{1,5,6\}$	0	4	2	4	0	0
x^4	$\{1\},\{2,3,4,5,6\})$	0	4	2	4	3	0
x^5	$\{1,3,4,5,6\},\{2\}$	3	0	2	4	3	0
x^6	$\{1,2,3,5,6\},\{4\}$	3	4	2	0	3	0
x^7	$\{1,2,3,4,6\},\{5\}$	3	4	2	4	0	0

TABLE 1. Farsighted Stable Set for Example 2.

[22]For instance, $\Pi^0(\{1,2,3\}) = \{(\{1\},\{2\},\{3\}),(\{1,2,3\}),(\{1\},\{2,3\})\}$.

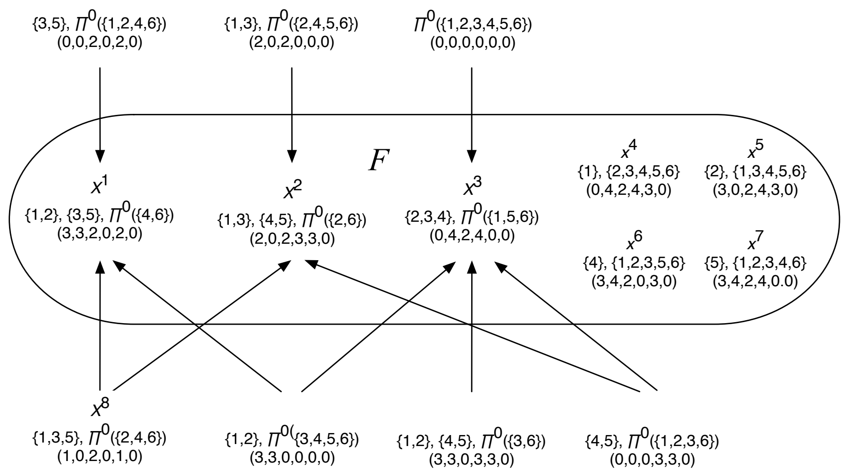

FIGURE 1. External Farsighted Stability of F in Example 2

We first show that F is a farsighted stable set. Figure 1 shows all the payoff equivalent states, with arrows indicating the states in F that farsightedly dominate a state not in F.

To see that F satisfies external farsighted stability: A state with payoff $(3,3,0,3,3,0)$ is dominated by one in X^3 through coalition $\{2,3,4\}$. The state with payoff $(3,3,0,0,0,0)$ is directly dominated by one in X^1 through $\{3,5\}$ and by one in X^3 through $\{2,3,4\}$. A state in X^8, with payoff profile $(1,0,2,0,1,0)$, is farsightedly dominated by a state in X^1 through the formation of coalition $\{1,2\}$, and also by a state in X^2 through coalition $\{4,5\}$. It is easy to see from Figure 1 that other states not in F are also farsightedly dominated by some state(s) in F.

To see that F satisfies internal farsighted stability: First observe that states x^4, x^5, x^6 and x^7 cannot dominate any other state (these states are in F only because they cannot be dominated by a state in $X^1 \cup X^2 \cup X^3$). This is so because such a state can emerge in only one of two ways: either a singleton precipitates it by leaving the grand coalition or it involves the active participation of player 6. Either case is inconsistent with farsighted dominance because both the singleton as well as player 6 receive 0. Secondly, none of these states can be farsightedly dominated by any other state. All players except for the excluded singleton are receiving the maximum possible payoff. Only the singleton has an incentive to change the state, but on her own she is powerless to do so. Thus, in checking internal stability we only need to compare states in X^1, X^2 and X^3.

From X^1 the only players who could gain by ending up at a state in X^2 are players 4 and 5. They can't move there directly. They could form a coalition of their own, resulting in payoffs $(3,3,0,3,3,0)$, but that can only be dominated by a state in X^3, not one in X^2, resulting in a payoff of 0 to player 5, which is of course not a farsighted improvement for $\{4,5\}$. Player 5 could exit coalition $\{3,5\}$ resulting in payoffs $(3,3,0,0,0,0)$ but from there the only possible moves are into X^1 or X^3, again making it impossible for player 5 to gain.

A state in X^1 cannot be farsightedly dominated by one in X^3 because any such move must begin by player 2 leaving coalition $\{1, 2\}$ which results in payoffs $(0, 0, 2, 0, 2, 0)$ from which the only further move that is possible is to X^1 or to X^2, not X^3, because players 3 or 5 the only ones who could initiate a move to X^3 have no interest in doing so. A similar argument shows that no state in X^2 can be farsightedly dominated by another state in F. Finally, note that at a state in X^3, all the non-zero payoff players are getting the highest possible amount and they together belong to one coalition, so no profitable deviation is possible.

This completes the proof that F satisfies farsighted internal stability.

Next, we show that F satisfies Property A and fails Property B. To see the former, notice, for instance, that Player 1 has varying payoffs in F but her worst payoff is attained at the state x^4. At that state each of the other players is getting their *maximum* possible payoff. We can make a parallel argument for Players 2, 4, and 5. Players 3 and 6 have payoffs that are invariant in F. So Property A is fully verified.

However, F does not satisfy Property B. Coalition $\{4, 5\}$ prefers a state in X^2 to a state in X^1 — the payoffs are $(3, 3)$ in the former, compared to (2,2) in the latter — and it can achieve the payoff $(3, 3)$ on its own.

Finally, we can show that F is not absolutely maximal; that is, it cannot be embedded in a absorbing, coalitionally acceptable, absolutely maximal process. This argument relies crucially on the fact that from a state in X^8 the *only* possible farsighted blocking chain runs to states in X^1 or X^2, not to a state in X^3. This is so because the only players who would prefer to have a state in X^8 replaced by one in X^3 are 2 and 4, but without the active participation of player 3 they are unable to carry out such a move.

Now consider any process that satisfies embedding of blocking chains and absorption into F. Take the history consisting of a single state in X^8. Since the only blocking chains from such a state are into X^1 or X^2, the continuation must be a single step into X^1 (through coalition $\{1, 2\}$) or into X^2 (through coalition $\{4, 5\}$). In the former case, coalition $\{4, 5\}$ has a profitable deviation into X^2 and in the latter coalition $\{1, 2\}$ has a profitable deviation into X^1. Thus, F is not absolutely maximal, which also shows that Property B cannot be dispensed with in our main theorem.

These examples also demonstrate that full history dependence (and zero discounting, as implicitly assumed) does not mean that anything goes. It is not the case that *any* farsighted stable set can be embedded in a coalitionally rational and absolutely maximal process. In short, a folk theorem is not to be had in the current context, particularly when we view the solution concept as pertaining to a *set* of states, which — in the spirit of von Neumann and Morgenstern stability — is the right thing to do.

Another kind of folk theorem might assert that a process satisfying coalitional rationality and absolute maximality can be constructed so that its absorbing states span the entire feasible set of payoffs, or something approximating it. But we seek to embed a farsighted stable set in an absorbing process. So this is not true simply because the internal stability of a (farsighted) stable set precludes it from being "too large." What if we don't insist on embedding a farsighted stable set in the process but focus on the absorbing states of a coalitionally rational and absolutely maximal process? Might that span the entire set of feasible payoffs? In general, in our model, the answer is still no. That follows from absolute maximality and our notion of an absorbing state which requires once such a state is reached, regardless of the history, it does not change. Together, these two properties imply that a non-core state and a state that (myopically) dominates it can't both be absorbing states.[23] Thus, in general, the absorbing states cannot span the entire set of feasible payoffs.

A loosening of these restrictions could lead to outcomes in which the entire set of payoffs is supportable. Dutta and Vartiainen (2017) consider a weaker notion of absorption or stationarity. A stable outcome in their sense may be stationary for some histories but not for others. The set of such states need not satisfy internal stability even if the process is maximal, and the set of "stable" outcomes can be large. Indeed, they find that in a strictly superadditive game, the set of all strictly positive, feasible payoffs is a "farsighted stable set" in their sense.

References

Aumann, Robert and Michael Maschler (1964), "The bargaining set for cooperative games", in Advances in Game Theory (M. Dresher, L. S. Shapley, and A. W. Tucker, Eds.), Annals of Mathematical Studies No. 52, Princeton Univ. Press, Princeton, NJ.

Aumann, Robert and Roger 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: Cambridge University Press.

Austen-Smith, David and Jeffrey Banks (1999), *Positive Political Theory I: Collective Preference*, Ann Arbor, MI: University of Michigan Press.

Baron, David and John Ferejohn (1989), "Bargaining in Legislatures," *American Political Science Review*, **83**, 1181–1206.

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

Bloch, Francis and Anne van den Nouweland (2017), "Farsighted stability with heterogeneous expectations", mimeo.

Chwe, Michael (1994), "Farsighted Coalitional Stability," *Journal of Economic Theory*, **63**, 299–325.

Diamantoudi, Effrosyni and Licun Xue (2003), "Farsighted Stability in Hedonic Games," *Social Choice and Welfare* **21**, 39–61.

Dutta, Bhaskar and Rajiv Vohra, "Rational Expectations and Farsighted Stability", (2017) *Theoretical Economics* **12**, 1191–1227.

Dutta, Bhaskar and Hannu Vartiainen, "Coalition Formation and History Dependence", (2017), mimeo.

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

[23] In other words, the set of absorbing states must satisfy myopic internal stability.

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

——— (2017), "Matching with Myopic and Farsighted Players", mimeo.

Harsanyi, John (1974), "An Equilibrium-Point Interpretation of Stable Sets and a Proposed Alternative Definition," *Management Science* **20**, 1472–1495.

Jordan, James (2006), "Pillage and Property," *Journal of Economic Theory* **131**, 26–44.

Kimya, Mert (2015),"Equilibrium Coalitional Behavior", mimeo, Brown University.

Konishi, Hideo and Debraj Ray (2003), "Coalition Formation as a Dynamic Process," *Journal of Economic Theory* **110**, 1–41.

Lucas, William (1992), "von Neumann-Morgenstern Stable Sets," in *Handbook of Game Theory, Volume 1*, ed. by Robert Aumann, and Sergiu Hart, 543–590. North Holland: Elsevier.

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

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

Ray, Debraj and Rajiv Vohra (1997), "Equilibrium Binding Agreements," *Journal of Economic Theory* **73**, 30–78.

——— (1999), "A Theory of Endogenous Coalition Structures," *Games and Economic Behavior* **26**, 286–336.

——— (2014), "Coalition Formation," in *Handbook of Game Theory, Volume 4*, ed. by H. Peyton Young and Shmuel Zamir, 239–326. North Holland: Elsevier.

——— (2015), "The Farsighted Stable Set", *Econometrica*, **83**, 977–1011.

von Neumann, John and Oskar Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press.

Winter, Eyal (1996), "Voting and Vetoing," *American Political Science Review* **90**, 813–823.

Xue, Licun (1998), "Coalitional Stability under Perfect Foresight," *Economic Theory*, **11**, 603–627.

ONLINE APPENDIX [NOT FOR PUBLICATION]

APPENDIX A.1. DROPPING PROPERTY B

Property B is known to be automatically satisfied in some sub-classes of games. The following Corollary is an immediate consequence of Theorem 1.

COROLLARY **2.** *If a farsighted stable satisfies Property A, then it is absolutely maximal in any of the following circumstances:*

(i) F is a single-payoff farsighted stable set (in which case even Property A is redundant),

(ii) If the game is a simple game, or

(iii) For every $x \in F$, $\pi(x) = N$.

Property B can also be dispensed with in a parallel situation described in the next section.

Appendix A.2. Strengthening Coalitional Acceptability

As we argued in Section 3, we restrict off-path histories to satisfy coalitional acceptability, which is a weaker requirement than the strict improvements required by a blocking chain. What if we were to strengthen coalitional acceptability to the requirement that the process define a blocking chain from *every* history ending at a state outside the farsighted stable set? Treating off-path and on-path histories in the same way does have the benefit of simplifying the definition of absolute maximality.

A farsighted stable set F is said to be *strictly embedded* in an absorbing process σ if

(i) F is the set of all absorbing states of σ.

(ii)′ At any initial history h with $x(h) \notin F$, the continuation chain from h is a blocking chain terminating in F.

Condition (ii)′ subsumes our earlier condition (ii) of embedding. It also ensures that σ automatically satisfies coalitional acceptability. We now explore the possibility of extending our result to show that a farsighted stable set can be strictly embedded in a process satisfying absolute maximality.

To strengthen Theorem 1 in this direction, it should be clear from its proof that it suffices to strengthen Lemma 2 so that the coalitionally acceptable chain constructed to deter deviations is in fact a blocking chain. At a minimum, this will require that when we dissuade an off-path deviation by by finding a coalitionally acceptable chain from y to $z \in F$, all players involved in this chain must receive a strictly positive payoff at z. Indeed, this must be a feature of any state in the farsighted stable that dominates some other state. In a general coalitional game, a state x is said to be *regular* if $u_i(x) > 0$ for every i such that $i \in S \in \pi(x)$ and $u_S(x) > 0$. In a simple game this reduces to the condition that $u_i(x) > 0$ for all $i \in W(x)$, as in Ray and Vohra (2015). That suggests that Property A must be modified to refer to regular states.

Property A′. *Suppose there are two regular states a and b in F such that $u_j(b) > u_j(a)$ for some j. Then there exists a regular state $z \in F$ such that $u_j(z) \leq u_j(a)$, and $u_i(z) \geq u_i(b)$ for all $i \neq j$.*

Clearly, every discriminatory farsighted stable set satisfies Property A′.

Modulo the replacement of Property A with A′, our next result shows that Corollary 2 can be strengthened to require strict embedding.

Proposition **1.** *If a farsighted stable set satisfies Property A′, then it can be strictly embedded in a process satisfying absolute maximality in any of the following circumstances:*

(i) F is a single-payoff farsighted stable set (in which case even Property A′ is redundant),

(ii) If the game is a simple game, or

(iii) For every $y \notin F$ there is a blocking chain from y to $x \in F$ with $\pi(x) = N$.[24]

Proof. (i) By Theorem 1 in Dutta and Vohra (2017) a single-payoff farsighted stable set can be strictly embedded in a history-independent (stationary) process satisfying maximality. Clearly *no* coalition can find a profitable deviation when the final payoff is unique, so in this case absolute maximality is also satisfied.

[24] This is weaker than the corresponding assumption stated in Corollary 1, namely that $\pi(x) = N$ for every $x \in F$.

(ii) Suppose F is a farsighted stable set of a simple game. For every state $x \notin F$, fix any blocking chain, $\mathbf{c}(x)$, with $\Psi(x) \in F$ as the terminal state. We can now modify Lemma 2 as follows:

LEMMA **3.** *Consider a farsighted stable set F of a simple game satisfying Property A$'$. Suppose T moves from state $x \notin F$ to state y, $\Psi(x) = a$ and $\Psi(y) = b$. Then there is a state $z \in F$ and a blocking chain from y to z such that $u_j(z) \leq u_j(a)$ for some $j \in T$.*

Proof. Fix states x, y, a, b and a coalition T as in the statement of the lemma. Note that as the terminal states of a blocking chain both a and b are regular states. If $u_j(b) \leq u_j(a)$ for some $j \in T$ there is nothing to prove, so assume that $u_T(b) \gg u_T(a)$. This means that T is not a veto coalition. Otherwise, we could find a farsighted chain from a to b (with T first moving to the zero state followed by $W(b)$ moving to b), contradicting the farsighted internal stability of F. Since T is not a veto coalition, and therefore not a winning coalition, $u_T(y) = 0$, which clearly means that $y \neq b$ and $y \notin F$. Pick any $j \in T$. By Property A', there exists a regular state $z \in F$ such that $u_j(z) \leq u_j(a)$ and $u_i(z) \geq u_i(b)$ for all $i \neq j$. To complete the proof we will now construct a farsighted chain from y to z.

Since T is losing, $u_T(y) = 0$ and $T \cap W(y) = \emptyset$. Given that there is a blocking chain from y to b, by Lemma 1 of Ray and Vohra (2015) there are two possibilities: (i) $W(b)$ moves directly from y to b or (ii) some $S \subset W(y)$ first moves from y to the zero state, followed by a move by $W(b)$ to b. In case (i) consider a move by $W(z)$ to the zero state (by breaking up into singletons) followed by a move by $W(z)$ to z. In case (ii) consider S moving to the zero state followed by $W(z)$ moving to z. All players except possibly j who gained by having y replaced with b, will also gain if y is replaced with z. If player $j \notin W(z)$ we have clearly found a blocking chain from y to z. If $j \in W(z)$, since z is regular, $u_j(z) > 0$ while $u_j(y) = 0$, so player j also gains in moving from y to z, and again we shown that there is a blocking chain from y to z. ■

The rest of the proof of Theorem 1 remains unchanged.

(iii) Consider a farsighted stable set F of a coalitional game such that for every $y \notin F$ there is a blocking chain from y to $x \in F$ with $\pi(x) = N$. Suppose F satisfies Property A$'$. Again, it will suffice to prove an appropriately modified version of Lemma 2 that provides a blocking chain rather than an acceptable chain.

For every state $x \notin F$, fix any blocking chain with $\Psi(x) \in F$ as the terminal state such that $\pi(\Psi(x)) = N$.

LEMMA **4.** *Consider a farsighted stable set F satisfying Property A$'$ and condition (iii). Suppose T moves from state $x \notin F$ to state y, $\Psi(x) = a$ and $\Psi(y) = b$, with $\pi(b) = N$. Then there is a state $z \in F$ and a blocking chain from y to z such that $u_j(z) \leq u_j(a)$ for some $j \in T$.*

Proof. Fix states x, y, a, b and a coalition T and assume that $u_T(b) \gg u_T(a)$. Of course, by farsighted internal stability, $T \neq N$, so $y \neq b$ and $y \notin F$. Fix one canonical blocking chain from y to b, say, $\{y, y^1, \ldots, y^{m-1}, y^m\}, \{S^1, \ldots, S^m\}$, where $y^m = b$. Proceeding exactly as in the proof of Lemma 2, we can find a player $j \in T$ and a new blocking chain $\{y, y^1, \ldots, y^{m-1}, y', y^m\}, \{S^1, \ldots, S^{m-1}, W - j, S^m\}$, such that at state y' player j has yet to move and $u_j(y') = 0$. Because $\pi(b) = N$, $S^m = N$: all players must be involved in the final move of the blocking chain from y to b and $u(b) \gg 0$.

Pick z as given by Property A$'$. We will now construct a blocking chain from y to z. (This is not surprising given Footnote 14). Consider the two cases corresponding to whether $u_j(z) > 0$:

(i) If $u_j(z) > 0$ consider the chain $\{y, y^1, \ldots, y^{m-1}, y', z\}$, $\{S^1, \ldots, S^{m-1}, W - j, N\}$. In the original blocking chain all players were active movers, so all players in $N - j$ are also strictly better off at z compared to the first step at which they made a move in the original blocking chain. The move to z is also strictly profitable for j. This means that we have a found a blocking chain from y to z.

(ii) If $u_j(z) = 0$ consider the chain $\{y, y^1, \ldots, y^{m-1}, y', z\}$, $\{S^1, \ldots, S^{m-1}, W - j, N - j\}$. The last step is feasible for $N - j$ because z is a regular state with $u_i(z) \geq u_i(b) > 0$ for all $j \neq i$ and $u_j(z) = 0$. It is also clearly a blocking chain. ■

■

APPENDIX A.3. WEAKENING ABSOLUTE MAXIMALITY

A farsighted stable set is *maximal* if it can be embedded in an absorbing, coalitionally acceptable process σ satisfying conditions (i) and (ii) in the main text, and:

[Maximality] At no history h does there exist a state y with $S(h) \in E(x(h), y)$, such that $u_{S(h)}(x^\sigma(h, y, S(h))) \gg u_{S(h)}(x^\sigma(h))$.

The difference from absolute maximality is that maximality requires *only* the moving coalition at any history to not have a better move. Under absolute maximality *no* coalition can have a better move.

If we weaken absolute maximality to maximality, Property B can be dropped from Theorem 1. Moreover, in simple games even Property A can be dropped:

PROPOSITION **2.** *If a farsighted stable set satisfies Property A then it is maximal.*

Proof. The proof relies on a version of Lemma 2 that does not invoke Property B. Consider any farsighted stable set F. If $x \in F$, define $\Psi(x) = x$. For each state $x \notin F$, fix any blocking chain, $\mathbf{c}(x)$, that reaches F in a *minimal* number of steps. In particular, if there is a coalition T such that $T \in E(x, a)$ with $a \in F$ and $u_T(a) \gg u_T(x)$, then $\mathbf{c}(x)$ must reach F in one step. Let $\Psi(x) \in F$ be the terminal state for this minimal blocking chain.

LEMMA **5.** *Consider a farsighted stable set F satisfying Property A. Suppose that in a minimal blocking chain* $\mathbf{c}$*, coalition T is supposed to make the first move from $x \notin F$ to y'. Suppose T moves instead from x to $y \neq y'$. Let $\Psi(x) = a$ and $\Psi(y) = b$. Then there is a state $z \in F$ and an acceptable chain from y to z such that $u_j(z) \leq u_j(a)$ for some $j \in T$.*

Proof. Fix states x, y, a, b and a coalition T as in the statement of the lemma and suppose that $u_T(b) \gg u_T(a)$. If $y \in F$, because $\mathbf{c}$ is a blocking chain initiated by T that leads to F in a minimal number of steps, a must also be the result of a one-step move by coalition T, i.e., $y' = a$. Notice that all players not in T are in the same coalition in a as in b and their payoffs are also the same in a and b. The only difference between states a and b is that T may be organized in a different partition and $u_T(b) \gg u_T(a)$. Clearly, $T \in E(a, b)$, but this contradicts internal stability of F. Thus, $y \notin F$ (and so $b \neq y$). The rest of the proof is the same as that of Lemma 2. ■

Lemma 5 can be used along with the rest of the proof of Theorem 1 to complete the proof of this proposition. ■

In simple games even Property A (or A′) can be dropped from Lemma 5. Fix states x, y, a, b and a coalition T as in the statement of Lemma 5 and suppose that $u_T(b) \gg u_T(a)$. In a simple game T must be a losing coalition; if it is a winning coalition, it could have gone from a to b on its own, contradicting internal stability. But a losing coalition that was supposed to move from x to y' cannot move anywhere other than y' (because of coalitional sovereignty), so in fact it has no available "deviation", and the conclusion of Lemma 5 follows. As in the proof of Proposition 1, we can strictly embed a farsighted stable set in a process satisfying maximality. It is for absolute maximality that we rely on Property A′. We therefore have:

PROPOSITION **3.** *(i) Every farsighted stable set of a simple game can be strictly embedded in a process satisfying maximality.*

(ii) Moreover, if it satisfies Property A′, then it can be strictly embedded in a process satisfying absolute maximality.

Dutta and Vartiainen (2017) show that the conclusion in (i) holds for strong maximality (not just maximality). The difference being that strong maximality allows deviations from any coalition that has a non-empty intersection with the one that is supposed to move. Whether our framework allows us to extend the previous proposition to strong maximality is not known to us. Absolute maximality, however, is a much stronger property, and as Example 1 shows, even in a simple game, it cannot be guaranteed without additional conditions such as Property A′.

Proposition 3 (as well as our earlier results) depend crucially on allowing for history dependent processes. To see this, consider the three–player simple game where $N = \{1,2,3\}$, $v(N) = v\{1,2\} = v\{1,3\} = 1$ and $v(S) = 0$ for all other S. Ray and Vohra (2015) show that every farsighted stable set of this game is a discriminatory set in which player 1 receives a fixed payoff $a \in (0,1)$ and the remaining surplus is divided in any arbitrary way between players 2 and 3. Of course, this set satisfies Property A′ and by Proposition 3 (ii) it can be strictly embedded in an absolutely maximal process. History dependence is important for this result. As Dutta and Vohra (2016) point out, a farsighted stable set of this form cannot be supported by a history independent process.

This example can also be used to show that the gap between absolute maximality and (strong) maximality persists even if we focus on the absorbing states of a history independent process, which need not be a farsighted stable set. Dutta and Vohra (2016) construct a stationary process that supports (as a SREFS) the set of states with payoffs $(\{a+b,b,0),(a+b,0,b),(a,b,b)\})$, where $b = (1-a)/2$. From the zero state the process proceeds to the last of these states. From a state where player 1 gets less than $a+b$ and 2 gets less than b it proceeds directly to the state where they get $a+b$ and b respectively. From a state where both 2 and 3 get less than 3, it leads to the zero.

But this process cannot satisfy absolute maximality. The reasoning is similar to the one used in Example 1 above. Consider the state x with $\pi(x) = N$ and $u(x) = (a+b-1/3\epsilon, b-2/3\epsilon, \epsilon)$. Suppose that a departure by player 3 results in the other two sharing the extra surplus equally. Now, if player 3 leaves the grand coalition, the new state leaves player 2 with a payoff strictly less than b. If players 1 and 2 share the surplus equally, player 2 gets less than b. This will then lead to the zero state, followed by N moving to the stationary state with payoffs (a, b, b). Thus, player 3, has a profitable deviation at state x.