

Ray, Debraj; Vohra, Rajiv

**Working Paper**

## The farsighted stable set

Working Paper, No. 2013-11

**Provided in Cooperation with:**

Department of Economics, Brown University

*Suggested Citation:* Ray, Debraj; Vohra, Rajiv (2013) : The farsighted stable set, Working Paper, No. 2013-11, Brown University, Department of Economics, Providence, RI

This Version is available at:

<https://hdl.handle.net/10419/102634>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# The Farsighted Stable Set

Debraj Ray (New York University)

and

Rajiv Vohra (Brown University)

October 2013

**Abstract.** Harsanyi (1974) criticized the von Neumann-Morgenstern notion of a stable set on the grounds that it implicitly assumes coalitions to be shortsighted in evaluating their prospects. He proposed a modification of the dominance relation to incorporate farsightedness. In doing so, however, Harsanyi retained another feature of the stable set: that a coalition  $S$  can impose any imputation as long as its restriction to  $S$  is feasible for  $S$ . This implicitly gives an objecting coalition complete power to arrange the payoffs of players elsewhere, which is clearly unsatisfactory. While this assumption is absolutely innocuous for the classical stable set, it is of crucial significance for farsighted dominance. Our proposed modification of the Harsanyi set respects “coalitional sovereignty.” The resulting farsighted stable set is very different, both from that of Harsanyi or of von Neumann and Morgenstern. We provide a necessary and sufficient condition for the existence of a farsighted stable set containing just a single payoff allocation. This condition is weaker than assuming that the relative interior of the core is non-empty, but roughly establishes an equivalence between core allocations and the union of allocations over all single-payoff farsighted stable sets. We state two conjectures: that farsighted stable sets exist in all transferable-utility games, and that when a single-payoff farsighted stable set exists, there are no farsighted stable sets containing multiple payoff allocations.

## 1. INTRODUCTION

In formulating a theory of binding agreements, von Neumann and Morgenstern (1944) proposed a “solution” for cooperative games, an equilibrium concept that is often referred to as the *von Neumann-Morgenstern stable set*. It is based on the concept of coalitional dominance. A feasible payoff profile is dominated by another if some coalition prefers the latter profile (all its members receive a higher payoff) and can unilaterally precipitate that profile. A set of feasible outcomes  $Z$  is a stable set if it satisfies two properties:

*Internal Stability.* If  $x \in Z$ , it is not dominated by  $y \in Z$ .

*External Stability.* If  $x \notin Z$ , then there exists  $y \in Z$  which dominates  $x$ .

Thus the elements of  $Z$  are precisely those outcomes (and only those) which are undominated by any other outcome in  $Z$ . von Neumann and Morgenstern interpreted a stable set as a “standard of behavior”. Once accepted, no allocation satisfying the standard can be overturned by another allocation also satisfying the standard (internal stability), and these allocations jointly overrule all outcomes that do not satisfy the standard (external stability).

Of course, a stable set must include the core, the set of all undominated payoff profiles. But it could have other members:  $x \in Z$  may well be dominated by  $y$  as long as  $y \notin Z$ . To be sure,

external stability guarantees that  $y$  in turn can be “blocked” by some other profile  $z \in Z$ . The presumption, then, is that  $x$  should still be considered “stable,” because  $y$  does not represent a lasting benefit.

Harsanyi (1974) took issue with this presumption. He observed that this argument is only valid if  $z$  isn’t preferred by the coalition which caused  $x$  to be replaced by  $y$ . After all, perhaps  $y$  was only a ruse to induce  $z$  in the first place. The vNM stable set is based on a myopic notion of dominance, and does not address this concern. Harsanyi went on to propose a modification of the dominance concept to incorporate farsighted behavior. A formal definition of a stable set based on Harsanyi’s notion, which we henceforth refer to as the Harsanyi stable set, appears in Chwe (1994).

In this paper, we argue that Harsanyi’s suggested modification of the stable set is problematic because it retains certain features of the original vNM concept that are fundamentally ill-suited for farsightedness. The problem arises from a seemingly innocuous device adopted by von Neumann and Morgenstern. They defined dominance and stability over the domain of *imputations*: efficient and individually rational payoff profiles over the *full* set of players. More precisely, an imputation  $y$  dominates another imputation  $x$  if there is a coalition  $S$  for which  $y_S$  (the restriction of  $y$  to  $S$ ) is feasible for  $S$ , and  $y_i > x_i$  for all  $i \in S$ . The interest, of course, lies in the restriction of  $y$  to  $S$ , because that is where the “dominance” occurs. The remainder of the dominating imputation  $y$  only ensures that all allocations live in the same full-dimensional space, making for simpler and more elegant exposition.

But, as we shall argue here, it does matter (and crucially so) when farsighted stability is involved. The use of imputations effectively grants a coalition  $S$  the power to replace imputation  $x$  with imputation  $y$  as long as  $y_S$  is feasible for  $S$ . That is, the objecting coalition dictates the complementary allocation  $y_{N-S}$ . That allocation need not even be feasible for the complementary set of players, but it is presumed that  $S$  can somehow engineer society-wide changes to make this happen. Moreover, even if it is feasible, the implicit presumption is that  $S$  has unlimited power in rearranging payoffs for  $N - S$ . In effect, then, the Harsanyi definition denies the “coalitional sovereignty” of players outside  $S$ , and taken literally, it grants a coalition extraordinary power in the affairs of outsiders.

None of this really matters for the vNM stable set. Whether or not we carry the extra irrelevant components of  $y$ , or whether or not we account properly for feasibility and coalitional sovereignty, makes no difference at all to the definition of vNM stability. As already noted, the “imputational approach” is only a convenience that allows all allocations to dwell in the same  $n$ -dimensional space. But with farsighted dominance matters aren’t quite as harmless. If coalition  $S$  can replace  $x$  with  $y$  (where  $y_S$  is feasible for  $S$ ), what transpires thereafter — for instance, whether another coalition  $T$  further replaces  $y$  with  $z$  — depends crucially on  $y_{N-S}$ . In particular, the specification of  $y_{N-S}$  affects the ability of  $S$  to trigger a farsighted objection. That forces us to look more closely at feasibility and coalitional sovereignty.

We propose a definition of a farsighted stable set that does just this, in effect extending Harsanyi’s suggested modification to the vNM stable set. As it turns out, this is not just a conceptual issue. It has a profound effect on the nature of farsighted stable sets.

A recent literature on Harsanyi stable sets provides existence results under fairly weak conditions; see Béal et al. (2008) for TU games, and Bhattacharya and Brosi (2011) for NTU games. This is somewhat surprising given the difficulties that were encountered in settling the existence of vNM stable sets, but as we shall see, the existence results here do depend to some degree on the neglect of feasibility and coalitional sovereignty in the Harsanyi definition. As far as our solution is concerned, our existence results are not as comprehensive. We provide a condition, called separability, that characterizes all farsighted stable sets that contain a *single* payoff allocation (Theorem 2). The separability condition is satisfied in all superadditive games in which the interior of the core is non-empty.

It turns out that all such single-payoff sets are core allocations, and every payoff allocation in the interior of the core is a (single-payoff) farsighted stable set. This is an intriguing connection and suggests that the core of a game — even though it is defined by the property of not being “myopically blocked” — has powerful farsighted stability properties that have not been hitherto explored.

Farsighted stable sets have also been studied in the class of hedonic games, where each coalition has a unique payoff profile. The focus here is on coalition structures rather imputations for the grand coalition. Since in these games there is no ambiguity about the payoff profile of  $N - S$  when coalition  $S$  makes a move, coalitional sovereignty is no longer an issue. The existence results of Diamantoudi and Xue (2003) for hedonic games and of Mauleon et al (2011) for matching games (a special class of hedonic games) are therefore not subject to our general criticism of Harsanyi stable sets. At the same time, our framework can be applied directly to hedonic games. The existence theorems of Diamantoudi and Xue (2003) and Mauleon et al (2011) in the context of hedonic games follow from our main result; see Section 5.3.

Section 5 contains a number of examples that explore the existence and structure of stable sets. Example 1 uses the celebrated counterexample of Lucas to show how the separability condition guarantees the existence of a farsighted stable set even when the vNM stable set fails to exist. However, Example 2 invokes the well-known “roommate problem” to show that a farsighted stable set does not exist in every non-transferable utility game (the vNM stable set does not exist either in this example). The harder question of existence in a transferable utility game remains open (this is the question that Lucas settled in the negative for vNM stable sets). Example 3 underlines the dramatic difference between the farsighted stable set and the Harsanyi stable set. In this Example, payoffs in the interior of the core corresponds precisely to the set of all separable allocations, and therefore to the union of all single payoff farsighted stable sets. In contrast, the union of all Harsanyi stable sets is the set of imputations *not* in the interior of the core. Indeed, under the mild conditions identified in Béal et al. (2008), this is a general truth: Harsanyi stable sets are always single-payoff sets, and they are *always* disjoint from the interior of the core.

The remaining examples attempt to explore the existence of farsighted stable sets in transferable-utility games when the separability condition fails. Examples 4 and 5 show that farsighted stable sets exist in a transferable utility game when the separability condition fails. These sets contain multiple payoffs, and there may be many such sets. Example 6 shows, however, that in a perturbed version of Example 5, *all* these sets conveniently disappear exactly when the separability condition is restored, and the only sets that are farsighted stable are precisely the single-payoff

sets uncovered in Theorem 2. Whether these properties are generally true of transferable-utility games remains a matter of conjecture.

## 2. STABILITY AND THE HARSANYI CRITIQUE

**2.1. Preliminaries.** A characteristic function game is denoted by  $(N, V)$  where  $N = \{1, \dots, n\}$  is the finite set of players and for each coalition  $S \subseteq N$ , the set of feasible utility vectors is  $V(S) \subseteq \mathbb{R}^S$ , the  $S$ -dimensional Euclidean space with coordinates indexed by the players in  $S$ . The generic notation  $u$  will be taken without comment to refer to a payoff vector for the entire set of players. We reserve the generic notation  $v$  for payoffs to a particular coalition.

For all  $S \subseteq N$ ,  $V(S)$  is assumed to be *comprehensive*: if  $u \in V(S)$  then  $u' \in V(S)$  for all  $u' \leq u$ . Normalize the game so that singletons obtain zero:  $V(\{i\}) = \mathbb{R}_-^{\{i\}}$  for all  $i \in N$ . Assume all coalitions can get nonnegative but bounded payoffs:  $V(S) \cap \mathbb{R}_+^S$  is nonempty and compact.

A payoff vector  $v \in V(S)$  is *efficient for  $S$*  if  $v \in \bar{V}(S) = \{v' \in V(S) \mid \text{there is no } v'' \in V(S) \text{ with } v'' > v'\}$ .<sup>1</sup> A game is *strictly comprehensive* if the set of efficient payoffs for  $S$  coincides with the boundary of  $V(S)$ , i.e.,  $\bar{V}(S) = \{v' \in V(S) \mid \text{there is no } v'' \in V(S) \text{ with } v'' \gg v'\}$ .

A transferable utility (TU) game is one in which each coalition  $S$  has a number (its worth),  $v(S)$ , such that  $V(S) = \{v \in \mathbb{R}^S \mid \sum_{i \in S} v_i \leq v(S)\}$ . Note that a TU game is strictly comprehensive.

A game is *superadditive* if for any  $S, T \subseteq N$  such that  $S \cap T = \emptyset$ ,  $V(S) \times V(T) \subseteq V(S \cup T)$ .

An *imputation* is any payoff vector that is feasible and efficient for the grand coalition, and individually rational; that is, nonnegative. Denote the set of all imputations by  $I(N, V) = \bar{V}(N) \cap \mathbb{R}_+^N$ .

**2.2. The Core and the Stable Set.**  $(S, v)$  is an *objection* to  $u \in V(N)$  if  $v \in V(S)$  and  $v \gg u_S$ . The *core* of  $(N, V)$ , denoted  $C(N, V)$ , is the set of all payoff profiles in  $V(N)$  to which there is no objection:

$$C(N, V) = \{u \in V(N) \mid \text{there is no objection to } u\}.$$

It will be useful to present an alternative definition of the core based on imputations. Say that an imputation  $u'$  *dominates* imputation  $u$  if there exists a coalition  $S$  such that  $u'_S \in V(S)$  and  $u'_S \gg u_S$ . It is easy to see that for superadditive games the core can be expressed equivalently as follows:<sup>2</sup>

$$C(N, V) = \{u \in I(N, V) \mid u \text{ is not dominated by any } u' \in I(N, V)\}.$$

For  $A \subseteq I(N, V)$ , let

$$\text{dom}(A) = \{u \in I(N, V) \mid u \text{ is dominated by some } u' \in A\}.$$

<sup>1</sup>We use the convention  $\geq, \gg$  to order vectors in  $\mathbb{R}^N$ .

<sup>2</sup>See Shapley and Shubik (1969).

The core can then be written as

$$C(N, V) = I(N, V) - \text{dom}(I(N, V)).$$

A set of imputations  $Z$  is said to be a *vNM stable set* of  $(N, V)$  if it satisfies:

*Internal Stability.* No imputation in  $Z$  is dominated by any other imputation in  $Z$ .

*External Stability.* Every imputation *not* in  $Z$  is dominated by some imputation in  $Z$ .

In other words  $Z$  is a vNM stable set if

$$Z = I(N, V) - \text{dom}(Z).$$

The definition of a stable set, unlike that of the core, is circular. While any imputation can be tested for core stability, the vNM stability notion applies to a *set* of imputations.

**2.3. The Harsanyi Critique.** If a vNM stable set exists it must contain the core. But it generally contains other imputations as well. These imputations have objections: if  $u$  is such an imputation, there exists another imputation  $u'$  that dominates it via some coalition  $S$ . However, internal stability assures us that  $u' \notin Z$ . Moreover, by external stability,  $u'$  must itself be dominated by some imputation  $u'' \in Z$ . That is, the stability of  $u$  is based on the fact that while  $S$  has the power to replace  $u$  with  $u'$ , where  $u'_S \gg u_S$ , this does not represent a “permanent gain” to  $S$  because  $u'$  will be replaced by  $u''$ .

But as Harsanyi (1974) correctly noted, this argument is flawed. Whether  $S$  should replace  $u$  with  $u'$  depends on how  $S$  will fare thereafter. For instance, if  $u'_S \gg u_S$ ,  $S$  may well replace  $u$  with  $u'$ , anticipating that  $u'$  will in turn be replaced by  $u''$ , which is stable and yields a permanent gain. All that matters is that  $u'_S \gg u_S$ , and not how  $u'_S$  compares with  $u_S$ . Harsanyi went on to suggest a notion of farsighted dominance that takes such considerations into account:

An imputation  $u'$  *farsightedly dominates*  $u$  if there are imputations  $u^0, u^1, \dots, u^m$  and a corresponding collection of coalitions,  $S^1, \dots, S^m$ , where  $u^0 = u$  and  $u^m = u'$ , such that:

- (i)  $u^k \in V(S^k)$  for all  $k = 1, \dots, m$ ; and
- (ii)  $u'_{S^k} \gg u(u^{k-1})_{S^k}$  for all  $k = 1, \dots, m$ .

Thus, there could be several steps in moving from  $u$  to  $u'$ . Farsighted dominance requires that each coalition that is called upon to make a (feasible) move gains at the *end* of the process. What matters to each coalition involved in farsighted dominance is the “final outcome”; what transpires along the “intermediate steps” is irrelevant.<sup>3</sup>

---

<sup>3</sup>This is actually the second of two dominance notions proposed by Harsanyi (1974), and the one we shall concentrate on. The other one he proposed required, in addition, that each coalition also make an instantaneous gain, i.e., in addition to (i) and (ii) the condition that  $u'_{S^k} \gg u_{S^k}^{k-1}$  for all  $k$ . We will say that  $u'$  *strictly farsightedly dominates*  $u$  if (i), (ii) and (iii) are satisfied. But if payoffs along the chain are considered important a more satisfactory approach would be to account for *all* the payoffs along the chain, in effect making for a model in which payoffs are received in real time. This is the approach taken in Ray and Vohra (2013), but we shall not pursue it here.

The new dominance relation leads to the following modification of the vNM stable set; see Chwe (1994). For  $A \subseteq I(N, V)$ , let

$$\text{dom}_H(A) = \{u \in I(N, V) \mid u \text{ is farsightedly dominated by some } u' \in A\}.$$

A set of imputations  $H \subseteq I(N, V)$  is a *Harsanyi stable set* if

$$H = I(N, V) - \text{dom}_H(Z).$$

Observe how this construction takes care of the Harsanyi critique. If  $S$  replaces  $u$  with  $u'$ , anticipating a string of moves to some stable final outcome  $u''$ , then  $u''$  farsightedly dominates  $u'$ . But  $S$  benefits as well, then in addition,  $u''$  also farsightedly dominates  $u$ . But that contradicts internal stability.

**2.4. Feasibility, Effectivity and Coalitional Sovereignty.** The Harsanyi stable set represents a minimal modification of the vNM stable set to account for farsightedness. In particular, the notion of coalitional objections or domination continues to be defined over imputations. What this means is that a coalitional move from the status-quo defines not only the payoffs to members of the objecting coalition *but also to all those outside this coalition*. Indeed, a coalition  $S$  is permitted to move to a new imputation  $u'$  whenever  $u'_S \in V(S)$ . This implicitly gives  $S$  enormous latitude in choosing  $u'_{N-S}$ . The ability of  $S$  to determine how the payoff is distributed across players *not* in  $S$  clearly violates the basic notion of coalitional sovereignty.

That said, the use of imputations and its implicit violation of coalitional sovereignty has no substantive implications for the definition of the core or the vNM stable set. If coalition  $S$  blocks  $u$  with  $u'$  (where  $u'_S \in V(S)$ ), all that matters to  $S$  is  $u'_S$ . How we specify the remaining entries of  $u'$ , and whether or not we respect coalitional sovereignty in the process, makes no difference at all to these solution concepts, which rely on one-shot, myopic blocking.<sup>4</sup>

But if — as in Harsanyi — a coalitional move is followed by other moves, and players are farsighted, then the distribution of payoffs among players not in an objecting coalition will have a profound effect on where things end up. We therefore need to squarely address the degree of power a coalition have over the payoffs of outsiders. The use of imputations implicitly gives an objecting coalition complete power in arranging the distribution of the payoff to outsiders. That is, the objecting coalition  $S$  dictates the complementary allocation  $y_{N-S}$ . That allocation need not even be feasible for the complementary set of players, but it is presumed that  $S$  can somehow engineer society-wide changes to make this happen. In effect, then, the Harsanyi definition grants a coalition extraordinary power in the affairs of outsiders.

This can be consequential in strange ways. For instance, it's possible that a dummy player may be assigned a positive payoff in a “stable” imputation; see Ray and Vohra (2013) for an example. This is not a property shared by the vNM stable set, or indeed by most solution concepts. But

---

<sup>4</sup>At the same time, myopic solution concepts that rely on *ongoing* blocking — see, e.g., Feldman (1974), Green (1974) and Sengupta and Sengupta (1996) — will be affected by the degree to which a deviating coalition can choose payoffs for its complement. Koczy and Lauwers (2004) discuss the importance of this issue in the context of this literature and consider a model with coalition structures in which a coalition's move leaves undisturbed the payoffs of those coalitions that are disjoint from it.

the Harsanyi stable set has even stranger implications, as we can infer from the following result of Béal et al. (2008).

**THEOREM 1.** (Béal et al. (2008)) *Suppose  $(N, v)$  is a TU game in which  $v(T) > 0$  for some  $T \subset N$ .  $H$  is a Harsanyi stable set if and only if it is a singleton imputation,  $u$ , such that  $u_S \gg 0$  and  $\sum_{i \in S} u_i \leq v(S)$  for some  $S \subset N$ .*

In light of this result, it is instructive to ask precisely which imputations are *not* stable in the Harsanyi sense. Consider a TU game with  $v(T) > 0$  for some  $T \subset N$  and suppose  $u$  belongs to the interior of the core in the sense that  $\sum_{i \in S} u_i > v(S)$  for any  $S \subset N$ . According to Theorem 1,  $u$  cannot be (part of) a Harsanyi stable set. Every such interior core allocation is excluded. This is in sharp contrast to the traditional vNM stable set, which must include the core whenever the core is nonempty. In particular, no vNM stable set is a Harsanyi stable set when the interior of the core is nonempty.<sup>5</sup>

For more details on just what drives this peculiar result, see the discussion following Example 3 in Section 5.2.

One might imagine that a simple modification of the definition that restricts coalitional power would only lead to a nested change — shrinkage or expansion — relative to the Harsanyi stable set. But the restrictions apply equally to “initial objections” and “later counterobjections,” and so change the set completely, as we shall see below.

### 3. A NEW DEFINITION OF FARSIGHTED STABILITY

It should be apparent by now that we need to impose some reasonable restrictions on what a deviating coalition is allowed to do. We proceed by specifying more explicitly what happens when a coalition forms. Depending on the specific context, the members of the formed coalition will be drawn from the group as a whole (the grand coalition), or perhaps from other existing subcoalitions. So the entire coalition structure will be affected. The payoffs that result must respect this structure. Specifically, not only must  $S$  be restricted to a payoff choice from  $V(S)$  but the remaining players must similarly abide by the payoff constraints imposed by the structure. (Of course, in the subsequent periods the players are free to adjust their coalitional membership as well.) Moreover, it may be unreasonable — or impossible — for  $S$  to dictate the division of payoffs among  $N - S$ .

To track these constraints, it will be useful to extend the concept of an outcome to a *state*, which refers to a coalition structure and a utility profile feasible for that structure. A typical state  $x$  is therefore a pair  $(u, \pi)$  (or  $\{u(x), \pi(x)\}$  when we need to be explicit), where  $u_S$  is feasible and efficient for  $S$ , i.e.,  $u_S \in \bar{V}(S)$  for each  $S \in \pi(x)$ . Let  $X$  denote the set of all states. Now introduce an *effectivity correspondence*,  $E(x, y)$ , that specifies the collection of coalitions —

---

<sup>5</sup>Of course, this observation also applies to three-person games, in which, according to Theorem 1 of Harsanyi (1974), strict farsighted dominance is equivalent to (myopic) dominance. Thus, contrary to Harsanyi’s (1974) assertion in the last paragraph of his paper, his Theorem 1 does not remain valid if strict farsighted dominance is changed to farsighted dominance. See Example 3 below for a comparison of the Harsanyi stable set and vNM stable set in a three-player game.



possibly empty — that have the power to change  $x$  to  $y$ , for every pair of states  $x$  and  $y$ . Denote by  $\Gamma = (N, V, E)$  the (extended) characteristic function game.

The previously defined classical concepts can be easily recast in this extended model. First, state  $y$  *dominates* state  $x$  with respect to  $E$  if there exists a coalition  $S \in E(x, y)$  with  $u(y)_S \gg u(x)_S$ . For  $A \subseteq X$ , let

$$\text{dom}_E(A) = \{x \in X \mid x \text{ is dominated by some } y \in A \text{ with respect to } E\}.$$

The *core* of  $(N, V, E)$  is

$$C(N, V, E) = X - \text{dom}_E(X).$$

A set  $Z \subseteq X$  is a *vNM stable set* of  $(N, V, E)$  if

$$Z = X - \text{dom}_E(Z).$$

Nothing of substance has changed by adopting these definitions over the standard concepts presented in the previous Section, provided the effectivity correspondence allows every coalition  $S$  to freely choose payoffs in  $\bar{V}(S)$ . We shall impose this basic property below. Given this, it should be clear that if  $x = (u, \pi) \in C(N, V, E)$ , then  $u$  is in the *coalition structure core* of  $(N, V)$  (see, e.g., Greenberg (1994), Owen (1995)). Moreover, if  $\pi = \{N\}$ , then  $u$  is in the core as defined earlier. Indeed, in superadditive games,  $x = (u, \pi) \in C(N, V, E)$  implies that there exists  $x' = (u, N) \in C(N, V, E)$ , so that  $u \in C(N, V)$ .<sup>6</sup>

Now we move on to the concept of farsighted domination. State  $y$  is said to *farsightedly dominate*  $x$  (with respect to  $E$ ) if there is a collection of states  $y^0, y^1, \dots, y^m$  (with  $y^0 = x$  and  $y^m = y$ ) and a corresponding collection of coalitions,  $S^1, \dots, S^m$ , such that for all  $k = 1, \dots, m$ :

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

and

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

For  $A \subseteq X$ , let

$$\text{dom}_E(A) = \{x \in X \mid x \text{ is farsightedly dominated with respect to } E \text{ by some } y \in A\}.$$

A set of states  $Z \subseteq X$  is said to be a *farsighted stable set* if

$$Z = X - \text{dom}_E(Z).$$

We now turn to a discussion of minimal, reasonable restrictions to be placed on the effectivity correspondence. When coalition  $T$  changes state  $x$  to  $y$  it will typically induce a change in the coalition structure  $\pi(x)$ . We remain silent on whether or not  $T \in \pi(y)$  (that is,  $T$  might deliberately fragment itself), but we must surely allow  $T$  to have the option of remaining intact, as also the option to choose from its own set of feasible payoffs. Next, if  $T$  intersects  $S \in \pi(x)$ , we are similarly silent on whether the “residual”  $S - T$  remains a coalition in  $\pi(y)$ , though in the sequel we will not permit  $T$  to dictate their composition or payoff. What we *do* insist on is the coalitional sovereignty of “untouched coalitions” in  $\pi(x)$  that had no overlap with  $T$ : they are presumed to remain in  $\pi(y)$  and their payoffs are assumed to be unchanged. To be sure,  $T$ 's

<sup>6</sup>The extended model is motivated by our attempt to capture farsightedness, but it serendipitously allows us to dispense with superadditivity as well.

move may be followed by further coalitional moves and deliberate payoff reallocations, but we refer here only to the *immediate* impact of  $T$ 's departure.

More formally, we assume that the effectivity correspondence satisfies the following properties.

- (i) If  $T \in E(x, y)$ ,  $S \in \pi(x)$  and  $T \cap S = \emptyset$ , then  $S \in \pi(y)$  and  $u(x)_S = u(y)_S$ .
- (ii) For every state  $x \in X$ ,  $T \subseteq N$  and  $v \in \bar{V}(T)$ , there is  $y \in X$  such that  $T \in E(x, y)$ ,  $T \in \pi(y)$  and  $u(y)_T = v$ .

Condition (i) grants coalitional sovereignty to the untouched coalitions: the formation of  $T$  cannot influence the membership of coalitions that are entirely unrelated to  $T$  in the original coalition structure, nor can it influence the going payoffs to such coalitions. Condition (ii) grants coalitional sovereignty to the deviating coalition: it can choose not to break up, and it can freely choose its *own* payoff allocation from its feasible set.

Condition (i) acquires its present force because the situation in hand is described by a characteristic function:  $T$  truly influences neither the composition of an “untouched” coalition nor the payoffs it can achieve. (In games with externalities, the condition would need to be suitably modified.) We will want to go further and apply similar considerations of sovereignty to *every* coalition, even those that are left as residuals when  $T$  forms. In the sequel, we introduce a “default function” which maps the move of  $T$  to a unique coalition structure which leaves not just the untouched coalitions but also the residuals intact, and assigns every non-member of  $T$  a payoff. For now, the reader is free to mentally impose (or not) these additional restrictions; it will make no difference to Theorem 2 below.

Conditions (i) and (ii) are similar to those imposed by Konishi and Ray (2003). In addition, they assume that the residuals may organize themselves into other coalitions (but through an exogenously given rule, not determined by the deviating coalition). Similarly, Conditions (i) and (ii) also appear in Kóczy and Lauwers (2004), who study myopic coalition formation. In addition, they assume that the deviating coalition can choose both the way in which residuals are organized as well as their payoffs. These additional considerations can be viewed as special cases of our formulation.

In the class of hedonic games there is no ambiguity about the payoffs to untouched coalitions or residuals. After all, a hedonic game is one in which there is a unique payoff allocation to each coalition, so states can be identified with coalition structures.<sup>7</sup>

#### 4. THE CORE AND SINGLE-PAYOFF FARSIGHTED STABILITY

Of special interest are farsighted stable sets which consist of a *single* payoff allocation. These are not, in general, singleton sets, because several coalition structures might generate the same payoff allocation, and some or all of them may need to be included in the set. In any case, such sets trivially pass the internal stability criterion: there aren't two distinct payoff vectors in the set and consequently no “internal threat” of any kind. But they must work much harder on the

---

<sup>7</sup>Strictly speaking, such games do not satisfy comprehensiveness, but nothing substantive changes if we define the utility set of each coalition to be the comprehensive hull of it (unique) efficient payoff vector.

external stability front: the payoff allocation, coupled with an accompanying coalition structure, must single-handedly serve as a farsighted objection to *every* alternative state.

Do such sets exist, and is it possible to characterize the payoff allocations they contain?

Single-payoff stability is certainly too demanding as far as the vNM solution concept is concerned. It would require one distinguished imputation to have the ability to block (via different coalitions to be sure) every other imputation. That is a condition that will not be satisfied in all but the most trivial and uninteresting games.

Yet exactly the opposite appears to be true of Harsanyi stability. In TU games, Theorem 1 informs us that under the mild restriction that some coalition has strictly positive worth, all Harsanyi stable sets *must* contain a single imputation. There are no other Harsanyi stable sets. This is a pretty dramatic contrast from vNM stability. The problem, of course, is that the payoff allocations in question have questionable properties: they can give positive payoffs to dummy players, and they cannot belong to the interior of the core. (This is quite apart from our criticism of the solution concept itself.)

For the solution concept we espouse, matters lie somewhere in between. It turns out that a single-payoff farsighted stable set does exist in a large class of games (though not as large as in the Harsanyi case), and that we can exactly characterize such sets. That is what we turn to now.

A collection of pairwise disjoint coalitions  $\mathcal{T}$  is a *strict subpartition* of  $N$  if  $N - \cup_{T \in \mathcal{T}} T$  is nonempty; these are the players “not covered” by  $\mathcal{T}$ .

A payoff allocation  $u$  is *efficient* if there does not exist another allocation  $u'$  such that  $u' > u$ . An efficient allocation  $u$  is *separable* if whenever  $u_T \in V(T)$  for every  $T$  in some strict subpartition  $\mathcal{T}$ , then  $u_S \in V(S)$  for some  $S \subseteq N - \cup_{T \in \mathcal{T}} T$ .

Separability has close (but not exact) links to the core. If  $u$  is separable, then  $u$  must belong to  $C(N, V, E)$ , the coalition structure core of  $(N, V)$ . For if this were false, there exists  $T \subset N$  and  $v \in V(T)$  such that  $v \gg u_T$ . By comprehensiveness,  $u_T \in V(T)$ . By separability there exists  $S \subseteq N - T$  such that  $u_S \in V(S)$ . If  $S \neq N - T$ , by another application of separability, there is  $S' \subseteq N - T - S$  with  $u_{S'} \in V(S')$ . By a repeated application of separability, if necessary, we can find a partition  $\pi' = (T, S, S' \dots)$  such that  $(v, u_{N-T})$  is feasible for  $\pi'$ . Since  $v \gg u_T$ , we have  $(v, u_{N-T}) > u$ , but this contradicts the efficiency of  $u$ .

The converse of this statement is not true. There are core allocations that are not separable; see Example 5 in Section 5.4 below. However, if  $u$  lies in the *interior* of  $C(N, V, E)$ , in the sense that  $u_S \notin V(S)$  for any  $S \subset N$ , then it is easy to see that  $u$  is separable.

In a superadditive game, the separability of  $u$  is equivalent to the statement that whenever  $u_T \in V(T)$  for some  $T \subset N$ , then  $u_{N-T} \in V(N - T)$ . Moreover, if  $u$  is separable, then it belongs to the core, and if  $u$  is in the interior of the core then it is separable. In a *strictly* superadditive game,<sup>8</sup> the separability of  $u$  is equivalent to the statement that  $u$  lies in the interior of the core.

---

<sup>8</sup>A superadditive game is strictly superadditive if for any pair of disjoint coalitions  $S$  and  $T$ ,  $V(S) \times V(T)$  is in the interior of  $V(S \cup T)$ .

In general, however, it is possible for the interior of the core is empty, but for there to exist a separable allocation. As we will discuss below, this is the case in Lucas's (1968) example of a TU game in which the core is non-empty but there is no vNM stable set.

Given a payoff allocation  $u$ , let  $[u]$  denote the collection of all *states* that are equivalent to  $u$  in terms of payoffs, i.e.,

$$[u] = \{y \in X \mid u(y) = u\}.$$

**THEOREM 2.** (a) *If  $u$  is separable, then  $[u]$  is a single-payoff farsighted stable set with respect to any effectivity correspondence satisfying Conditions (i) and (ii).*

(b) *If  $u$  is not separable, no subset of  $[u]$  is a single-payoff farsighted stable set with respect to any effectivity correspondence satisfying Conditions (i) and (ii).*

One implication of this result is that if the interior of the core (or of the core of a coalition structure) is non-empty, then a farsighted stable set exists for every effectivity correspondence satisfying Conditions (i) and (ii).

*Proof.* (a) Suppose that  $u$  is separable. Consider a state  $y^0 = (u^0, \pi^0)$  where  $u^0 \neq u$ . We will construct a farsighted objection from  $y^0$  to a state in  $[u]$  through a collection of coalitions  $T^1, \dots, T^M$  and states  $y^1, \dots, y^M$ , where  $y^M \in [u]$ . That is, each coalition  $T_k$  will lie in  $E(T^{k-1}, T^k)$ , with

$$y^0 \rightarrow_{T^1} y^1 \rightarrow \dots \rightarrow_{T^{M-1}} y^{M-1} \rightarrow_{T^M} y^M = (u, \pi) \in [u]$$

and

$$u_{T^k} \gg u_{T^k}^{k-1} \text{ for all } k = 1, \dots, M.$$

Our construction is in two stages. The first stage involves the formation of singletons, and the second stage involves a final move to  $u$  via a suitable aggregation of the singletons at the end of the first stage.

STAGE 1.

**Step 0.** Define  $B^0 \equiv \{i \in N \mid u_i > u_i^0\}$ . Because  $y$  is efficient and  $u^0 \neq u$ ,  $B^0$  is nonempty. If each  $i \in B^0$  is a singleton in  $\pi^0$ , i.e., if  $\{i\} \in \pi^0$  for every  $i \in B^0$ , go to Stage 2 below. Otherwise, there exists  $i \in B^0$  such that  $i \in S$  for some  $S \in \pi^0$  with  $|S| \geq 2$ . Let  $i_1$  be any player for whom this property holds. Let  $T_1 = \{i_1\}$ . Move to any state  $y^1 = (u^1, \pi^1)$  such that  $T_1 \in E(y^0, y^1)$ ; such a state exists by Condition (ii). If  $u^1 = u$ , stop; otherwise move to Step 2.

Recursively, suppose we have arrived at state  $y^k = (u^k, \pi^k)$ , for  $k \geq 1$ , and that  $u^k \neq u$ .

**Step  $k$ .** Define  $B^k \equiv \{i \in N \mid u_i > u_i^k\}$ . Because  $y$  is efficient and  $u^k \neq u$ ,  $B^k$  is nonempty. If each  $i \in B^k$  is a singleton in  $\pi^k$ , go to Stage 2. Otherwise, there exists  $i \in B^k$  such that  $i \in S$  for some  $S \in \pi^k$  with  $|S| \geq 2$ . Let  $i_k$  be any player for whom this property holds. Let  $T_k = \{i_k\}$ . Move to any state  $y^{k+1} = (u^{k+1}, \pi^{k+1})$  such that  $T_k \in E(y^k, y^{k+1})$ . If  $u^{k+1} = u$ , stop; otherwise move to Step  $k + 1$ .

Note that at each step the partition gets refined with the formation of one new singleton coalition. It is therefore trivial to see that this process must either “stop,” or that a move to Stage 2 will be

called for at some step. In the former case the proof is complete, as the entire chain is a farsighted objection terminating at some  $(u, \pi) \in [u]$ . We therefore turn to

STAGE 2.

Suppose that at the initiation of stage 2, we are in state  $y^m$  (where  $m$  is the step at which Stage 2 was invoked). Note that  $u^m \neq u$ , so that  $B^m \equiv \{i \in N \mid u_i > u_i^m\} \neq \emptyset$ . Moreover, every  $i \in B^m$  is in a singleton coalition; i.e.,  $\{i\} \in \pi^m$  for every such  $i$ .

There are now two cases to consider. In case 1,  $B^m = N$ . Pick any  $\pi$  such that  $y = (u, \pi) \in [u]$ . Form each of the coalitions in  $\pi$  in any order, with each coalition  $S$  achieving  $u_S$ . Note that later coalitions cannot upset the payoffs to earlier coalitions by coalitional sovereignty (Condition (i)). Combining this with Stage 1, it is easy to see that we have constructed a farsighted objection leading from  $y^0$  to  $y$ .

Otherwise, because  $B^m$  is nonempty,  $\pi^m$  restricted to the complement of  $B^m$  must be a strict subpartition. Because no one in the complement strictly prefers  $u$  to  $u^m$ , it must be that  $u_T \in V(T)$  for every  $T$  in that strict subpartition. By the separability of  $u$ , there exists a coalition  $S(1)$  in  $B^m$  such that  $u_{S(1)} \in V(S(1))$ . If  $B^m - S_1 \neq \emptyset$ , we can repeat the argument, again applying separability, until all members of  $B^m$  have been gathered into coalitions  $\{S(1), \dots, S(\ell)\}$ . (Again, Condition (i) is used to ensure that these moves do not affect the payoffs to untouched coalitions.) Note that at the end of this process we arrive at a state of the form  $y^* = (u^*, \pi^*)$  with  $u^* \geq u$ . However,  $u$  is separable and so efficient; therefore  $u^* = u$ , i.e.,  $y^* \in [u]$ . It is easy to see that the above procedure combined with Stage 1 yields a farsighted objection leading from  $y^0$  to  $y^*$ .

Because no two elements in  $[u]$  can dominate each other, this completes the proof that  $[u]$  is a farsighted stable set.

(b) Given  $u$ , suppose that  $[u]$  is a farsighted stable set. Then it is immediate that  $u$  must be efficient. If not, consider any state  $z$  with payoff  $u(z) > u$ . There cannot be a farsighted objection running from  $z$  to any member of  $[u]$ , a contradiction.

With that settled, suppose on the contrary that  $[u]$  is a farsighted stable set but  $u$  isn't separable. Then there exists a strict subpartition  $\mathcal{T}$  with  $u_T \in V(T)$  for every  $T \in \mathcal{T}$ , and with  $u_S \notin V(S)$  for every  $S \subseteq R(\mathcal{T}) \equiv N - \cup_{T \in \mathcal{T}} T$ . Define a state  $x$  in the following way. Construct a partition  $\pi'$  by appending the subpartition  $\mathcal{T}$  to the collection of singletons from  $R(\mathcal{T})$ , and let  $u'$  be any efficient allocation for this partition such that  $u'_T \geq u_T$  for every  $T \in \mathcal{T}$ . (Because  $u_T \in V(T)$  for every  $T \in \mathcal{T}$ , this is clearly possible.) Let  $x = (u', \pi')$ .

Note that  $x$  cannot be an element of  $[u]$ . Therefore, since  $[u]$  is a farsighted stable set, there must be a farsighted objection running from  $x$  to  $z \in [u]$ . Since no player in any  $T \in \mathcal{T}$  gains from such a move, no such player can be part of the first coalition in the objection. By Condition (i), the payoffs to these players must remain unchanged following the first move, which proves that they cannot participate in any *later* move as well. On the other hand, by the assumed absence of separability, no coalition  $S$  in  $R(\mathcal{T})$  can "implement"  $u_S$ . That contradicts the presumption that  $z \in [u]$ . ■

Theorem 2 establishes a close connection between the coalition structure core of a game and the payoff allocation associated with single-payoff stable sets. It is a separable allocation, and *only* a separable allocation, that can act as a farsighted stable set. As we've already seen, all separable allocations are core allocations and all interior core allocations are separable. So this shows (somewhat loosely speaking) that “almost every” allocation in the core of a game, while defined by an entirely myopic blocking concept, can *farsightedly* dominate every other state, and that non-core allocations do not possess this property.

This result is related to other observations made in the literature. Feldman (1974), Green (1974), Sengupta and Sengupta (1996) and Koczy and Lauwers (2004) explored whether chains of *myopic* objections found their limit in the core. Clearly, once at the core, no such chain can begin; the question is whether all such chains end there. Ray (1989) proved that the core is immune to farsighted objections provided that such objections are “nested,” in the sense of coming from progressively smaller subsets of coalitions. No such restriction is imposed here. Konishi and Ray (2003, Theorems 4.1 and 4.2) proved that the core of a game can be described as the limit of a real-time dynamic process of a coalition formation, provided that the discount factor is close enough to 1. For more on real-time processes and these connections, see Ray and Vohra (2013).

Note again the striking contrast with Harsanyi stable sets. They are almost the very antithesis of the stable sets described here, in that no Harsanyi set contains any allocation in the interior of the core. Matters are very different, both conceptually and in terms of substantive findings.

Finally, it is of interest to note that single-payoff stable sets lend more weight to the dominance relation that we use, in the following sense. Notice that a farsighted dominance relation involves a sequence of objections, with the property that each coalition is better off pushing things forward relative to staying where they are. But this raises two problems. First, what if the coalition in question has an even better end-point to look forward to by precipitating a different state? Second, if a coalition does not move, why is the implicit counterfactual the going state: why would the state not be pushed somewhere else? These and related matters are discussed in Ray and Vohra (2013), in which we argue for real-time definitions of coalition formation that free us from such ambiguities. But it is worth noting that within the ambit of single-payoff stability, no such ambiguities exist: all roads lead to Rome.

## 5. DISCUSSION

**5.1. Existence, With or Without Separability.** The existence question was one that von Neumann and Morgenstern (1944) regarded as vitally important:

There can, of course, be no concessions as regards existence. If it should turn out that our requirements concerning a solution... are, in any special case, unfulfillable, — this would certainly necessitate a fundamental change in the theory. Thus a general proof of the existence of solutions... for all particular cases<sup>9</sup> is most desirable. It will appear from our subsequent investigations that this proof

---

<sup>9</sup>In the terminology of games: for all numbers of participants and for all possible rules of the game.

has not yet been carried out in full generality but that in all cases considered so far solutions were found. (Section 4.6.3.)

One presumes that von Neumann and Morgenstern were referring to transferable-utility games, because it is quite easy to show that vNM stable sets do not exist over the entire domain of all characteristic functions. Stearns (1964) constructed such an example, and we do something similar in Example 2 below. But the TU case proved to be much harder to resolve, and the question was not settled until Lucas (1968) provided a ten-person example of a TU game without a vNM stable set. We reproduce this celebrated example here, as it will provide a noteworthy instance of the role played by separability.

**EXAMPLE 1.** *A ten-player TU game (Lucas, 1968):*  $v(N) = 5$ ,  $v(1, 3, 5, 7, 9) = 4$ ,  $v(3, 5, 7, 9) = v(1, 5, 7, 9) = v(1, 3, 7, 9) = 3$ ,  $v(3, 5, 7) = v(1, 5, 7) = v(1, 3, 7) = v(3, 5, 9) = v(1, 5, 9) = v(1, 3, 9) = v(1, 4, 7, 9) = v(3, 6, 7, 9) = v(2, 5, 7, 9) = 2$ ,  $v(1, 2) = v(3, 4) = v(5, 6) = v(6, 7) = v(7, 8) = v(9, 10) = 1$ , and  $v(S) = 0$  for all other coalitions  $S$ .

*This is a game with a non-empty core but without a vNM stable set. Yet a farsighted stable set exists.*

Although this game is not superadditive, it has the property that any efficient payoff can be achieved through the grand coalition.<sup>10</sup> An payoff allocation  $u$  belongs to the core of this game if and only if it satisfies the following two conditions:

$$(1) \quad u_1 + u_2 = u_3 + u_4 = u_5 + u_6 = u_7 + u_8 = u_9 + u_{10} = 1,$$

and

$$(2) \quad u_1 + u_3 + u_5 + u_7 + u_9 \geq 4.$$

As Lucas observed, this game admits allocations that solve (1) and (2), so the core is nonempty. However, because  $v(S) = 1$  for  $S$  consisting of adjacent pairs beginning with odd indices, e.g.,  $\{1, 2\}$  or  $\{5, 6\}$ , it follows from (1) that the interior of the core is empty. At the same time, there are core allocations that are separable. Consider any  $u$  in the core such that every odd player receives  $u_i > 0.8$ . For instance consider  $u^*$  where

$$u_i^* = \begin{cases} 0.9 & \text{if } i \text{ is odd} \\ 0.1 & \text{if } i \text{ is even} \end{cases}$$

It can be shown that the only coalitions that can achieve  $u^*$  are coalitions that consist of pairs of adjacent players starting with odd indices. Thus, if  $u^*$  is feasible for a strict subpartition it must be one in which each coalition consists of such adjacent pairs. But then, by the fact that the subpartition is strict, the complement must contain at least one more such adjacent pair. For this pair  $u^*$  is feasible. Therefore  $u^*$  is separable. By Theorem 2,  $[u^*]$  is a farsighted stable set.<sup>11</sup>

<sup>10</sup>Indeed, as Lucas (1968) points out, the superadditive cover of this game also does not possess a stable set.

<sup>11</sup>Harsanyi stable sets exist as well, though these are very different from the farsighted stable set; more on this contrast in Section 5.2. Define  $\bar{u}$  by  $\bar{u}_i = 0.8$  if  $i$  is odd, and by  $\bar{u}_i = 0.2$  if  $i$  is even.  $\bar{u}$  is part of a Harsanyi stable set as it satisfies the conditions of Béal et al. (2008); see also Diamantoudi and Xue (2005). Notice also that  $\bar{u}$  is feasible for the five-player coalition consisting of all the odd numbered players but not for any coalition in the complement. The state  $x = (\bar{u}, N)$  It is therefore not separable, and so by Theorem 2, it cannot be a farsighted stable set for any effectivity correspondence satisfying Conditions (i) and (ii).

Yet we know from Lucas (1968) that this game admits no vNM stable set.

Separability plays a key role here in establishing the existence of a single-payoff farsighted stable set, even when there is no vNM stable set.<sup>12</sup> How general, then, is the existence property: what can we say when separable allocations do not exist? Certainly, one cannot hope that the existence of a farsighted stable set is invariably guaranteed:

**EXAMPLE 2.** *A three-player NTU “roommate game” (Lucas 1992, Banerjee et al. 2001, Diamantoudi and Xue 2003, and Herings et al. 2011):  $V(S) = \{v \leq a^S\}$  for all  $S \subset N$ . With some abuse of notation, we write  $(i,j)$  for coalition  $\{i, j\}$ .*

$$a^{12} = (3, 2), a^{23} = (3, 2), a^{13} = (2, 3)$$

and  $V(N)$  is the comprehensive hull of  $\{a^{12}, a^{23}, a^{13}\}$ .

It is easy to see that this example has an empty core and does not possess either a vNM stable set or a farsighted stable set.<sup>13</sup>

But this example is not entirely definitive, for the following reason. This very example is well-known as an easy counterexample to the existence of the traditional vNM set for general NTU games. But that did not prevent a search for a counterexample in the TU context, a question that was finally settled after many years with the counterexamples of Lucas (see one of them in Example 1). For the same reason, we would like to know if there are corresponding counterexamples to the existence of the farsighted stable set in TU games. Or does every TU game necessarily admit at least one such set? We continue this discussion in Section 5.4.

In light of this history it is quite remarkable that there are several positive results on the existence of the Harsanyi stable set. Diamantoudi and Xue (2005) showed existence in Lucas’s example. Béal et al. (2008) proved existence for TU games under very mild conditions and this was subsequently generalized to NTU games in Bhattacharya and Brosi (2011). But as we have argued in detail above, the Harsanyi set is conceptually problematic, and the outcomes that are generated are problematic as well, a topic to which we return in the very next subsection.

**5.2. The Contrast With Harsanyi Stability.** The stable sets of Theorem 2 are essentially singletons, as in the case of the Harsanyi sets. But the two sets of collections couldn’t be more different. Our stable sets are entirely “compatible” with the core, in the sense that every farsighted stable set of Theorem 2 contains a payoff allocation that belongs to the coalition structure core. (Some non-separable core allocations are excluded, but these are boundary exceptions: every *interior* core allocation is separable and therefore compatible with far-sighted stability.) In

<sup>12</sup>It should be noted that the nonemptiness of the core isn’t enough for the existence of a single-payoff stable set. In Example 5 below, the core is also nonempty, but there are no separable allocations and consequently no single-payoff farsighted stable sets.

<sup>13</sup>It follows from Bhattacharya and Brosi (2011) that there do exist Harsanyi farsighted stable sets in this example if imputations are defined as individually rational payoffs on the boundary of  $V(N)$ . For example,  $(3, 2, 0)$  is a singleton Harsanyi stable set. The farsighted objection from  $(0, 3, 2)$  to this imputation involves player 1 leaving the grand coalition and assigning 0 to player 2 (moving to the imputation  $(0, 0, 2)$ ), which leads 1 and 2 to move to  $(3, 2, 0)$ . This argument obviously does not work if the departure of 1 from the grand coalition results in  $(0, 3, 2)$ .



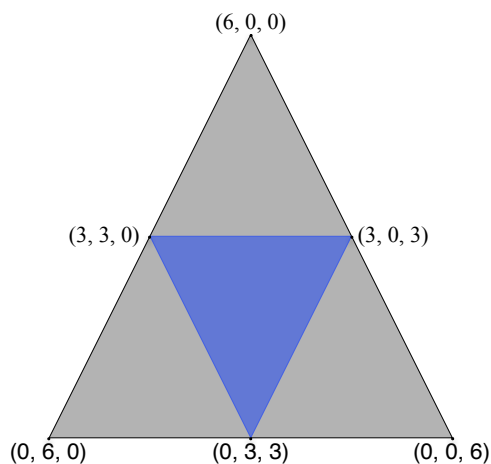


FIGURE 1. COMPARISON OF STABLE SETS

contrast, the Harsanyi sets *cannot contain any allocation in the interior of the core*. We emphasize that our critique of the Harsanyi approach is not based on this outcome, but rather on the conceptual underpinnings of that approach, as already discussed. Yet this could be grounds for additional misgiving.

The following example reiterates this point, but takes it a step further.

**EXAMPLE 3.** A three-player TU game:  $v(S) = 3$  for  $S$  such that  $|S| = 2$ , and  $v(N) = 6$ . The set of efficient allocations is depicted in Figure 1. The core is the convex hull of  $(3, 3, 0)$ ,  $(0, 3, 3)$  and  $(3, 0, 0)$ , shown as the inverted central triangle in Figure 1.

Since this is a convex game, the coalition structure core coincides with the unique vNM stable set. It is easy to check that the set of separable payoff profiles coincides with the interior of the core. By Theorem 2, every state with one of these associated separable payoff profiles is a farsighted stable set. The Harsanyi stable sets are starkly different. By Béal et al. (2008), they have as payoff vectors all strictly positive imputations which are feasible for some two-player coalition. In Figure 1, these are all points in the *complement* of the inverted central triangle.

Actually, the contrast between the two stable sets is even sharper, because under mild restrictions on the effectivity correspondence that we describe below, there are no other farsighted stable sets in this example, single-payoff or otherwise. The Appendix contains a formal statement and proof of this assertion.

It is worth using the example to understand just why the Harsanyi concept yields such unpalatable outcomes, removing every interior core allocation in particular. Consider the imputation  $(1, 1, 4)$ , which is not in the core. Start from any other imputation in which some player gets less; say player 3 for the sake of concreteness. To construct a domination chain, have player 3 “block” and get 0, and assign 6 to player 1 and 0 to player 2. Note how the definition of an objection is satisfied: the new allocation is an imputation, and the piece of it accruing to player 3 (namely, 0)

is something that player 3 can guarantee on his own. Now continue by having player 2 move; she gives herself 0, player 1 zero and gives the entire surplus of 6 to player 3. Finally, 1 and 2 jointly implement the imputation  $(1, 1, 4)$ . We can construct such chains from any starting imputation to  $(1, 1, 4)$ .

But we cannot do this with an interior core allocation. Because the domain consists of imputations, there can never be an objection from the grand coalition that myopically improves payoffs for everyone, so the last coalition to move must be a *strict* subcoalition. That is why no domination chain can terminate in the interior of the core. For instance,  $(2, 2, 2)$  cannot be the final point of any farsighted Harsanyi-style objection, because there is no two-player coalition for which it is feasible. Therefore,  $(2, 2, 2)$  is not in the Harsanyi stable set.

It is worth reiterating that we've departed from the Harsanyi approach in two ways: (a) our domain is the set of feasible utility profiles for coalition structures, and (b) we've imposed coalitional sovereignty (recall Conditions (i) and (ii) on the effectivity correspondence). Both departures are important. If we were to impose only (a), *every* efficient state with strictly positive payoff would form a stable set, an unsatisfactory conclusion.<sup>14</sup> And, as already discussed in detail, imposing (b) alone makes no sense in the absence of (a): we would be unable to entertain ongoing chains of deviations.

**5.3. A Remark on Hedonic Games.** Diamantoudi and Xue (2003) show that in a hedonic game with strict preferences every allocation in the core is a single-payoff farsighted stable set.<sup>15</sup> This result can be derived as a corollary of Theorem 2. Suppose  $(N, V)$  is a hedonic game with efficient payoffs in  $S$  denoted  $a^S$ . For every  $S \subseteq N$ , let  $V(S) = \{v \in \mathbb{R}^S \mid v \leq a^S\}$ , so that  $\bar{V}(S) = \{a^S\}$ . Suppose further that players have strict preferences across coalitions in the sense that no player is indifferent between belonging to distinct coalitions, i.e., if  $i \in S \cap T$ , then  $a_i^S \neq a_i^T$ . Suppose  $y = (u, \pi)$  is in the core of  $(N, V, E)$ . By efficiency  $u = ((a^S)_{S \in \pi})$ . We claim that  $y$  is separable. Suppose  $u_T \in V(T)$  for every  $T$  in some strict subpartition  $\mathcal{T}$ . If every  $T \in \mathcal{T}$  is in  $\pi$  separability is trivially satisfied. Otherwise, there is a coalition  $T \in \mathcal{T}$  such all players in  $T$  were in a different coalition in  $\pi$ . Let  $S_i$  denote the coalition in  $\pi$  that contains  $i$ . We can now write  $u = (a_i^{S_i})_{i \in N}$ . Since  $u_T \in V(T)$ ,  $a^T \geq u_T$ , or  $a_j^T \geq a_j^{S_j}$  for all  $j \in T$ . By strict preferences, the last inequality must be strict. But this contradicts the hypothesis that  $u$  is in the core of  $(N, V, E)$ , and completes the proof.

**5.4. Other Farsighted Stable Sets.** Theorem 2 completely characterizes farsighted stable sets that are associated with a unique payoff vector. But we don't have a characterization of *all* farsighted stable sets. We do know that sets that are *not* single-payoff cannot generically contain

<sup>14</sup>To see this, consider an efficient allocation  $u \gg 0$ . Let  $y$  be any state not in  $[u]$ . There must be  $i \in N$  such that  $u_i > u(y)_i$ . Disregarding coalitional sovereignty, suppose this player induces the state in which he stands alone, and for every nonsingleton coalition that remains, assigns coalitional surplus to the member with the lowest index, all other members getting 0. Now continue the process so that each of the players with 0 leave sequentially, until all players are in singletons. Then make a final move by the grand coalition to  $u$ .

<sup>15</sup>Mauleon et al. (2011) also show that in a matching game with strict preferences *every* farsighted stable set is of this form. If preferences are not strict, it is possible that a core allocation in a hedonic game does not satisfy coalition structure separability (see Example 3 in Diamantoudi and Xue (2003)), and is therefore not a single-payoff farsighted stable set.

any state  $x$  with a separable payoff vector  $u$ . That is because  $u$ , being efficient, cannot generically belong to two distinct coalition structures; i.e.,  $[u] = \{x\}$ . But then,  $x$  can farsightedly dominate every other state, and so cannot co-exist in a stable set with any other payoff vector.

Therefore separable payoff allocations must generically stand on their own, but there are deeper questions that remain open:

1. Are there farsighted stable sets other than those characterized in Theorem 2?
2. Do such sets invariably make their appearance when no allocation satisfies separability?
3. Do such sets disappear when some payoff allocation is separable?

We do not have complete answers to these questions, but the examples that follow suggests some interesting connections. To discuss these in a reasonable setting we will place some additional structure on the effectivity correspondence. Recall that our existing Conditions (i) and (ii) allow for both deviating and untouched coalitions to have coalitional sovereignty. But we remained silent on the residuals left behind by deviating coalitions. In what follows, we extend a measure of sovereignty to such coalitions.

Let  $T$  move. Consider a residual coalition  $W$ ; i.e.,  $W = S - T$  for some  $S \in \pi(x)$ , where  $S \cap T \neq \emptyset$ . We want to assign a “default payoff” to  $W$  immediately following  $T$ ’s departure (as emphasized earlier, this can change in subsequent stages). Formally, a *default function*  $f^W$  maps payoffs  $v$  in  $R^W$  to fresh payoffs  $f^W(v)$  on the coalitional frontier  $\bar{V}(W)$ . The interpretation is that  $W$  is a residual after some members from a larger coalition have left.  $W$  was enjoying a payoff of  $v$  just before that move (i.e.,  $v = u_W(x)$ , where  $x$  was the “earlier” state), and  $f^W(v)$  is the payoff vector that it receives just after the move.<sup>16</sup>

A default function  $f^W$  is *monotonic* if for every  $v \in R^W$ , either  $f^W(x)_i > v_i$  for all  $i \in W$ , or  $0 < f^W(x)_i < v_i$  for all  $i \in W$  with  $v_i > 0$ , or  $f^W(x)_i = v_i$  for all  $i \in W$ . In words, immediately following the move, the payoffs to every residual member either uniformly go up, or go down relative to what they were getting before.

Now we can formally state our additional restriction on the effectivity correspondence:

(iii) There exists a monotonic, continuous default function  $f^W$  defined for every subcoalition  $W$  such that whenever there is a move from  $x$  to  $y$ ,  $u(y)_W = f^W(u(x)_W)$  for every residual coalition  $W$ .

Our first example shows that a far-sighted stable set consisting of two or more distinct payoff allocations can exist, when there are no separable allocations:

**EXAMPLE 4.** A three-player TU game:  $v(S) = v(N) = 1$  for all two-player coalitions  $S$ , and  $v(S) = 0$  otherwise.

Let  $y^i$  denote the state in which the coalition structure is the grand coalition, player  $i$  receives zero and the remaining players receive 0.5 each. Under Conditions (i)–(iii) on the default function,

<sup>16</sup>The default could, in general, depend both on the pre-move state as well as the identity of the moving coalition. It simplifies the exposition to assume that the default payoff depends only on the pre-move payoff profile to  $W$ .

the collection

$$A = \{[y^1], [y^2], [y^3]\}$$

is a (farsighted) stable set. For any alternative state  $x$ , assume without loss of generality that  $x_1 \leq x_2 \leq x_3$ . Since  $x_1 \leq 1/3$  and  $x_2 \leq 0.5$ , if  $x \notin A$ , it must be the case that  $x_2 < 0.5$ , and there is an objection from 1 and 2 to  $(0.5, 0.5, 0)$ , along with the coalition structure  $\{1, 2\}, \{3\}$ , which is an element of  $[y^1]$ . It is also easy to see that  $A$  satisfies internal stability if untouched coalitions have sovereignty (Condition ii) and if the default function is monotonic (Condition iii). The only possible farsighted objection to any state in  $[y^1]$  must be initiated by  $\{3\}$  but, by these assumptions, the payoffs to 1 and 2 are unchanged and there can be no further change.

The remaining examples in this section suggest that stable sets not of the form  $[u]$  appear precisely when there are no separable allocations, and conveniently disappear when there are separable allocations. Whether this behavior is more general remains a conjecture at this stage.

**EXAMPLE 5.** *A three player TU “veto game”:  $N = \{1, 2, 3\}$ ,  $v(N) = v(\{1, 2\}) = v(\{1, 3\}) = 1$  and  $v(S) = 0$  for all other  $S$ . Player 1 can be viewed as the veto player. The coalition structure core has the single payoff  $(1, 0, 0)$ .*

There are many vNM stable sets in this game; see Lucas (1992) for details. A typical vNM set is depicted in Figure ??: it is a continuous curve that begins at  $(1, 0, 0)$  and continues to the opposite edge.<sup>17</sup>

The core consists of a single allocation  $(1, 0, 0)$ . Because separable allocations in superadditive games must be core allocations, the only possible candidate for separability is the allocation  $u = (1, 0, 0)$ . But  $u_{23} = (0, 0)$  is feasible for coalition  $\{2, 3\}$ , yet 1 cannot get  $u_1 = 1$  on her own. By Theorem 2, there is no farsighted stable set with a singleton payoff allocation, for any effectivity correspondence satisfying Conditions (i) and (ii).

However, farsighted stable sets do exist.

**OBSERVATION 1.** *Under Conditions (i)–(iii), we claim every payoff set of the form  $A = \{x \in \Delta \mid x_1 = c\}$ , for any fixed  $c \in (0, 1)$ , describes the payoffs from some farsighted stable set. Moreover, such sets describe all the farsighted stable sets of the veto game.*

The horizontal line in Figure 2 depicts one set of the sort described in Observation 1. More precisely, each farsighted stable set is given by the pairs of the form  $(u, \pi)$ , where  $u \in A$ , and  $\pi$  can be taken to be the grand coalition structure in each instance, except at the extremities  $C$  and  $D$ , where we also append the relevant coalition structure  $\{1i\}, \{j\}$ . With this understood, we will loosely refer to payoff allocations in  $A$  as elements of the stable set.

Observe that  $A$  does not include the (unique) core allocation. Note, too, that each  $A$  is very different from a vNM set, which as we’ve explained runs “vertically” in the simplex.

We prove the first part of the Observation here: that sets of the form  $A$  are farsighted stable.

<sup>17</sup>In contrast, Harsanyi stability doesn’t make much of a prediction in this example. Every payoff allocation, except for  $(1, 0, 0)$  and except for any  $x$  with  $x_1 = 0$ , is a singleton Harsanyi stable set. This follows from the characterization result of Béal et al. (2008), and shows that even in three-player games the vNM and Harsanyi stable sets do not coincide.

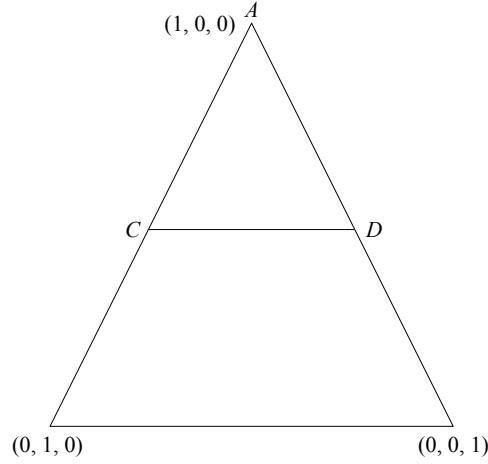


FIGURE 2. FARSIGHTED STABLE SETS IN THE VETO GAME

To verify internal stability, consider  $u \in A$ . Suppose that  $S$  starts a farsighted objection leading to  $u' \in A$ . Player 1 has nothing to gain, so  $1 \notin S$ . Furthermore,  $S$  cannot contain *both* 2 and 3, as both cannot simultaneously gain in the move to  $u'$ . So, without loss of generality,  $S = \{2\}$ , which initially precipitates the coalition structure  $\{2\}, \{13\}$ . Because the default function picks out a point in  $\bar{V}(13)$  and is monotonic,  $\{13\}$  now enjoys a payoff that weakly dominates  $u$ . Therefore 1 cannot gain by moving to  $u'$ . Nor can 3, since  $u'_2 > u_2$  and  $u'_1 = u_1$  imply that  $u'_3 < u_3$ . This shows that the farsighted goal of 2 — to attain  $u'$  — is impossible to accomplish.

To verify external stability, consider any state  $x' = (u', \pi')$  with  $u' \notin A$ . If player 1 stands alone in  $\pi'$ , then  $u' = (0, 0, 0)$ , and a single step to any  $u \gg 0$  in  $A$  constitutes a farsighted objection. So assume that 1 is part of a nonsingleton coalition. Now, if  $u'_1 < u_1$ , 1 can move by standing alone, thereafter precipitating  $(0, 0, 0)$ . The grand coalition can then move to any  $u \in A$  with  $u \gg 0$ , competing the farsighted objection. Finally, if  $u'_1 > u_1$ , players 2 and 3 can join forces by precipitating the structure  $\{1\}, \{23\}$  with payoff  $(0, 0, 0)$ . This is followed by a grand coalitional move to any  $u \in A$  with  $u_1 > 0, u_2 > u'_2$  and  $u_3 > u'_3$ .

The proof that there is no other farsighted stable set is relegated to the Appendix. The intuition for why any such set must be “horizontal” is simple: if there are two states  $x$  and  $y$  with  $u_1(x) > u_1(y)$ , then player 1 can engineer a farsighted deviation from the state from  $y$  to  $x$  by unilaterally precipitating a zero payoff allocation, and then inducing the other players to move to state  $x$ . (The full argument needs additional detail, especially concerning the strict positivity of the components of  $x$  and  $y$ , and we provide these in the Appendix.)

But not only must every farsighted set be horizontal, it must span the entire width of the payoff simplex. That is, for *every* nonnegative payoff vector  $u$  with  $u_1$  equal to the common value for player 1 on the horizontal set, there is a stable state  $y \in A$  with  $u(y) = u$ . For if this were false for some  $y$ , it is so because  $y$  is blocked by a farsighted objection from some  $z \in A$ . Observe that the interests of players 2 and 3 are opposed in this move; both cannot gain. So  $\{23\}$  cannot be the *joint* first mover, of course, nor can 1 as his payoff is unaltered between  $y$  and  $z$ . Therefore,

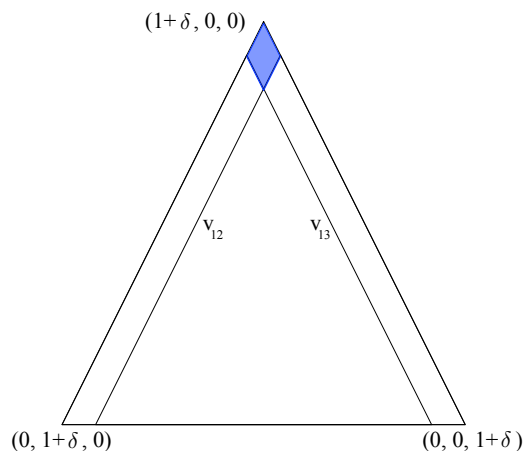


FIGURE 3. PERTURBED VETO GAME

2 or 3 is the first mover. Say it is 2; then  $u_2(z) > u_2(y)$ . But then  $u_3(z) < u_3(y)$ . Moreover, because  $u_2(y) \geq 0$  to begin with, the departure of player 2 does not decrease the aggregate payoff available to  $\{13\}$ . By monotonicity, both players 1 and 3 must then enjoy an intermediate payoff no smaller than what was available under  $y$ . So neither 1 nor 3 will participate in further moves. We've shown, in other words, that we have a contradiction:  $y$  must be stable to begin with.

Yet *all* the sets described in Observation 1 disappear in the face of the following small perturbation, in which we create a variant that possesses a core with nonempty interior.

**EXAMPLE 6.** *Change Example 5 by setting  $v(N) = 1 + \delta$  for any  $\delta > 0$ . (Think of  $\delta$  as small.) Now the interior of the core is non-empty; see the shaded area in Figure 3.*

Now Theorem 2 is back in force again, because every interior core allocation is separable. Every interior core allocation, partnered with the grand coalition, is a farsighted stable set.

At the same time:

**OBSERVATION 2.** *Under Conditions (i)–(iii) on the effectivity function, there is no other farsighted stable set other than the ones described by Theorem 2. In particular, the continuum of stable sets described in Observation 1 must entirely disappear.*

A formal proof of the Observation is in the Appendix, but here is the main argument. It is possible to show, in a series of steps that parallel those made for Observation 1, that the only candidates for farsighted stability are collections of states that all generate the same payoff for individual 1. However, now these sets cannot stretch from one end of the payoff simplex to the other, as they did for the pure veto game. In particular, neither player 2 nor 3 can be given a payoff lower than their marginal contribution to the grand coalition, which is  $\delta$ . See Figure 4 for a depiction of the possible sets that remain as potential candidates.

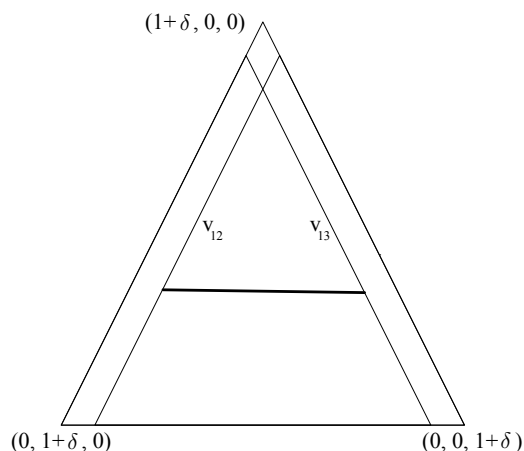


FIGURE 4. POSSIBLE CANDIDATE FOR FAR-SIGHTED STABLE SET

To see why such a claim is true, suppose that player 3 is pushed below  $\delta$  in some farsighted stable state  $y$  containing all three players,<sup>18</sup> and unilaterally leaves the arrangement. The overall payoff available to players 1 and 2 is now strictly less than  $u_1(y) + u_2(y)$ . By monotonicity, the resulting default payoff must be lower for both 1 and 2. Next, suppose 1 leaves  $\{12\}$ , precipitating the singletons with zero payoff. In the final step the grand coalition can make an improvement by moving to the “edge” of the farsighted set, represented by the payoff  $(u_1(x), 1 - u_1(x), \delta)$ . We have constructed a farsighted objection to  $y$ , starting with a departure by player 3, followed by player 1, and then the grand coalition.

But such “truncated” segments cannot be powerful enough to meet the external stability requirement. Consider the left-most extremity of the line segment in Figure 4, which is a payoff allocation of the form  $p = (c, 1 - c, \delta)$ . Using continuity and monotonicity, we can find a strictly positive payoff allocation to the “northwest” of  $p$  (consult Figure 4) such that the default payoff configuration for  $\{12\}$  — following  $s$ ’s departure — is precisely  $(c, 1 - c)$ . See the Appendix for details. However, by external stability,  $w$  must be displaced by a farsighted objection ending inside the horizontal line segment that forms our putative stable set.

But player 1 cannot be part of any coalition that starts such a move: he cannot be better off on any point on the line segment compared to  $w$ . Since there is no point to the left of  $p$  in the stable set, the same is true for player 2. But player 3 cannot gain because by leaving he will precipitate  $p$ , at which he gets 0 and from which there is no further move. Therefore, external stability is violated.

## REFERENCES

- BÉAL, S., J. DURIEU AND P. SOLAL (2008), “Farsighted Coalitional Stability in TU-Games,” *Mathematical Social Sciences*, **56**, 303–313.

<sup>18</sup>Lemma 2 in the Appendix proves that the starting state must involve the grand coalition of all three players.

- BHATTACHARYA, A. AND V. BROSI (2011), “An Existence Result for Farsighted Stable Sets of Games in Characteristic Function Form,” *International Journal of Game Theory*, **40**, 393–401.
- CHWE, M. (1994), “Farsighted Coalitional Stability,” *Journal of Economic Theory*, **63**, 299–325.
- DIAMANTOUDI, E. AND L. XUE (2003), “Farsighted stability in hedonic games,” *Social Choice and Welfare*, **21**, 39–61.
- (2005), “Lucas’ Counter Example Revisited,” Discussion paper, McGill University.
- FELDMAN, A. (1974), “Recontracting Stability,” *Econometrica*, **42**, 35–44.
- GREEN, J. (1974), “The Stability of Edgeworth’s Contracting Process,” *Econometrica*, **42**, 21–34.
- GREENBERG, J. (1994), “Coalition structures,” in *Handbook of Game Theory, Volume 2*, ed. by R. J. Aumann, and S. Hart, 1305–1337.
- HARSANYI, J. (1974), “An Equilibrium-Point Interpretation of Stable Sets and a Proposed Alternative Definition,” *Management Science*, **20**, 1472–1495.
- KONISHI, H. AND D. RAY (2003), “Coalition Formation as a Dynamic Process,” *Journal of Economic Theory*, **110**, 1–41.
- KÓCZY, L. AND L. LAUWERS, (2004), “The Coalition Structure Core is Accessible”, *Games and Economic Behavior*, **48**(1), 86–93.
- LUCAS, W. (1968), “A Game with No Solution,” *Bulletin of the American Mathematical Society*, **74**, 237–239.
- LUCAS, W. F. (1992), “von Neumann-Morgenstern Stable Sets,” in *Handbook of Game Theory, Volume 1*, ed. by R. J. Aumann, and S. Hart, 543–590. North-Holland.
- MAULEON, A., V. VANNETELBOSCH AND W. VERGOTE (2011), “von Neumann-Morgenstern farsighted stable sets in two-sided matching,” *Theoretical Economics*, **6**, 499–521.
- OWEN, G. (1995), *Game Theory*, Academic Press.
- RAY, D. (1989), “Credible Coalitions and the Core,” *International Journal of Game Theory*, **18**, 185–187.
- RAY, D. AND R. VOHRA (2013), “Coalition Formation”, Working Paper, Brown University.
- SENGUPTA, A. AND K. SENGUPTA (1996), “A Property of the Core,” *Games and Economic Behavior*, **12**, 266–73.
- SHAPLEY, L. AND M. SHUBIK (1969), “On Market Games”, *Journal of Economic Theory*, **1**, 9–25.
- STEARNS, R. (1964), “On the Axioms of a Cooperative Game without Sidepayments”, *Proceedings of the American Mathematical Society*, **15**, 82–86.
- VON NEUMANN, J. AND O. MORGENSTERN (1944), *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press.

## APPENDIX

**LEMMA 1.** *Suppose  $(N, V)$  is superadditive and the default function is monotonic. If there is a state  $x$  with  $u(x) \gg 0$ , then no state  $x'$  with  $u(x') = 0$  (such as the one with all players in singletons) can belong to a farsighted stable set.*

*Proof.* Let  $Z$  be a farsighted stable set. Consider a state  $x' = (\pi', u')$  with  $u' = 0$  and suppose there is a state  $x = (\pi, u)$  is such that  $u \gg 0$ . By superadditivity, we can assume without loss of generality that  $\pi = N$ . Thus the grand coalition has an objection from  $x'$  to  $x$ . If  $x \in Z$ , then clearly  $x' \notin Z$  (by internal stability) and the proof is complete. Suppose, therefore, that  $x \notin Z$ .



Since  $x \notin Z$ , there is a collection of states  $x^0, x^1, \dots, x^m$ , where  $x^0 = x$  and a corresponding collection of coalitions,  $S^1, \dots, S^m$ , such that for all  $k = 1, \dots, m$ :

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

and

$$u(x^m)_{S^k} \gg u(x^{k-1})_{S^k}.$$

We will now show that there is  $\bar{x} \in [x^m]$  which farsightedly dominates  $x'$ , to complete the proof that  $x' \notin Z$ . From  $x'$  consider a sequence of moves by each of the coalitions  $S^1, \dots, S^m$ , choosing the same payoffs at each move as they did in the move from  $x$  to  $x^m$ . This induces a sequence of states  $x'^1, \dots, x'^m$  with the property that for every  $k = 1, \dots, m$ ,

$$u(x'^k)_{S^k} = u(x^k)_{S^k}.$$

While the formation of coalitions is the same as in the previous sequence, the coalition structure at each step may be different because  $\pi$  may be different from  $\pi'$ .<sup>19</sup> Suppose the formation of  $S^k$  leaves behind a residual coalition  $W^k$  such that none of the players in  $W^k$  have been part of any previous move. This means that they were receiving 0 before the move by  $S^k$  and must therefore have belonged to a coalition with a zero worth. By superadditivity then, they must continue to receive 0 in state  $x'^k$ .

Thus, for any  $k$ , if  $i \notin \cup_{j=1}^k S^j$ , then  $u(x'^k)_i = 0$  and if  $i \in \cup_{j=1}^k S^j$ , then  $u_i(x'^k) = u_i(x^k)$ . In particular, this means that  $u_i(x^k) \geq u_i(x'^k)$  for all  $i \in N$ . And all players in  $S \equiv \cup_{j=1}^m S^j$  belong to the same coalition in  $\pi^m$  as in  $\pi$  and receive the same payoffs. If  $S = N$ , this means that  $x'^m = x^m$  and we have constructed a farsighted objection from  $x'$  to  $x^m$ . If  $N - S \neq \emptyset$ , given that  $x^m$  originated from  $x^1$  with  $\pi(x^1) = N$ , it follows that  $N - S \in \pi(x^m)$ . At  $x'^m$ , all players in  $N - S$  are either in residuals left by the formation of  $S^1, \dots, S^m$  or in coalitions that are untouched by the formation of these coalitions. In either case, players in  $N - S$  receive 0 in all states  $x', x'^1, \dots, x'^m$ . The only possible difference between  $\pi$  and  $\pi'$  therefore is that players in  $N - S$  are together in one single coalition in  $\pi$ , while in  $\pi'$  they may be in a sub-partition of  $N - S$ . If  $\bar{V}(N - S) = \{0\}$ , clearly  $x'^m \in [x^m]$ . Otherwise, by monotonicity,  $u(x'^m)_{N-S} \gg 0$ , while  $u(x^m)_{N-S} = 0$ . In this case, a final move by all players in  $N - S$  to merge into one coalition with payoff  $u(x^m)_{N-S}$ , moving to a new state  $x'^{m+1} \in [x^m]$ , results in a farsighted objection from  $x'$  to a state in  $Z$ . This completes the proof that  $x' \notin Z$ . ■

**LEMMA 2.** *Suppose  $(N, v)$  is a strictly superadditive three-person TU game. If default functions are continuous and monotonic, then any state in a farsighted stable set must consist of the grand coalition and an imputation.*

*Proof.* Let  $Z$  be a farsighted stable set. Since the game is strictly superadditive, by Lemma 1, a state with the finest coalition structure cannot be in  $Z$ . We will now show that there does not exist  $x^i = (\pi^i, u^i) \in Z$ , where  $\pi = (jk, i)$ . Suppose not. There are two cases to consider:

(i)  $x^i = (\pi^i, u^i) \in Z$ , where  $\pi = (jk, i)$  and  $(u_j^i, u_k^i) \gg 0$ . Consider  $x = (N, u)$  where  $u \gg u^i$  and  $f_{jk}(x) = (u_j^i, u_k^i)$ . This is possible because the default function is monotonic and

<sup>19</sup>It is also possible that for some  $j = 1, \dots, m$ ,  $S^j \in \pi'$ , in which case  $x'^{j-1} = x'^j$ .

continuous.<sup>20</sup> By internal stability,  $x \notin Z$  and there exists a farsighted objection by coalition  $S$  from  $x^i$  to  $x'$ , leading eventually to  $x^m \in Z$ . If  $S = (i)$ , then, since  $f_{jk}(x) = (u_j^i, u_k^i)$ ,  $x' = x^i \in Z$  and there cannot be any further change. But this is not a farsighted objection since  $S$  receives 0 by making this move. If  $S \neq (i)$ , we can follow the argument in the proof of Lemma 1 to construct a farsighted objection to  $x^i$  ending at  $x^m \in Z$ , contradicting the supposition that  $x^i \in Z$ .

(ii)  $x^i = (\pi^i, u^i) \in Z$ , where  $\pi = (jk, i)$ ,  $u_j^i > 0$ ,  $u_k^i = u_i^i = 0$ . Note that  $k$  can induce the finest coalition structure. As we have already argued, this state cannot be stable; there must be a farsighted objection leading to a state in  $Z$ . If  $k$  is part of any coalition in such a move it must end up with more than 0. But then  $k$  could precipitate such a farsighted move by inducing the finest coalition structure, which contradicts the hypothesis that  $x^i \in Z$ . Thus, the only farsighted move from the finest coalition structure must involve players  $i$  and  $j$ , both of whom must end up with strictly positive amounts in the end. But this is not possible as we have already ruled out, in the previous paragraph, the possibility of there being a state in  $Z$  consisting of a double coalition with strictly positive payoffs. ■

The following Observation formalizes the assertion in Example 3.

**OBSERVATION 3.** *Suppose all default functions are continuous and monotonic.  $Z$  is a farsighted stable set of the game in Example 3 if and only if  $Z = \{(N, u)\}$ , where  $u$  is an imputation in the interior of the core.*

*Proof.* It follows from Theorem 2 that  $\{(N, u)\}$  is a farsighted stable set for every imputation  $u$  in the interior of the core. We now show that there cannot be a farsighted stable set that is not of this form. From Lemma 2 we know that every state in a farsighted stable set must consist of the grand coalition and an imputation. Suppose there is a farsighted stable set  $Z$  which is not a singleton with an imputation in the interior of the core. Since no farsighted stable set can contain another,  $Z$  must consist only of imputations that are *not* in the interior of the core. This means, in particular, that if  $x = (N, u) \in Z$ ,  $u_i \geq 3$  for some  $i \in N$ . Without loss of generality, suppose  $u_1 \geq 3$ .

**Claim.** There does not exist  $z' = (N, u') \in Z$ , such that  $u'_1 \geq 3$ ,  $u' \neq u$ .

Suppose there is  $u' \in Z$  such that  $u'_1 \geq 3$  and  $u'_2 < u_2$ . This implies that  $u'_2 < 3$  and  $u'_1 + u'_3 > 3$ . By strict monotonicity, if player 2 leaves to form a singleton, player 1 will get less than 3. Now player 2 can precipitate a farsighted move from  $z'$  to  $z$  by leaving the grand coalition, resulting in player 1 getting less than 3, followed player 1 leaving 13 and the grand coalition moving to  $z$ . But this contradicts internal stability and proves the claim. Thus  $u'_2 = u_2$ . A similar argument shows that  $u'_3 = u_3$ , and this completes the proof of the claim.

Of course,  $Z$  cannot be a singleton as we have already remarked. In fact,  $z$  cannot dominate  $y$  where  $y$  consists of the coalition structure (1), (23) and 2 and 3 get at least as much as  $z$ .

<sup>20</sup>The argument is the following. Let  $u'$  be an imputation where  $u'_j > u_j^i$  and  $u'_k = u_k^i$ . Since  $u'_j + u'_k > 3$ , and  $u'_j, u'_k > 0$ , by monotonicity,  $f_{jk}(u')_k < u_k^i$ . Similarly, for an imputation  $u''$  with  $u''_j = u_j^i$  and  $u''_k > u_k^i$ , monotonicity implies that  $f_{jk}(u'')_j < u_j^i$ . It now follows from the continuity of the default function that there is some imputation,  $u$ , on the line segment between  $u'$  and  $u''$  such that  $f_{jk}(u) = u_{jk}^i$ .

This means that we must have  $z' \in Z$ ,  $z_1 < 3$ . Without loss of generality, since  $z'$  is not in the interior of the core,  $z'_2 \geq 3$ .

Suppose  $z_3 > z'_3$ . Then we have a farsighted objection from  $z'$  to  $z$  as follows. Player 3 leaves the grand coalition at  $z'$ . Player 1 is then left with less than 3. He leaves, and the grand coalition moves to  $z$ . This contradicts internal stability. By redoing this argument in the other direction we can claim that  $z_3 = z'_3$ .

But now consider a move by player 1 to go from  $z'$  to  $z$ . When he leaves, player 3 must get less than  $z_3 = z'_3$ . So player 3 is willing to leave next, and the grand coalition moves to  $z$ . This contradicts internal stability and completes the proof. ■

*Proof of Observation 1.* We have already proved the first part in the main text. Here, we show that there is no other farsighted stable set.

Claim I. In every farsighted stable set, there is a state  $x$  with  $u(x) \gg 0$ .

Let  $A$  be a farsighted stable set. Consider states of the form  $x(\epsilon)$ , where  $u(x(\epsilon)) = (1 - 2\epsilon, \epsilon, \epsilon)$ . If our claim is false for set  $A$ , then  $x(\epsilon) \notin A$  for every  $\epsilon \in (0, 1/4)$ . For  $\epsilon_1 \in (0, 1/4)$ , there is a farsighted objection which must end in a stable state  $y^1$  such that  $u_1(y^1) > 1 - 2\epsilon_1$  and  $u_i(y^1) > 0$ , but  $u_j(y^1) = 0$ , where  $i, j = 2, 3$  and  $j \neq i$ . Notice that for this objection to work, it *must* be that  $\pi(y^1) = \{1i\}, \{i\}$ .<sup>21</sup>

Now pick  $\epsilon_2 \in (0, 1/4)$  such that  $u_1(x(\epsilon_2)) > u_1(y^1)$ . Once again, there is a farsighted objection ending in a stable state  $y^2$  such that  $u_1(y^2) > 1 - 2\epsilon_2$  and  $u_k(y^2) > 0$ , but  $u_\ell(y^2) = 0$ , where  $k, \ell = 2, 3$  and  $k \neq \ell$ . As before, note that  $\pi(y^2) = \{1k\}, \{k\}$ .

But now we have a contradiction, for it is possible to construct a farsighted objection leading from  $y^1$  to  $y^2$ . This is done by having player 1 leave, precipitating the state with zero payoff vector and a coalition structure of singletons, followed by the formation of coalition  $\{1, k\}$  leading to state  $y^2$ . This contradiction establishes the claim.

Claim II. Fix a stable set  $A$  and  $x$  as in Claim I. For every  $y \in A$ ,  $u_1(y) \geq u_1(x)$ , and if  $(u_2(y), u_3(y)) \gg 0$ ,  $u_1(x) = u_1(y)$ .

The proof is the same as that of Step 2 in the proof of Observation 2 below.

Claim III. Fix stable set  $A$  and  $x \in A$  as in Claim I. Then for every payoff vector  $u \in \Delta$  with  $u_1 = u_1(x)$ , there is a stable state  $y \in A$  with  $u(y) = u$ .

Follow exactly the same proof as Step 3 in the proof of Observation 2, setting  $\delta = 0$  throughout.

The next claim strengthens Claim II.

Claim IV. Fix a stable set  $A$  and  $x$  as in Claim I. Then no state  $y$  with  $u_1(y) \neq u_1(x)$  can belong to  $A$ .

<sup>21</sup>This is because there is no strict improvement in the final step of the objection for player  $j$ .

The case  $u_1(y) < u_1(x)$  has already been dealt with in Claim II. So suppose that  $u_1(y) > u_1(x)$ . Then by Claim III, there must exist  $z \in A$  with  $(u_2(z), u_3(z)) \gg (u_2(y), u_3(y))$  and  $u_1(z) = u_1(x) > 0$ . It is now easy to construct a farsighted objection leading from  $y$  to  $z$  in which  $\{23\}$  moves first from  $y$ , precipitating the zero payoff vector, and then the grand coalition moves to secure  $z$ .<sup>22</sup>

Claims III and IV together complete the proof. ■

*Proof of Observation 2.* We proceed in a series of steps.

Step 1. In every farsighted stable set, there is a state  $x$  with  $u(x) \gg 0$ .

This parallels Claim 1 in the proof of Observation 1 but the proof is far simpler because of strict superadditivity. By Lemma 2, every state in the stable set must have the grand coalition as the accompanying coalition structure. So the absence of a strictly positive stable payoff will make it impossible for a farsighted stable state to block any strictly positive payoff allocation.<sup>23</sup> Step 1 follows.

Step 2. Fix a stable set  $A$  and  $x$  as in Step 1. For any other  $z \in A$ ,  $u_1(z) \geq u_1(x)$ , and if  $(u_2(z), u_3(z)) \gg 0$ ,  $u_1(z) = u_1(x)$ .

Suppose that  $z$  is a state with  $u_1(z) < u_1(x)$ . Construct a farsighted objection leading to  $x$  by having 1 leave the state  $z$ , precipitating the zero payoff vector, followed by the grand coalition moving to  $x$ . Therefore  $z \notin A$ .

If  $(u_2(z), u_3(z)) \gg 0$ , then in addition, it cannot be that  $u_1(z) > u_1(x)$ . For if this were the case, construct a farsighted objection leading from  $x$  to  $z$  by having 1 leave the state  $x$ , precipitating the zero payoff vector, followed by the grand coalition moving to  $z$ . But then  $x$  cannot be in  $A$ , a contradiction.

Step 3. Fix stable set  $A$  and  $x \in A$  as in Step 1. Then for every payoff vector  $u \in \Delta$  with  $u_1 = u_1(x)$  and  $\min\{u_2, u_3\} \geq \delta$ , there is a stable state  $y \in A$  with  $u(y) = u$ .

Suppose not. Then there is  $u \in \Delta$  with  $u_1 = u_1(x)$  and  $\min\{u_2, u_3\} \geq \delta$ , such that every state  $y$  with  $u(y) = u$  is blocked by a farsighted objection from  $z \in A$ . Because  $u_1(y) = u_1(x)$ , it follows from Step 2 that  $u_1(z) \geq u_1(y)$ .

Consider the possibility that 1 is *not* the first mover in the objection. Because  $u_1(z) \geq u_1(y)$ , the interests of 2 and 3 are opposed in the move from  $y$  to  $z$ ; both cannot gain. So  $\{23\}$  cannot be the *joint* first mover. Therefore, it is either 2 or 3 who must be the first mover. Say it is 2; then  $u_2(z) > u_2(y)$ . But then  $u_3(z) < u_3(y)$ . Moreover, because  $u_2(y) \geq \delta$  to begin with, the departure of player 2 does not decrease the aggregate payoff available to  $\{13\}$ . By monotonicity, both players 1 and 3 must then enjoy an intermediate payoff no smaller than what was available under  $y$ . So neither 1 nor 3 will participate in further moves. That eliminates this possibility.

<sup>22</sup>Note that the coalition structure associated with  $z$  *must* be the grand coalition, because  $(u_2(z), u_3(z)) \gg 0$ .

<sup>23</sup>After all, each player must find it strictly profitable to participate in the final move.

So 1 is the first mover in the objection, and in particular this implies that  $u_1(z) > u_1(y) = u_1(x)$ . Now it is easy to see that  $z$  can farsightedly object to  $x$  as well: construct exactly the same moves that were used to block  $y$ , beginning with the move of 1. This contradicts the fact that  $x \in A$ .

Step 4. Fix a stable set  $A$  and  $x \in A$  as in Step 1. Then no state  $y$  with  $u_1(y) = u_1(x)$ ,  $u_2(y) > 1 - u_1(x)$  and  $0 \leq u_3(y) < \delta$ , can be in  $A$ . (The same assertion is true if we flip 2 and 3.)

Suppose 3 leaves the grand coalition. The aggregate payoff for players 1 and 2 is now 1, which is strictly less than  $u_1(y) + u_2(y)$ . By monotonicity, the resulting default payoff must be lower for both 1 and 2. Next, suppose 1 leaves  $\{12\}$ , precipitating the singletons with zero payoff. In the final step the grand coalition can make an improvement by moving to the “edge” of the segment, represented by the payoff  $(u_1(x), 1 - u_1(x), \delta)$ , which we already know to be in  $A$ . We have constructed a farsighted objection to  $y$ , starting with a departure by player 3, followed by player 1, and then the grand coalition.

We now complete the proof of nonexistence. Consider the left-most extremity of the line segment in Figure 4, which is a payoff allocation of the form  $p = (c, 1 - c, \delta)$ . Using continuity and monotonicity, we can find a strictly positive payoff allocation to the “northwest” of  $p$  such that the default payoff configuration for  $\{12\}$  — following  $s$ ’s departure — is precisely  $(c, 1 - c)$ . (The argument is the same as in footnote 20).

By external stability,  $w$  must be displaced by a farsighted objection ending inside  $A$ . Because the possible additional points along the edges of the simplex cannot serve as objections,<sup>24</sup> the farsighted objection must result in a payoff allocation that is somewhere on the horizontal line segment.

It follows that player 1 cannot be part of any coalition that starts such a move: he cannot be better off on any point on the line segment compared to  $w$ . Since there is no point to the left of  $p$  in the stable set, the same is true for player 2. But player 3 cannot gain because by leaving he will precipitate  $p$ , at which he gets 0 and from which there is no further move. Therefore, external stability is violated. ■

---

<sup>24</sup>Recall that a stable coalition structure must always be a singleton consisting of the grand coalition. Therefore the edge points cannot serve as objections, as either player 2 or 3 will be unable to strictly gain in the final move.