ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 83, No. 3 (May, 2015), 977–1011

THE FARSIGHTED STABLE SET

DEBRAJ RAY
New York University, New York, NY 10012, U.S.A.

RAJIV VOHRA
Brown University, Providence, RI 02912, U.S.A.

Econometrica, Vol. 83, No. 3 (May, 2015), 977–1011

THE FARSIGHTED STABLE SET

BY DEBRAJ RAY AND RAJIV VOHRA[1]

Harsanyi (1974) criticized the von Neumann–Morgenstern (vNM) stable set for its presumption that coalitions are myopic about their prospects. He proposed a new dominance relation incorporating farsightedness, but 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 it. This implicitly gives an objecting coalition complete power to arrange the payoffs of players elsewhere, which is clearly unsatisfactory. While this assumption is largely innocuous for myopic dominance, it is of crucial significance for its farsighted counterpart. Our modification of the Harsanyi set respects "coalitional sovereignty." The resulting farsighted stable set is very different from both the Harsanyi and the vNM sets. We provide a necessary and sufficient condition for the existence of a farsighted stable set containing just a single-payoff allocation. This condition roughly establishes an equivalence between core allocations and the union of allocations over all single-payoff farsighted stable sets. We then conduct a comprehensive analysis of the existence and structure of farsighted stable sets in simple games. This last exercise throws light on both single-payoff and multi-payoff stable sets, and suggests that they do not coexist.

KEYWORDS: Core, stable set, farsightedness, coalition formation, simple games, veto players.

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 (vNM) stable set*. It is based on the concept of coalitional dominance. A feasible payoff profile is dominated by another profile 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 $u \in Z$, it is not dominated by $u' \in Z$.

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

The elements of Z are those outcomes (and only those) that 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, which is the set of all undominated payoff profiles. But it could have other members: $u \in Z$ may well be

[1]We are extremely grateful to Salvatore Nunnari for discussions on simple games with veto. We are also grateful to a co-editor and three anonymous referees, whose comments have served to substantially improve the paper. Ray acknowledges funding from the National Science Foundation under Grant SES-1261560.

 DOI: 10.3982/ECTA12022

dominated by u' as long as $u' \notin Z$. To be sure, external stability guarantees that u' in turn can be "blocked" by some other profile $u'' \in Z$. The presumption, then, is that u should still be considered "stable," because u' does not represent a lasting benefit.

Harsanyi (1974) took issue with this presumption. He observed that this argument is only valid if u'' is not preferred by the coalition that caused u to be replaced by u'. After all, perhaps u' was only a ruse to induce u'' 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).[2]

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*, which are efficient and individually rational payoff profiles for the full set of players. More precisely, an imputation u' dominates another imputation u if there is a coalition S for which u'_S (the restriction of u' to S) is feasible for S and $u'_i > u_i$ for all $i \in S$. The interest, of course, lies in the restriction of u' to S, because that is where the dominance occurs. The remainder of the dominating imputation u' only ensures that all allocations live in the same full-dimensional space, making for simpler and more elegant exposition.

This use of imputations has crucial implications when farsighted stability is involved. It grants a coalition S the power to replace imputation u with imputation *u' as long as u'_S is feasible for S*. In effect, the objecting coalition dictates the complementary allocation u'_{N-S}, which is problematic. For one thing, that allocation need not be feasible for the complementary set of players.[3] But feasibility aside, even the distribution of the payoffs to players outside the objecting coalition becomes important. The implicit presumption is that S can freely rearrange payoffs for $N - S$ and can somehow engineer society-wide changes to make these happen. 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.[4]

[2]Chwe's objective was to introduce another solution concept, the *largest consistent set*, which contains the Harsanyi stable set. Some of our comments on the latter will therefore also apply to the largest consistent set. See Section 6 for a more detailed discussion.

[3]The feasibility issue with the vNM stable set becomes apparent when the characteristic function is derived from an underlying economic model such as an exchange economy. When a coalition strikes out on its own, withdrawing its endowment from the rest of the economy, it may become impossible to sustain efficiency, regardless of how the rest of the players allocate their endowment.

[4]We are not the first to note this problem. Bhattacharya and Brosi (2011, p. 395) write that "we do not consider [the Harsanyi definition] to be very satisfactory because this implies that

While considerations of feasibility naturally matter for the vNM stable set as well,[5] the precise specification of the complementary allocation u'_{N-S} makes little difference to the definition of vNM stability. But with farsighted dominance, matters are not quite as harmless. If coalition S can replace u with u' (where u'_S is feasible for S), what transpires thereafter—for instance, whether another coalition T further replaces u' with u''—depends crucially on u'_{N-S}. In particular, the specification of u'_{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. As it turns out, this is not just a conceptual issue: our definition yields outcomes that depart significantly from their Harsanyi or vNM counterparts. Recent literature on Harsanyi stable sets provides existence results under fairly weak conditions; see Béal, Durieu, and Solal (2008) for transferable utility (TU) games, and Bhattacharya and Brosi (2011) for nontransferable utility (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. What is even more interesting is the form of the Harsanyi stable sets. They are always sets containing a single-payoff allocation, and what is more, *no such allocation is ever in the interior of the core*. Stranger still, every strictly positive allocation *not* in the core *must* be a singleton Harsanyi stable set.

This is odd. After all, the core interior can never be blocked—even weakly—by any subcoalition. That may be justifiably thought of as too demanding a requirement, and a solution concept that admits allocations other than just these is certainly worth entertaining. But a solution concept that invariably *excludes* all interior core allocations (while admitting all noncore allocations) needs critical examination.

It turns out that the imposition of coalitional sovereignty effectively overturns this result. All farsighted stable sets containing just a single-payoff allocation are core allocations under the solution we propose, and every payoff allocation in the interior of the core is a (single-payoff) farsighted stable set. Quite apart from the contrast with the Harsanyi set, this is an intriguing connection and suggests that the core of a game—even though it is defined by the

---

a coalition can directly enforce pay-offs even for the members outside itself. However, we are not aware of any simple and immediate way of resolving this while remaining within this environment of characteristic functions. So, we stick to this."

[5]As Greenberg, Luo, Oladi, and Shitovitz (2002) show, the stable set in allocation space may not be the same as the one in imputation space. They go on to show that this difference disappears if Harsanyi's notion of farsightedness is used. However, they continue the practice of allowing a deviating coalition to choose *any* (feasible) payoff for the complementary set of players. For an illustration of these issues in a matching model, see Example 3 in Mauleon, Vannetelbosch, and Vergote (2011).

property of not being "myopically blocked"—has powerful farsighted stability properties.[6] We are also able to provide a necessary and sufficient condition for the existence of a single-payoff farsighted stable set. In fact, we prove that a payoff allocation describing any single-payoff farsighted stable set must satisfy a particular property called *separability*, and as a converse, that every separable allocation is supportable as a single-payoff farsighted stable set (Theorem 2). The separability condition is satisfied in all superadditive games in which the interior of the core is nonempty.

Coalitional stability has also been studied in the class of so-called *hedonic games*, in which each coalition has a unique (efficient) payoff profile. For instance, matching models without transfers yield hedonic games. For such games, core stability has been studied in Banerjee, Konishi, and Sönmez (2001) and Bogomolnaia and Jackson (2002), and farsighted stability has been studied in Diamantoudi and Xue (2003) and Mauleon, Vannetelbosch, and Vergote (2011). Because there is no ambiguity about the payoff profile of $N - S$ when coalition S makes a move, coalitional sovereignty is not an issue in these games. The analysis of farsighted stability in Diamantoudi and Xue (2003) for hedonic games and in Mauleon, Vannetelbosch, and Vergote (2011) for matching games (a special class of hedonic games) is, therefore, not subject to our general criticism of Harsanyi stable sets. Indeed, their theorems on the existence of a farsighted stable set follow from our main result; see the Supplemental Material (Ray and Vohra (2015)).

As already mentioned, a Harsanyi stable set must be a single-payoff set. That is not true of the farsighted stable set studied here. After all, a single-payoff farsighted stable set exists if and only if separable allocations do, and separable allocations do not always exist. So the following questions arise:

Do (possibly multiple-payoff) farsighted stable sets generally exist?

What is the payoff structure of such sets?

Do all multi-payoff farsighted stable sets disappear when some allocation is separable?

We do not have general answers to these questions, but in Section 5, we study a broad class of situations, known as *simple games*, for which we can provide a definitive treatment. In simple games, either a coalition of players is winning, in which case it can collectively assure itself a payoff of 1, or it is losing, in which case it obtains zero. Depending on the questions to be asked, simple games provide a useful description of a parliament, or bargaining institution, or a committee. Simple games are rich for our purposes, in that they may or may not possess empty cores, and they may or may not have separable allocations, and we know the exact conditions under when these situations occur.

[6]While to our knowledge, this connection has not been explored before, Ray (1989) does show that the core has strong "internal consistency" properties, while Konishi and Ray (2003) argue that core allocations form rest points of a dynamic process of coalition formation with farsighted players.

It turns out that for this class of games, we can provide clear answers to the questions posed above. Farsighted stable sets with multiple payoffs do indeed exist, and they appear precisely when there are no separable allocations. In addition, we are able to describe the structure of such stable sets, and we show (among other things) that each stable set must provide a constant payoff to every *veto player*; that is, individuals who must be part of any winning coalition. This is in contrast to the structure of vNM stable sets for such games. We also identify some important special cases in which the farsighted stable set coincides with the vNM stable set.

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 $V(S) \subseteq R^S$, the S-dimensional Euclidean space with coordinates indexed by the players in S.

For all $S \subseteq N$, $V(S)$ is assumed to be *comprehensive*: if $v \in V(S)$, then $v' \in V(S)$ for all $v' \leq v$. Normalize the game so that singletons obtain zero.[7] Assume all coalitions can get nonnegative but bounded payoffs: $V(S) \cap R^S_+$ is nonempty and compact.

A payoff vector $v \in V(S)$ is *efficient for S* if $v \in \bar{V}(S) = \{v' \in V(S) \mid$ there is no $v'' \in V(S)$ with $v'' > v'\}$.[8]

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

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 $I(N, V) = \bar{V}(N) \cap R^N_+$.

2.2. *The Core and the Stable Set*

A pair (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:

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

[7]Formally, $V(\{i\}) = R^{\{i\}}_-$ for all $i \in N$.

[8]We use the convention $\geq, >\gg$ to order vectors in $\mathbb{R}^N$.

The *interior of the core* is defined as

$$\mathring{C}(N, V) = \{u \in I(N, V) \mid u_S \notin V(S) \text{ for any } S \subset N\}.$$

Note that $\mathring{C}(N, V) \neq \emptyset$ if and only if $C(N, V)$ is full dimensional, that is, of dimension $n - 1$.

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[9]

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

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

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

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 the following types of stability:

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''.

[9]See Shapley and Shubik (1969).

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.[10] 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 the following statements hold:

(i) We have $u^k_{S^k} \in V(S^k)$ for all $k = 1, \dots, m$.

(ii) We have $u'_{S^k} \gg u^{k-1}_{S^k}$ for all $k = 1, \dots, m$.

That is, 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.[11]

The new dominance relation leads to the following modification of the vNM stable set. A set of imputations H is said to be a *Harsanyi stable set* if it satisfies the following types of stability:

Internal Stability. No imputation in H is farsightedly dominated by any other imputation in H.

External Stability. Every imputation *not* in H is farsightedly dominated by some imputation in H.

In other words, if we define a new domination relationship by

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

for $A \subseteq I(N, V)$, then a Harsanyi stable set H is given by

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

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

[10]Note that S may not be able to directly block using u'', because u''_S may not be feasible for S.

[11]This is actually the second of two dominance notions proposed by Harsanyi (1974), and the one we shall concentrate on. The other notion he proposed required, in addition, that each coalition also make an instantaneous gain, that is, it addition to (i) and (ii), the condition that (iii) $u^k_{S^k} \gg u^{k-1}_{S^k}$ 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 (2014), but we shall not pursue it here.

Notice that in the formulations of both vNM and Harsanyi, the counterobjection u'' is taken to be a final rest point, by virtue of the fact that u'' is a member of the solution set (Z or H), which is taken to be "stable." Of course, the solution itself is a contextual object that may also be challenged, but that is not the issue to be addressed here. The presumption is that there may well be many solutions, but each of them, if in place, successfully addresses potential threats to its internal and external stability.

2.4. *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 a fundamental 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 (modulo the feasibility issue). 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 one-shot, myopic blocking.[12]

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. The use of imputations in the Harsanyi definition implicitly grants a coalition extraordinary power in the affairs of outsiders.

This can be consequential in strange ways. For instance, it is possible that a dummy player may be assigned a positive payoff in a stable imputation; see Ray and Vohra (2014) for an example. This is not a property shared by the vNM stable set, or indeed by most solution concepts. But the Harsanyi stable set has even stranger implications, as we can infer from the following result of Béal, Durieu, and Solal (2008).

[12]At the same time, myopic solution concepts that rely on *ongoing* blocking (see, for example, 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. Kóczy 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.

THEOREM 1—Béal, Durieu, and Solal (2008): *Suppose* (N, v) *is a TU game in which* $v(T) > 0$ *for some* $T \subset N$. *A set of imputations* H *is a Harsanyi stable set if and only if it contains a single 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 $u \in \mathring{C}(N, v)$, that is, $\sum_{i \in S} u_i > v(S)$ for every $S \subset N$. According to Theorem 1, u *cannot* be (part of) a Harsanyi stable set. Every 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.[13]

For more details on just what drives this peculiar result, see the discussion following Example 1 in Section 4.

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 S forms. Depending on the specific context, the members of S 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 relevant payoff constraints imposed on them. (Of course, in the subsequent periods, the players are free to adjust their coalitional membership.) Moreover, it may be unreasonable—or impossible—for S to dictate the division of payoffs among $N - S$.

To track these constraints, we 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, π) (or $(u(x), \pi(x))$ when we

[13]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 1 below for a comparison of the Harsanyi stable set and the vNM stable set in a three-player game.

need to be explicit), where u_S is feasible and efficient for S, that is, $u_S \in \bar{V}(S)$ 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—possibly empty—that have the power to change x to y for each pair of states x and y. The collection (N, V, E) goes beyond the characteristic function, in that it incorporates effectivity.

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

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

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) and 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 $u \in C(N, V)$.[14]

Now we move on to the concept of farsighted domination. State y *farsightedly dominates* x (under 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 $k = 1, \dots, m$,

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

and

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

For $A \subseteq X$, let

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

[14]The extended model is motivated by our attempt to capture farsightedness, but it serendipitously allows us to dispense with superadditivity as well.

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

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

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 $T \in \pi(y)$ (T might deliberately fragment itself), but we must surely allow T to have the option of remaining intact, as well as the option to choose from its own set of feasible payoffs. If T intersects $S \in \pi(x)$, there is also a question about whether the residual $S - T$ remains a coalition in $\pi(y)$. While our main result in the next section does not depend on the composition of the residual, in Section 5 we will assume that the residual does indeed remain as a coalition. For now, 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 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:

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

Condition (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; neither 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 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 (see Condition (iii) in Section 5.3 and Condition (iv) in Section O.3 of the Supplemental Material), we introduce a default function that maps the move of T to a unique coalition structure that leaves not just the untouched coalitions, but also the residuals, intact, and assigns every nonmember 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 and 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.[15]

4. THE CORE AND SINGLE-PAYOFF FARSIGHTED STABILITY

It is of special interest to consider farsighted stable sets that consist of a *single*-payoff allocation. In the space of states, these cannot generally be singleton sets, because several coalition structures might generate the same payoff allocation, and so will also need to be included in the set. Do such single-payoff sets exist, and is it possible to characterize the payoff allocations they contain?

Notice that single-payoff sets trivially pass the internal stability criterion: there are not two distinct payoff vectors in the set and, consequently, no "internal threat" of any kind. But they must work much harder on the external stability front: the payoff allocation, coupled with an accompanying coalition structure, must single-handedly serve as a farsighted objection to *every* alternative state. That happens to be too demanding as far as the vNM solution concept is concerned. That one distinguished imputation has the ability to block *every* other imputation is a requirement 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. This is what we turn to now.

[15]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 its (unique) efficient payoff vector.

A payoff allocation u is *efficient* if there does not exist another allocation u' (feasible for some coalition structure) such that $u' > u$. A collection of pairwise disjoint coalitions $\mathcal{T}$ is a *strict subpartition* of N if $N - \bigcup_{T\in\mathcal{T}} T$ is nonempty; these are the players not covered by $\mathcal{T}$. 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 $S \subseteq N - \bigcup_{T\in\mathcal{T}} T$.

Separability has close (but not exact) links to the core. If u is separable, then u must belong to the coalition structure core of (N, V). For if this were false, there would exist $T \subset N$ 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 $\pi' = (T, S, S', \ldots)$ such that (v, u_{N-T}) is feasible for π'. 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 2 in Section 5 below. However, if $u \in \mathring{C}(N, V)$, then it is easy to see that it 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 $u_{N-T} \in V(N-T)$. It is possible for a superadditive game to possess a separable allocation even though the interior of the core is empty. This is the case in Lucas's (1968) example of a TU game in which the core is nonempty but there is no vNM stable set; see Example O.1 in the Supplemental Material. However, in a *strictly* superadditive game,[16] the separability of u is equivalent to the statement that u lies in the interior of the core.

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

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

THEOREM 2: *Consider* (N, V, E) *such that* E *satisfies Conditions* (i) *and* (ii). *Then* $[u]$ *is a single-payoff farsighted stable set if and only if* u *is separable.*

PROOF: To prove the "if" part, 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 $T^1, \ldots, T^M$ and states $y^1, \ldots, y^M$, where $y^M \in [u]$. That is, each coalition T^k will lie in $E(y^{k-1}, y^k)$, with $u_{T^k} \gg u^{k-1}_{T^k}$ for all $k = 1, \ldots, M$.

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

[16]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 $V(S \cup T)$.

Since u is efficient, for every state $z \notin [u]$, $B(z) \equiv \{i \in N \mid u_i > u_i(z)\} \neq \emptyset$.

Stage 1. Since $y^0 \notin [u]$, $B(y^0) \neq \emptyset$. If each $i \in B(y^0)$ is a singleton in π^0, that is, if $\{i\} \in \pi^0$ for every $i \in B(y^0)$, go to Stage 2 below. Otherwise, pick a player $i_1 \in B(y^0)$ such that $i \in S$ for some $S \in \pi^0$ with $|S| \geq 2$. 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). Since i_1 is in a singleton coalition, $u_{i_1}(y^1) = 0 < u_{i_1}$. Thus, $y^1 \notin [u]$ and $B(y^1) \neq \emptyset$. If each $i \in B(y^1)$ is a singleton, go to Stage 2. Otherwise, form a new singleton coalition from $B(y^1)$ to move to y^2, where $y^2 \notin [u]$. Repeat this process as long as there is some player in a nonsingleton coalition who prefers u. 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 lead, in a finite number of steps, to a state in which all players who prefer u are in singleton coalitions. At this point we invoke Stage 2.

Stage 2. At the initiation of this stage we are in state y^m, where $B(y^m) \neq \emptyset$ and every $i \in B(y^m)$ is in a singleton coalition.

There are now two cases to consider. In case 1, $B(y^m) = N$. Pick any π such that $y = (u, \pi) \in [u]$. Form each of the coalitions in π 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 $B(y^m)$ is nonempty, π^m restricted to the complement of $B(y^m)$ 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 $B(y^m)$ such that $u_{S(1)} \in V(S(1))$. By Condition (ii), $S(1)$ can move to a state that it weakly prefers to u. 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 $\{S(1), \ldots, S(\ell)\}$. (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 $u^* = u$, that is, $y^* \in [u]$. 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.

We now turn to a proof of the "only if" part of the theorem. 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 to the contrary that $[u]$ is a farsighted stable set but u is not 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 $S \subseteq R(\mathcal{T}) \equiv N - \bigcup_{T \in \mathcal{T}} T$. Define a state x in the following way. Construct a partition π' 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 $z \in [u]$. *Q.E.D.*

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 have 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 noncore 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 Kóczy 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 (2014).

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. While our critique of the Harsanyi approach is not based on this outcome, but rather on the conceptual underpinnings of that approach, such odd outcomes are grounds for additional misgiving. The following example reiterates this point.

EXAMPLE 1: Given a three-player TU convex 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 $(3, 0, 3)$, shown as the inverted central triangle in Figure 1.

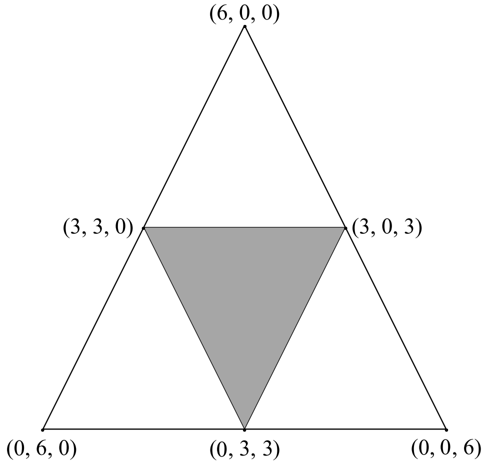

FIGURE 1.—Comparison of stable sets.

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, Durieu, and Solal (2008), they have as payoff vectors all strictly positive imputations that are feasible for some two-player coalition. In Figure 1, these are all points in the *complement* of the inverted central triangle.

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, while assigning 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. (No other restriction is placed, thus allowing the division $(6, 0)$ between players 1 and 2 to be implemented.) Now continue by having player 2 move; she gives herself 0, gives 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 no interior core allocation has the same power to dominate. 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 have departed from the Harsanyi approach in two ways: (a) our domain is the set of feasible utility profiles for coalition structures, and (b) we have 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.[17] 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.

Actually, the contrast between the two stable sets is even sharper. Under a weak additional condition on the effectivity correspondence, there are no other farsighted stable sets in this example, single payoff or otherwise. The Supplemental Material contains the details.

By Theorem 2, the presence of a separable allocation is sufficient for the existence of a farsighted stable set. In particular, existence is assured if the interior of the core is nonempty. And, as we show in the Supplemental Material, in hedonic games with strict preferences, nonemptiness of the core suffices for existence. More generally, while separability is necessary and sufficient for the existence of a single-payoff farsighted stable set, multiple-payoff farsighted stable sets often exist when there are no separable allocations. This is the case for the class of TU games that we study in the next section and, indeed, we have yet to encounter an example of a TU game without a farsighted stable set. However, existence is not guaranteed in every NTU game; see Example O.2 of a three-player roommate game in the Supplemental Material.

We end this section by pointing out a conceptual issue with farsighted dominance. In our framework, as well as in Harsanyi's, the initial deviating coalition is optimistic in presuming that other coalitions will abide by their anticipated moves as long as these are profitable. After all, profitability does not rule out the possibility that other participating coalitions may have gained *even* more by doing something else, which may not necessarily have been in the interest of the initial deviator. The difficulty of dealing with multiple continuation paths following an initial move also crops up in Greenberg's (1990) theory of social situations, where he distinguishes between "optimistic" and "conservative" notions of dominance as alternative ways of refining the spectrum of continu-

[17]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, he 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 leaves sequentially, until all players are in singletons. Then make a final move by the grand coalition to u.

ations. Conservative farsighted dominance requires the initial deviating coalition to be better off at all possible final outcomes.[18]

In general, as argued in Ray and Vohra (2014), optimism and pessimism are both ad hoc assumptions, and one may turn out to be more reasonable than the other, depending on the context. A more nuanced approach would assume that each coalition along a dominance chain chooses *maximally* among all possible feasible moves. These and related matters are discussed in Ray and Vohra (2014), in which we argue for real-time definitions of coalition formation that free us from such ambiguities.

However, it *is* true that in the important special case of single-payoff farsighted stability this critique has no bite. A single-payoff farsighted stable set is also a farsighted conservative stable set. This is so because all moves eventually lead to the *same* outcome. No coalition can benefit by choosing an alternative path: wherever the coalition goes, every continuation thereafter will terminate in the same payoff outcome. Within the ambit of single-payoff stability therefore, no such ambiguities exist: all roads lead to Rome.

5. BEYOND SINGLE-PAYOFF STABILITY: SIMPLE GAMES

This section is motivated by two considerations. First, we wish to examine farsighted stable sets beyond the realm of Theorem 2, that is, in games in which the interior of the core is empty or, more generally, in games that do not possess separable allocations. Second, we provide some insights into the nature of farsighted stable sets in such environments and compare them to vNM stable sets. Our inquiry is motivated by questions such as the following:

Are there farsighted stable sets with multiple payoffs?

What is the payoff structure of such sets?

Do such sets invariably appear when no allocation is separable?

Do such sets disappear when some allocation is separable?

We do not have general answers to these questions. But in this section, we analyze these issues in the class of (proper) *simple games* and attempt to provide a definitive treatment. In simple games, either a coalition of players is winning, in which it can collectively assure itself a payoff of 1, or it is losing, in which case it obtains zero.[19]

Depending on the context and the questions asked, simple games can provide a useful abstract description of a parliament, or a bargaining institution, or a committee. Such games have been extensively analyzed in the context of

[18]For a definition of the corresponding farsighted conservative stable set and its close connection with Chwe's largest consistent set, see Diamantoudi and Xue (2003). For related solution concepts, see Herings, Mauleon, and Vannetelbosch (2004) and Mauleon and Vannetelbosch (2004). We discuss the connection between the farsighted stable set and the largest consistent set in Section 6.

[19]The qualification "proper" refers to the additional qualification that if a coalition is winning, its complement is losing; see below.

the vNM stable set (see, for example, Lucas (1992)) or used in theories of bargaining with majority voting Baron and Ferejohn (1989), and have played a significant role in the analysis of political institutions; see, for example, Austen-Smith and Banks (1999) and Winter (1996). We place significant (though not exclusive) emphasis on an important subclass of such games: those with "veto players," who are individuals who must be included in every winning coalition. These players might together form a winning coalition in their own right (an "oligarchy") or they might not. The former case has obvious applications, but the latter forms an equally important framework in political science with wide applications. For instance, under the decision rule employed in the U.N. Security Council, the five permanent members are each veto players, but together do not form an oligarchy, as the affirmation of at least 4 out of the remaining 10 nonpermanent members is also required. A nonoligarchic setup with veto players also represents the relationship between the U.S. Congress and the U.S. President, as neither of them can unilaterally pass a reform of the status quo (see Winter (1996, pp. 818–820) for more examples).[20]

Quite apart from intrinsic appeal and applicability, simple games are rich for our purposes, in that they may or may not possess empty cores and they may or may not have separable allocations, and we know the exact conditions under when these situations occur.

It turns out that for this class of games, we can provide fairly complete answers to the questions posed above. Farsighted stable sets with multiple payoffs do indeed exist, and they appear precisely when there are no separable allocations. In addition, we are able to describe the structure of such stable sets, and we show (among other things) that each stable set must provide a constant payoff to every *veto player*; that is, individuals who must be part of any winning coalition. This is in contrast to the structure of vNM stable sets for such games, which we also describe for the purposes of comparison.

5.1. *Proper Simple Games*

More formally, we study superadditive TU games with the property that for every coalition S, $v(S) = 1$ or $v(S) = 0$, and if $v(S) = 1$, then $v(N - S) = 0$. Coalition S is a *winning coalition* if $v(S) = 1$ and a *losing coalition* if $v(S) = 0$. Let $\mathcal{W}$ denote the set of all winning coalitions.

A coalition S is a *veto coalition* if its complement is losing, that is, if $v(N - S) = 0$.[21] A *minimal veto coalition* is a veto coalition such that no strict

[20]For more on veto players, see, for example, Matthews (1989), Cameron (2000), Diermeier and Myerson (1999), Tsebelis (2002), Gehlbach and Malesky (2010), Diermeier, Egorov, and Sonin (2013), and Nunnari (2014).

[21]Note that a winning coalition is necessarily a veto coalition. Shapley (1962) defines a blocking coalition as a losing coalition such that its complement is also losing. In our terminology these are veto coalitions that are losing.

subset of it is a veto coalition. A singleton veto coalition—if it exists—will be referred to as a *veto player*; note that every veto player must belong to every winning coalition. The collection of all veto players, also known as the *collegium*, is denoted $S^* = \bigcap_{S \in \mathcal{W}} S$. A *collegial game* is one in which $S^* \neq \emptyset$. The collegium (and the corresponding game) will be called *oligarchic* if S^* is itself a winning coalition.

The set of imputations of a simple game is just the nonnegative n-dimensional unit simplex $\triangle$. The core of a simple game is nonempty if and only if the collegium, S^*, is nonempty, and in that case it takes the form

$$C(N, v) = \left\{ u \in \triangle \,\middle|\, \sum_{i \in S^*} u_i = 1 \right\}.$$

It is easy to see that a separable allocation u is a core allocation with the additional property that $\{i \in N \mid u_i > 0\}$ is a winning coalition. This implies (and is implied by the fact) that the game is oligarchic and $u_i > 0$ for all $i \in S^*$. Thus, oligarchic games yield the precise subclass of simple games for which Theorem 2 applies, while in all nonoligarchic games, a single-payoff farsighted stable set cannot exist. For this reason, simple games provide a fertile field of inquiry for farsighted stable sets with multiple payoffs.

We will continue to assume that the effectivity correspondence satisfies Conditions (i) and (ii). These coalitional sovereignty conditions imply that all states with the same winning coalition and the same payoff allocation are, in effect, equivalent: the payoffs to winning coalitions are completely unaffected by changes in the structure elsewhere. To make the exposition far simpler without sacrificing anything of substance, we refer to this entire equivalence class as a (single) state. So a state x will be fully described by its winning coalition $W(x)$ (if any) and the payoff allocation $u(x)$ among members of this coalition, where we take for granted that $u_i(x) = 0$ for all i not in $W(x)$. It follows that there is only "one state" without any winning coalition, which we refer to as the *zero state*.

A state x is *regular* if $W(x) \neq \emptyset$ and $u_i(x) > 0$ for all $i \in W(x)$. Every farsighted objection must culminate in a regular state, because the last coalition to move must be a winning coalition and all its members must gain. The task of ensuring the external stability of a farsighted stable set, F, must therefore rest on regular states in F. In particular, every farsighted stable set must contain at least one regular state. On the other hand, consider the following lemma.

LEMMA 1: *Suppose that x and y both belong to a farsighted stable set of a collegial game and that $u_i(y) > u_i(x)$ for some $i \in S^*$. Then y cannot be regular.*

PROOF: If x contains a winning coalition, then $i \in W(x)$ since i is a veto player. Now i can stand alone, creating the zero state. If y is regular, the coalition $W(y)$ can thereafter precipitate a move to y, making all its members better off. So y farsightedly dominates x, which contradicts internal stability. (The argument holds a fortiori if x is the zero state and has no winning coalition.) *Q.E.D.*

Lemma 1 takes a step toward a central conclusion of this section, which is that the payoff of a veto player must be fully pinned down in any farsighted stable set. See Section 5.3.

5.2. *Oligarchic Games and the Core*

In oligarchic games, separable allocations exist: these are core allocations with strictly positive payoffs to every member of the collegium. So Theorem 2 applies. In fact, we can sharpen that theorem in this case to assert that the *only* farsighted stable sets are those identified in Theorem 2.

THEOREM 3: *Assume Conditions* (i) *and* (ii). *If the game admits an oligarchic collegium, then F is a farsighted stable set if and only if $F = [u]$, where $u \in C(N, v)$ and $u_{S^*} \gg 0$.*

PROOF: By Theorem 2, $[u]$ is a single-payoff farsighted stable set if and only if $u \in C(N, v)$ and $u_{S^*} \gg 0$. Now we prove that there are no other farsighted stable sets. Suppose, on the contrary, that F is a farsighted stable set and is not of the form $[u]$ where u is separable. Since no farsighted stable set can contain another, this means that F does not contain any state with a separable allocation. Of course, F contains a regular state; fix some such state $z \in F$. Since it is not separable, it cannot be in the core. Therefore, there is a core state x with $W(x) = S^*$ and $u_j(x) > u_j(z)$ for every $j \in S^*$. Because $z \in F$, we have $x \notin F$, and so x must be farsightedly dominated by some regular state y in F. This must mean that $u_i(y) > u_i(x)$ for some $i \in S^*$. Otherwise, by Condition (i), there cannot be a farsighted move from x. But we now have $z, y \in F$, with y regular and $u_i(y) > u_i(z)$ for some $i \in S^*$. This contradicts Lemma 1 and completes the proof. *Q.E.D.*

Note that a pure bargaining game, one in which the only winning coalition is N, is an example of an oligarchic game: every player is a veto player. In fact, an oligarchic collegial game is nothing other than a pure bargaining game with the possible addition of one or more dummy players. It is, therefore, not surprising that in Theorem 2 we have a characterization of farsighted stable sets in which dummy players receive 0.[22]

With the connection to separable allocations and to the previous section out of the way, the remainder of this section is devoted to a study of simple games

[22]In contrast, any $u \in \triangle$ such that $u_{S^*} \gg 0$ (even one assigning strictly positive payoffs to dummy players) is a singleton Harsanyi set.

that do not possess any separable allocation, that is, nonoligarchic games. We will show, under some mild conditions, that farsighted stable sets continue to exist and they have a special structure with connections to both the single-payoff farsighted stable sets we have already studied and to vNM stable sets.

5.3. *Nonoligarchic Games and the Structure of Multi-Payoff Stable Sets*

When a game is nonoligarchic, then there is no separable allocation and, consequently, no farsighted stable set with a single payoff. But farsighted stable sets do exist. Our objective is to establish the existence of such sets and to describe their payoff structure. It turns out that farsighted stable sets continue to exhibit payoff constancy, but for a particular subclass of players.

In their analysis of simple games, von Neumann and Morgenstern (1944) showed that certain collections of imputations, which they referred to as *discriminatory sets*, play a distinguished role. A discriminatory set takes the form $D(K, \mathbf{a}) = \{u \in \triangle \mid u_i = a_i \text{ for } i \in K\}$, where $a_i \geq 0$ for all $i \in K$. The members of K are the "fixed-payoff" players, each receiving a constant amount. The rest of the surplus is divided in an arbitrary way among the remaining agents—the "bargaining players." It turns out that for any minimal winning coalition S and with some restrictions on $\mathbf{a}$, the fixed payoff vector, discriminatory vNM stable sets of the form $D(N - S, \mathbf{a})$ exist. In particular, the bargaining players are members of the minimal winning coalition.

For instance, in a three-person majority game, where each two-player grouping is a minimal winning coalition, there are discriminatory stable sets corresponding to each such coalition in which there is a single fixed-payoff player (the excluded member from the winning coalition), who receives less than 0.5.[23] In other words, for every $i \in N$, $D(\{i\}, \mathbf{a})$ is a stable set for $a < 0.5$. For a general characterization of discriminatory stable sets, see Owen (1965).

An extreme version of a discriminatory set is one in which *every* player gets a fixed payoff. That is just a single-payoff set. While a vNM stable set typically cannot be of that form (due to the demands of external stability), we have already seen farsighted stable sets of this form in Theorems 2 and 3. In nonoligarchic games, discriminatory farsighted stable sets abound, as we will show below. However, they are not singleton payoff sets. They appear in this sense to be similar to discriminatory vNM stable sets. But there are significant differences as well.

To begin with, discriminatory sets are even more central to the farsighted theory than they are in the vNM setting. In the latter case, discriminatory sets are one of several kinds of vNM stable sets, even in simple games. We have already seen that in oligarchic simple games, all farsighted stable sets must be single payoff. In a nonoligarchic game, every veto player must receive the same

[23] These are not the only stable sets; see the Supplemental Material for a discussion of the main simple solutions.

amount over all states in a farsighted stable set. Put another way, *all* farsighted stable sets must have payoffs that are subsets of $D(S^*, \mathbf{a})$. If $\sum_{i\in S^*} a_i$ is close to 1, this is almost a single-payoff set. However, in contrast to the case in which the collegium is oligarchic, the stable set must leave some surplus for the other players, which implies that a farsighted stable set must be disjoint from the core. In comparing farsighted stable sets to vNM stable sets, we will sometimes abuse terminology in referring to the payoffs of a farsighted stable set, without reference to the corresponding winning coalition(s), as a farsighted stable set.

To develop these ideas formally, we will impose an additional, intuitive restriction on the effectivity correspondence. Recall that Conditions (i) and (ii) on effectivity allow for both deviating and untouched coalitions to have coalitional sovereignty. But they are silent on the residuals left behind by deviating coalitions. When those residuals are not winning, there is no condition to impose. But when some residual coalition is winning, it must remain intact. Moreover, the deviants who broke away from it were previously adding no value, so their departure is of no overall loss to the winning residual coalition and, indeed, no *individual* member of that residual should suffer any immediate loss either.[24] Formally we impose the following monotonicity condition.

Condition (iii). Suppose $S \in W(x)$ and $T \in E(x, y)$. If $S - T$ is a winning coalition, then $S - T = W(y)$ and $u_i(y) \geq u_i(x)$ for all $i \in W(y)$.

We can now state the following theorem.

THEOREM 4: *Assume Conditions* (i)–(iii). *Suppose that the game is a nonoligarchic collegial game*. *If F is a farsighted stable set*, *then for every $x \in F$*, *the following statements hold*:

(a) *We have $u(x) \in D(S^*, \mathbf{a})$*, *where $a \gg 0$*. *Every veto player gets a fixed*, *positive payoff*: *$u_i(x) = u_i(y) \equiv a_i > 0$ for every $i \in S^*$ and any pair of states $x, y \in F$*.

(b) *We have $u(x) \notin C(N, v) : \sum_{i\in S^*} a_i < 1$*.

PROOF: Begin by assuming, to the contrary, that there exist x and y in F with $u_i(y) > u_i(x)$ for some $i \in S^*$. By Lemma 1, y is not regular, and $J \equiv \{j \in W(y) \mid u_j(y) = 0\} \neq \emptyset$. Consider a regular state y' with (i) $\pi(y') = \pi(y)$ (so, in particular, $W(y) = W(y')$), (ii) $u_i(x) < u_i(y') < u_i(y)$, (iii) $u_j(y') > 0$ for all $j \in J$, and (iv) $u_k(y') = u_k(y)$ for all $j \notin J \cup \{i\}$. (It is trivial to see that such a state exists.) Because y' is regular and $u_i(y') > u_i(x)$, Lemma 1 applies again and $y' \notin F$. This means that there is a farsighted objection from y' to a (regular) state $z \in F$.

[24]For coalition sovereignty, what is important is that the deviating coalition not be allowed to choose how the residuals organize themselves or how they distribute their surplus among themselves; these decisions must be taken as exogenously given by the deviating coalition. The assumption that the residual, if it is winning, remains intact and none of its members lose is a reasonable rule with this property. An analogue of this condition for more general games is Condition (iv) in Section O.3 of the Supplemental Material.

Since $x, y, z \in F$ and z is regular, it follows from Lemma 1 that $u_{S^*}(z) \leq u_{S^*}(y)$ and $u_{S^*}(z) \leq u_{S^*}(x)$. Given (ii) and (iv) in the construction of y', it follows that

$$u_{S^*}(z) \leq u_{S^*}(y').$$

Since there is a farsighted objection leading from y' to z, Lemma 2 (stated and proved in the Appendix) asserts that there is also a farsighted objection in two steps. The first involves a coalition $S \subseteq W(y') - S^*$ breaking away to implement the zero state and then moving with $W(z)$ to z. Moreover,

$$u_S(z) \gg u_S(y') \geq u_S(y),$$

where the second inequality uses the fact that $S \subseteq W(y') - S^*$. But this means that S can generate the same kind of objection leading from y to z (recall that $W(y) = W(y')$). Since $y, z \in F$, this contradicts internal stability and completes the proof that for any $x, y \in F$, $u_{S^*}(x) = u_{S^*}(y) \equiv a$.

Of course, F must include a regular state. But that is impossible if $a_i = 0$ for any $i \in S^*$ or if $\sum_{i \in S^*} a_i = 1$ (the latter because S^* is not winning). This completes the proof. *Q.E.D.*

The "fixity" of payoffs for veto players is an intuitive implication of farsighted stability. If multiple payoffs were available for a veto player given the standard of behavior implicit in a farsighted stable set, our player would try to negotiate for the best of these. With vNM stability, this is not possible: the veto player cannot get to the desired payoff by a single objection; he will need assistance from a winning coalition to achieve that payoff, but the other members of the winning coalition may well be worse off in the proposed move and so not cooperate.

However, under farsighted stability, our veto player can paralyze all negotiations by withdrawing from them and precipitating the zero state. Now a winning coalition will be willing to move to the state that generates our veto player's best payoff, if it gives all its members a positive payoff. (The formal proof takes care of this last technical requirement.)

On the other hand, part (b) of the theorem states that the collective of all veto players cannot arrogate the entire surplus: they must relinquish some of it to the nonveto players if the game is nonoligarchic. This is because the complement of the set of veto players is, in fact, a veto coalition, and they can also bring all negotiations to a standstill. The formal proof of the theorem shows that this ability must result in some strictly positive payoff for the complement of the collegium. This is an extremely important point, as it states that a nonoligarchic collegium S^* must fail to extract the entire surplus, and the outcome is no longer in the core.

The feature that in a farsighted stable set, the veto players *must* collectively release some surplus to other players, is not generally true of all vNM stable sets. This is an additional difference between the farsighted stable set and the vNM stable set, despite their common reference to discriminatory sets.

A final important difference is that the set of fixed-payoff players in the farsighted case *includes* all the veto players, whereas in a vNM stable set, the set of fixed-payoff players consists of the complement of a minimal winning coalition and must, therefore, *exclude* all veto players. We illustrate this contrast in the next example:

EXAMPLE 2: Given a three-player 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 is the veto player, but is not winning. The core has the single payoff $(1, 0, 0)$.

The core of this game has player 1 absorbing the entire surplus of the game. This is also the equilibrium outcome in an extensive bargaining game without discounting; see Winter (1996). With discounting, however, the expected payoff to all players is positive (see Baron and Ferejohn (1989), Chatterjee, Dutta, Ray, and Sengupta (1993), and Winter (1996)). Now, full surplus absorption is not a necessary feature of a vNM set, but it *must* include that possibility. A typical vNM set is as depicted in Figure 2(a): it is a continuous curve that begins at $(1, 0, 0)$ and continues to the opposite edge; see for example Lucas (1992). There are also two discriminatory vNM stable sets, $D(\{2\}, 0)$ and $D(\{3\}, 0)$.

In contrast, a farsighted stable set, according to Theorem 4, is disjoint from the core, and cannot exhibit a varying payoff for player 1. In fact, in this example, it has the additional property that the remaining surplus can be shared in any way among the nonveto players (see Corollary 1 below). In other words, it

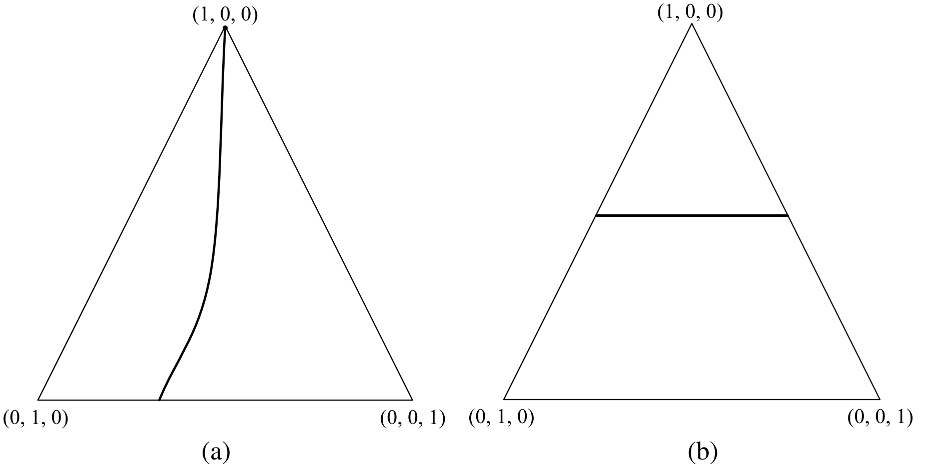

FIGURE 2.—(a) von Neumann–Morgenstern sets, and (b) farsighted stable sets in the veto game.

must be a discriminatory set of the form $D(\{1\}, \mathbf{a})$. Figure 2(b) depicts a typical farsighted stable set.

To summarize, in this example, there are two discriminatory vNM stable sets, but there are also many other (nondiscriminatory) vNM stable sets. Farsighted stable sets, however, must all be discriminatory. Moreover, in a discriminatory vNM stable set, either player 2 or player 3 is the fixed-payoff player, while in a farsighted stable set, it is player 1, the veto player, who is the fixed-payoff player. This too is not accidental, as shown by Theorem 4; see also Theorem 5 and its corollary below.

Theorem 4 identifies a necessary condition for stability, but does not actually show that stable sets exist. That will require a more detailed account of how agents other than veto players are treated. To approach this question, recall that no agent not in the collegium S^* is a veto player (by definition of S^*). But the complementary set $N - S^*$ does have veto *coalitions*. Because the collegium is not an oligarchy, it is certainly true, for instance, that $N - S^*$ itself is a veto coalition.

This is, of course, true of a wide class of institutions. Recall the example of the U.N. Security Council, which has a collegium made up of the set of five permanent members. But a decision also needs the affirmative vote of 4 of the 10 nonpermanent members. So every collection of members with at least 7 nonpermanent members forms a veto coalition, and every collection with exactly 7 nonpermanent members is a *minimal veto coalition*, which is a veto coalition with the property that no subcoalition of it is a veto coalition.

In what follows, we impose the following mild additional restriction. Say that a veto coalition M is *nonelitist* if it is minimal and $(M - \{i\}) \cup \{j\}$ is also a veto coalition for every $i \in M$ and every $j \notin M$. Such a coalition is (loosely speaking) purely built on the principle of numbers. It cannot be smaller (it is minimal), but it can replace every individual member with any outsider and still retain its veto power; hence the term "non-elitist."[25] Note that veto players *are* elitist in this sense as long as everyone is not a veto player. Our restriction is that there exists a nonelitist veto coalition.

This is a weak requirement. It does not require symmetry for the game as a whole and, in particular, the existence of a nonelitist veto coalition implies neither the presence nor absence of veto players. The U.N. Security Council satisfies this condition and, indeed, so do most nonoligarchic institutions we can think of.[26]

[25]However, even a nonelitist veto coalition may not afford to lose two or more of its members, so in this sense the definition is weak.

[26]When there are dummy players who add value to no coalition, the nonelitist condition fails, but the our next theorem extends easily as long as there is a veto coalition that is nonelitist with respect to the nondummy players. There is a farsighted stable set in which nondummy players get payoffs as described in Theorem 5 ,while dummy players receive 0. We conjecture, in fact, that the nonelitism restriction can be dropped without cost, but have not proved this.

The existence of a nonelitist veto coalition is sufficient for the existence of a farsighted stable set.

THEOREM 5: *Assume Conditions* (i)–(iii). *Suppose there exists a nonelitist veto coalition* M. *Then* $[D(N-M,\mathbf{a})] \equiv \{x \in X \mid u(x) \in D(N-M,\mathbf{a})\}$ *is a farsighted stable set*, *where the following conditions hold*:

(a) *We have* $a_i > 0$ *for all* $i \in N - M$.

(b) *We have* $a(S^*) \equiv \sum_{i \in S^*} a_i < 1$, *as required by Theorem* 4.[27]

(c) *We have* $a(M) \equiv \sum_{j \in M} a_j > \frac{(m-1)}{m}(1 - a(S^*))$, *where* m *is the cardinality of* M.

PROOF: Let $F = [D(N-M,\mathbf{a})]$, where $\mathbf{a}$ satisfies (a), (b), and (c). We first verify the internal stability of F. Suppose $y, z \in F$ and there is a farsighted objection leading from y to z. Since $u_i(y) = u_i(z)$ for all $i \notin M$, by Lemma 2, there is a coalition $S \subseteq W(y) \cap M$ such that $W(y) - S$ is losing and $u_S(z) \gg u_S(y)$. If $S = M$, then it cannot be that $u_S(z) \gg u_S(y)$, because the sum of payoffs to M is constant across all states in F. Otherwise, S is a strict subset of M. Define $Q \equiv N - W(y)$; then Q is also a subset of M.[28] Because no subset of M is a veto coalition, and yet $W(y) - S$ is losing, it must be that $Q \cup S = M$. But the aggregate payoff to Q under y was zero, so the aggregate payoff to S must have been $a(M)$. This is an upper bound to the aggregate payoff to S under z, which again contradicts $u_S(z) \gg u_S(y)$.

To verify external stability, consider any state $y \notin F$. We need to show that there is a farsighted objection to y from some $z \in F$. If $u(y) = 0$, there is an objection by the grand coalition to $z \in F$, where $u(z) \gg 0$. So suppose that $u(y) > 0$, which also means that $W(y)$ is well defined. There are now three cases to consider.

(i) Suppose $u_i(y) < a_i$ for some $i \in S^*$. Then player i, who must be in $W(y)$, can unilaterally precipitate the zero state and thereafter lead the grand coalition to $z \in F$, where $u(z) \gg 0$.

(ii) Suppose $\sum_{i \in M} u_i(y) < 1 - \bar{a}$. The coalition $W(y) \cap M$ (which must be nonempty, because M is a veto coalition) can then construct a farsighted objection by leaving $W(y)$—thus precipitating the zero state—and then forming the grand coalition to reach a regular state in F such that all players in $W(y) \cap M$ gain by doing so.

(iii) Finally, consider y such that $u_i(y) \geq a_i$ for all $i \in S^*$ and $\sum_{i \in M} u_i(y) \geq a(M)$. If $y \notin F$, this must mean that $u_j(y) < a_j$ for some $j \in N - S^* - M$. Pick any such j and form the coalition $M' \equiv (M - \{k\}) \cup \{j\}$, where k is a player with the highest payoff in M (under y). Our assumption that M is nonelitist implies that M' is a veto coalition. Of course, $M' \cap W(y) \neq \emptyset$, so these players can leave $W(y)$, causing everyone to receive 0. We will now argue that M' has

[27]Use the convention that $a(S^*) = 0$ if $S^* = \emptyset$.

[28]The members of $N - M$ earn strictly positive payoffs and, therefore, cannot be in Q.

a farsighted objection by going through this step and then moving, with the grand coalition, to a regular state in F.

Denote by c the aggregate of payoffs to M under the state y. The player in M who is excluded from M' has the highest payoff under y, which is, therefore, at least c/m. It follows that $\sum_{i\in M-\{k\}} u_i(y) \leq \frac{m-1}{m}c$. Therefore,

$$\sum_{i\in M-\{k\}} u_i(y) \leq \frac{m-1}{m}c \leq \frac{m-1}{m}\big(1-a\big(S^*\big)\big) < a(M).$$

It follows that there is a regular state in F such that all players in $M-\{k\}$ get more than they do at y, player j gets $a_j > u_j(y)$, and all other players receive a positive payoff. This completes the proof that $(M-\{k\})\cup\{j\}$ can—by precipitating the zero state—engineer a farsighted objection starting at y and ending at a regular state in F. *Q.E.D.*

Theorem 5 shows the existence of (discriminatory) farsighted stable sets in which the set of bargaining players is the nonelitist veto coalition, while the set of fixed-payoff players includes all veto players (if any). As we saw in Example 2, discriminatory vNM stable sets can be fundamentally different because they specify the bargaining players to be a minimal winning coalition, rather than a nonelitist veto coalition. In the case of the U.N. Security Council, a discriminatory vNM stable set will treat the five permanent members and any four nonpermanent members (a minimal winning coalition) as the bargaining players. The farsighted stable set identified here does the opposite: it will have seven nonpermanent members (a nonelitist veto coalition) as the bargaining players, while the remainder obtain a fixed payoff.

There is an important special case in which this stark difference vanishes.

EXAMPLE 3: Given a symmetric, simple majority game with an odd number of players n. That is, $v(S)=1$ if and only if $|S|>n/2$.

Minimal winning coalitions are all coalitions of size $(n+1)/2$. But these are precisely the nonelitist veto coalitions as well. The structure of discriminatory stable sets is, therefore, the same in both myopic and farsighted cases: there must be $(n+1)/2$ bargaining players. The two sets may differ, however, because of the restrictions they place on the payoffs to fixed-payoff players. For example, when $n=3$, the fixed-payoff player in a vNM stable set gets $a\in[0,0.5)$. For the farsighted stable set, condition (c) of Theorem 5 also yields the same upper bound, but the payoff must be positive: $a\in(0,0.5)$.

Must every farsighted stable set be a discriminatory stable set? Sometimes this is the case. For instance, suppose that $N-S^*$ is a minimal veto coalition. Put another way, for every $i\in N-S^*$, $S^*\cup\{i\}$ is a winning coalition. In this class of games, farsighted stable sets *must be* discriminatory sets with S^* as the fixed-payoff players.

COROLLARY 1: *Assume Conditions* (i)–(iii), *and suppose that* $N - S^*$ *is a minimal veto coalition. Then* F *is a farsighted stable set if and only if* $F = [D(S^*, \mathbf{a})]$, *where* $a \gg 0$ *and* $\sum_{i \in S^*} a_i < 1$.

PROOF: Note that $N - S^*$ must be nonelitist. Therefore, sufficiency follows directly from Theorem 5.

To prove necessity, first observe from Theorem 4 that if F is a farsighted stable set, then $F \subseteq [D(S^*, \mathbf{a})]$, where $a \gg 0$ and $\sum_{i \in S^*} a_i < 1$. We will now show the reverse inclusion: every $y \in [D(S^*, \mathbf{a})]$ belongs to F. Suppose, to the contrary, that $y \in [D(S^*, \mathbf{a})]$ but is not in F. Then there must be a farsighted objection to y that ends in some $z \in F$. Because no member of S^* can gain, Lemma 2 applies, and we can presume that the first coalition to move from y to z is some $S \subseteq W(y) - S^*$, precipitating the zero state, while in the second (and final) step, $W(z)$ forms, with $u_S(z) \gg u_S(y)$. However, the fact that the zero state is created in the first step means, by the assumptions of this corollary, that S must *equal* $W(y) - S^*$. But all members of $W(y) - S^*$ cannot gain, because they obtain an aggregate of $1 - a(S^*)$ at y, and no more than this amount at z. *Q.E.D.*

While we conjecture that Corollary 1 can be substantially extended, there are examples of farsighted stable sets that are not discriminatory, though of course every veto player must continue to receive a fixed payoff.

EXAMPLE 4: Given a four-player veto game: $N = \{1, 2, 3, 4\}$, $v(N) = v(\{1, i, j\}) = 1$ for all $i, j \in \{2, 3, 4\}$ and $v(S) = 0$ for all other S. Player 1 is the veto player, but needs at least two partners to win. Every two-player coalition from $\{2, 3, 4\}$ is nonelitist, so all the conditions of Theorem 5 are satisfied.

In this example, there exists a farsighted stable set F that is not discriminatory and contains precisely six elements in it. Player 1 gets some fixed amount $a \in (0, 1)$ in each of them, in accordance with Theorem 3. The remainder, $1 - a$, is divided equally between a pair $\{i, j\}$ of players drawn from $\{2, 3, 4\}$. To this payoff allocation append two coalition structures, one with winning coalition $\{1, i, j\}$ and the other with winning coalition N, thus completing the description of two states. (By varying the pairs, that makes six states in all.) Clearly, F is internally stable. To verify external stability, pick any state $y \notin F$. If $u_1(y) < a$, then player 1 can initiate a domination chain by standing alone. On the other hand, if $u_1(y) \geq a$, then, because it is distinct from any element of F, it must be that two players i and j from $\{2, 3, 4\}$ obtain *strictly* less than $(1 - a)/2$ each. But now the coalition $\{i, j\}$ can start a domination chain by precipitating the zero state, moving thereafter to the state with winning coalition $\{1, i, j\}$.

The last example is suggestive of a close connection between farsighted stability and discrete solution sets of the kind that von Neumann and Morgenstern call a *main simple solution*. We explore this connection further in the

Supplemental Material. Via the concept of a main simple solution, we also draw a connection between our solution concept and the *demand bargaining set* of Morelli and Montero (2003). The case of "pillage games," introduced in Jordan (2006), provides another interesting situation in which farsightedness does not modify the myopic notion of a stable set; again see the Supplemental Material.

6. A REMARK ON THE LARGEST CONSISTENT SET

A leading and influential example of a farsighted solution concept based on conservative behavior is the largest consistent set proposed by Chwe (1994). Given some abstract set of outcomes Y and an effectivity correspondence E, a set $K \subseteq Y$ is *consistent* if

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

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

When discussing TU coalitional games, Chwe (1994) follows von Neumann and Morgenstern (1944) and Harsanyi (1974) by taking Y to be the set of imputations and letting $S \in E(u, w)$ if and only if $w_S \in V(S)$. In this setting, Béal, Durieu, and Solal (2008) show that in strictly superadditive games with at least four players, the largest consistent set is the entire collection of imputations, and, therefore, it lacks any predictive power. To see why, note that in a strictly superadditive game with at least four players, for every player i, there is a coalition S with $v(S) > 0$ and another player j such that $i, j \notin S$. By Theorem 1, there is a singleton farsighted stable set consisting of an imputation u such that $u_S \in V(S)$, with $u_S \gg 0$ and $u_i = 0$. Hence, from every other imputation, there is a farsighted objection leading to u. This is sufficient to deter player i from joining any coalitional move. As this argument holds for every player, all imputations are "stable."[29]

As in Harsanyi's set, is this lack of predictive power due to Chwe's problematic use of the imputation-based approach? The answer appears to be no,

[29] A similar argument shows that in all games, including three-player games, all *strictly positive* imputations belong to the largest consistent set provided every player belongs to *some* strict subset of N with positive worth (as in Example 1). Imputations that are not strictly positive could, however, be excluded by the largest consistent set. In Example 1, imputations such as $(3, 3, 0)$ and $(6, 0, 0)$ are not in the largest consistent set; see also Example 1 in Béal, Durieu, and Solal (2008).

at least for some classes of games. We could restrict the effectivity correspondence to satisfy coalitional sovereignty; that is, Conditions (i) and (ii). Nevertheless, the set of single-payoff farsighted stable sets could be large enough to deter a move from *any* strictly positive imputation. More precisely, suppose for every player i and $\varepsilon > 0$, there exists a separable allocation u such that $u_i < \varepsilon$. Then, by Theorem 2, a proposed move by i to any state exposes her to the risk that a subsequent farsighted objection will assign her a payoff arbitrarily close to 0. All states with strictly positive payoffs would then belong to the largest consistent set.

Might the set of separable allocations be large in this sense? The answer is yes. Example 1 has this property; see Figure 1. Indeed, the answer is in the affirmative for all strictly convex games. A TU game is *convex* if for all S, T, $v(S) + v(T) \leq v(S \cup T) + v(S \cap T)$. It is *strictly convex* if the inequality is strict for S and T that are not nested. It follows from Theorems 1 and 6 in Shapley (1971) that $\mathring{C}(N, v) \neq \emptyset$ in all strictly convex games.[30] At the same time, we know from Shapley (1971) that in a convex game, for each player i, there is an extreme point u^i of the core with $u^i_i = 0$. Convex combinations of u^i_i and some point in $\mathring{C}(N, v)$ then yield interior core allocations (hence separable) that bring the payoff of i arbitrarily close to 0.

The argument just made is sufficient, but it is not necessary. The simple games of Section 5 often have no separable allocations. But the union of all farsighted stable sets contains elements that are capable of driving any one player's payoff as close to zero as we wish. By Chwe (1994), the largest consistent set contains every farsighted stable set (defined relative to the same effectivity correspondence). It follows that *every* strictly positive imputation belongs to the largest consistent set.

One might respond that in this last case, farsighted stable sets also span a large range of outcomes. However, the predictions of all theories of stable sets are *collections* of outcomes—the stable sets—and not the outcomes per se. Under the identifying assumption that the norms governing the establishment of a particular stable set are persistent, our theory would be falsified if two outcomes were to be observed from two different farsighted stable sets, each of which makes a very particular set of predictions. In contrast, if a stable set (such as the largest consistent set) contains every positive imputation, it would be near impossible to falsify that theory.

7. CONCLUSION

von Neumann and Morgenstern (1944) formulated the concept of a stable set, based on the concept of coalitional dominance. No two outcomes in the

[30] Reny, Winter, and Wooders (2012) show this to be true in the wider class of *quasi-strictly convex* games, which require the strict inequality to hold only for nonnested S and T such that $S \cup T = N$.

stable set dominate each other, and any outcome not in the stable set is dominated by some outcome in the stable set. This represented an elegant and powerful approach to the problem of consistency in dominance: an objection is tested in just the same way as it tests the original proposal.

Harsanyi (1974) modified the stable set, arguing that it took inadequate care of farsighted coalitional objections. He replaced the one-step dominance relationship in vNM with a chain of objections, each step in that chain being driven by the prospect of gain at the terminal outcome of the chain. In this paper, we argue that the suggested modification is problematic because it retains certain features of the original vNM concept that are fundamentally ill-suited for farsightedness. Our central point is that the Harsanyi definition denies the coalitional sovereignty of players, and taken literally, it grants a coalition unlimited power in the affairs of outsiders.

We propose a definition of a farsighted stable set that respects coalitional sovereignty. Our definition has a profound effect on the nature of farsighted stable sets. Harsanyi stable sets are always singletons and turn out to be entirely disjoint from the interior of the core. Indeed, every allocation *not* in the core is a singleton Harsanyi stable set. Our formulation overturns this unsatisfactory property. All single-payoff farsighted stable sets are core allocations under the solution we propose, and every payoff allocation in the interior of the core (along with appropriate coalition structures) forms a farsighted stable set.

Quite apart from the contrast with the Harsanyi set, this result suggests that the core of a game has powerful farsighted stability properties. We are also able to fully characterize single-payoff farsighted stable sets. These must all employ "separable allocations," a concept that is closely related to (though not identical with) that of a core allocation.

We then turn to multi-payoff farsighted stable sets and are able to provide a fairly complete description of them in a broad subclass of games called simple games. Such games embody succinct abstractions of a parliament, or a bargaining institution, or a committee such as the U.N. Security Council. Simple games are rich for our purposes, in that they may or may not possess empty cores, or they may or may not have separable allocations, and we know the exact conditions under which these situations occur.

For this class of games, we show (under mild conditions) that farsighted stable sets with multiple payoffs do indeed exist, and they appear precisely when there are no separable allocations. (When there are separable allocations, the only farsighted stable sets are single payoff.) In addition, we are able to describe the structure of such stable sets, and we show that each stable set must provide a constant payoff to every veto player. There are intriguing contrasts, as well as some similarities, to the well studied concept of vNM stability for the very same class of games.

These are comprehensive results, but they leave several questions open, at least for general games. We conjecture that farsighted stable sets exist for all transferable utility characteristic function games (some discussion of the existence question is in the Supplemental Material). An equally important question from an applied perspective has to do with whether our single-payoff solution concept needs to coexist with other farsighted stable sets when the existence condition for the former is met. Put another way, do multi-payoff sets make an appearance if and only single-payoff sets fail to exist? Our analysis of simple games strongly suggests that the answer to this important question is in the affirmative, but a general exploration of this question must be left open as a future research project.

APPENDIX

LEMMA 2: *Consider a proper simple game, and assume Conditions* (i), (ii), *and* (iii) *on the effectivity correspondence. Consider a farsighted objection leading from y to z. Then either of the following statements holds*:

(a) *We have $u_i(z) > u_i(y)$ for all $i \in W(z)$, in which case $W(z)$ can move directly from y to z.*

(b) *We have $u_i(z) \leq u_i(y)$ for some $i \in W(z)$. In this case, $W^+ = \{i \in W(y) \mid u_i(z) > u_i(y)\} \neq \emptyset$ and the first stage in the farsighted objection involves a coalition $S \subseteq W^+$ breaking away* (*perhaps in several steps*) *from $W(y)$ and precipitating the zero state* (*so $W(y) - S$ is a losing coalition*). *If W^+ is a losing coalition, then there is no loss of generality in assuming that the farsighted objection has two steps. The first step consists of coalition W^+ breaking away from $W(y)$ to precipitate the zero state. In the second, and final, step $W(z)$ moves to z.*

PROOF: The first part of the lemma is obvious. Suppose, therefore, that $u_i(z) \leq u_i(y)$ for some $i \in W(z)$. Since is z is a regular state, this means that $u_i(y) > 0$ and $W(y) \cap W(z) \neq \emptyset$. Partition $W(y)$ into W^+ and W^-, where $W^+ = \{i \in W(y) \mid u_i(z) > u_i(y)\}$ and $W^- = \{i \in W(y) \mid u_i(z) \leq u_i(y)\}$. A farsighted move from y to z must, at some stage, involve, a change in $W(y)$. (Otherwise, by Condition (i), the payoff vector cannot change.) None of the players in W^- will be part of the first coalition that disrupts $W(y)$ since they have nothing to gain by having y replaced with z. Thus, the first coalition to form, say S^1, which causes the payoffs to change, must include a subset of W^+. Either $W(y) - S^1$ is a losing coalition or, by Condition (iii), all the players left as residuals by the formation of S^1 remain intact and (weakly) gain as a result of S^1's departure. In the latter case, the next coalition, say S^2, which causes $W(y) - S^1$ to change, must also include a subset of W^+. Eventually, enough players in W^+ must leave the (shrinking) winning coalition until it is no longer winning. In other words, at some stage, the coalition $W(y)$ must shrink to $W(y) - S$, where $S \subseteq W^+$ and $W(y) - S$ is losing. Of course, the final step in any farsighted move from y to z is the formation of $W(z)$. In the event that W^+ is a losing coalition, the farsighted move from y to z can be achieved in two steps. First, W^+ breaks away from $W(y)$ to precipitate the zero state; in the second step, $W(z)$ moves from the zero state to z. *Q.E.D.*

REFERENCES

AUSTEN-SMITH, D., AND J. BANKS (1999): *Positive Political Theory I: Collective Preference*. Ann Arbor, MI: University of Michigan Press. [995]

BANERJEE, S., H. KONISHI, AND T. SÖNMEZ (2001): "Core in a Simple Coalition Formation Game," *Social Choice and Welfare*, 18, 135–153. [980]

BARON, D., AND J. FEREJOHN (1989): "Bargaining in Legislatures," *American Political Science Review*, 83, 1181–1206. [995,1001]

BÉAL, S., J. DURIEU, AND P. SOLAL (2008): "Farsighted Coalitional Stability in TU-Games," *Mathematical Social Sciences*, 56, 303–313. [979,984,985,992,1006]

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. [978,979]

BOGOMOLNAIA, A., AND M. O. JACKSON (2002): "The Stability of Hedonic Coalition Structures," *Games and Economic Behavior*, 38, 201–230. [980]

CAMERON, C. (2000): *Veto Bargaining: Presidents and the Politics of Negative Power*. Cambridge: Cambridge University Press. [995]

CHATTERJEE, K., B. DUTTA, D. RAY, AND K. SENGUPTA (1993): "A Noncooperative Theory of Coalitional Bargaining," *Review of Economic Studies*, 60, 463–477. [1001]

CHWE, M. (1994): "Farsighted Coalitional Stability," *Journal of Economic Theory*, 63, 299–325. [978,1006,1007]

DIAMANTOUDI, E., AND L. XUE (2003): "Farsighted Stability in Hedonic Games," *Social Choice and Welfare*, 21, 39–61. [980,994]

DIERMEIER, D., AND R. MYERSON (1999): "Bicameralism and Its Consequences for the Internal Organization of Legislatures," *American Economic Review*, 89, 1182–1196. [995]

DIERMEIER, D., G. EGOROV, AND K. SONIN (2013): "Endogenous Property Rights," available at http://ssrn.com/abstract=2327412. [995]

FELDMAN, A. (1974): "Recontracting Stability," *Econometrica*, 42, 35–44. [984,991]

GEHLBACH, S., AND E. MALESKY (2010): "The Contribution of Veto Players to Economic Reform," *Journal of Politics*, 72, 957–975. [995]

GREEN, J. (1974): "The Stability of Edgeworth's Recontracting Process," *Econometrica*, 42, 21–34. [984,991]

GREENBERG, J. (1990): *The Theory of Social Situations*. Cambridge, MA: Cambridge University Press. [993]

——— (1994): "Coalition Structures," in *Handbook of Game Theory*, Vol. 2, ed. by R. J. Aumann and S. Hart. Amsterdam: North-Holland, 1305–1337. [986]

GREENBERG, J., X. LUO, R. OLADI, AND B. SHITOVITZ (2002): "(Sophisticated) Stable Sets in Exchange Economies," *Games and Economic Behavior*, 39, 54–70. [979]

HARSANYI, J. (1974): "An Equilibrium-Point Interpretation of Stable Sets and a Proposed Alternative Definition," *Management Science*, 20, 1472–1495. [977,978,983,985,1006,1008]

HERINGS, P., A. MAULEON, AND V. VANNETELBOSCH (2004): "Rationalizability for Social Environments," *Games and Economic Behavior*, 49, 135–156. [994]

JORDAN, J. (2006): "Pillage and Property," *Journal of Economic Theory*, 131, 26–44. [1006]

KONISHI, H., AND D. RAY (2003): "Coalition Formation as a Dynamic Process," *Journal of Economic Theory*, 110, 1–41. [980,987,991]

KÓCZY, L., AND L. LAUWERS (2004): "The Coalition Structure Core Is Accessible," *Games and Economic Behavior*, 48, 86–93. [984,988,991]

LUCAS, W. (1968): "A Game With no Solution," *Bulletin of the American Mathematical Society*, 74, 237–239. [989]

——— (1992): "von Neumann–Morgenstern Stable Sets," in *Handbook of Game Theory*, Vol. 1, ed. by R. J. Aumann and S. Hart. Amsterdam: North-Holland, 543–590. [995,1001]

MATTHEWS, S. (1989): "Veto Threats: Rhetoric in a Bargaining Game," *Quarterly Journal of Economics*, 104, 347–369. [995]

MAULEON, A., AND V. VANNETELBOSCH (2004): "Farsightedness and Cautiousness in Coalition Formation Games With Positive Spillovers," *Theory and Decision*, 56, 291–324. [994]

MAULEON, A., V. VANNETELBOSCH, AND W. VERGOTE (2011): "von Neumann–Morgenstern Farsighted Stable Sets in Two-Sided Matching," *Theoretical Economics*, 6, 499–521. [979,980]

MORELLI, M., AND M. MONTERO (2003): "The Demand Bargaining Set: General Characterization and Application to Majority Games," *Games and Economic Behavior*, 42, 137–155. [1006]

NUNNARI, S. (2014): "Dynamic Legislative Bargaining With Veto Power," Report, Department of Political Science, Columbia University. [995]

OWEN, G. (1965): "A Class of Discriminatory Solutions to Simple N-Person Games," *Duke Mathematical Journal*, 32, 545–553. [998]

——— (1995): *Game Theory*. San Diego: Academic Press. [986]

RAY, D. (1989): "Credible Coalitions and the Core," *International Journal of Game Theory*, 18, 185–187. [980,991]

RAY, D., AND R. VOHRA (2014): "Coalition Formation," in *Handbook of Game Theory*, Vol. 4, ed. by H. P. Young and S. Zamir. Amsterdam: North-Holland, 239–326. [983,984,991,994]

——— (2015): "Supplement to 'The Farsighted Stable Set'," *Econometrica Supplemental Material*, 83, http://dx.doi.org/10.3982/ECTA12022. [980]

RENY, P., E. WINTER, AND M. WOODERS (2012): "The Partnered Core of a Game With Side Payments," *Social Choice and Welfare*, 39, 521–536. [1007]

SENGUPTA, A., AND K. SENGUPTA (1996): "A Property of the Core," *Games and Economic Behavior*, 12, 266–273. [984,991]

SHAPLEY, L. (1962): "Simple Games: An Outline of the Descriptive Theory," *Behavioral Science*, 7, 59–66. [995]

——— (1971): "Cores of Convex Games," *International Journal of Game Theory*, 1, 11–26. [1007]

SHAPLEY, L., AND M. SHUBIK (1969): "On Market Games," *Journal of Economic Theory*, 1, 9–25. [982]

TSEBELIS, G. (2002): *Veto Players: How Political Institutions Work*. Princeton, NJ: Princeton University Press. [995]

VON NEUMANN, J., AND O. MORGENSTERN (1944): *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press. [977,998,1006,1007]

WINTER, E. (1996): "Voting and Vetoing," *American Political Science Review*, 90, 813–823. [995, 1001]

Dept. of Economics, New York University, New York, NY 10012, U.S.A.; debraj.ray@nyu.edu

and

Dept. of Economics, Brown University, Providence, RI 02912, U.S.A.; rajiv_vohra@brown.edu.

Manuscript received November, 2013; final revision received October, 2014.