WARWICK

ECONOMICS

CRETA

Centre for Research in Economic Theory and its Applications

Discussion Paper Series

Coalition Formation and History Dependence

Bhaskar Dutta & Hannu Vartiainen

July 2017 No: 33

CRETA
Centre for Research in Economic Theory and its Applications
Department of Economics
University of Warwick, Coventry,
CV4 7AL, United Kingdom

warwick.ac.uk/fac/soc/economics/research/centres/creta/papers

Coalition Formation and History Dependence *

Bhaskar Dutta

University of Warwick and Ashoka University

Hannu Vartiainen

University of Helsinki and Helsinki Centre of Economic Research

July 2017

Abstract. Farsighted formulations of coalitional formation, for instance by Harsanyi (1974) and Ray and Vohra(2015), have typically been based on the von Neumann-Morgenstern (1944) stable set. These farsighted stable sets use a notion of indirect dominance in which an outcome can be dominated by a chain of coalitional 'moves' in which each coalition that is involved in the sequence *eventually* stands to gain. Dutta and Vohra(2016) point out that these solution concepts do not require coalitions to make optimal moves. Hence, these solution concepts can yield unreasonable predictions. Dutta and Vohra (2016) restricted coalitions to hold common, history independent expectations that incorporate optimality regarding the continuation path. This paper extends the Dutta-Vohra analysis by allowing for history dependent expectations. The paper provides characterization results for two solution concepts corresponding to two versions of optimality. It demonstrates the power of history dependence by establishing non-emptyness results for all finite games as well as transferable utility partition function games. The paper also provides partial comparisons of the solution concepts to other solutions.

*The first author takes this opportunity of recording his intellectual debt to Debraj Ray and Rajiv Vohra for many years of most helpful discussions. The second author gratefully acknowledges the hospitality of Clare Hall and INET at the University of Cambridge. The paper has also benefited from discussions with Francis Bloch, Hamid Sabourian as well as comments from seminar participants in the University of Montreal, the annual Coalitions and Networks Workshop in University of Glasgow, and the Paris School of Economics.

1. INTRODUCTION

The von Neumann-Morgenstern (vNM) stable set has had a distinguished standing as a solution concept in cooperative game theory. It is based on the notion of coalitional dominance, with one social state y dominating state x if some coalition has the power or ability to change the state from x to y and all members of the coalition prefer y to x. von Neumann and Morgenstern identified a stable set as one which satisfied two properties properties: (1) *internal stability* in the sense that no stable outcome dominates any other stable outcome; (2) *external stability* in the sense that every outcome not in the stable set is dominated by some stable outcome. Of course, the *core*, the set of states which are not dominated by any other state, must be contained in any stable set. The predominant position of the vNM stable set is evident from the large literature on this solution concept.[2]

Both the core and the stable set are *myopic* solution concepts in the sense that a deviating coalition only cares about the *immediate* consequence of a deviation. But if coalition S decides to change x to y because the latter gives strictly higher payoffs to each member of S, it does not ask itself whether y itself is a stable outcome. Conversely, the implicit rationale of the vNM set is that if x is not dominated by any coalition, then x must be in the solution set since no coalition objects to it. Harsanyi (1974) criticised the underlying logic by pointing out the following. Suppose coalition S has the power to enforce y from x. Suppose also that at least one member of S does not gain from the move to y. Then, myopic solution concepts would decree that S will not in fact effect the move from x to y. But now suppose that some state z which is deemed stable dominates y *and* all members of S strictly prefer z to x. Harsanyi argued that S should in fact move the state from x to y expecting the "final" outcome to be z. In other words, a non-myopic or *farsighted* approach to coalitional stability negates the logic underlying solution concepts such as the vNM stable set.

Following Harsanyi, there has been a large literature on solution concepts that are based on "farsighted" individuals who base their decisions on whether to deviate from the current status not on the immediate consequence of the deviation, but on how they will fare at the "final" outcome following further deviations by other coalitions. [3] A common feature in much of this literature is the absence of any extensive form specifying the order in which players or coalitions move as well as any pre-specified set of terminal states.

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

[3]See, for instance, Chwe (1994), Bloch (1996), Ray and Vohra (1997, 1999), Xue (1998), Diamantoudi and Xue (2003), Konishi and Ray (2003), Herings *et al.* (2004, 2009), Anesi (2010), Mauleon *et al.* (2011), Vartiainen (2011), Anesi and Seidmann (2014), Ray and Vohra (2015), Chander (2015), Kimya (2015), Dutta and Vohra (2016). Aumann and Myerson (1988) also modeled farsighted behavior, but from a different perspective. Ray and Vohra (2014) provide an insightful survey of this literature.

So, farsighted or forward looking behavior cannot be captured through the use of any reasoning analogous to backwards induction.

Clearly, this approach requires the specification of the "final" outcome of any sequence of coalitional deviations. Since pre-specified terminal outcomes do not exist in this approach, the final outcome must be one from which no coalition wants to deviate. This suggests that the final outcome is one which is "stable". Then, farsightedness essentially requires that a coalition compares the payoffs of its members at the current status quo to what it expects will be their payoffs at the stable outcome that will be reached if the coalition does deviate. But this implies that deciding on the stability of a particular outcome against a sequence of moves requires us to know which other outcomes are stable! This makes the notion of stability circular and suggests the use of a solution concept based on the principles of internal and external stability that underlie the original vNM stable set. Indeed, Harsanyi (1974) and much of the literature in this area after him have modified the stable set by allowing for sequences of coalitional moves, so that both internal and external stability are replaced by their farsighted counterparts.

Dutta and Vohra (2016) (henceforth DV) raise two issues with this approach. First, the Harsanyi stable set and other variants do not restrict coalitions to make optimal moves. That is, suppose x is the current status quo and coalition S is contemplating a deviation. Then, if S has two possible deviations, with one deviation Pareto-dominating the other, then it should not take the latter move. Moreover, all coalitions that have deviated before S should also assume that S will only take Pareto-undominated or *maximal* moves.[4] DV also point out that farsighted objections as typically modelled also permit coalitions to hold different beliefs about the continuation path of coalitional moves. That is, x may not be in the farsighted stable set because coalition S^1 replaces it with y, anticipating a second, and final, move to z. At the same time, another coalition S^2 may deviate from x' to y in the belief that the next (and final move) twill be to z' (not z). That is coalitions S^1 and S^2 hold different beliefs about the continuation from state y. DV refer to this issue as one of holding *consistent beliefs*, although they point out that such seemingly inconsistent beliefs may arise because coalitional moves are *history-dependent*.[5]

DV incorporated maximality and consistency (or history independence) of beliefs in the notion of farsighted stability. They use the tool of an *expectation function*, a concept borrowed from Jordan (2006). In this framework, the expectation function describes the transition from one state to another, as well as the coalition which is supposed to effect the move. Thus, the expectation function represented the *commonly held beliefs* of all agents about the sequence of coalitional moves, if any, from every state.[6] The use of a single expectation function immediately incorporates consistency.

[4] Examples 1 and 2 demonstrate this issue.

[5] Notice that in this example, the state y is reached along different histories of past coalitional moves.

[6] Although there is no extensive form in our framework, the imposition of commonly held beliefs about continuation paths is analogous to that of such beliefs in non-cooperative equilibria such as subgame

Importantly, DV assumes that the transition from any state x to another state y only depends on the current state. Together with a the expectation function, each state is then identified with a *stationary* outcome that is eventually reached from this state. Using this correspondence, DV define the notion of Maximality of an expectation: it is a move that a coalition cannot improve given the consequences of the deviation. DV defined two versions of Maximality, one demanding that the move is maximal for the active coalition and the other that the move is maximal for *any* relevant coalition. The latter condition implies strong robustness but also leads to existence problems. The sets of stationary points of an expectation function satisfying one or the other notion of maximality as well as farsighted versions of internal and external stability then gave two different solution concepts. DV showed that these solution concepts are very different from the ones defined earlier.

The point of departure in this paper is to incorporate *history dependence* into the DV framework. Formally, this extension implies that a coalitional move may depend on the past history of coalitional moves and not only on the current state. So, history dependence permits coalitions to remember which coalitions or individuals have been active and potentially condition their future behavior on the past experiences.

The dependence of coalitional moves on past history is intuitively appealing. For instance, we are more likely to join groups of individuals with whom we have had a pleasant experience in the past. Correspondingly, we are less likely to associate with individuals who have lost our trust. Allowing agents to have memory is also standard in non-cooperative games.

Notice that since history independence is a special case of history dependence, the DV solutions remain solutions in our framework. However, as is standard in the non-cooperative framework, the introduction of history dependence expands the sets of stable outcomes quite dramatically. In particular, it allows us to prove powerful non-emptyness results - we show that the set of stable outcomes is non-empty in all finite games as well as in all transferable utility partition function games. What is more, the latter result is derived under the strong Maximality property of an expectation, implying remarkable robustness of the solution. To the best of our knowledge, this result has no analogue in the cooperative game theory literature.

Apart from expectation functions, a key tool in the paper will be *objection paths*. An objection is a finite sequence of coalitional deviations starting from an initial state and ending up in a terminal state, with the property that each coalition in the sequence strictly prefers the terminal state to the state from which it is deviating. In other words, it represents a farsighted objection. We will characterise our solution concepts in terms of collections of such objection paths- the terminal states in the appropriate collection will

perfection. For an alternative approach, see Bloch and van den Nouweland (2017) who allow individuals to hold *different* beliefs about the path of future actions.

constitute a solution in our framework. While these are not "direct" characterisations since the necessary and sufficient conditions are not stated in terms of sets of states,[7] we show subsequently that even the "indirect" characterisation is remarkably useful - they are used extensively in the proofs of the nonemptyness results as well in yielding very transparent result on the structure of the solution(s).

The plan of the paper is the following. In the next section, we introduce some key concepts. In section 3, we describe formally the Ray-Vohra farsighted stable set and the largest consistent set of Chwe (1994), present examples to illustrate the importance of maximality and consistency, and then go on to introduce our solution concepts. Section 4 contains our main characterisation results in terms of objection paths, while section 5 contains the characterisation for simple games. An important bye-product of the analysis for simple games is that notions of maximality are rendered irrelevant, in a sense to be explained in section 5. In section 6, we discuss the structural properties of our solution concepts. We go on to present the nonemptyness results in section 7. Section 8 discusses the relationship of our solution concepts to the Ray-Vohra farsighted stable set and the largest consistent set. In particular, we show that our solutions are refinements of the largest consistent set. This is particularly interesting in view of the usual criticism of the largest consistent set as being too permissive.

2. The Background

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

The set of outcomes as well as the effectivity correspondence will depend on the specific model that is being studied. For instance, in a *partition function game*, (N, v), the function v will specify a real number for each *embedded coalition* (S, π) where π denotes the coalition structure with $S \in \pi$ being one of the coalitions in the partition π. Feasibility will imply that an embedded coalition (S, π) can distribute at most $v(S, \pi)$ to individuals in S. A state for partition function games will refer to a coalition structure π and a corresponding payoff allocation which is feasible and efficient for each embedded coalition corresponding to π. Much of traditional cooperative game theory has focused on the simpler but more restrictive transferable utility characteristic function games in which a

[7]We also provide an alternative characterisation in terms of sets of states for the special class of *simple games.*

coalition can assure itself of a minimum aggregate utility $v(S)$. The dominant tradition in the literature has treated the set of states to be the set of *imputations*, the Pareto efficient utility profiles in $v(N)$, implicitly assumed that $S \in E(x, y)$ iff $y_S \in v(S)$. Ray and Vohra (2015) provide a convincing critique of why this assumption is unsatisfactory for studying farsightedness. We return to this issue below.

State y *dominates* x if there is $S \in E(x, y)$ such that $u_S(y) \gg u_S(x)$.[8] In this case we also say that (S, y) is an *objection* to x.

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

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

- (*Internal stability*) For any $x \in K$, there is no $y \in K$ such that y dominates x,
- (*External stability*) For any $x \notin K$, there is $y \in K$ such that y dominates x.

The core and vNM stable set are myopic solution concepts since they are based on single rounds of deviations. In order to introduce farsighted solutions, it is convenient to introduce the concept of *objection paths*.

DEFINITION **1.** *An objection path is a finite sequence* $(y_0, S_1, y_1, \ldots, S_m, y_m)$*, such that, for all* $k = 1, \ldots m$*,* $S_k \in E(y_{k-1}, y_k)$ *and* $u_{S_k}(y_m) \gg u_{S_k}(y_{k-1})$.

Given the abstract game $(N, X, E, u_i(.))$, we denote the set of all objection paths by P^*. We will often use $P \subseteq P^*$ to denote a subset of objection paths, and P_x to denote the set of objection paths in P with initial element x. We will use p_x to denote a typical objection path in P_x, and $\mu(p)$ to denote the terminal state in the objection path p.

State y *farsightedly dominates* x if there is an objection path p_x such that $y = \mu(p_x)$.

Farsighted or indirect domination takes into account forward looking behaviour because at each point in the objection path, the deviating coalition takes into account the utility profile not at the next state in the sequence but at the “final ” state in the objection path. Of course, this leaves open the question of how the terminal state is determined. This is going to be a central issue of the paper.

The relation of dominance or farsighted dominance depends on the specification of the effectivity function. Ray and Vohra (2015) point out the importance of imposing appropriate restrictions on the effectivity function in the construction of farsighted solution concepts. In the context of characteristic games, the standard practice allowed a coalition S complete freedom to choose even the payoffs to individuals in the complementary coalition $N - S$. Notice that this does not matter for solution concepts like the core or the vNM stable set since these are based on myopic deviations - the deviating coalition simply compares its *own* payoff allocations at the current state and the state following

[8]We write $u_S(y) \gg u_S(x)$ if $u_i(y) > u_i(x)$, for all $i \in S$.

immediately after the deviation.[9] But why or how can coalition S dictate either the payoffs accruing to the complementary coalition or how $N - S$ organises itself after S deviates? Of course, this does matter even in characteristic function games since it may influence what coalitions form along the sequence. Ray and Vohra (2015) demonstrate that this assumption can significantly alter the nature of the farsighted version of the vNM stable set. They show that imposing reasonable restrictions on the effectivity correspondence results in a farsighted stable set that is very different from that of Harsanyi (1974).

We will impose the appropriate restrictions on the effectivity function when we apply our solution concept to partition function games and simple games later on.

3. RATIONAL EXPECTATIONS AND FARSIGHTED SOLUTION CONCEPTS

Virtually all farsighted solution concepts are either implicitly or explicitly based on notions of sequences of objections or paths as we have defined here. Suppose that the "current" state is x and coalition S is contemplating whether to deviate from x to y. In a farsighted solution concept, it has to look ahead to the terminal state of the sequence of deviations that will take place *after* y. Obviously, there can be many objection paths from y, and typically S itself has no control over which one will actually take place. The multiplicity of such paths has resulted in a multiplicity of different solution concepts. We illustrate this point by describing two such solution concepts below.

DEFINITION **2.** *A set $F \subseteq X$ is a farsighted stable set if:*

- *(Farsighted internal stability) For any $x \in F$, there is no $y \in F$ such that y farsightedly dominates x,*
- *(Farsighted external stability) For any $x \notin F$, there is $y \in F$ such that y farsightedly dominates x.*

The farsighted stable set is based on an optimistic view of the coalitions involved in a farsighted objection - a state is dominated if there exists *some* path that leads to a better outcome. Chwe (1994) proposed a farsighted solution concept based on conservative or pessimistic behavior; this is good at identifying states that cannot possibly be considered stable.

DEFINITION **3.** *A set $K \subseteq X$ is consistent if*

$$K = \left\{ x \in X : \begin{array}{l} \textit{for all } y \textit{ and } S \textit{ with } S \in E(x,y), \textit{ there is } z \in K \textit{ such that } z = y, \\ \textit{or } z \textit{ farsightedly dominates } y \textit{ and } u_S(z) \not\gg u_S(x) \end{array} \right\}.$$

[9]Note that this aspect of the effectivity function is important even for myopic solution concepts of partition function games since the deviating coalition has to "predict what coalition structure will prevail immediately after the deviation since its aggregate utility depends on what partition forms.

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 which contains all other consistent sets, and defines this to be the *largest consistent set* (LCS). The largest consistent set has also received a lot of attention in the literature.

DV point out that such implicit assumptions of optimistic or pessimistic expectations are unsatisfactory because they are ad hoc. Instead, coalitions should assume that all subsequent coalitions will behave in an *optimal* manner. The following examples from DV illustrate the importance of this point.

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

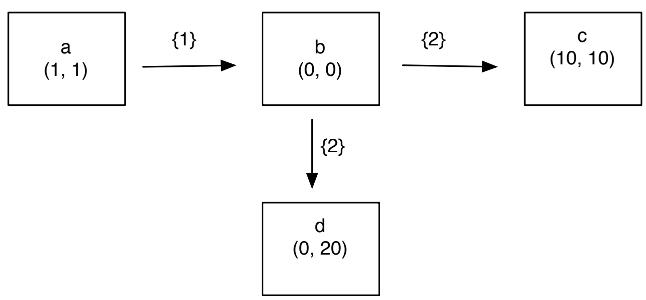

FIGURE 1

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

EXAMPLE **2.** *Figure 2 shows a modification of Example 1 as shown in Figure 2.*

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

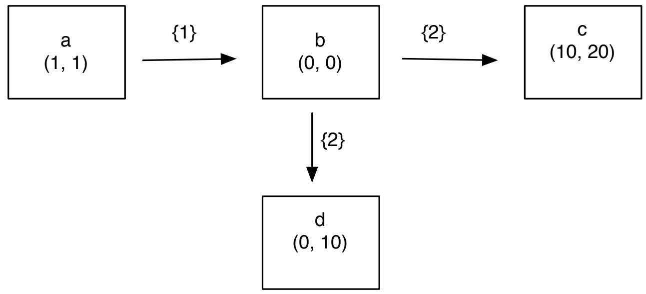

FIGURE 2

These examples show that both the LCS and the farsighted stable set suffer from the problem that they do not require coalitions (in these examples,$\{2\}$) to make moves that are *optimal* among all profitable moves.

As we have mentioned earlier, DV address this issue by using an *expectations function* to model the transition from one state to another, as well as the coalition which is supposed to effect the move. The use of an expectation function to represent the transition from one state to another is adapted from Jordan (2006) who used such a function to represent commonly held beliefs about the transition from any state to the final outcome. The expectation function represents the *commonly held beliefs of all agents* about the sequence of coalitional moves, if any, from every state. One can then choose to impose restrictions on the expectation function in order to make the function reasonable. An obvious restriction is that the expectation function must be consistent with the underlying game and hence with the effectivity function associated with the game - it cannot specify a move from state x to state y by coalition S if $S \notin E(x, y)$. Another restriction which is desirable can be that the expectation function specify moves that are optimal. We will describe below slightly different notions or degrees of optimality - each will give rise to a specific restriction on the expectation function.

DV assumed that the process of transition is *myopic* or *history independent* ; that is, if the expectation function specifies a transition from state x to state y, then it must do so irrespective of how state x is reached.[10] The essential purpose of this paper is to show how the DV analysis can be extended to incorporate *history dependence* into this transition process. Allowing for history dependence obviously results in a more general framework in which future coalitional moves can in principle depend on the evolution of past coalitional moves. There at least two reasons why this is an interesting exercise. We have mentioned earlier that there are a variety of contexts where history does matter. Moreover, from a purely formal perspective, it is well known that history dependence

[10]Note that, in our framework, we cannot interpret states as nodes of an extensive form game since a state can be reached along several different objection paths.

enlarges the set of noncooperative equilibria. In principle, this logic may carry forward to the present context. Indeed, we show in the next section that this is indeed true. We show that even in the case of a Condorcet cycle over three states when virtually all solution concepts fail to provide a nonempty solution, our history dependent solution will pass the non-emptyness test.

With this in mind, we define *histories* more formally. Let x_0 be an initial status quo. At period $t = 0, 1, ...,$ coalition S can challenge the current state x_t by demanding an outcome x_{t+1} such that $S \in E(x_t, x_{t+1})$. We allow only *one* coalition to be active at any time, without describing any explicit protocol which chooses the active coalition. In such a case, x_{t+1} becomes the new status quo at period $t + 1$. If no coalition challenges some state x in period t, then the game terminates and x is implemented. A *history* is a sequence $(x_0, S_1, x_1, ..., S_m, x_m)$ that specifies the past play path and coalitions that have been active till x_m has been reached. Let H represent the set of all (finite) histories. Note that for any $h \in H$, (h, S, x) is also a history where the state x is induced by S after history h has been reached.

Expectation Function

An expectation is a function $F : H \to X \times \mathcal{N}$, specifying the active coalition and its move for all possible current states and past histories. Denote $F(h) = (S(h), f(h))$, where $f(h)$ is the state that is expected to follow at history h, and $S(h)$ is the coalition expected to induce the next state. If $S(h) = \emptyset$, then no coalition wants to change the state and the final state of the history h will be implemented. [11]

As usual, history independence is a special case of history dependence. Consider any two histories $h, h' \in H$, and any state $x \in X$. Then, the DV expectation function satisfied $F(h, x) = F(h', x)$ In other words, the continuation path only depends on the state x and not on whether the state was reached via history h or history h'.

Given an expectation $F = (S, f)$, let F^k denote the k-fold composition of F such that $F^0(h) = F(h)$ and $F^{k+1}(h) = F(h, F^k(h)))$ for all $k = 0, 1, 2, ...$. Similarly, denote by $S^k(h)$ and $f^k(h)$ the first and second components of $F^k(h)$, respectively, so that $F^k(h) = (S^k(h), f^k(h))$, for any k.

We say that history h is *stationary* if $S(h) = \emptyset$.

An expectation F is *absorbing* if, for every $h \in H$, there exists k such that $S^k(h) = \emptyset$.

For any history $h = (x_0, S_1, x_1, ..., S_k, x_k)$, we will use $\mu(h)$ to denote the terminal state of h. In this example, $\mu(h) = x_k$. Notice that *all* finite histories have well-defined terminal states. Of course, a stationary state is terminal, but not all terminal states are stationary.

[11]For history dependent solutions in related contexts, see Vartiainen (2011, 2014, and 2015).

A history $(h, S_1, y_1, ..., S_m, y_m)$ is a *farsighted objection to* h if

$$(\mu(h), S_1, y_1, \ldots, S_m, y_m) \in P_{\mu(h)}.$$

That is, the new history is formed from h by appending an objection path to it.

For an absorbing F, the path $\overline{F}(h)$ generated by F from history h, i.e.,

$$\overline{F}(h) = (F(h), F^1(h), F^2(h), ...)$$

has a finite length, and $\mu(\overline{F}(h))$ is well defined for any h.

Let $\overline{F}(H) = \cup_{h \in H}\{\overline{F}(h)\}$ denote the sets of possible paths that is generated by an absorbing F, by varying the initial history, and $\mu(\overline{F}(H))$ the stationary states associated with these paths. Hence, assuming that expectation F is played in the continuation game, $\mu(\overline{F}(H))$ is the set of states that can be eventually reached by starting from any initial history. So, it makes sense to view $\mu(\overline{F}(H))$ as a farsighted solution when F is the function describing the transition from state to state.

We now turn to the issue of describing "reasonable" restrictions on F keeping in mind that these translate into restrictions on $\mu(\overline{F}(H))$, the set of stationary points.

We first describe two restrictions on the expectation F that are the farsighted analogues of internal and external stability.

: (**I**) If h is a stationary history, then there does not exist $y \in X$ and $S \in E(\mu(h), y)$ such that $u_S(\mu(\overline{F}(h, S, y))) \gg u_S(\mu(h))$.
: (**E**) If h is a nonstationary history, then $(h, \overline{F}(h))$ is a farsighted objection to h.

If Condition I is not satisfied, then for some stationary state x, there is a coalition S which can deviate anticipating that the resulting sequence of transitions according to F will lead to another stationary state that all members of S prefer. Clearly, this is a violation of internal stability. Condition E states that if $\mu(h)$ is not a stationary state, then some farsighted objection will result in a stationary state - this is an obvious requirement of farsighted External Stability.

Notice that nothing has been said so far about the optimality of coalitional deviations involved in any farsighted objection implicit in Condition E. We now describe two different versions of optimality or maximality.

: (**M**) If h is a nonstationary history, then there does not exist $y \in X$ such that $S(h) \in E(\mu(h), y)$ and $u_{S(h)}(\mu(\overline{F}(h, S, y))) \gg u_{S(h)}(\mu(\overline{F}(h)))$.
: (**M***) If h is a nonstationary history, then there does not exist $y \in X$ and $S \in E(\mu(h), y)$ such that $S(h) \cap S \neq \emptyset$ and $u_S(\mu(\overline{F}(h, S, y))) \gg u_S(\mu(\overline{F}(h)))$.

Maximality assumes that at a nonstationary history x, some coalition $S(h)$ is the coalition that has the floor. Then, $S(h)$ should not be able to deviate to another path that all $i \in S(x)$ prefer. Condition M* (Strong Maximality) is stronger. This allows for the possibility that more than one coalition may be able to move at state x. For instance, there may be some coalition S such that $i \in S \cap S(h)$, $y \in X$ with $S \in E(\mu(h), y$ and $u_i(\mu_i(\overline{F}(h, S, y))) > u_i(\mu(\overline{F}(h)))$. Then, a "rational" i should join coalition S instead of $S(h)$. Condition M* precludes this possibility.

We will say that history dependent and absorbing expectation F is *(strongly) rational,* abbreviated HRE (resp. HSRE), if it satisfies Properties I, E, and M (resp. M*).

Our farsighted solution concepts are defined below.

DEFINITION **4.** *The set of stationary points,* $\mu(\overline{F}(H))$ *of an HRE (HSRE)* F *is* history dependent (strongly) rational expectation farsighted stable set, *abbreviated HREFS (HSREFS).*

Of course, every HSRE is a HRE, and hence a HSREFS is a HREFS. But the converse is not true.

REMARK **1.** *DV defined Conditions I, E, M and M* for expectation functions which are history independent. They called their solution concepts REFS and SREFS. Of course, any set which is REFS is also HREFS and similarly any SREFS is HSREFS.*

4. CHARACTERIZATION

In this section, we provide characterization results for HREFS and HSREFS of abstract games. Our characterization exercises are not directly in terms of sets of *states*, but in terms of the terminal states of sets of objection paths. That is, we provide necessary and sufficient conditions so that the terminal states corresponding to any set P of objection paths will be HREFS (or HSREFS) iff P satisfies these conditions.

While we are aware that it may be difficult to check whether a specific subset of states satisfies the necessary and sufficient condition, it is very handy in proving general nonemptiness results - we provide constructive proofs of nonempty HREFS in all finite abstract games as well as a nonempty HSREFS in all superadditive partition function games. The characterization results also throws light on the logical structure of sets of HREFS, including the fact that a *largest* HREFS exists. Finally, the characterization is employed when we analyze the relationship of HREFS and HSREFS to other solution concepts.

DEFINITION **5.** *Let* P *be a collection of objection paths.*

- *An objection path* $p = (x_0, S_1, x_1, ...)$ *is* S_1-dominated *in* P *via* y *if* $S_1 \in E(x_0, y)$ *and* $u_{S_1}(\mu(p_y)) \gg u_{S_1}(\mu(p_{x_0}))$, *for all* $p_y \in P$.
- *An objection path* (x) *is* S-dominated *in* P *via* y *if* $S \in E(x, y)$ *and* $u_S(\mu(p_y)) \gg u_S(x)$ *for all* $p_y \in P$.

That is, an objection path p is dominated via node y in the set P of paths if the members of the *first* active coalition profit by directing the play to node y rather than continuing along the path p to the terminal state. Notice that the definition requires that once S_1 deviates to y, it takes into account the possibility that *any* objection path in P with y as the original state may be followed in future. Clearly, if this condition is satisfied, and S_1 believes that only the set of paths P are "possible" paths that can be followed, then it cannot be optimal for S_1 to deviate to x_1. Part (ii) stipulates that if x is not followed by any other state, then any coalition can dominate it via some y if an analogous condition is satisfied.

DEFINITION **6.** *A collection of objection paths* P *is* coherent *if:*

(1) *The set* P_x *is nonempty, for all* $x \in X$,
(2) *If* $(x_0, S_1, ...) \in P$, *then* $(x_k, S_{k+1}, ...) \in P$, *for all* $k = 0, 1, ...$,
(3) *If* $(x_0, S_1, ...) \in P$, *then* $(x_0, S_1, ...)$ *is not* S_1-*dominated in* P *(via any* y*),*
(4) *If* $(x) \in P$, *then* (x) *is not* $S-$*dominated in* P *(via any* y*), by any* S.

REMARK **2.** *Suppose* $\mu(p) = x$ *for some* $p \in P$, *where* P *is a coherent collection of paths. Then, by part (2) of Definition 6,* $(x) \in P$.

Our first theorem shows that any HREFS must be the set of terminal states of a coherent collection of objection paths.

THEOREM **1.** *A set* $Y \subseteq X$ *is HREFS if and only if* $Y \equiv \mu(P)$ *for some coherent collection of objection paths* P.

The proof of the theorem will follow from two lemmas.

LEMMA **1.** *Let* F *be a history dependent, absorbing expectation function satisfying conditions I, E and M. Then,* $\overline{F}(H)$ *is a coherent collection of objection paths.*

Proof. Since F is absorbing, $\overline{F}$ consists of finitely long paths. Moreover, for any non-stationary configuration (h), $\overline{F}(h)$ is an objection path by Property E.

We now check the defining conditions of a coherent collection of paths. Take any history h such that $\mu(h) = x$.

First, if $S(h) = \emptyset$, then $\overline{F}(h) = (x)$. If $S(h) \neq \emptyset$, then, by Property E, there is $S \in E(x, y)$ such that $u_S(\mu(\overline{F}(h, S, y))) \gg u_S(x)$. By construction, $\overline{F}(h, S, y) \in \overline{F}(H)$. Thus, in all cases, $\overline{F}(h) \in \overline{F}(H)_x$ for all $x \in X$.

Second, since $\overline{F}(h) = (F(h), \overline{F}(h, F(h)))$, and $(F(h), \overline{F}(h, F(h)) \in \overline{F}(H)$ it follows by induction that if $(x_0, S_1, x_1, ...) \in \overline{F}(H)$, then $(x_k, S_{k+1}, ...) \in \overline{F}(H)$ for all $k = 0, 1, ...$

Next, suppose that $\overline{F}(h)$ is $S(h)$-dominated in $\overline{F}(H)$ via y. Then $u_{S(h)}(\mu(\overline{F}(h, S(h), y))) \gg u_{S(h)}(\mu(\overline{F}(h)))$. But this violates Property M.

Finally, suppose that $\overline{F}(h)$ is S-dominated in $\overline{F}(H)$ via y. Then $u_S(\mu(\overline{F}(h, S, y))) \gg u_S(\mu(h))$ for some S such that $S \in E(\mu(h), y)$. But this violates Property I.

This shows that $\overline{F}(H)$ satisfies all the four requirements defining a coherent set of objection paths. ■

We now want to prove the converse result; if P is a coherent collection of objection paths, then the terminal sates associated with P is HREFS. The proof of the claim is constructive - given any coherent set P, we specify an absorbing expectations function satisfying Properties I, E and M. We now illustrate the construction in the canonical "difficult" case - the Condorcet cycle involving three states and three individuals.

An Example

Let $N = \{1, 2, 3\}$, and $X = \{1, 2, 3\}$. The utility profile is

- $u_1(x) > u_1(y) > u_1(z)$.
- $u_2(y) > u_2(z) > u_2(x)$
- $u_3(z) > u_3(x) > u_3(y)$

Any two individuals constitutes a majority. The effectivity function E specifies that for all $w, w' \in X$, $S \in E(w, w')$ iff $|S| \geq 2$. The domination relation is cyclic, and virtually all solutions (REFS, farsighted stable set, vNM set, etc) are empty. However, we show below that X itself is HREFS.

Identify a coherent set of objections paths

(1) $$P = \{(x), (y), (z), (x, \{2, 3\}, z), (y, \{1, 3\}, x), (z, \{1, 2\}, y)\}.$$

It is trivial to check that Conditions 1-3 of Coherence are satisfied. To check 4, it suffices (given the symmetry in the example) to check that x is not S-dominated in P for any S via any $w \in X$. The only possible S is $\{2, 3\}$ and $w = z$. But, then if S moves to z, there is $p_z = (z, \{1, 2\}, y) \in P$ and $u_3(x) > u_3(y)$.

Partition the set of histories H into six *phases* $H_{p\in P}$ recursively. Each phase contains all the information required for construction of F. That is, if $h, h' \in H_p$ for any $p \in P$, then the constructed F will have the property that $F^k(h) = F^k(h')$ for all $k = 1, \ldots$.

The initial step in the recursion is: choose $w \in X$ and let $w \in H_{(w)}$. The inductive step is: for any $h \in H$,

(1) if $h \in H_{(x)}$, then $(h, \{2,3\}, z) \in H_{(z,\{1,2\},y)}$ and $(h, S, w) \in H_{(w)}$ for all $(S, w) \neq (\{2,3\}, z)$,
(2) if $h \in H_{(y)}$, then $(h, \{1,3\}, x) \in H_{(x,\{2,3\},z)}$, and $(h, S, w) \in H_{(w)}$ for all $(S, w) \neq (\{1,3\}, z)$,
(3) if $h \in H_{(z)}$, then $(h, \{1,2\}, y) \in H_{(y,\{1,3\},x)}$, and $(h, S, w) \in H_{(w)}$ for all $(S, w) \neq (\{1,2\}, x)$,
(4) if $h \in H_{(x,\{2,3\},z)}$, then $(h, \{2,3\}, x) \in H_{(x)}$, $(h, \{2,3\}, y) \in H_{(y)}$, $(h, \{2,3\}, z) \in H_{(z)}$,
(5) if $h \in H_{(y,\{1,3\},x)}$, then $(h, \{1,3\}, x) \in H_{(x)}$, $(h, \{1,3\}, y) \in H_{(y)}$, $(h, \{1,3\}, z) \in H_{(z)}$,
(6) if $h \in H_{(z,\{1,2\},y)}$, then $(h, \{1,2\}, x) \in H_{(x)}$, $(h, \{1,2\}, y) \in H_{(y)}$, $(h, \{1,2\}, z) \in H_{(z)}$.

Notice that $h \in H_{(x)} \cup H_{(x,\{2,3\},z)}$ implies $\mu(h) = x$, and so on.

Construct an expectation F^P such that, for any $h \in H$,

(1) if $h \in H_{(w)}$, then $F^P(h) = (\emptyset, w)$, for all $w \in X$,
(2) if $h \in H_{(x,\{2,3\},z)}$, then $F^P(h) = (\{2,3\}, z)$,
(3) if $h \in H_{(y,\{1,3\},x)}$, then $F^P(h) = (\{1,3\}, x)$,
(4) if $h \in H_{(z,\{1,2\},y)}$, then $F^P(h) = (\{1,2\}, y)$.

The intuitive reason why this construction works is the following. Histories in each H_w are designed to be stationary. Consider any deviation from say $h \in H_x$. The deviation by $\{2,3\}$ to z leads to the non-stationary history $(h, \{2,3\}, z) \in H_{z,\{1,2\},y}$. But, (x) is not $\{2,3\}$-dominated in P. And this allows $\{2,3\}$ to be "punished" since $F^P(h, \{2,3\}, z) = (\{1,2\}, y)$ resulting in the terminal state y.

We now proceed to a more formal description of why the construction provides an HREFS.

First, that F is absorbing follows from the steps below.

- if $h \in H_{(w)}$, then $F^P(h) = (\emptyset, w)$, for all $w \in X$.
- if $h \in H_{(x,\{2,3\},z)}$, then $F^P(h, F^P(h)) = F^2(h) = (\emptyset, z)$,
- if $h \in H_{(y,\{1,3\},x)}$, then $F^P(h, F^P(h)) = F^2(h) = (\emptyset, x)$,

- if $h \in H_{(z,\{1,2\},y)}$, then $F^P(h, F^P(h)) = F^2(h) = (\emptyset, y)$.

To check ***Property I***, suppose that h is a terminal history with $h \in H_{(x)}$. By construction, $(h, \{2,3\}, z) \in H_{(z,\{1,2\},y)}$ and $(h, S, w) \in H_{(w)}$ for all $(S, w) \neq (\{2,3\}, z)$. Thus $F^P(h, \{2,3\}, z) = (\{1,2\}, y)$ and $F^P(h, S, w) = (\emptyset, w)$. Hence, $\mu[\overline{F}^P(h, \{2,3\}, z)) = y$ and $\mu[\overline{F}^P(h, S, w)) = w$ for all $(S, w) \neq (\{2,3\}, z)$. Thus a deviation by {2,3} to z is not profitable for agent 3. No other deviation is profitable for the deviating coalition either.

To check ***Property E***, suppose that h is a nonterminal history, with $h \in H_{(x,\{2,3\},z)}$. By construction, $\overline{F}^P(h) = (h, \{2,3\}, z)$. Since $\mu[h) = x$, this is a farsighted objection to h.

Finally, consider ***Property M***. Suppose that h is a nonterminal history. Then, say, $h \in H_{(x,\{2,3\},z)}$. By construction, $(h, \{2,3\}, w) \in H_{(w)}$, for all $w \in X$. Thus $F^P(h, \{2,3\}, w) = (\emptyset, w)$, and $\overline{F}((h, \{2,3\}, w) = w$ for all $w \in X$. No deviation is profitable for $\{2,3\}$ relative to the proposed choice z.

The $HREFS$ associated to F^P is $\{x, y, z\}$.

Now we turn back to the general case. The next lemma builds on the logic of the previous example.

LEMMA **2.** *If P is any coherent collection of objection paths, then $\mu(P)$ is HREFS.*

Proof. Fix a coherent collection of objection paths P for the rest of the proof.

Take any path $p_x = (x, S_1, \dots) \in P$ and pair (S, y) such that $S \in E(x, y)$ with $S = S_1$ if $p_x \neq (x)$. Define a function ξ with the property that $\xi(p_x, (S, y)) \in P$ and $u_S(\mu(\xi(p_x, (S, y)))) \not\gg u_S(\mu(p_x))$ Such a function ξ must exist for each such $(p_x, (S, y))$ from Conditions 3 and 4 of Definition 6.

Given a coherent collection of objection paths P, we now construct a history dependent and absorbing expectation function F^P such that $\mu(\overline{F}(H)) = \mu(P)$.

Interpret P as an index set and let $\{H_p\}_{p \in P}$ be a partition of the set of histories H. We construct F^P that is measurable with respect to this partition so that for each $p \in P$, and histories $h, h' \in H_p$, $F(h) = F(h')$. So, each element H_p of the partition of H contains all the relevant information concerning the past coalition actions.

We specify the partition of H recursively. For each $x \in \mu(P)$, from Remark 2, we know that $(x) \in P$. For each such x, let $(x) \in H_{(x)}$. Recursively, take any $p_{x_0} = (x_0, S_1, x_1, \dots) \in P$ and $h \in H_{p_{x_0}}$. Let $S \in E(x_0, y)$ be such that $S = S_1$ if $S_1 \neq \emptyset$, and let

$$(2) \qquad (h,S,y) \in \begin{cases} H_{(x_1,S_2,...)}, & \text{if } (S,y) = (S_1,x_1), \\ H_{\xi(p_{x_0},(S,y))}, & \text{if } (S,y) \neq (S_1,x_1). \end{cases}$$

Proceeding from the initial history $\emptyset$, each element in the set of histories H is allocated into exactly one element of $\{H_{(x_0,S_1,...)}\}_{(x_0,S_1,...)\in P}$. Note that if $h \in H_{(x_0,S_1,...)}$, then $\mu(h) = x_0$.

Construct now an expectation F^P such that, for any $h \in H_{(x_0,S_1,...)}$,

$$(3) \qquad F^P(h) = \begin{cases} (S_1,x_1), & \text{if } S_1 \neq \emptyset, \\ (\emptyset,x_0), & \text{if } S_1 = \emptyset. \end{cases}$$

First, we check that F^P is absorbing.

Take any $(x_0,S_1,...) \in P$ and any $h \in H_{(x_0,S_1,...)}$. Then $F^P(h) = F^1(h) = (S_1,x_1)$, $F^P(h,F^P(h)) = F^2(h) = (S_2,x_2)$, and so on. Thus F^P continues along the path $(x_0,S_1,x_1,S_2,...) \in P$ until a stationary state is reached. Since any objection path is finitely long, F is absorbing.

We now verify the three properties of a rational expectation.

Property I: Suppose that h is a terminal history. Then $h \in H_{(\mu(h))}$. Consider y such that $S \in E(\mu(h),y)$. Then, $(h,S,y) \in H_{\xi((\mu(h),(S,y))}$. By the construction of F^P, $(y,F^1(h,S,y),F^2(h,S,y),...) = \xi((\mu(h)),(S,y))$. By the definition of ξ, $u_S(\mu(\xi(\mu(h),y))) \not\gg u_S(\mu(h))$.

Property E: Suppose that h is a nonterminal history. Find the path $(x_0,S_1,...) \in P$ be such that $h \in H_{(x_0,S_1,...)}$. By the construction of F^P, $(F^1(h),F^2(h),...) = (S_1,x_1,S_2,...)$. Since $(x_0,S_1,x_1,S_2,...)$ is a finitely long objection path, the contination play leads a terminal history $(h,F^1(h),F^2(h),...) = (h,x_0,S_1,...)$, which is a farsighted objection to h.

Property M: Suppose that h is a nonterminal history. Find the path $(x_0,S_1,...) \in P$ be such that $h \in H_{(x_0,S_1,...)}$. Then $x_0 = \mu(h)$. Take any y such that $S_1 \in E(x_0,y)$. By the construction of F^P, $(y,F^1(h,S,y),F^2(h,S,y),...) = \xi((x_0,S_1,...),(S,y))$. By the definition of ξ, $u_{S_1}(\mu(\xi((x_0,S_1,...),(S,y)))) \not\gg u_{S_1}(\mu((x_0,S_1,...)))$.

This completes the proof of the lemma. ■

Lemmas 1 and 2 prove Theorem 1.

Of course, neither the theorem nor the lemmas throw any light on the existence of a coherent collection of paths, nor how such a set can be identified if it exists. The following example demonstrates that the rudimentary structure of the abstract game does not itself guarantee the existence of a coherent collection of paths, and hence a HREFS.

Consider a one agent $N = \{1\}$ decision problem with $X = (-1, 0)$, and where $\{1\} \in E(x, y)$ if and only if $y = x/2$. Let $u_1(x) = x$ for all $x \in X$. Now any (trivial) objection path (x) except (0) is dominated via $x/2$. Hence the only candidate for the HREFS is $\{0\}$. But there is no finite objection path that initiates from any x and ends in 0. Hence Condition 6.1 is violated by any collection of paths, and there cannot be any HREFS.

Our objective is to prove the existence of HREFS in a large and natural class of games. We will, in fact, provide a sufficient condition for a stronger version of the solution, HSREFS. To this end, we will define a stronger version of Coherence.

DEFINITION **7.** *A collection of objection paths P is* strongly coherent *if,*

(1) *P_x is nonempty, for all $x \in X$,*
(2) *If $(x_0, S_1, ...) \in P$, then $(x_k, S_{k+1}, ...) \in P$ for all $k = 0, 1, ...,$*
(3) *If $(x_0, S_1, ...) \in P$, then $(x_0, S_1, ...)$ is not $S-$dominated in P (via any y), for any S such that $S_1 \cap S \neq \emptyset$,*
(4) *If $(x) \in P$, then (x) is not $S-$dominated in P (via any y), for any S.*

So, strong coherence strengthens Condition 6.3, all other requirements being the same as for coherence.

THEOREM **2.** *If P is a strongly coherent collection of objection paths, then $\mu(P)$ is HSREFS.*

Proof. Let P be some strongly coherent collection of objection paths. We construct an HSRE F^P such that $\overline{F}(H) = P$.

Identify a function ξ that is defined for each pair $((x_0, S_1, ...), (S, y))$ such that $(x_0, S_1, ...) \in P$ and $S \in E(x_0, y)$ with $S_1 \cap S \neq \emptyset$ if $S_1 \neq \emptyset$. Then ξ is defined by the property that $\xi((x_0, S_1, ...), (S, y)) \in P_y$ and

$$\begin{aligned} u_S(\mu(\xi((x_0, S_1, ...), (S, y)))) \quad &\not\gg \quad u_S(\mu((x_0, S_1, ...))), \\ &\text{for any pair } ((x_0, S_1, ...)(S, y))). \end{aligned}$$

Since P satisfies Definition 7, such a function ξ does exist.

As before, interpret a Strong coherent path structure P as an index set and let $\{H_p\}_{p \in P}$ be a partition of the set of histories H. We construct F that is measurable with respect to this partition.

We specify the partition of H recursively. As before, let $(x, \emptyset) \in H_x$ for all $x \in \mu(P)$. For any history h, find $(x_0, S_1, ...) \in P$ such that $h \in H_{(x_0,S_1,...)}$. For any S and y such that $S \in E(x_0, y)$ and such that $S_1 \cap S \neq \emptyset$ if $S_1 \neq \emptyset$, let

$$(h, S, y) \in \begin{cases} H_{(x_1,S_2,...)}, & \text{if } (S, y) = (S_1, x_1), \\ H_{\xi((x_0,S_1,...),(S,y))}, & \text{if } (S, y) \neq (S_1, x_1). \end{cases} \tag{4}$$

Then, each element in the set of histories H is allocated into exactly one component of the partition $\{H_p\}_{p\in P}$. Note that, by construction, $\mu(h) = x_0$ for all $h \in H_{(x_0,S_1,...)}$.

Construct now an expectation F such that, for any $h \in H_{(x_0,S_1,...)}$,

$$F^P(h) = \begin{cases} (S_1, x_1), & \text{if } S_1 \neq \emptyset, \\ (\emptyset, x_0), & \text{if } S_1 = \emptyset. \end{cases} \tag{5}$$

It suffices to verify Property M* since the rest of the proof is identical to that of Lemma 2. Suppose that h is a nonstationary history. Find the path $(x_0, S_1, ...) \in P$ be such that $h \in H_{(x_0,S_1,...)}$. Then $x_0 = \mu(h)$. Take any S and y such that $S \in E(x_0, y)$ and such that $S_1 \cap S \neq \emptyset$. By the construction of F^P, $(y, F^1(h, S, y), F^2(h, S, y), ...) = \xi((x_0, S_1, ...),(S, y))$. By the definition of ξ, $u_S(\mu(\xi((x_0, S_1, ...),(S, y)))) \not\gg u_S(\mu((x_0, S_1, ...)))$. ■

We will use these characterisation theorems repeatedly in subsequent sections. In particular, we will use Theorem2 to construct nonempty HSREFS in all superadditive transferable utility partition games, as well as non-empty HREFS in all finite games.

5. SIMPLE GAMES

In this section, the focus is on the class of *NTU simple games*, which we formalise by a non-empty set $\mathcal{W}$ of *winning coalitions* and the set of states X. von Neumann and Morgenstern (1944) described simple games by a characteristic function v such that $v(S) = 1$ if $S \in \mathcal{W}$ and $v(S) = 0$ otherwise.[12] Of course, this assumes that utility is transferable and winning brings the same aggregate benefit to the winning coalition. We describe a wider class of games where, in particular, utility is not transferable. Consider for instance a legislature which has to choose whether to pass a bill along with a set of possible amendments. This is clearly one example of a context which is more suitably modelled as a non-transferable utility game.

[12] Farsightedness for this class of simple games was studied by both Ray and Vohra(2015) as well as DV(2016).

Our focus is on monotonic and proper simple games, such that:

(i) If $S \in \mathcal{W}$, and $S \subset T$, then $T \in \mathcal{W}$.[13]

(ii) If $S \in \mathcal{W}$, then $N - S \notin \mathcal{W}$ for all $S \subseteq N$.

Given $\mathcal{W}$, a coalition B is a *blocking* coalition if $N - B$ is not a winning coalition. Let $\mathcal{B}$ denote the set of blocking coalitions.

In a simple game, only winning and blocking coalitions have any power to change outcomes. This is captured by an appropriate description of the effectivity function.

Assumption 1. *For each $S \in \mathcal{W}$, there is a nonempty set $X^S \subset X$ such that for all $T \in \mathcal{W}, S \subset T$ implies that, for all $i \in S$ and $x \in X^T$, there exists $y \in X^S$ s.t. $u_i(y) \geq u_i(x)$.*

Notice that this restriction on the sets $\{X^S\}_{S \in \mathcal{W}}$ is trivially satisfied in the case of transferable utility simple games since $v(S) = v(T) = 1$ when the two are winning coalitions. It also makes eminent sense since winning coalitions have complete power to change outcomes. Furthermore,

$$X = \bigcup_{S \in \mathcal{W}} X^S \cup X^0$$

where X^0 is the set of states where no winning coalition has formed. With some abuse of notation, we will group together all states in X^0 and label the group a "zero state", to be denoted x^0. We will normalise utility functions so that $u_i(x^0) = 0$ for all $i \in N$. Moreover, we assume

Assumption 2. *For each $S \in \mathcal{W}$, there is $x \in X^S$ such that $u_i(x) > 0$ for all $i \in S$.*

We can now describe effectivity function.

Assumption 3. *The effectivity function E satisfies the following*

(1) *For all $S \in \mathcal{W}$, for all $x \in X$, $S \in E(x, y)$ iff $y \in X^S$.*
(2) *For all $B \in \mathcal{B}$, $B \in E(x, x^0)$ for all $x \in X$.*
(3) *For all $S, T \in \mathcal{W}$ if $S \subset T$, then for all $x \in X^T$, $(T - S) \in E(x, y)$ only if $u_i(y) \geq u_i(x)$ for all $i \in S$.*
(4) *For all $S \in \mathcal{W}$, for all $R \subset N - S$, for all $x \in X^S$, $T \notin E(x, y)$ if $x \neq y$.*

The first condition states that a winning coalition S has the power to change any "initial" state to any state in X^S, while the second part states that a blocking coalition can change any state to a zero state. The third condition states that if some members of a winning coalition deviate but the remaining members remain a (new) winning coalition, then no

[13] So, $N \in \mathcal{W}$.

member of the new winning coalition is worse-off. Finally, the fourth condition states that if a winning coalition has formed, then no subset of the complementary coalition has any power to change the outcome. Since a state also describes a partition of N, it may seem as if winning or blocking coalitions are also deciding on the partition of $N - S$. This would of course run foul of the Ray-Vohra critique on the appropriate specification of effectivity functions. However, the structure of simple games a implies that the only relevant characteristic of a partition is whether it contains a winning coalition, and so the specification above is consistent with the spirit of the Ray-Vohra critique.

In this setting, we derive a transparent necessary and sufficient condition for HSREFS and HREFS in terms of sets of outcomes rather than sets of objection paths. As we have mentioned earlier, the advantage of this more direct approach is that it is easier to check whether a given set of social states Y can be supported as a solution. The intuitive reason why it is possible to derive this direct characterisation is because of the special structure of simple games - the only coalitions with some power are winning coalitions, or blocking coalitions that have the power to prevent the complementary coalition from winning. Importantly, we are also able to show that this stark distribution of power implies that any absorbing expectation function satisfying Conditions I and E is an HSRE. That is, *neither version of maximality plays a role for NTU simple games in the presence of history dependence.*

For any $x \in X$ and $S \in \mathcal{N}$, denote $D_S(x) = \{y \in X : u_S(y) \gg u_S(x)\}$. Our characterization is in terms of the system of sets $\{D_S(x)\}_{S \in \mathcal{N}, x \in X}$.

DEFINITION **8.** *A set $Y \subseteq X$ satisfies Condition C if for any $y \in Y$, for any $S \in \mathcal{N}$, $z \in X^S$ such that $z \in (Y \cap D_S(y)) \cup (X^S - Y)$, there are $B \in \mathcal{B}$, $W \in \mathcal{W}$ and $x \in X^W$ such that $x \in Y \cap D_B(z) \cap (D_W(x^0) - D_S(y))$.*

REMARK **3.** *Note that this definition allows for the possibility that $B = W$. This will be the case if there is $T \in \mathcal{W}$ and $x \in Y \cap X^T \cap D_T(z) - D_S(y)$.*

The following Fact will be used in what follows.

Fact 1: For all $x \in X$, if there is an objection path $p_x \neq (x)$, with $\mu(p_x) = y$, then either there is a winning coalition S such that $p'_x = (x, S, y)$ or there is a pair $(B, S) \in \mathcal{B} \times \mathcal{W}$ such that $p''_x = (x, B, x^0, S, y)$.[14]

Our main result of this section follows.

[14]Since y is the terminal state of an objection path, some winning coalition must have formed and $u_i(y) > 0$ for every member of the winning coalition. Let S be this coalition. Suppose y_K is the penultimate state in the objection path p_x. If a winning coalition has formed in y_K, then $y_K = x$ and then S can move directly from x to y. Otherwise, y_K is a zero state and then S can move from x to y with the help of some blocking coalition.

THEOREM **3.** *In all proper simple games, the following statements are equivalent for any set $Y \subset X$.*

(1) $Y = \mu(\bar{F})$ *where F is an absorbing expectation function satisfying Conditions I and E.*
(2) Y *is HSREFS.*
(3) Y *satisfies Condition C.*

Proof. Since (2) obviously implies (1), it is sufficient to show that (3) implies (2) and (1) implies (3).

Step 1: We first show that (3) implies (2).

Suppose Y satisfies Condition C. Define a function ϕ such that, for any $y \in Y$, for any $z \in (X - Y) \cup (Y \cap D_S(y))$ and $S \in E(y, z)$,

$$\phi(y, S, z) \;=\; (B, x^0, W, x) \text{ s.t. } B \in \mathcal{B}, W \in \mathcal{W} \text{ and } x \in Y \cap X^W \cap D_B(z) \cap (D_W(x^0) - D_S(y)).$$

Since Y satisfies Condition C, such a function ϕ exists. By construction, $(z, \phi(y, S, z)) = (z, B, x^0, W, x)$ is an objection path, for any such specified (y, S, z).

We show that there is a strongly coherent collection of objection paths P with $\mu(P) = Y$.

Let $\bar{P} = \{(z, \phi(y, S, z)) : y \in Y, S \in \mathcal{N}, z \in X^S, z \in (X - Y) \cup (Y \cap D_S(y))\}$, and construct P by

$$P = \{(y)\}_{y \in Y} \cup \bar{P}.$$

We show that P satisfies parts 1-4 of Definition 7.

To check part 1 of the definition, note that $(y) \in P$ for each $y \in Y$. Take any $x \notin Y$. If $x \in X^0$, then $B \in E(y, x)$ for all $y \in Y, B \in \mathcal{B}$. So, choose some $y \in Y, B \in \mathcal{B}$, and note that $p_x = (x, \phi(y, B, x)) \in P$. If $x \notin X^0$, then $x \in X^S$ for some $S \in \mathcal{W}$. Then, again $p_x = (x, \phi(y, S, x)) \in P$.

Part 2 follows immediately.

To check part 4, consider any path $(y) \in P$. Take any $S \in \mathcal{N}$, and $z \in X^S$. If $z \notin (Y - D_S(y))$, then $p_z = (z, \phi(y, S, z)) \in P$ and $\mu(p_z) \in Y - D_S(y)$. So, (y) is not S-dominated in P via z. If $z \in (Y - D_S(y))$, then again (y) is not S-dominated in P since $(z) \in P$.

For part 3, consider any path $p_y \in P$. Identify $\mu(p_y) = x$, Take any $S \in \mathcal{N}$ and $z \in X^S$. From Assumption 3, $S \in E(x, z)$. Either $z \in Y - D_S(x)$ and $(z) \in Y$ or there is $p_z \in \bar{P} \subset P$. Noting that $\mu(p_z) \in Y - D_S(x)$, path p_y is not S-dominated in P via z.

Hence, P is indeed a strongly coherent collection of objection paths. It follows from Theorem 2 that Y is HSREFS if Y satisfies Condition C.

This completes the proof of Step 1.

We now prove the other implication.

Step 2: We now show that (1) implies (3).

Take any absorbing F satisfying Conditions I and E, and suppose $Y = \mu(\bar{F})$.

Take any $y \in Y$. Then, there must be a stationary history h such that $F(h) = y$. Take any $S \in \mathcal{N}$ and $z \in Y \cap D_S(y)$. Since F satisfies Property I, (h, S, z) is not stationary. So, using Fact 1, there is $p_z = (z, B, x^0, T, x)$ such that $x \in Y - D_S(y)$ and $(h, y, S, z, B, x^0, T, x)$ is stationary.

Next, suppose $z \in X^S - Y$. Since $z \notin Y$, (h, S, z) is not stationary. Then, Property E and Fact 1 again imply the existence of $p_z = (z, B, x^0, T, x)$ such that $x \in Y - D_S(y)$ and $(h, y, S, z, B, x^0, T, x)$ is stationary.

This shows that Condition C is satisfied.

■

Thus we conclude that in the class of simple games, any HREFS can also be supported by the more robust mode of coalitional behavior depicted by HSREFS. Conversely, a strengthening the solution from HREFS to HSREFS does not bring any cutting power.

6. Structure of HREFS

In this section, we describe some results on the structure of HREFS. We point out at the end of the section that analogous results also go through for HSREFS.

PROPOSITION **1.** *Let P^1 and P^2 be coherent collections of objection paths. Then, $P^1 \cup P^2$ is also a coherent collection of objection paths.*

Proof. Let P^1 and P^2 be coherent collections of objection paths. Let $\bar{P} = P^1 \cup P^2$. We show that $\bar{P}$ satisfies all the conditions specified in Definition 6.

Clearly, $\bar{P}_x$ is nonempty since P^1_x and P^2_x are both nonempty, implying Definition 6.1.

Take any $\bar{p} = (x_0, S_1, x_1, \ldots) \in \bar{P}$. Without loss of generality, $\bar{p} \in P^1$. Then, by Definition 6.2, $(x_k, S_{k+1}, \ldots) \in P^1 \subset \bar{P}$, for any k.

Finally, notice that if S does not dominate some p in a set P, then S does not dominate p in P' with $P \subset P'$. This shows that $\bar{P}$ satisfies Definitions 6.3 and 6.4

So, $\bar{P}$ is a coherent collection of paths. ■

The following is immediate.

COROLLARY **1.** *If Y^1 and Y^2 are both HREFS, then so is $Y^1 \cup Y^2$.*

The corollary suggests the possibility that there may be a "largest " coherent set.

The Ultimate Undominated Set

For any set of objection paths P, define

$ud(P) = \{(x_0, S_1, x_1, ...) \in P :$ for all k, $(x_k, S_{k+1}, x_{k+1}, ...)$ is not S_{k+1}-dominated in $P\}$.

LEMMA **3.** *Let $P \subseteq P'$. Then $ud(P) \subseteq ud(P')$.*

Proof. For any $(x_0, S_1, x_1, ...)$ and any $k = 0, 1, ...$, if $(x_k, S_{k+1}, x_{k+1}, ...)$ is covered in P' via y, then it is dominated in P via y. Conversely, if $(x_k, S_k, x_{k+1}...)$ is *not* dominated in P via any y and for any k, then it is not dominated in P' via any y and for any k. ■

Recall that P^* denotes the set of all objection paths. Define $UD^0 \equiv P^*$, and $UD^t \equiv ud(UD^{t-1})$, for all $t = 0, 1, 2, ...$. By Lemma 3, $UD^{t+1} \subseteq UD^t$. Denote by

$$UUD = \cap_t UD^t$$

the *ultimate undominated set* associated to the problem. So, the ultimate undominated set is the limit set, obtained by recursively eliminating dominated objection paths. Notice that if X is a finite set, then only finitely many elimination rounds are needed.

The next theorem and corollary provides a condition under which UUD is the largest coherent collection.

THEOREM **4.** *Let $P = UUD$. If P_x is nonempty for all x, then P is the largest coherent collection of paths and so $\mu(P)$ is the largest HREFS.*

Proof. It is clear that P satisfies Definition 6.2-4. So, if P_x is nonempty for all x, then P is a coherent collection of objection paths.

Let P be any other coherent collection of objection paths. We show by induction that $P \subseteq UD^t$, for all $t = 0, 1, ...$

It is clear that $P = ud(P)$ since no path is P is dominated because of Definitions 6.3-4.

By assumption $P \subseteq P^* = UC^0$. Let $P \subseteq UD^t$. Then, by Lemma 3, $P = ud(P) \subseteq ud(UD^t) = UD^{t+1}$. Hence UUD contains all coherent collections, and $\mu(P)$ is the largest HREFS. ■

It is straightforward to adapt the proofs of Proposition 1 and Theorem 4 to prove identical results for strong coherence and HSREFS, after appropriately modifying the definition of $ud(.)$.

7. Nonemptyness Results

In this section, we show that a non-empty HREFS exists both when the set of social states X is finite as well as in the case of transferable utility partition function games. In fact, we prove a stronger result in the latter case by constructing a non-empty HSREFS.

7.1. **The Finite Case.** Suppose X, the set of social states, is finite. Since we make no other assumptions about the abstract game, this covers a wide variety of cases such as hedonic games, social network games without monetary transfers, etc.

Given finiteness of X, the set of acyclic objection paths is finite. This implies that the ultimate undominated set is, at each elimination round t, non-empty and well defined. The difficult part is to show that that UD^t contains a path p_x with initial state x, for arbitrary $x \in X$, as required by Coherence. The proof of the next lemma, which does this, is relegated to the Appendix.

LEMMA **4.** *For all $\in X$, there is p_x such that $p_x \in UUD$.*

The proof of the next theorem follows immediately from Lemma 4 and Theorem 4.

THEOREM **5.** *If X is finite, there is a non-empty HREFS.*

7.2. **Non-empty HSREFS for Partition Function Games.** In this section, we prove an existence result for HSREFS for the large class of games represented by partition function games. In view of the demanding nature of HSREFS, this non-emptyness result demonstrates the power of history dependence.

Let Π be the set of all *partitions* of N. An *embedded coalition* is a pair (S, π) where $\pi \in \Pi$ and $S \in \pi$. With some abuse of notation, we will use $v(N)$ to denote the embedded coalition $v(N, \{N\})$.

A *TU partition function game* is a mapping v specifying a real number $v(S, \pi)$ for each embedded coalition (S, π). That is, $v(S, \pi)$ is the sum of utilities that coalition S can achieve if the partition π forms. This formulation allows for externalities -what S can get depends on the entire coalition structure.

For any coalition $S \subseteq N$, we let π_S denote a partition of S, while Π_S denotes the set of all partitions of S. Also, Π_{-S} is the set of all partitions of $N - S$, with typical element denoted π_{-S}.

For any π and $S, T \in \pi$, we use $\pi_{-S\cup T}$ to denote the partition of $N - S \cup T$ obtained from π. That is, $R \in \pi_{-S\cup T}$ iff $R \in \pi$ and $R \notin \{S, T\}$.

Henceforth, we assume that v satisfies:

Superadditivity : For all $\pi \in \Pi$, for all $S, T \in \pi$, $v(S, \pi) + v(T, \pi) \leq v(S \cup T, \{S \cup T, \pi_{-S\cup T}\}$

Note that superadditivity ensures that for all $\pi \in \Pi$, $v(N) \geq \sum_{S\in\pi} v(S, \pi)$

Throughout, we will also assume that the partition function v is 0-normalized so that $v(\{i\}, \pi) = 0$ for all $i \in N$ and all $\pi \in \Pi$ with $\{i\} \in \pi$.[15]

We now have to specify the effectivity function associated with a partition function v. Consider any initial state x. Suppose coalition S deviates from x. It makes sense to assume that the coalition S can choose to form *any* partition $\pi_S \in \Pi$. Of course, a specific partition - say $\{S\}$ itself- may give it the best *short-term* payoff given the partition of $N - S$ that will "result" if indeed S forms. But, it may be in the long-term interest of S to form a different partition, and we allow that to happen. Next, we have to consider the partition of $N - S$ that will result once S deviates and forms some π_S. Of course, S cannot influence what partition of $N - S$ will form and we need to allow for this possibility. Since the worth of each embedded coalition depends on the partition structure, it is notationally complicated to explicitly formalise the effectivity function. Fortunately, we need to consider only very specific deviations and so do not need to describe the effectivity function in full detail.

Let $x^0 \in X$ be the *zero state* such that $u_i(x^0) = 0$ for all i and $\pi(x^0) = \{\{1\}, \ldots, \{n\}\}$. That is, the partition formed in the zero state is one in which each element of the partition of N consists of a single individual, and all corresponding embedded coalitions get zero utility.[16]

We assume the following.

Assumption 4. *For all $i \in N$, $N - \{i\} \in E(x, x^0)$ for all $x \in X$.*

This is straightforward since $N - \{i\}$ can always decide to break up into singletons. We will use this assumption repeatedly in the proof of a crucial lemma.

DEFINITION **9.** *Player i is essential iff $v(N) > v(N - \{i\}, \{N - \{i\}, \{i\}\})$.*

So, player i is essential if she adds positive value added to coalition $N - \{i\}$. Let

$$Z = \{x \in X : \sum_{i\in N} u_i(x) = v(N), u_i(x) > 0, \text{ if } i \text{ is essential}\}.$$

[15] This is without loss of generality.

[16] The latter follows since v is 0–normalized.

LEMMA **5.** *For all* $(x, y, k) \in Z \times X \times N$*, there is* p_y *such that* $u_k(\mu(p_y)) \le u_k(x)$*.*

Proof. Choose any triple $(x, y, k) \in Z \times X \times N$. We consider two cases.

Case 1: $u_k(y) > 0$.

Since $y \in X$, superadditivity implies that $\sum_{i\in N} u_i(y) \le v(N)$. So, there is $y' \in X$ (possibly $y = y'$) such that $\sum_{i\in N} u_i(y') = v(N), u_i(y') \ge u_i(y)$ for all $i \in N$

Suppose $u_k(x) > 0$. Since $u_k(y') > 0$, this implies that there is $z \in X$ such that

$$\begin{aligned}\sum_{i\in N} u_i(z) &= v(N),\\ u_i(z) &> u_i(y'), \text{ for all } i \neq k,\\ u_k(x) &\ge u_k(z) > 0.\end{aligned}$$

Then, define $p_y = (y, N - \{k\}, x^0, N, z)$. Clearly, p_y satisfies all the requirements of the lemma.

Next, suppose $u_k(x) = 0$. Since $x \in Z$, i is not essential. So, $v(N - \{k\}, \{N - \{k\}, \{k\}\}) = v(N)$ Clearly, this allows us to choose $z \in X$ such that $z(\pi) = \{N - \{k\}, \{k\}\}$, $N - \{k\} \in E(y, z)$ and

$$\begin{aligned}u_i(z) &> u_i(y), \text{ for all } i \neq k,\\ \sum_{i\neq k} u_i(z) &= v(N - \{k\}, \{N - \{k\}, \{k\}\}) = v(N),\\ u_k(z) &= 0.\end{aligned}$$

Then, let $p_y = (y, N - \{k\}, z)$. Again, p_y satisfies the requirements of the lemma.

Case 2: $u_k(y) = 0$.

Suppose k is essential, so that $u_k(x) > 0$. Let $\{k\} \in E(y, w)$ where $\{k\} \in \pi(w)$. Then, $u_k(w) = 0$. Note that we do not make any other assumption about $\pi(w)$ or $u_i(w)$ for $i \neq k$.

Since k is essential, $\sum_{i\neq k} u_i(w) < v(N)$ Since $u_k(w) = 0$, we can choose $z \in X$ such that

$$\begin{aligned}\sum_{i\in N} u_i(z) &= v(N),\\ u_i(z) &> u_i(w), \text{ for all } i \in N,\\ u_k(x) &\ge u_k(w).\end{aligned}$$

Then, $p_y = (y, \{k\}, w, N, z)$ satisfies the requirements of the lemma.

Suppose k is not essential. If $y \in Z$, then $p_y = (y)$ satisfies the requirements of the lemma. If $y \notin Z$, then either

(i) $\sum_{i \in N} u_i(y) < v(N) = v(N - \{k\}, N - \{k\}, \{k\}\})$, or

(ii) $i \neq k$ is essential, but $u_i(y) = 0$.

If (i) holds, then let $p_y = (y, N - \{k\}, z)$ where $\sum_{i \neq k} u_i(z) = v(N - \{k\}, N - \{k\}, \{k\}) = v(N)$, and $u_i(z) > u_i(y)$ for all $i \neq k, u_k(z) = 0$ Clearly, such $z \in Z$ exists and so p_y satisfies the requirements of the lemma.

If (ii) holds, then let i be essential, and $u_i(y) = 0$. Then, let $\{i\} \in E(y, w)$ where $\{i\} \in \pi(w)$. Using the fact that $\sum_{j \neq i} u_j(w) < v(N)$, we can choose $p_y = (y, \{i\}, w, N - \{k\}, z)$ such that

$$\begin{aligned}
\sum_{j \neq k} u_j(z) &= v(N - \{k\}, \{N - \{k\}, \{k\}\}) = v(N), \text{ (since } k \text{ is not essential)} \\
u_j(z) &> u_j(w), \text{ for all } j \neq k, \\
u_k(z) &= 0.
\end{aligned}$$

This completes the proof of the lemma. ■

Let P^Z is the collection of objection paths terminating in Z:

$$P^Z = \{p \in P^* : \mu(p) \in Z\}$$

We will prove that Z is HSREFS by showing that P^Z constitutes a strongly coherent collection of objection paths.

THEOREM **6.** *Z is an HSREFS.*

Proof. Take any $y \in X$. Choose arbitrary $x \in Z$ and $k \in N$. Lemma 5 implies that there is $p_y \in P^Z$ such that $u_k(\mu(p_y)) \leq u_k(x)$. Hence, $P_y \cap P^Z$ is nonempty and Condition 1 is satisfied.

For any objection path in P^Z, a subpath that begins from a state in the middle is also an objection path of blocking coalitions with a terminal element in Z, and hence a member of P^Z. That is, Condition 2 is satisfied.

Next, take any $p_z \in P^Z$ with $x = \mu(p_z)$. Suppose that p_z is $S-$covered via y for some S. Choose some $k \in S$. By Lemma 5, there is an objection path $p_y \in P^Z$ such that $u_k(\mu(p_y)) \leq u_k(x)$, contradicting the assumption that p_z is $S-$covered via y. Hence, Condition 3 is satisfied.

Now, take any $(z) \in P^Z$. Suppose that (z) is $S-$covered via y for some S. Choose some $k \in S$. By Lemma 5, there is an objection path $p_y \in P^Z$ such that $u_k(\mu(p_y)) \leq$

$u_k(x)$, contradicting the assumption that (z) is $S-$covered via y. So, Condition 4 is also satisfied and so P^Z is indeed strongly coherent.

This shows that Z is HSREFS. ■

HSREFS need not be unique. We leave it to the reader to check that

$$W = \{w \in X : \sum_{i \in N} u_i(x) \leq v(N), u_i(x) > 0, \text{if } i \text{ is essntial}\}$$

is also HSREFS. Of course, $Z \subseteq W$.[17]

Thus we conclude that, the in class of simple games, any HREFS can also be supported by the more robust mode of coalitional behavior depicted by HSRE. Conversely, by stregthening the solution from HREFS to HSREFS does not bring any cutting power. Also, the solutions are identifiable by objections paths of length 2 (at most).

8. Relationship to Other Solution Concepts

The core does have a strong stability property in the sense that all myopic solution concepts will typically contain the core. The stark difference between myopic and far-sighted concepts is brought out by the next example which shows that there may be abstract games where the core is *disjoint* from HREFS.

EXAMPLE **3.** *Let* $N = \{1, 2, 3\}, X = \{a, b, c, d\}$ *be the set of social states, where each state is a utility vector* (v_1, v_2, v_3) *showing the utility derived by each individual* i.

$a = (1.5, 2, 2), b = (1, 3, 2), c = (2, 1, 3), d = (3, 2, 1)$ *The effectivity relations are* $1 \in E(a, b), \{1, 3\} \in E(b, c), \{1, 2\} \in E(c, d), \{2, 3\} \in E(d, b)$ *Then, only individual* 1 *has the power to change the state* a *and she can only replace it with* b*. But she is worse-off doing so. Hence,* a *must be in the core. In fact, the core is* $\{a\}$ *since the other three alternatives are involved in a cycle.*

One HREFS is $\{b, c, d\}$*, disjoint from the core, However, there is a second HREFS which is the entire set.*

We now show that HREFS is a refinement of consistent sets. Even the largest (in terms of set inclusion) HREFS is a subset of the largest consistent set. Since the usual criticism of the LCS is that it is too permissive this makes HREFS a more attractive solution concept.

PROPOSITION **2.** *If* P *is a coherent collection paths, then* $\mu(P)$ *is a consistent set.*

[17] The proof that W is HSREFS is almost identical.

Proof. Suppose that $\mu(P)$ is not a consistent set.

Then there is an $x \in \mu(P)$, a y and an $S \in E(x, y)$ such that $u_S(z) > u_S(x)$, for all objection paths $(z_0, S_1, ..., S_m, z_m)$ with $z_0 = y$ and $z_m = z \in \mu(P)$.

But since P is a subset of all objection paths, this contradicts the assumption that (x) is not $S-$covered in P' via y. ■

The Farsighted Stable Set

The farsighted stable set (Definition 2) is not necessarily be HREFS in abstract games since domination chains may violate maximality.

However, as we have demonstrated in Section 5, the problem of maximality disappears in simple games. This essentially yields the following.[18]

PROPOSITION **3.** *If V is a farsighted stable set in a simple game, then V is HREFS.*

It is trivial that V must satisfy Conditions I and E. So, this result follows from our characterization result on simple games.

Covering Sets and Tournament Relations

In this subsection, we describe HSREFS for what have sometimes been called *tournament games* in the literature on voting games.[19] This class of games corresponds to the following situation. There is a finite set of alternatives A, and individuals in N have *strict* (ordinal) preferences over A, so that for all distinct $a, b \in A$, $u_i(a) \neq u_i(b)$. Moreover, the distribution of power is derived from a *strong* proper simple game :

$$\text{for all } S \subset N, S \in \mathcal{W} \text{ or } N - S \in \mathcal{W}$$

That is, a coalition is either winning or its complement in N is winning. Notice that $\mathcal{B} \equiv \mathcal{W}$ in strong proper simple games. An example of such games would be the majority voting game with an odd number of voters where any strict majority of votes constitutes a winning coalition.

We will (slightly) abuse terminology by identifying A with X, although a *state* should strictly specify an alternative from A along with a partition π of N. This will not create any confusion in what follows.

[18]See Ray and Vohra (2017) for a related result.

[19]See Laslier(1997) for an illuminating account of the literature on solution concepts for tournament games.

Denote by $M \subset X \times X$ the resulting *tournament* (total, antisymmetric) relation such that for all $x, y \in X$,

$$xMy \text{ iff } \{i \in N | u_i(x) > u_i(y)\} \in \mathcal{W}$$

Note that M is total because individuals are never indifferent between distinct alternatives and we are restricting attention to strong proper simple games.

Fix the underlying strong proper simple game and utility profile u. This determines M. Let

$$\text{for all } x \in X, M(x) = \{y \in X | yMx\}$$

Consider any subset Y of X, and $x, y \in Y$. Then, x *covers* y in Y if $M(x) \cap Y \subseteq M(y)$. The uncovered set of Y, denoted $uc(Y)$, is the set of alternatives in Y that are not covered in Y.

Furthermore, a subset of alternatives Y is a *covering set* (Dutta 1988) if it satisfies the following property:

$$\text{for all } y \in X, uc(Y \cup \{y\}) = Y$$

Our principal result is to establish that any covering set can be supported as stable states of a strong history dependent expectation function.

PROPOSITION **4.** *Let Z be a covering set. Then Z is HSREFS.*

Proof. We will construct a strongly coherent collection of paths P such that $\mu(P) = Z$.

Clearly, if Z contains only a single alternative, say x, then xMy for all $y \neq x$, and the proposition follows. Henceforth, assume that Z contains more than one alternative.

*Claim 1***:** For any $x \in Z$ and $y \in X - \{x\}$ there is $z \in Z$ such that $z \in M(y) - M(x)$.

*Proof***:** Fix $x \in Z$ and $y \in X - \{x\}$. If $y \notin M(x)$, then $x \in M(y)$ and so $z = x$ qualifies as such an element.

Thus let $y \in M(x)$. If $y \in Z$, then, since $x \in uc(Z \cup \{y\})$, it follows that there is $z \in Z$ such that zMy and xMz. If $y \notin Z$, then, since Z is a covering set, there is $z \in Z$ such that $M(z) \cap Z \subseteq M(y)$. Since $z \in M(y)$, we are done if $z \notin M(x)$. Suppose that $z \in M(x)$. Since $x, z \in Z$, and Z is a covering set, there is $w \in Z$ such that wMz and xMw. Since $M(z) \cap Z \subseteq M(y)$, $w \in M(y)$. So, $w \in M(y) - M(x)$. $\square$

Construct a set of paths P as follows:

$$P = \{(z) : z \in Z\} \cup \{(y, S, z) : S \in \mathcal{W},\ u_S(z) \gg u_S(y),\ y \in X, \text{ and } z \in Z\}.$$

Then $\mu(P) = Z$. We argue that P is a strongly coherent collection of objection paths.

Claim 2: P meets properties 1-4 of of Definition 7.

It suffices to check properties 1, 3, and 4.

Property 1: P_y is nonempty for any $y \in X$ since, by Claim 1 there is, for any $x \in Z$, $(y, S, z) \in P$ such that $z \in Z \cap M(y) - M(x)$.

Property 3: Let $(x, S, y) \in P$. Take any $z \in X$. By Claim 1, there is $w \in Z$ such that $w \in M(z) - M(y)$. Since $(z, T, w) \in P$ for some winning coalition T, and $w \notin M(y)$, there is no $S' \in \mathcal{W}$ such that (x, S, y) is $S'-$dominated in P.

Property 4: Let $(y) \in P$. Take any $z \in X$. Again, by Claim 1, there is $w \in Z$ such that $w \in M(z) - M(y)$. Since $(z, T, w) \in P$ for some majority T, and $w \notin M(y)$, there is no majority S' such that (y) is $S-$dominated in P. $\square$ ∎

While the converse is not true, we now show that the maximal HSREFS is also a covering set.

Let $X = UC^0$ and $UC^{t+1} = uc(UC^t)$, for all $t = 0, 1, 2, \ldots$. Since the covering relation is transitive, and the set of states is finite, UC^t is nonempty and $UC^{t+1} \subseteq UC^t$ for all t. Denote by $UUC = \cap_t UC^t$ the *ultimate uncovered set*.

From Dutta (1988), it is known that UUC is a covering set. Thus, by Proposition 4, UUC is HSREFS. The next result shows that UUC is the largest (in terms of set inclusion) HSREFS.

PROPOSITION **5.** *Let Z be HSREFS. Then $Z \subseteq UUC$.*

Proof. Clearly $Z \subseteq UC^0$. Suppose that $Z \subseteq UC^t$ but $y \in Z - uc(UC^t)$, for some t. Then y is covered in $UC^t \cup \{y\}$ by some $x \in UC^t$. Then, $M(x) \cap Z \subseteq M(y)$. But then (y) is $S-$dominated by some winning coalition S via x, contradicting the hypothesis that Z is a HSREFS. ∎

9. CONCLUDING REMARKS

This paper studies the consequences of memory on coalition formation. To this end, we extend the rational expectation stable set solution of Dutta and Vohra (2016) by allowing coalitions to condition their behavior on the *history* of blockings. The resulting solution satisfies the same stringent stability properties as the Dutta-Vohra solution but has an extra degree of freedom because of history dependence.

History dependence turns out to have very powerful implications. We show that a history dependent rational expectation solution exists under very general conditions, for example whenever the set of states is finite. What is more, we demonstrate that even the more stringent version of the solution, which requires that the current coalitional move is optimal also for non-active coalitions, exists and is nonempty in all partition function games (containing, for example, all TU-games). We are not aware of prior existence results in the literature with similar robustness and existence properties. Our results suggests that the introduction of history dependence in the study of coalition formation is a fruitful avenue for further research.

10. APPENDIX

In this Appendix, we prove Lemma 4: for all $x \in X$, UUD contains some objection path originating from x.

Proof of Lemma 4

Since $UD^0 = \mathbf{P}$ and hence contains objection paths originating from x, it suffices to prove that for all $x \in X$, for all $t = 0, \ldots$, if $UD_x^t \neq \emptyset$, then UD_x^{t+1} is nonempty as well.

Choose some set P of objection paths. Find, for any x such that $(x) \notin ud(P)$, a coalition $S(x)$ such that (x) is $S(x)-$dominated in P.

For any x, identify a set $C(x, P)$ such that

$$C(x, P) = \{y : (x) \text{ is } S(x) - \text{covered in } P \text{ via } y\}. \tag{6}$$

Further, denote by $C^*(x, P)$ the subset of $C(x, P)$ that contains any y that induces the maxmin payoff to coalition $S(x)$ in $C(x, P)$. That is,

$$C^*(x, P) = \{y \in C(x, P) : \max_{z \in C(x,P)} \min_{p \in P_z} u_{S(x)}(\mu[p)) \not\gg \min_{p \in P_y} u_{S(x)}(\mu[p))\}. \tag{7}$$

Note that $(x) \in ud(P)$ if and only if $C(x, P) = C^*(x, P) = \emptyset$.

We say that $(x_0, S(x_0), ..., x_J)$ is a $C^*(\cdot, P)-$sequence that originates from x if $x = x_0$ and $x_{j+1} \in C^*(x_j, P)$ for all $j = 0, ..., J-1$.

Denote by $\overline{C}^*(\cdot, P)$ the transitive closure of $C^*(\cdot, P)$. [20] Denote the set of maximal elements of $\overline{C}^*(\cdot, P)$ by $V(P) = \{x \in X : y \in \overline{C}^*(x, P) \text{ implies } x \in \overline{C}^*(y, P), \text{ for all } y\}$.

LEMMA **6.** *Let $y \in C^*(x, P)$. Then, for any $p_y \in P_y$, the sequence $(x, S(x), p_y)$ is an objection path and it is not dominated in P' if $P \subseteq P'$.*

[20]That is, $y \in \overline{C}^*(x, P)$ if and only if there is a $C^*(\cdot, P)-$sequence originating from x and ending in y.

Proof. Since p_y is a member of P_y, and x is $S(x)$ covered in P via y, $(x, S(x), p_y)$ is an objection path.

If $(x, S(x), p_y)$ is dominated in P', and $P \subseteq P'$, then there is z such that

$$\min_{p_z \in P_z} u_{S(x)}(\mu[p_z)) \geq \min_{p_z \in P'_z} u_{S(x)}(\mu[p_z)) \gg u_{S(x)}(\mu[p)) \gg u_{S(x)}(x).$$

The third inequality, which implies that $(x, S(x), p_y)$ is an objection path, follows from the assumption that $y \in C(x, P)$. Thus the first inequality implies that also $z \in C(x, P)$. But together with (6) this contradicts the assumption that $y \in C^*(x, P)$. ■

LEMMA **7.** *For any* $t = 0, 1, ...,$ *for any* $x_0 \in X$*, let* $(x_0, S(x_0), x_1, ..., x_J)$ *be a* $C^*(\cdot, UD^t)-$*sequence with* $(x_J) \in UD^{t+1}$*. Then* $(x_0, S(x_0), x_1, ..., x_J) \in UD^{t+1}$.

Proof. Of course, $UD^{t+1} \subseteq UD^t$ for all τ. So, Lemma 6 implies that the sequence $(x_j, S(x_j), x_j, ..., x_J)$ is not dominated in UD^t, for any $j = 0, 1, ..., J-1$. Since, in addition, (x_J) is not dominated in UD^t, we have $(x_0, S(x_0), x_1, ..., x_J) \in UD^{t+1}$. ■

LEMMA **8.** *For any* $t = 0, 1, ...,$ *for any* $x \in X$*, there is* $y \in \overline{C}^*(x, UD^t)$ *such that* $(y) \in UD^{t+1}$.

Proof. Claim 1: For any t, if $x \in V(UD^t)$ and $(x) \in UD^t$, then $(x) \in UD^{t+1}$.

Proof: Suppose that $(x) \in UD^t - UD^{t+1}$ and $x \in V(UD^t)$. Since X is a finite set, there is a $C^*(\cdot, UD^t)-$ sequence $(x_0, S(x_0), x_1, ..., x_L)$ such that $x = x_0 = x_L$. By Lemma 7, $(x_1, S(x_1), x_2, ..., x_L) \in UD^t$. But then, since $x_L = v_0$, x_0 is not dominated via x_1 in UD^t, a contradiction to the hypothesis that $x_1 \in \overline{C}^*(x_0, UD^t)$. □

Claim 2: For any t, $C^*(x, UD^t) = C(x, UD^t)$, for all $x \in V(UD^t)$.

Proof: Fix any $x \in V(UD^t)$. It suffices to show the direction $C(x, UD^t) \subseteq C^*(x, UD^t)$. If $(x) \in UD^{t+1})$, then $C(x, UD^t) = C^*(x, UD^t) = \emptyset$.

Suppose that $(x) \notin UD^{t+1}$.Since $x \in V(UD^t)$, there is a $C^*(\cdot, UD^t)-$ sequence $(x_0, S(x_0), x_1, ..., x_L)$ such that $x = x_0 = v_L$. Choose any $x' \in C(x_0, UD^t)$. By Lemma 6, $(x_0, S(x_0), p') \in UD^t$ for any $p' \in UD^t_{x'}$. Iterating backwards on $j = L-1, L-2, ..., 2$ it follows that

$$(x_1, S(x_1), ..., x_{L-1}, S(x_{L-1}), x_0, S(x_0), p') \in UD^t, \text{ for any } p' \in UD^t_{x'}.$$

Thus $\cup_{p \in UD^t_{x'}} \mu[p) \subseteq \cup_{p \in UD^t_{x_1}} \mu[p)$ implying, by (7), that $x' \in C^*(x_0, UD^t)$. Since x' is an arbitrary element of $C(x_0, UD^t)$, we conclude that $C(x_0, UD^t) = C^*(x_0, UD^t)$.

Claim 3: For any t, for any $x \in V(UD^t)$ there is $x' \in \overline{C}^*(x, UD^t)$ such that $(x') \in UD^{t+1}$.

Proof: Initial step: $t = 0$. Then $(x') \in UD^0$ for all $x' \in X$. By Claim 1, $(x') \in UD^1$, for all $x' \in V(UD^0)$.

Inductive step: $t > 0$. Let the claim hold for $t - 1$. We show it holds for t. By the definition of V, $\overline{C}^*(x, UD^t) \subseteq V(UD^t)$ for all $x \in V(UD^t)$. Thus, by Claim 2,

$$\overline{C}(x, UD^t) \subseteq V(UD^t), \text{ for all } x \in V(UD^t). \tag{8}$$

By the maintained assumption, there is a $x' \in \overline{C}^*(x, UD^{t-1})$ such that $(x') \in UD^t$. Since $C^*(\cdot, UD^{t-1}) \subseteq C(\cdot, UD^{t-1}) \subseteq C(\cdot, UD^t)$, also $v' \in \overline{C}(v, UD^t)$. By (8), $v' \in V(UD^t)$. By Claim 1, $(v') \in UD^{t+1}$. □

Claim 4: For any $x \in X$, there is $y \in \overline{C}^*(x, UD^t)$ such that $(y) \in UD^{t+1}$.

Proof: If $x \notin V(UD^t)$, then there is $y \in \overline{C}^*(x, UD^t) \cap V(UD^t)$. By Claim 3, there is $z \in \overline{C}^*(y, UD^t)$ such that $(z) \in UD^{t+1}$. By transtitivity, $z \in \overline{C}^*(x, UD^t)$.□ ■

It is now follows by Lemmata 7 and 8 that:

LEMMA **9.** *For any $t = 0, 1, ...,$ for any $x \in X$, there is a $C^*(\cdot, UD^t)-$sequence $(x_0, S(x_0), ..., x_J)$ such that $(x_0, S(x_0), ..., x_J) \in UD_x^{t+1}$.*

This completes the proof of Lemma 4.

REFERENCES

Anesi, Vincent (2010), "Noncooperative Foundations of Stable Sets in Voting Games," *Games and Economic Behavior*, **70**, 488–493.

Anesi, Vincent and Daniel Seidmann (2014), "Bargaining over an Endogenous Agenda," *Theoretical Economics*, **9**, 445–482.

Aumann, Robert and Roger Myerson (1988), "Endogenous Formation of Links Between Players and of Coalitions, An Application of the Shapley Value," in *The Shapley Value: Essays in Honor of Lloyd Shapley*, Alvin Roth, ed., 175–191. Cambridge: Cambridge University Press.

Bloch, Francis and Anne van den Nouweland (2017), " Farsighted Stability with Heterogenous Expectations," mimeo.

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

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

Chander, Parkash (2015), "An Infinitely Farsighted Stable Set", mimeo, Jindal Global University.

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

Dutta, Bhaskar (1988), "*Covering Sets and a New Condorcet Choice Function*, *Journal of Economic Theory* **44**, 63–88.

Dutta, Bhaskar and Rajiv Vohra (2016), "Rational Expectations and Farsighted Stability", to appear in *Theoretical Economics*.

Gomes, Armando and Philippe Jehiel (2005), " Dynamic Processes of Social and Economic Interactions: On the Persistence of Inefficiencies", *Journal of Political Economy*, **113**, 626–667.

Greenberg, Joseph (1990), *The Theory of Social Situations*, Cambridge, MA: Cambridge University Press.

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

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

——— (2009),"Farsightedly Stable Networks," *Games and Economic Behavior*, **67**, 526–541.

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

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

Laslier, Jean-Francois, (1997), *Tournament Solutions and Majority Voting*, Springer-Verlag, Berlin.

Lucas, William (1968), "A Game with No Solution," *Bulletin of the American Mathematical Society*, **74**, 237–239.

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

Mauleon, Ana and Vincent Vannetelbosch (2004), "Farsightedness and Cautiousness in Coalition Formation Games with Positive Spillovers," *Theory and Decision*, **56**, 291–324.

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

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

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

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

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

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

——— (2017), "Maximality in the Farsighted Stable Set", mimeo.

Vartiainen, Hannu (2011), "Dynamic coalitional equilibrium," *Journal of Economic Theory*, **143**, 672–698.

——— (2014), "Endogenous agenda formation processes with the one-deviation property," *Theoretical Economics*, **9**, 187216.

——— (2015), "Dynamic stable set as a tournament solution," *Social Choice and Welfare*, **42**, 309-327.

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

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