Coalition Formation and History Dependence *

Bhaskar Dutta

University of Warwick and Ashoka University

Hannu Vartiainen

University of Helsinki and Helsinki Graduate School of Economics

May, 2019

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 (2017) 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 (2017) 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 second author gratefully acknowledges the hospitality of Clare Hall and INET at the University of Cambridge. The paper has benefited from comments and discussions with Francis Bloch, Mert Kimya, Debraj Ray, Rajiv Vohra, Hamid Sabourian, three anonymous referees, the Co-editor Dilip Mookherjee, 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 : (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 (2017). 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.

Ray and Vohra (2014) raised an important issue with much of the literature. They pointed out that 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] The following comments of Ray and Vohra (2015) about the existing farsighted solution concepts are instructive :

"Stable outcomes can be modeled either with optimistic beliefs or conservative beliefs or perhaps some combination of the two. However, this is serious drawback of the blocking approach".

They go on to add : "A key requirement that is missing in the notion of farsighted blocking, is that of constraining objecting coalitions to make *maximal moves* (our emphasis) among their profitable alternatives."

Dutta and Vohra (2017) (henceforth 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) will be to z' (not z). That is coalitions S^1 and S^2 hold different beliefs about

[4]See Examples 1 and 2 in Section 4.

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.

Importantly, DV assumes that the transition from any state x to another state y only depends on the current state. Together with the expectation function, each state is then identified with a *terminal* or *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 upon 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 may also lead 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.

[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 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.

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 nonemptyness 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.

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 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" characterisations are 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). In particular, we are able to show that our solution is always contained in Chwe's largest consistent set. Since the largest consistent set is viewed as being too permissive, this inclusion result is of some interest.

The plan of the paper is the following. In the next section, we introduce some key concepts. In section 3, we describe formally the framework introduced by DV, and then go on to introduce our solution concepts. We discuss related solution concepts in Section 4. Section 5 contains our main characterisation results in terms of objection paths, while section 6 contains the characterisation for simple games. An important byproduct of the analysis for simple games is that notions of maximality are rendered irrelevant, in a sense to be explained in section 5. We present two applications of our solution concept in section 7. These applications demonstrate the importance of history dependence. In section 8, we discuss some properties of our solution concepts. We go on to present the nonemptyness results in section 8. We conclude in Section 9.

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

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

non-empty 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 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 y_m in the objection path $p = (y_0, S_1, y_1, \ldots, S_m, y_m)$.

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

State y *indirectly 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 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

As we have mentioned earlier, DV incorporate both Maximality as well as common beliefs about continuation paths (of coalitional deviations) in their analysis. They use 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

[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.

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 is that the expectation function specify moves that are optimal. We will describe below slightly different notions of degrees of optimality - each will give rise to a specific restriction on the expectation function.

DV assumed that the process of transition is *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 enlarges the set of noncooperative equilibria. In principle, this logic may carry forward to the present context. Indeed, the applications later on illustrate the instrumental importance of history dependence.

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 $S \in E(x_t, x_{t+1})$. 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, with a typical element h.

For any history $h = (x_0, S_1, x_1, ..., S_k, x_k)$, we will use $\mu(h)$ to denote the *terminal* state x_k of h. Notice that *all* finite histories have well-defined terminal states.

We use the following notation on concatenation of path. For any history h, (h, S, x) is the history reached by adding to h, the state x that is induced by coalition S from the final state $\mu(h)$ of h. Note that $S \in E(\mu(h), x)$ for this to be valid.

Expectation Function

An expectation is a function $F : H \to \mathcal{N} \times X$, specifying the active coalition and its move for all possible current states and past histories.

[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.

The expectation function "predicts" that *one* coalition is going to be active at any history, without describing any explicit protocol which chooses the active coalition. [11]

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.[12]

As usual, history independence is a special case of history dependence. Consider any two histories $h, h' \in H$. Then, the DV expectation function satisfied $F(h) = F(h')$ whenever $\mu(h) = \mu(h')$. So, the continuation path once a state x is reached does not depend on whether the state was reached via history h or history h'.

Given an expectation $F = (S, f)$, note that $(h, S(h), f(h))$ is also a history. We denote $F^0(h) = h$, $F^1(h) = F(h)$, and generally $F^{k+1}(h) = F(h, F^0(h), ..., F^k(h))$ 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 $S(h) = \emptyset$. If h is stationary, then we will also denote by $\mu(h)$ the terminal point of h as a *stationary point*.

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

A history $(h, S_1, y_1, ..., S_m, y_m)$ is an *indirect 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.,

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

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

[11] A referee has questioned why only one coalition is assumed to move at any point. Consider an extensive form or game tree which represents a specific protocol that describes the player who moves at any particular node in the tree. The tree also describes the possible paths that may be followed from any given node. Here, we have no explicit protocol. The expectation function is supposed to be a formalisation of the commonly held beliefs of players about the continuation path from any given state, including the coalition that is supposed to move. Notice that this assumption is implicit in all solution concepts based on objection paths. However, the stronger version of Optimality - Condition M* to be defined later - does allow for the possibility that a deviation can come from a coalition that is different from the one specified by the expectation function.

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

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 $(\mu(h), S, y, \bar{F}(h, S, y))$ is an objection path.
: (**E**) If h is a nonstationary history, then $(\mu(h), \overline{F}(h))$ is an objection path.

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 farsighted 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 indirect 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 $u_{S(h)}(\mu(\overline{F}(h, S(h), y))) \gg u_{S(h)}(\mu(\overline{F}(h)))$.
: (**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 $i \in S(h)$ 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 **2.** *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).*

In what sense is a set like HREFS (or HSREFS) a *stable* set? Given any history h, the (absorbing) expectation function specifies that the terminal state $\mu(h)$ will be reached. The property of being absorbing implies that no coalition wants to deviate once the state $\mu(h)$ is reached. Of course, the terminal states of *any* absorbing expectation will not constitute a reasonable stable set. This is where the conditions I, E, M and M* come in by eliminating ad hoc absorbing expectation functions whose terminal points do not satisfy intuitive notions of stability. It is worth pointing out that while *any* given history will have only one terminal state, HREFS and HSREFS will in general be *set-valued* since we have to consider the union of the terminal states of all histories - that is, $\mu(\overline{F}(H))$. Indeed, like any vNM-type solution set, individual elements of $\mu(\overline{F}(H))$ are not stable.

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. RELATED SOLUTION CONCEPTS

Following Harsanyi (1974), there have been several papers modelling farsighted cooperative solution concepts, several of them being based on indirect domination.

Many of these 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, a coalition can consider further deviations and indeed several coalitions may contemplate deviations. So, there can be multiple 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.[13] We discuss here some of these solution concepts as well as how they compare to HREFS and HSREFS.

[13]Many of the solution concepts based on farsightedness can also be viewed through Greenberg's (1990) theory of social situations.

An "obvious" way of introducing farsightedness - and one suggested by Harsanyi himself - is to modify the original vNM solution by replacing the direct domination relation by the indirect domination relation. Chwe (1994) was the first to define this formally. His definition is given below.

DEFINITION **3.** *A set $F \subseteq X$ is a farsighted stable set if it satisfies: :*

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

The farsighted stable set is based on an *optimistic* view of the coalitions involved in an indirect 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.

DEFINITION **4.** *A set $K \subseteq X$ is consistent if*

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

Thus, any potential move from a point in a consistent set is deterred by *some* indirect 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). Chwe himself points out that the LCS is a weak solution concept that is "not so good at picking out, but ruling out with confidence".[14] For example, the LCS may contain elements not in the core. Nevertheless, the largest consistent set is an important concept that has received much attention in the literature.

Several papers have used either the farsighted stable set or its variants in specific contexts. For instance, Beal et al (2008) study farsighted stability in transferable utility games, while Bhattacharya and Brosi (2011) extend the analysis to NTU games. Some authors have also analysed farsighted stability in hedonic games. These include Diamountoudi and Xue (2003, 2007), Mauleon, et al (2011). The latter analyse one-to-one matching markets and show, surprisingly, that a set of matchings is a farsighted stable set if and only if it is a singleton element of the core.

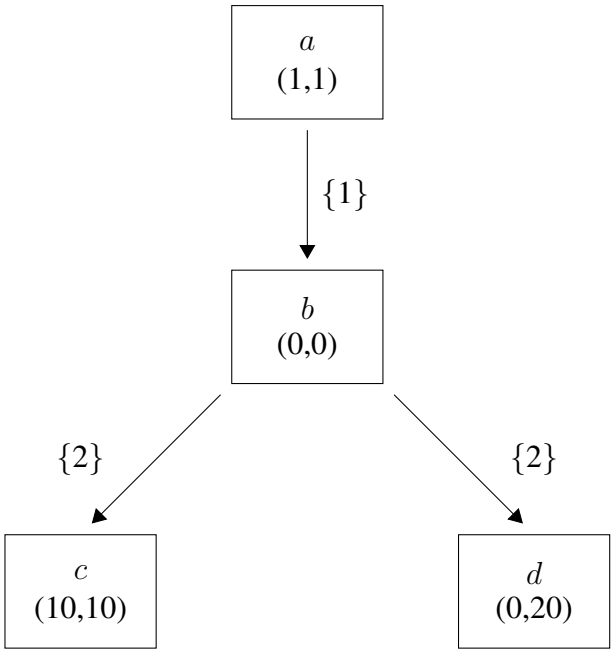
Herings et al (2009, 2010) are two notable recent solutions with connections to farsighted stable sets. Herings et al (2009) analyse farsighted stable networks in a model where, as in much of the literature on networks, only two-person coalitions can form and can change only one link at a time (these are assumptions on the effectivity function). Their solution called *pairwise farsighted stability* requires farsighted external stability

[14] See Chwe (1994), page 300.

but no version of internal stability. Instead, deviations from inside the stable set to anything outside are required to be deterred by the possibility of becoming (weakly) worse-off, and a minimality requirement is also imposed since (in the absence of internal stability), the entire set of networks satisfies the two conditions. Herings et al (2010) apply the same solution concept to hedonic games where outcomes are partitions of N and individual payoffs are defined directly on partitions. However, like Herings et al (2009), they do not impose internal stability. [15]

The next couple of simple examples, adapted from Xue (1998), Herings et al (2004), Kimya (2017) illustrate some drawbacks of these solution concepts.

Example 1

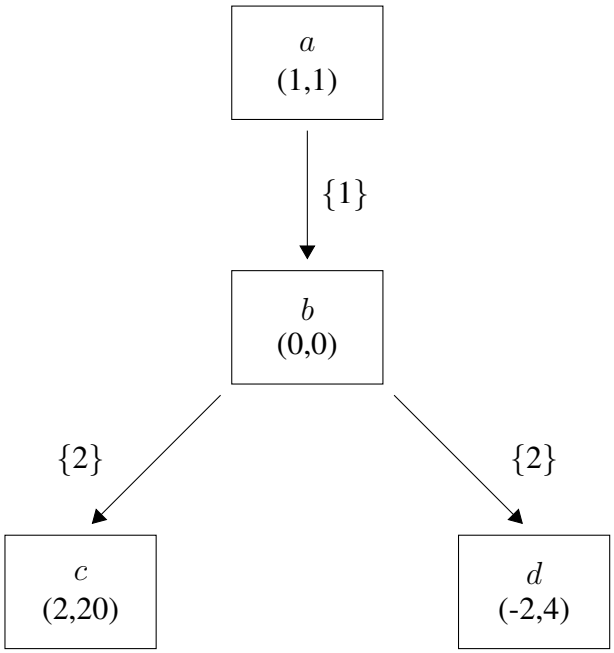


Both c and d belong to the farsighted stable set since they are terminal states. Since there is a farsighted objection from a to c, the former is not in the farsighted stable set. However, this is based on the expectation that player 2 will choose to replace b with c rather than d even though 2 prefers 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.

[15]The lack of internal stability allows Herings et al (2009, 2010) to prove a strong existence theorem. Most history-independent farsighted (and myopic) solution concepts that use both internal and external stability typically fail to guarantee a nonempty solution when the indirect domination relation is cyclic as in the voting paradox.

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

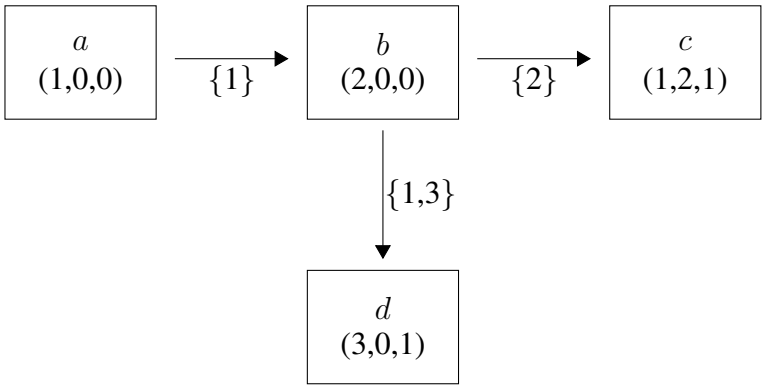


Now the optimal move for player 2 is to choose c rather than d. The LCS and farsighted stable set remain unchanged. But now it is the LCS which provides the wrong answer because player 1 should not fear that player 2 will (irrationally) choose d instead of c. In this example, the farsighted stable set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

We show in Proposition 2 below that HREFS is a subset of LCS. Another notable refinement of LCS is the *Largest Cautious Consistent Set* (LCCS) by Mauleon and Vannetelbosch (2004) which assumes "cautious" coalitional behavior. Cautiousness is a response to the criticism of LCS that a blocking of an element in LCS can be deterred by the mere possibility that a post blocking objection path may lead to an outcome in the LCS that makes the blocking coalition indifferent between blocking or not, *even if* the other objection paths leading back to the consistent set are strictly profitable for the blocking coalition. Notice that in this case, the coalition can only win by blocking. However, LCS will include the initial element. Mauleon and Vannetelbosch correct for this by imposing the requirement that a potential deviating coalition assigns strictly positive probability to all objection paths.

While the LCCS certainly does correct one problem inherent in the LCS, its basic logic is somewhat similar to that of LCS. In particular, it does not require optimal behaviour on the part of coalitions. That is, LCCS does not satisfy Maximality. Notice that in Example 2, the LCCS may coincide with the LCS and thus make the wrong prediction.[16] There is also no clear relationship between the LCCS and HREFS. For example, as demonstrated by Mauleon and Vannetelbosch (2004) LCCS can be empty even in finite games, whereas we show below that HREFS is always nonempty in this class of games. Furthermore, Example 3 below (Figure 2 in Mauleon and Vannetelbosch 2004) demonstrates that LCCS can be a strict subset of HREFS/SREFS. There, LCCS is $\{c, d\}$ whereas one HREFS (and hence LCS) is $\{a, c, d\}$. Since HREFS depends on the expectation function, there may not be a unique HREFS - the solution will depend on whether 2 or $\{1, 3\}$ is expected to move from b. In fact, one HREFS coincides with the LCCS. Of course, HREFS may also be contained in LCCS as shown in Example 2.

Example 3



Xue (1998) also argued that the LCS and farsighted stability did not really incorporate farsighted behaviour. He suggested that the focus should be on stability of *objection paths* rather than the terminal outcomes. He defined stable paths to be ones which satisfied analogues of Greenberg's optimistic and conservative standards of behavior. Notice that neither standard of behavior incorporates Maximality. Importantly, Herings et al

[16] Example 2 also shows that there may be a multiplicity of LCCS, depending on the probability distributions over objective paths. Of course, the imposition of Maximality rules out the move of 2 from b to d, and hence selects a unique continuation path from b to c.

(2004) and Ray and Vohra (2015) point out other problems with Xue's approach, arising essentially from any lack of a protocol specifying the order in which coalitions can move.

One way of interpreting Examples 1 and 2 is that Maximality imposes the type of logic inherent in backwards induction, or more generally subgame perfection. In a recent interesting paper, Kimya (2017) also formulates two solution concepts that integrates coalitional analysis with noncooperative logic. Kimya too focuses on paths with elements being triples of the form (x, y, S) with $S \in E(x, y)$ and $x, y \in X$.[17] A coalitional behavior ϕ (or coalitional strategy) is a complete plan of action at every state. A coalition S can deviate from the prescribed behavior ϕ at state x if S intersects T where T is the coalition that is supposed to move at x. The specification of a coalitional behavior ϕ with reference to which deviations can take place plays a role very similar to expectations functions and so incorporates consistency.

The coalitional deviation of S from ϕ is profitable if a deviation by S is *profitable* at every node at which an action changes; that is, everybody in S prefers the new path of play to the initially prescribed path of play. Kimya defines a coalitional behavior ϕ to be an *equilibrium coalitional behavior* (ECB) if there does not exist a profitable coalitional deviation from ϕ. ECB satisfies Strong Maximality; indeed, Kimya shows the coincidence between the terminal outcomes supported by ECB and SREFS in any abstract game.

He defines a *credible* set of coalitional behaviors to be a set of coalitional behaviors such that any profitable deviation from a coalitional behavior in the set is followed by further profitable deviations back into an element in the credible set that makes one of the initial deviators worse off. Kimya defines a coalitional behavior to be a *credible ECB* (CECB) if it is immune to profitable and credible deviations. Note that since coalitional strategies are strategies that are defined on states and not histories, these are essentially Markovian. So, CECB and ECB are *history independent* solutions analogous to those in Dutta and Vohra (2017). Moreover, due to different criteria for credible deviation, Kimya's solution CECB does not have a clear relationship with the solution of the current paper, or that of Dutta and Vohra (2017); neither solution implies the other.

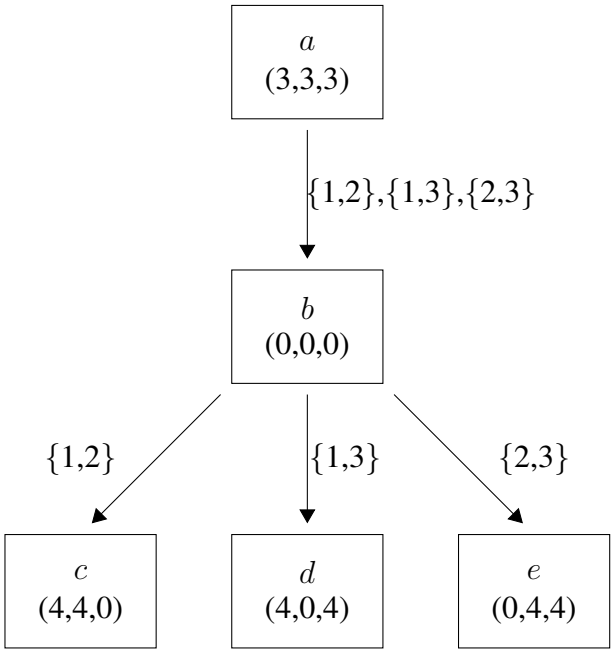
Kimya shows that CECB exists in all finite games, but has no characterization result. He also states that CECB is not a "satisfactory solution concept for characteristic function games" partly because CECB depends on a type of backwards induction that is not possible in this class of games.

[17] Mariotti (1996) restricts attention to normal form games and also focuses on objection paths. However, he assumes *optimistic* behavior on the part of deviating coalitions. Hence, his solution violates Maximality.

Herings et al. (2004) and Granot and Hanany (2016) propose solution concepts on the abstract game by constructing noncooperative extensive form games. Granot and Hanany (2016) assume that nature chooses the coalition that moves at any stage, with each coalition allowed to remain at the current state or select a new outcome according to the effectivity function. Their solution concept is the set of outcomes that can be supported as the subgame perfect equilibria of the extensive form game. They are able to show existence only for *finite* set of states.[18] They also need to impose the assumption that players are pessimistic in order to make the solution concept independent of nature's moves. This is of course a violation of Maximality. Herings et al (2004) define a notion of social rationalizability in their multi-stage games. Herings et al study rationalizable strategies and the resulting outcomes in a game form that specifies how coalitions are formed. Like the largest consistent set, social rationalizability of Herings et al is meant to be a "weak" concept that removes strategies with confidence but does not predict or explain which of them are eventually taken. In particular, it does not attempt to provide an equilibrium story of coalition formation. Interestingly, Herings et al(2004) construct an example in which the LCS excludes too much. Since HREFS is a subset of LCS, it too will exclude too much in this specific example.

The solutions discussed so far are all history independent. We close this section with a couple of examples illustrating the role of history dependence.

Example 4



[18]Herings et al (2004) also assume a finite set of states.

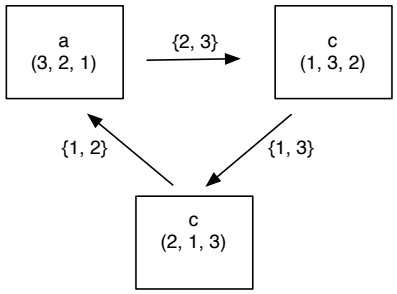
In this example, a is the surplus-maximising outcome, and hence the unique socially efficient outcome corresponding to the utilitarian rule. However, none of the history-independent solutions other than the LCS and LCCS can support a as an outcome. All the majority coalitions have profitable deviations from b. History independence means that which coalition moves is *independent* of previous movements. In particular, punishments are not possible. So, whichever coalition(s) is expected to move at b will also want to move from a to b. So, the prediction must be that $\{c, d, e\}$ (the set of terminal outcomes) constitute the set of stable outcomes. Consider, however, how this prediction changes when history dependence is introduced. Suppose, for instance, that $S = \{1, 2\}$ deviates from a to b. Then, one of the players in S, say 1, can be punished for the deviation by choosing the continuation path to be $\{2, 3\}$ moving to e. Clearly, any deviation from a can be punished in this way by making the continuation path from b dependent on the initial deviation from b. Hence, HSREFS and so HREFS will be $\{a, c, d, e\}$.

REMARK **2.** *In this example, the LCS coincides with HREFS. However, the permissiveness of the LCS can easily be demonstrated by embedding this example in a bigger one as follows. Consider two new alternatives f with utility vector $(1, 1, -0.25)$ and g with utility vector $(1, 1, -0.5)$. Let $\{3\} \in E(f, b)$ and $\{1, 2\} \in E(b, g)$. Then, 3 will not deviate from f to b under the pessimistic expectation that $\{1, 2\}$ might move to g from b. However, this is a violation of maximality - the optimal move for $\{1, 2\}$ from b is to c which is better than f for 3.*

Finally, History Dependence also helps in establishing existence of a nonempty solution set in situations where several history dependent solutions will be empty. One example in which this is the case is the three-player NTU 'roommate game' depicted below.

Example 5

From every state there is one two-player coalition that gains by moving to another state.



It is easy to see that this games possesses no vNM stable set, no farsighted stable stable, no ECB and no REFS.

In section 7 below, we describe two specific examples which illustrate the same general principles outlined here through abstract games.

5. 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 nonemptyness 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 for all finite games. Finally, the characterization is employed when we analyze the relationship of HREFS and HSREFS to other solution concepts. In particular, the characterization proves useful in showing that every HREFS is a subset of the LCS.

Recall that we will use p_x, p_y, etc. to denote objection paths with initial state x, y. Similarly, given any set of objection paths P, we will use P_x to denote the subset of objection paths in P with initial state x.

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 move to x_1. Part (ii) stipulates that if x is not followed by any other state, i.e. is stationary, 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 **3.** *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.[19]

The first two conditions are obvious. The first condition requires that the set P must contain at least one objection path with initial state x for every state. After all, we must be able to predict what happens starting from any initial state x. The second condition states that if an objection path p is in P, then any objection path which is a subpath of p must also be in P. Conditions 3 and 4 are in some sense the two crucial conditions. Suppose condition 3 is not satisfied by some set P. Then, P must include an objection path $p = (x_0, S_1, x_1, ..)$ that is S_1 dominated in P via some y. This would mean that S_1 can deviate to y and be assured that all paths in P_y make it strictly better off than following the path p. This implies that S_1 is not taking a maximal move if it moves from x_0 to x_1. Condition 4 requires that if (x) is in P, and hence x is a terminal state of a path in P, then not deviating from x must be a maximal move for every coalition. We are going to show that the terminal states of paths in a coherent collection P constitute an HREFS. Condition 4 is required to ensure that **I** is satisfied.

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 $x \in X$.

[19] Path based coalitional solutions include Xue (1998), Mariotti (1997), and Kimya (2017). However, as discussed in section 4, they are based on assumptions that are, in general, not compatible with HREFS and, hence, not directly comparable to Coherence.

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 states 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.

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 $p_x = (x, S_1, \ldots) \in P$ 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 3, we know that $(x) \in P$. For each such x, let $(x) \in H_{(x)}$. Recursively, take any $p_{x_0} = (x_0, S_1, x_1, \ldots) \in P$ 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

$$(h, S, y) \in \begin{cases} H_{(x_1, S_2, \ldots)}, & \text{if } (S, y) = (S_1, x_1), \\ H_{\xi(p_{x_0}, (S, y))}, & \text{if } (S, y) \neq (S_1, x_1). \end{cases} \tag{1}$$

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 $h \in H_{(x_0,S_1,...)}$,

$$F^P(h) = \begin{cases} (S_1, x_1), & \text{if } S_1 \neq \emptyset, \\ (\emptyset, x_0), & \text{if } S_1 = \emptyset. \end{cases} \tag{2}$$

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 continuation play leads a terminal history $(h, F^1(h), F^2(h), ...) = (h, x_0, S_1, ...)$, which is an indirect 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. The strengthening involves ensuring that *any coalition* with a nonempty intersection with S_1 should not want to deviate.

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

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

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

$$(h,S,y) \in \begin{cases} H_{(x_1,S_2,\ldots)}, & \text{if } (S,y) = (S_1,x_1), \\ H_{\xi((x_0,S_1,\ldots),(S,y))}, & \text{if } (S,y) \neq (S_1,x_1). \end{cases} \tag{3}$$

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 $h \in H_{(x_0,S_1,\ldots)}$.

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

$$F^P(h) = \begin{cases} (S_1,x_1), & \text{if } S_1 \neq \emptyset, \\ (\emptyset,x_0), & \text{if } S_1 = \emptyset. \end{cases} \tag{4}$$

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 $(x_0,S_1,\ldots) \in P$ be such that $h \in H_{(x_0,S_1,\ldots)}$. 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, $(y, F^1(h,S,y), F^2(h,S,y),\ldots) = \xi((x_0,S_1,\ldots),(S,y))$. By the definition of ξ, $u_S(\mu(\xi((x_0,S_1,\ldots),(S,y)))) \not\gg u_S(\mu((x_0,S_1,\ldots)))$. ■

We will use these characterisation theorems repeatedly in subsequent sections. In particular, we will use Theorem 2 to construct nonempty HSREFS in all superadditive transferable utility partition games, as well as non-empty HREFS in all finite games.

6. 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.[20] Of course, this assumes that utility is transferable and winning brings the same aggregate benefit to the winning coalition. We use a different formalisation of simple games, motivated at least partly by the kind of contexts that are typically mentioned as potential applications of simple games. Consider, for instance, a legislature which has to choose whether to pass a bill along with a set of possible amendments. Or consider a committee voting on an up-or-down decision. In such cases, the rules of the legislature or the committee specify what groups of individuals are *decisive* (or *winning*) in the sense of being able to take decisions, and so the simple game structure in terms of winning coalitions seems appropriate. However, it is somewhat inappropriate to assume either that utility is transferable or that the final

[20]Farsightedness for this class of simple games was studied by both Ray and Vohra (2015) as well as DV (2017).

decision or outcome brings the *same* aggregate benefit to the group whose vote wins the day. Our formulation preserves the essential structure of simple games, so that winning coalitions can enforce any outcome, but drops the assumption that aggregate benefits are equal no matter the outcome that is chosen, and also transferability of utility.

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

(i) If $S \in \mathcal{W}$, and $S \subset T$, then $T \in \mathcal{W}$.[21]

(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. A coalition is a *losing* coalition if its complement in N is a winning coalition.

In this section, we will explicitly assume that a *state* consists of an outcome a from some feasible set A (for instance, the set of legislative bills) as well as a partition π of N. That is, any state x is a pair (a, π) and $X = A \times \Pi$. Of course, a winning coalition may not form. In such cases, we will assume that the outcome will be a distinguished element a^0 (the status quo), which is also in A. We use X^0 to denote the set of "zero" states in which no winning coalition has formed, so that a typical element of X^0 will be $x^0 = (a^0, \pi)$, with no element of π being in $\mathcal{W}$.

We assume that each individual i has a utility function defined over A and that

$$u_i(a, \pi) = u_i(a) \text{ for all } i \in N, (a, \pi) \in X$$

That is, individuals care only about the bill that is passed or the decision that is taken by the committee and not about the partition that represents the voting choices.

In a simple game, a winning coalition has the power to choose any outcome in A, while a blocking coalition can ensure that the status quo a^0 is the resulting outcome. The only way in which a losing coalition T can change the utility allocation is if T leaves a winning coalition S and $S - T$ is not a winning coalition. So, for instance suppose $|N| = 5$ and any coalition of three or more is a winning coalition. Let $\pi = \{\{1, 2, 3\}, \{4\}, \{5\}\}$. Then, any $i \in \{1, 2, 3\}$ can leave the coalition and ensure that a zero state emerges. On the other hand, if $\pi = \{\{1, 2, 3, 4\}, \{5\}\}$, then no singleton has any power to change the outcome. This illustrates the limitations on the power of losing coalitions - if L is a losing coalition which is a subset of a winning coalition S, then L can change the outcome to a^0 iff $S - L$ is not winning.

In order to express these formally, we use the following notation. For any partition π and coalition $S \subset N$, let $(S, \pi_S) \in \Pi$ represent the partition where S is an element of the partition and $T \in \pi_S$ iff $T = R - S$ for some $R \in \pi$. That is, if a coalition S forms and deviates from π, then the new partition consists of S and all original elements of π

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

without members of S. We will also use (S, π_{-S}) to denote the partition with S as an element and some partition π_{-S} of $N - S$.

The power of winning, blocking and losing coalitions is captured in the following assumption, which describes two properties of an effectivity function for simple games.

Assumption 1. *The effectivity function E satisfies the following*

(1) *For all $S \in \mathcal{W}$, for all $x = (a, \pi) \in X$, $S \in E(x, y)$ if $y = (b, (S, \pi_S))$ for any $b \in A$.*
(2) *For all $B \in \mathcal{B}$, $B \in E(x, x^0)$ for all $x \in X$.*
(3) *For all $L \subset S \in \mathcal{W}$, for all $x = (a, (S, \pi_{-S})) \in X$, $(b, (L, S - L, \pi_{-S})) \in E(x, L)$ iff $a = b$ or $[(b = a^0)$ and $S - L \notin \mathcal{W}]$.*

Part 1 of the assumption implies that a winning coalition is eligible to induce state $(b, (S, \pi))$ from any state (a, π). Part 2 just says that a blocking coalition can always induce a zero state. Part 3 implies that a subcoalition L of a winning coalition S that is not winning nor blocking can change the outcome *only if* the residual coalition, i.e. $S - L$ ceases to be a winning coalition. In such a case, no winning coalition forms and so L enforces the status quo.

Given these restrictions on the power of coalitions and utility functions, the only relevant details of a state not in X^0 are given by the identity of the winning coalition S and the outcome chosen by S.

We normalise utility functions so that $u_i(a^0) = 0$, and make the following assumption.

Assumption 2. *For all $S \in \mathcal{W}$, there is $a \in A$ such that $u_i(a) > 0$ for all $i \in N$.*

Assumption 2 ensures that every winning coalition has at least one alternative that its members strictly prefer to the status quo outcome.

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 "powerful" coalitions 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 $S \in \mathcal{N}$, $z \in E(y, S)$, either $z \in Y - D_S(y)$, or there are $B \in \mathcal{B}$, $W \in \mathcal{W}$ and $x \in Y$ such that $x \in Y \cap D_B(z) \cap (D_W(x^0) - D_S(y))$.*

REMARK **4.** *Note that this definition allows for the possibility that $B = W$. This will be the case if there is $T \in \mathcal{W}$ and $x \in Y \cap D_T(z) - D_S(y)$.*

A set Y satisfies Condition C if the following is true. Take any y in Y and any S which can deviate to z. Suppose $z \in Y$, but all members of S do not strictly prefer z to y. In that case, one does not have to worry about the possibility of this deviation taking place. In all other cases, we have to ensure that this deviation is blocked. Condition C states that if S does deviate from y to z, then some blocking coalition B can precipitate the status quo and then some winning coalition W can make a further deviation to $x \in Y$. The state x has the property that all members of B strictly prefer x to z and all members of W strictly prefer x to the status quo. On the other hand, someone in S is not better off at x compared to y. So, if the expectation is that there will be a move to x following a move to z, then the initial deviation will not take place.

Our main result of this section follows.

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. Pick any $y \in Y$, coalition S and $z \in E(S, y)$ such that $z \in Y \cap D_S(y)$ or $z \in X - Y$. Define a function ϕ such that, for any $y \in Y$, for any $z \in (X - Y) \cup (Y \cap D_S(y))$ and $S \in E(y, z)$,

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

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 $P_1 = \{(z, \phi(y, S, z)) : y \in Y, S \in \mathcal{N}, z \in E(S, y), z \in (X - Y) \cup (Y \cap D_S(y))\}$.
Let $P_2 = \{(x^0, W, x) | (z, B, x^0, W, x) \in P_1\}$.

Construct P by

$$P = \{(y)\}_{y \in Y} \cup P_1 \cup P_2.$$

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 $p_x = (x, \phi(y, B, x)) \in P_1$. If $x \notin X^0$, then for some $S \in \mathcal{W}$, $S \in E(y, x)$, and $p_x = (x, \phi(y, S, x)) \in P_1$.

Part 2 follows immediately since P_2 is a subset of P.

To check part 4, consider any path $(y) \in P$. Take any $S \in \mathcal{N}$, and $z \in E(S, x)$. 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 $p_y \in P$. Suppose p_y is S-dominated in P via some S. Identify $\mu(p_y) = x$. Since p_y is S-dominated in P, there is $z \in E(S, y)$ such that

$$u_S(\mu(p_z)) > u_S(x) \text{ for all } p_z \in P \tag{5}$$

Suppose $z = (b, \pi) \notin X^0$. Then, $S \in \mathcal{W}$. Since S is a winning coalition, $S \in E(x, (b, \pi'))$ for some π' with $S \in \pi'$. Using equation 5, it follows that (x) is S-dominated in P. This is a contradiction since we have shown that part 4 of of Strong Consistency is satisfied.

Suppose $z \in X^0$. Now, $p_y \in P_1 \cup P_2$. Since $\mu(p_y) = x$, in either case, there is $W \in \mathcal{W}$ such that $(x^0, W, x) \in P_2$. Hence, there is $p_z \in P$ such that $\mu(p_z) = x$, contradicting equation 5.

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, 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 $z \in X - Y$. Since $z \notin Y$, (h, S, z) is not stationary. Then, Property E again implies 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, it is possible to provide a direct characterisation of HREFS (and HSREFS) in terms of sets of outcomes. Perhaps, more importantly, the theorem also demonstrates that maximality plays no role in simple games.[22]

7. THREE APPLICATIONS

In this section, we describe three applications which are designed to show the roles of history dependence and Maximality. We have remarked earlier that an important implication of history dependence is that punishments can now be history-specific. The first application is designed to show that such punishments can be used to support efficient outcomes. The final two applications illustrates the role of maximality by emphasising the permissiveness of solutions like LCS.

Why History Dependence Matters

This is a model of global pollution abatement. Each country can undertake industrial activities which result in *low* or *high* levels of pollution, labelled 0 and 1. Given other countries' choices, each country i strictly prefers choosing 1 to 0 because 0 involves high cost clean technology. On the other hand, other countries prefer that i chooses 0 rather than 1. So, let $x_i \in \{0, 1\}$ denote a typical strategy of i. So, this is an example of negative externalities. Payoff functions are given by :

$$u_i(x_1, x_2, x_3) = x_i - \frac{1}{2}(\sum_{j \neq i} x_j)^2 \tag{6}$$

Consider the normal form game in which each country chooses its level of pollution simultaneously. The only Nash equilibrium is $(1, 1, 1)$ while $O \equiv \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)\}$

[22] On this issue, see also Ray and Vohra (2017).

is the set of socially optimal allocations in the sense of maximising the *sum* of payoffs for the countries. Notice that for all $x \in O$,

$$\sum_{i \in N} u_i(x_1, x_2, x_3) = 0$$

We show that O is HREFS, though it cannot be sustained as a REFS. This shows the importance of history dependence.

REMARK **5.** *In this example, O is also the LCS.*

Let $x^0 = (0,0,0), x^1 = (1,0,0), x^2 = (0,1,0), x^3 = (0,0,1), x^4 = (1,1,0), x^5 = (1,0,1), x^6 = (0,1,1), x^7 = (1,1,1)$.

In order to keep the notation simple, we suppress the coalition structure and write a state just in terms of an allocation. So, for instance, we will write (x^0) instead of $(x^0, \{\{1\},\{2\},\{3\}\}$ and so on. Also, we will write an objection path by describing just the states figuring in the sequence and suppress mention of the coalitions that effect the move since there will be no ambiguity about the latter.

The effectivity function follows straightaway from the underlying normal form game. Any coalition $S \in C$ is effective in moving from x^k to $x^{k'}$ if $x_i^k = x_i^{k'}$ for $i \notin S$. That is, the transition from x^k to $x^{k'}$ cannot involve any change of strategy of those not in S. Note that since $\{1,3\}$ and $\{1,2,3\}$ are not permissible coalitions, there cannot be any immediate transition from x^0 to x^5 or from x^7 to x^0.

Let P be the set of objection paths given below *along with* all their subpaths so that P satisfies Condition 2 of Coherence.

- (x^0)
- (x^1, x^5, x^7, x^3)
- (x^2, x^4, x^7, x^1)
- $(x^3, x^6, x^7, x^4, x^2)$

We want to show that the set P forms a coherent set. Before we go into this demonstration, we explain why the path $(x^3, x^6, x^7, x^4, x^2)$ is qualitatively different from (x^1, x^5, x^7, x^3) and (x^2, x^4, x^7, x^1). Notice that in the latter two paths, coalitions $\{1,2\}$ and $\{2,3\}$ are deviating jointly at x^7. But, $\{1,3\}$ is not a permissible coalition and so this deviation is not possible at x^7. That is why we need 3 and 1 to deviate sequentially.

Clearly, P satisfies the first two conditions of Coherence. So, we only to check for Conditions 3 and 4. Consider, for instance, (x^1, x^5, x^7, x^3), where the first coalition to move (from x^1 to x^5) is country 3. Notice that $\mu(x^1, x^5, x^7, x^3) = x^3$. Since this is 3's most preferred outcome in O, (x^1, x^5, x^7, x^3) is not S-dominated by any S containing 3. Analogous arguments show that Condition 3 is satisfied.

We also need to show that x^0 is not S-dominated in P in order to demonstrate that Condition 4 of Coherence is satisfied. Consider any singleton coalition, say $\{1\}$. if 1 deviates to x^1 (its only possible deviation), then since (x^1, x^5, x^7, x^3) is in P, and $u_1(x^0) > u_1(x^3)$, x^0 is not 1- dominated in P. Analogous arguments hold for 2 and 3. Finally, note that there is no two-person coalition S and outcome in O that is preferred by S to x^0.

This shows that O is HSREFS. An obvious question is the role of history dependence. To see this, consider how x^0 is supported as part of the HSREFS. Suppose 1 deviates and chooses 1. Then, the objection path (x^1, x^5, x^7, x^3) is used to punish 1 for the initial deviation. Notice that this requires that $\{1, 2\}$ to deviate jointly at x^7. Now, suppose 2 deviates at x^0. Then, the objection path (x^2, x^4, x^7, x^1) is used to punish 2. This requires $\{2, 3\}$ to deviate at x^7. History independence (as in REFS or CECB) would not allow two different continuation paths at x^7, but is possible given history dependence.

Why Maximality Matters

This example illustrates the connection between Maximality and subgame perfection. While the specific model is based on how the Conservative party in the UK elects its leader, the same issues arise in other contexts.

The Conservative party uses a two-step procedure. In the first step, members of the Conservative party in the House of Commons select a panel of two candidates from the list of candidates seeking nomination if the latter set contains more than two candidates.[23] Successive rounds of voting are held if necessary, with the weakest candidate being eliminated in each round until only two candidates remain. In the second stage, all members of the party choose one candidate from the panel chosen by the members of parliament.

In our model, we assume that only 3 MPs -x, y and z- are seeking nomination. Hence, only one round of voting in the first stage is required to obtain a two-member panel. Let N_1 be the electorate in stage 1 and N_2 in stage 2. We assume $N_1 = \{1, 2, 3\}$ and $N_2 = \{L(eave), R(emain)\}$. The latter definition is meant to represent the fact that there are only two groups in N_2, with individuals in each group having unanimous preferences. While this is a simplifying assumption, it is not farfetched in the context of the 2016 election when Brexit seemed to be the only issue concerning voters.

Let $|L| > |R|$ so that L gets to decide the eventual winner.

[23]There were 5 candidates in the last contested election in 2016, and four in 2005.

For each $i \in N_1$, a strategy s_i is to eliminate one candidate.[24] A strategy for each $j \in N_2$ is a *function* s_j selecting a winner from each two-element subset of $\{x, y, z\}$.

Let preferences be as follows :

- xP_1yP_1z
- xP_2zP_2y
- zP_3xP_3y
- zP_LxP_Ly
- yP_RxP_Rz

Assume that players also have a ranking over all nonempty subsets of $\{x, y, z\}$, this ranking being consistent with preferences over singleton sets defined above and that any singleton set is preferred over any set containing more than one element.[25]

Clearly, the optimal or maximal strategy for L is to choose z from any panel containing z and to choose x otherwise. Given this, the subgame equilibrium path of play should be for $\{1, 2\}$ to eliminate z so as to ensure the eventual winner to be x. This must also be the unique HREFS. In fact, history dependence plays no role here and so several history-independent solutions like REFS and CECB that satisfy Maximality will coincide with HREFS.

However, the LCS can also be supported by $\{1, 2\}$ eliminating x or y, so that L chooses z. A deviation of $\{1, 2\}$ is deterred if $\{1, 2\}$ expect L to choose y from $\{x, y\}$. So, the LCS (and LCCS) will be $\{x, z\}$.

The next example uses Voting by Veto Procedures. Voting by Veto was first proposed by Mueller (1978) as a way of inducing preferences for public goods. Various forms of voting by veto procedures have been extensively analysed by H. Moulin[26] amongst many others. In the original procedure described by Mueller (1978), n individuals propose n alternatives. The proposed alternatives along with a status quo x_0 constitute the issue set. The voters then *sequentially* veto one alternative from the set that has not been vetoed already. Since there are $n + 1$ alternatives and n voters, exactly one alternative will escape a veto and this is declared the chosen outcome. In what follows, we will avoid the proposal stage since that does not change the nature of the result, and focus on the vetoing stage.

Let $N = \{1, 2, 3\}$ and $A = \{x_0, x, y.z\}$ be the issue set. Suppose the order of sequential vetoing is $1, 2, 3$. Then, individual 1's strategy is to pick one alternative from A (the

[24] Some tie-breaking rule, which is irrelevant for our purpose, is used to select the eliminated candidate if each i selects a different candidate.

[25] This is an "artificial assumption in order to fit the model with our general framework.

[26] See for instance Moulin (1983).

alternative vetoed by 1) , individual 2's strategy is to pick one alternative from each three-element subset of A, while 3 picks one alternative from each two-alternative subset of A.

Let individual preferences over A be

- $x_1 P_1 x_2 P_1 x_0 P_1 x_3$
- $x_0 P_2 x_1 P_2 x_2 P_2 x_3$
- $x_3 P_3 x_0 P_3 x_1 P_3 x_2$

Extend these preferences over all two-element and three-element subsets of A such that individual elements are strictly preferred to all supersets that can be constructed out of A. We do this because X will be the set of all non-empty subsets of A. Then, the unique HREFS will involve 1 vetoing x_0, 2 vetoing x_3 and 3 vetoing x_2 yielding the outcome x_1. However, the LCS (and LCCS) will also include x_0 if 1 vetoes x_3 under the pessimistic assumption 2 will not veto x_3. This is clearly a non-maximal move by 2 and so 1 should not expect this to take place.

It is easy to construct another preference profile where the farsighted stable set gives the wrong prediction because of a violation of maximality.

Applications 2 and 3 also illustrate the close connection between subgame perfection and maximality in our framework despite the cooperative nature of our solution concept. This point has also been emphasized by Kimya in the context of ECB and CECB.

8. PROPERTIES OF HREFS

In this section, we describe some results on the structure and properties 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 $Y^1 \cup Y^2$.*

Notice that this corollary establishes that there is a largest HREFS whenever X is finite.

We now show that HREFS is a refinement of Chwe's consistent sets. Even the largest (in terms of set inclusion) HREFS is a subset of the largest consistent set. Moreover, Example 2 demostrates that HREFS can be a strict subset of LCS. As we have remarked before, this makes HREFS a more attractive solution concept given the usual criticism of the LCS is that it is too permissive.

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

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

Then there is $x \in \mu(P)$, y and $S \in E(x, y)$ such that $u_S(z) \gg u_S(x)$, 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 (Definition 3) is not necessarily HREFS in abstract games since domination chains may violate maximality.

However, as we have demonstrated in Section 6, the problem of maximality disappears in simple games. This essentially yields the following.[27]

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.

9. 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.

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

9.1. **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.

We provide a constructive proof that HREFS is nonempty in all finite games.

For any set of objection paths P, define

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

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 dominated 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 lemma provides a condition under which UUD is a coherent collection.

LEMMA **4.** *Let $P = UUD$. If P_x is nonempty for all x, then P is a coherent collection of paths.*

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. ■

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 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 **5.** *Let X be finite. For all $x \in X$, there is p_x such that $p_x \in UUD$.*

The proof of the next theorem follows immediately from Lemma 4 and Lemma 5.

THEOREM **4.** *If X is finite, there is a non-empty HREFS.*

9.2. **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 superadditive partition function games. In view of the demanding nature of HSREFS, this nonemptyness 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 (N) to denote the embedded coalition $(N, \{N\})$.

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 a typical element 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 $\{i\} \in \pi$.[28]

We should specify the effectivity function associated with a partition function game. Take any initial state x and suppose some coalition S deviates from x. It makes sense

[28]This is without loss of generality.

to assume that S can choose any partition of Π_S, and that it cannot dictate how $N - S$ chooses a partition in Π_{N-S}. However, it is notationally complicated to explicitly formalise the effectivity function. Fortunately, for our purposes it suffices to consider only certain kinds of coalitional moves and so we 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.[29]

We assume the following.

Assumption 3. *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 to coalition $N - \{i\}$. Let

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

LEMMA **6.** *For all $(x, y, k) \in Z \times X \times N$, there is p_y such that $u_k(\mu(p_y)) \leq u_k(x)$.*

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 $\sum_{i \in N} u_i(y) \leq v(N)$. So, there is $y' \in X$ (possibly $y = y'$) such that $\sum_{i \in N} u_i(y') = v(N), u_i(y') \geq u_i(y)$ 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

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

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

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

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

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

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 **5.** *Z is an HSREFS.*

Proof. Take any $y \in X$. Choose arbitrary $x \in Z$ and $k \in N$. Lemma 6 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 6, 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 6, 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, $Z \subseteq W$.[30]

10. 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 (2017) 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 superadditive partition function 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.

11. Appendix

In this Appendix, we prove Lemma 5: for all $x \in X$, UUD contains some objection path originating from x.

Proof of Lemma 5

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

$$C(x, P) = \{y : (x) \text{ is } S(x) - \text{covered in } P \text{ via } y\}. \tag{7}$$

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,

$$C^*(x, P) = \{y \in C(x, P) : \max_{z \in C(x,P)} \min_{p \in P_z} u_{S(x)}(\mu[p)) \not\gg \min_{p \in P_y} u_{S(x)}(\mu[p))\}. \tag{8}$$

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

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 $C^*(\cdot, P)$. [31] Denote the set of maximal elements of $\overline{C}^*(\cdot, P)$ by $V(P) = \{x \in X : y \in \overline{C}^*(x, P)$ implies $x \in \overline{C}^*(y, P)$, for all $y\}$.

LEMMA **7.** *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 $P \subseteq P'$.*

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 (7) this contradicts the assumption that $y \in C^*(x,P)$. ■

LEMMA **8.** *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 7 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 **9.** *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 8, $(x_1, S(x_1), x_2, ..., x_L) \in UD^t$. 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 $x \in V(UD^t)$.

[31] 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: 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 7, $(x_0, S(x_0), p') \in UD^t$ 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 (8), 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,

$$\overline{C}(x, UD^t) \subseteq V(UD^t), \text{ for all } x \in V(UD^t). \tag{9}$$

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 (9), $v' \in V(UD^t)$. 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 8 and 9 that:

LEMMA **10.** *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 $(x_0, S(x_0), ..., x_J) \in UD^{t+1}_x$.*

This completes the proof of Lemma 5.

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.

Beal, S., Durieu, J., P. Solal (2008), "Farsighted Coalitional Stability in TU Games", *Mathematical Social Sciences*, **56**, 303–313.

Bhattacharya, A. and V. Brosi (2011), "An Existence Result for Farsighted Stable Sets of Games in Coalitional Form", *International Journal of Game Theory*, **40**, 393–401.

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

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

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

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

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

——— (2007), "Coalitions, Agreements and Efficiency", *Journal of Economic Theory*, **136**, 105–125.

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

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.

Granot, D. and E.Hanany (2016)," Subgame Perfect Farsighted Stability", mimeo.

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.

——— (2010), "Coalition Formation with Farsighted Agents", *Games*, **1**, 286–298.

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

Kimya, Mert (2017),"Equilibrium Coalitional Behavior", mimeo, Brown University.

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

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

Mariotti, M. (1997), " A Model of Agreements in Strategic Form Games", *Journal of Economic Theory*, **74** , 196–217.

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.

Mueller, Dennis (1978), "Voting by Veto", *Journal of Public Economics*, **10**, 57-75.

Moulin, H. (1983), *The Strategy of Social Choice*, North Holland : Elsevier.

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.