Maastricht University

Dominik Karos, Laura Kasper

Farsighted Rationality

RM/18/011

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O Box 616
NL- 6200 MD Maastricht
The Netherlands

Farsighted Rationality

Dominik Karos[1] and Laura Kasper[1,2]

[1] *School of Business and Economics, Maastricht University,*
d.karos@maastrichtuniversity.nl
[2] *Chair of Economic Theory, Saarland University, l.kasper@maastrichtuniversity.nl*

April 24, 2018

Abstract

Farsighted deviations are based on agents' abilities to compare the outcome of a farsighted deviation to the status quo. However, agents do not account for deviations by others in case they do *not* change the status quo; so, they are not fully farsighted. We use extended expectation functions to capture a coalition's belief about subsequent moves of other coalitions in both cases. We provide three stability and optimality axioms on coalition behavior and show that an expectation function satisfies these axioms if and only if it corresponds to an equilibrium of the abstract game that is stable with respect to coalitional deviations. We provide applications of our solution for games in characteristic function form and matching problems.

Keywords: abstract games, farsighted stability, expectation functions, coalition stable equilibrium

JEL: C71, C72

1 Introduction

An abstract game consists of a set of states (or outcomes), agents' utilities in each state, and an effectivity correspondence that specifies for any two states what coalitions can enforce a move from one to the other. In particular, it does not specify strategies for any player; specifically it abstracts away any strategic interaction. Put differently, abstract games specify *what* coalitions can achieve, but not *how*. Both cooperative and non-cooperative games can, therefore, be modeled as abstract games. Naturally, one can use both cooperative and non-cooperative instruments to solve these games.

Cooperative solutions of abstract games heavily rely on *dominance*: one state *dominates* another state if there is a coalition that (i) can implement a change from the latter to the former, and (ii) thereby achieves a strictly better outcome for all their members. The most prominent solutions based on dominance are, arguably, the stable set (von Neumann and Morgenstern, 1944) and the core (Gillies, 1959). Both these solutions make the same crucial assumptions about agents' reasoning, namely that agents are *myopic*: (a) whenever a coalition deviates to a new state, they expect to remain in that state; (b) whenever a coalition does not deviate to a new state, they expect to remain in the old state. That is, coalitions expect that no other coalition will ever implement a change. While assumption (a) has already been weakened in the literature on *farsighted* behavior, assumption (b) has been criticized (for instance by Chwe, 1994), but a convincing solution has not been offered yet. This paper attempts to do so, both in terms of a dominance relation and in terms of an equilibrium in a non-cooperative game that is stable with respect to coalitional deviations.

Myopic behavior is a severe restriction of agents' rationality, and the whole dilemma comes to light if one closely investigates the stable set of von Neumann and Morgenstern (1944). The stable set is a set of outcomes that satisfies *internal stability* – no outcome in the stable set dominates any other outcome in the stable set – and *external*

stability – every outcome that is not in the stable set is dominated by some outcome in the stable set. It is, however, possible that a state outside the stable set dominates a state inside the stable set. According to von Neumann and Morgenstern (1944) no coalition would implement this outside state because another coalition would implement a move back into the stable set right away. This argument is surely valid for *single-payoff* stable sets, i.e. stable sets in which all states have the same payoff vector. However, what if this is not the case? Harsanyi (1974) provided the following thought: a coalition S could deviate from a state x inside the stable set to a state y outside, in anticipation of another coalition's deviating from y to third state z inside. If z is preferred over x by all members of S, then S would actually be quite happy with this development, so there is an *indirect* dominance of z over x.

Chwe (1994) formalized this indirect dominance relation according to which coalitions only care about the final outcome of a chain of deviations. This takes care of criticism (a) above: coalitions do not expect to remain in the state they deviate to. Chwe's formulation was quite successful and in the sequel a whole branch of literature was based upon it: Xue (1998) endowed agents with expectations about each others' behavior based on whether they are optimistic or pessimistic, Diamantoudi and Xue (2003) consider farsightedness in hedonic games, Herings et al. (2009) apply farsightedness to problems of network formation, and Mauleon et al. (2011) are interested in stable sets of two-sided one-to-one matchings when players are farsighted.

All these stable sets rely on an indirect dominance relation according to which a state is dominated if there is some indirect dominance path to another state. However, none of them required that this final state is in fact not itself indirectly dominated. More recent approaches have explicitly tackled this problem. Jordan (2006) introduced *expectation functions* for abstract games. These function specify for each state what (unique) coalition moves to what (unique) state. If such a function is commonly known and accepted, then coalitions know exactly what will happen after any potential move, and a dominance relation can be based on this expectation. Of course

there should be some coherence between the expectation function itself and the dominance relation it creates, in the sense that moves should only be expected if they are beneficial. Dutta and Vohra (2017) introduce three attractive axioms an expectation function should satisfy, and the stationary points of such a function then constitute a form of stable set. Hence, as long as expectations are common knowledge, problem (a) above seems solved; the case of heterogeneous expectations has recently been explored by Bloch and van den Nouweland (2017).

Problem (b), however, persists: in the model of Dutta and Vohra (2017) agents compare the final outcome of a deviation to the status quo, making the implicit assumption that the status quo will remain if they do not deviate. So, agents are farsighted enough to consider what happens if they deviate, but not farsighted enough to consider what happens if they do not move (Chwe, 1994). The first goal of this paper is to adapt the model of Dutta and Vohra (2017) in order to solve this issue. An *extended expectation function* specifies for each state a list of coalitions and their moves, and each coalition knows that if they don't move, the next one will. This way coalitions can compare the result of their move to the scenario if they do not move. Of course, there might be different orders in which coalitions are allowed to move, and different orders might lead to different behavior and different stationary points. Yet, this non-uniqueness also emerges from specifying what coalition is allowed to move at a given state, if there are more than one coalition that could move according to the rules of the game. Hence, we might end up with different solutions, but each solution is supported by at least one specification of how the game is being played. We impose three axioms on these extended expectation functions and call a function *rational* if it satisfies all of them. First, the only reason for a coalition T not to move out of a state is that there is a different coalition S whose move out of this state is at least as beneficial for at least one member of T as T's move. This member would basically veto any such move and wait for S to move. Second, if a coalition does move, it must be strictly better for all members than not moving and leaving the floor to the next

coalition. Here, we explicitly address problem (b) by allowing a comparison between deviating and not deviating without the assumption that the latter would maintain the status quo. Third, if a coalition does move to some state, there is no other state they could move to which would be strictly better for all members. All these axioms are independent of the order in which coalitions move; however, they might be satisfied for some order, but not for others. Similar to Dutta and Vohra (2017) we are interested in the stationary points of these rational extended expectation functions.

Most criticism about dominance relations, including ours, is based on implied irrational behavior of coalitions or players. But this problem is very natural: dominance is defined in a very abstract setting without strategies, whereas the criticism is formulated in terms of behavior and, hence, strategies. It, therefore, seems natural to motivate a dominance relation by supporting it in a non-cooperative fashion. This shall be the second aim of our paper. An abstract game is, essentially, a coalition formation game, and the existing literature on non-cooperative coalition formation is quite rich (cf. Ray, 2007). The translation of an abstract game into a non-cooperative coalition formation game requires the specification of an extensive form, and, as Ray and Vohra (2017) point out, the solution might then very much depend on the choice of this extensive form.

We shall avoid this issue: instead of providing an extensive form in which players might have very complicated strategies, we endow *coalitions* with rather simple strategies. A coalition's strategy specifies for any state of the abstract game whether they will remain there or deviate to another state for which they are effective. This setup is very similar to Kimya (2015) who gives a non-cooperative foundation for the model of Dutta and Vohra (2017).

The translation between coalitions' strategies and extended expectation functions is straightforward. Suppose an order $\succeq$ in which coalitions are allowed to move in each state is given. Then for each state, one constructs a list of coalitions by selecting

all those coalitions whose strategies specify a move out of this state, and one orders this list according to $\succeq$. So, strategy profiles build a natural foundation of extended expectation functions.

We define a *best response* of a coalition S as a profile of strategies, one for each nonempty subcoalition of S, such that each proper subcoalition plays a best response, the strategy of S is such that it would not be vetoed by any player, and S could not implement any move that would be strictly better for all members. An *equilibrium* is then a strategy profile in which each coalition plays a best response, or equivalently, which is a best response for the grand coalition. Our main result is that an extended expectation function satisfies our axioms if and only if it corresponds to an equilibrium in the associated coalition formation game. Hence, we call the set of stationary points of an extended expectation function *equilibrium stable set*.

The structure of the paper is as follows. We introduce some necessary notation in Section 2. In Section 3 we describe rational expectations as proposed in Dutta and Vohra (2017) and introduce our extended expectation functions. In particular, we propose our axioms and illustrate how they relate to those in Dutta and Vohra (2017). In Section 4 we illustrate the difference between the equilibrium stable set, and other solutions proposed in the literature on abstract games. Section 5 develops our main result, namely the non-cooperative foundation of our axioms. In Section 6, we propose some simple applications of the equilibrium stable set.

2 Preliminaries

2.1 Notation

Let N be a finite set of players. Subsets $S \subseteq N$ are called *coalitions*. For $S \subseteq N$ write 2^S for the set of subsets of S, and $P(S)$ for the set of nonempty subsets. An *abstract game* is a tupel $\big(N, X, E, (U_i(\cdot))_{i\in N}\big)$, where X is the set of *outcomes* or *states*, $U_i : X \to \mathbb{R}$ is player i's utility function over states, and $E : X \times X \rightrightarrows 2^N$

is an *effectivity correspondence*. E specifies coalitions that have the power to replace one state by another one: for $x, y \in X$ the (possible empty) set $E(x, y)$ comprises all coalitions that can replace x with y. We assume that $E(x, x) = 2^N$, that is each coalition has the option not to change the status quo, and $\emptyset \in E(x, y)$ if and only if $x = y$.

2.2 Expectation Functions

Let $\left(N, X, E, (U_i(.))_{i \in N}\right)$ be an abstract game. An *expectation function* is a map $F : X \to X \times 2^N$ with $F(x) = (f(x), S(x))$ such that $S(x) \in E(x, f(x))$. F describes transitions from every state either to itself or to another state (if there is a coalition that is effective for such a transition): $f(x)$ is the state that is transitioned to and $S(x)$ is the coalition that implements the transition. We follow the standard convention and require $S(x) = \emptyset$ whenever $f(x) = x$. Given an expectation function F and a state $x \in X$, a *path* $P = P_F(x)$ is a finite or infinite sequence of states $(x^1, x^2, \ldots)$ with $x^1 = x$, $x^{k+1} = f\left(x^k\right)$ for all $k \in \mathbb{N}$, and $x^k \neq x^l$ for all $k \neq l$. P is *terminal* if there is $m < \infty$ such that $f(x^m) = x^m$.[1] In this case we write $t(P_F(x))$ to denote the *terminal node* of P_F, i.e. $t(P_F(x)) = x^m$. An expectation function F is *absorbing* if $P_F(x)$ is terminal for all $x \in X$.

3 Rational Expectations

3.1 Rational Expectation Functions

Let F be an expectation function. We assume that players only care about their utility in terminal nodes of F and that any terminal node is better than reaching no

[1] If P is not terminal, P is either infinite or cycling. In the latter case x^m is the last state before the cycle is closed.

terminal node at all (cf. Harsanyi, 1974). So, for $i \in N$ let

$$u_i(x, F) = \begin{cases} U_i(t(P_F(x))) & \text{if } P_F(x) \text{ is terminal} \\ -\infty & \text{otherwise.} \end{cases} \tag{1}$$

If F is absorbing then $u_i(x, F) = U_i(t(P_F(x)))$ for all $x \in X$. The following properties an absorbing expectation function may satisfy have been introduced by Dutta and Vohra (2017).

Internal Stability, I'. For all $x \in X$, if $f(x) = x$ then for all nonempty $T \subseteq N$ and all $y \in X$ with $T \in E(x, y)$ there is $i \in T$ such that $u_i(x, F) \geq u_i(y, F)$.

External Stability, E'. For all $x \in X$, if $f(x) \neq x$ then $u_i(x, F) > U_i(x)$ for all $i \in S(x)$.

Maximality, M'. For all $x \in X$, if $f(x) \neq x$ and there is $y \neq x$ with $S(x) \in E(x, y)$ then there is $i \in S(x)$ such that $u_i(x, F) \geq u_i(y, F)$.

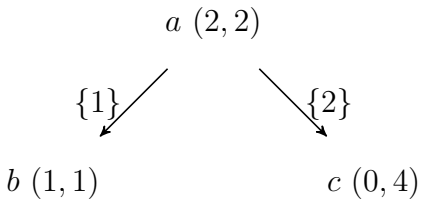
Strong Maximality, SM'. For all $x \in X$, if $f(x) \neq x$ and there are $y \neq f(x)$ and $T \subseteq N$ with $T \cap S(x) \neq \emptyset$ and $T \in E(x, y)$ then there is $i \in T$ such that $u_i(x, F) \geq u_i(y, F)$.

I' ensures that if no coalition moves out of a state x, then there is no coalition that could profit by doing so. **E'** ensures that a coalition only moves out of a state x if the terminal node is strictly preferred over x by all members.[2] **M'** requires that deviations are optimal; that is, the deviating coalition $S(x)$ has no better states to deviate to. Finally, **SM'** requires, additionally, that no subgroup of the deviating coalition $S(x)$ could improve by joining some other coalition instead of $S(x)$: such a subgroup would have no reason to support the move of $S(x)$.

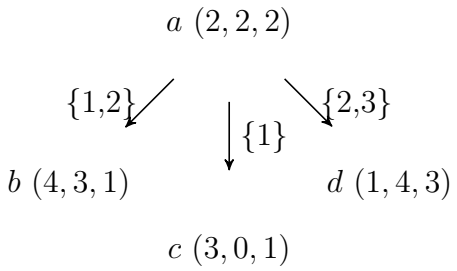
[2]Dutta and Vohra (2017) require that $P_F(x)$ is a farsighted objection against x. But since External Stability is required in all non-terminal nodes of the path $P_F(x)$, this is equivalent.

Figure 1: External stability and maximality revisited

(a) Should 1 move to b?



(b) Should 2 boycott a move to b?



An expectation function is called *rational* if it is absorbing and satisfies **I'**, **E'**, and **M'**; it is called *strongly rational* if it satisfies **I'**, **E'**, and **SM'**. The set of terminal nodes of a rational expectation function is called a *rational expectations farsighted stable set* (REFS), and the set of terminal nodes of a strongly rational expectation function is called a *strongly rational expectations farsighted stable set* (SREFS).

While **I'** and **M'** seem rather uncontroversial, **E'** might be more critical: the condition that $u_i(x, F) > U_i(x)$ for all $i \in S(x)$ makes the implicit assumption that the players in $S(x)$ expect the status quo to be sustained in case they don't move. But, putting it in the words of Chwe (1994):

> *This is clearly inconsistent. A coalition is farsighted enough to consider what further moves other coalitions will make once it moves, but does not consider what other coalitions will do if it does not move.*

Example 3.1. Let $N = \{1, 2\}$, $X = \{a, b, c\}$ and consider the effectivity correspondence and payoffs depicted in Figure 1a. Suppose that $F(a) = (b, \{1\})$. This clearly violates **E'**, as $\{1\}$ reduces her payoff from 2 to 1. Nevertheless, if she stayed in a, it is very likely that $\{2\}$ takes the opportunity and moves to b, reducing 1's payoff even further. Hence, $\{1\}$'s move from a to b is the choice of the lesser evil. □

In the previous example agent 1's optimal move crucially depends on her expectation about agent 2's behavior *if she did not move*. But these "counterfactuals" cannot be described by an expectation function.

Another, yet similar, issue arises when we inspect **SM'** more closely. A move by a coalition S is forbidden if there is a coalition T that intersects with S and that could move to a different state y which (eventually) is preferred over $f(x)$ by all members of T. So, in this instance the members of T are expected to compare their payoffs from y to their payoffs from $f(x)$ rather than x, in contrast to the expected reasoning behind **E'**.[3] Moreover, it is assumed that in this case the members of $S \cap T$ will boycott the move of S and instead try to implement, together with $T \setminus S$, the move to y. But is it clear that such a boycott brings the expected result?

Example 3.2. Let $N = \{1, 2, 3\}$, $X = \{a, b, c, d\}$ and consider the effectivity correspondence and payoffs depicted in Figure 1b. **SM'** requires $F(a) \neq (b, \{1, 2\})$ on the grounds that 2 could, together with 3, beneficially implement a move to d. However, if 2 boycotts b, she can by no means be sure that d will actually be implemented as $\{1\}$ might move to c instead. □

The optimal behavior in the previous example depends, again, on counterfactuals: what happens if $\{1, 2\}$ does not implement b? In the setup of Dutta and Vohra (2017) agent 2 pays no attention to this possibility, as she assumes that she will be able to implement d together with 3. An expectation function as defined above cannot incorporate such expectations. In order to analyze farsighted behavior it is, however, essential to have expectations about what happened if a coalition *did not* change the status quo. We shall therefore extend expectation functions so as to incorporate these counterfactuals in the next subsection.

[3] This issue of maximality has recently been tackled by Ray and Vohra (2017).

3.2 Extended Expectation Functions

An *extended expectation function* is a map F that maps each $x \in X$ to an ordered list $\left(F^1(x), \ldots, F^{k(x)}(x)\right)$ such that $F^l(x) = \left(f^l(x), S^l(x)\right)$ with $S^l(x) \in E\left(x, f^l(x)\right)$ for all $l = 1, \ldots, k(x)$, $f^l(x) \neq x$ for all $l \neq k(x)$, and $S^{k(x)}(x) = \emptyset$. (This implies $S^l(x) \neq \emptyset$ for $l \neq k(x)$ and $f^{k(x)}(x) = x$.) Both the length and the order of this list might depend on x. These ordered lists allow us to incorporate the counterfactuals mentioned above: if $S^1(x)$ did not move to $f^1(x)$ then $S^2(x)$ would move to $f^2(x)$.

Note that for any extended expectation function F the map F^1 is a (normal) expectation function, which reflects the "true behavior" in each node. In particular, F induces the path $P_{F^1}(x)$ for all $x \in X$. As there is no danger of confusion, we shall simply write $P_F(x)$ to refer to this path, even if F is an extended expectation function. In particular, for an extended expectation function F we shall write $u_i(x, F)$ for $u_i(x, F^1)$.

In the following we adapt the axioms of Dutta and Vohra (2017) so that they apply to extended expectation functions. It is worth mentioning, though, that with the definition of the utility function in Equation (1), these axioms are well defined for F even if F^1 is not absorbing.

Internal Stability, I. For all $x \in X$ and all coalitions $T \notin \left\{S^1(x), \ldots, S^{k(x)}\right\}$ there is $l \leq k(x)$ such that for each $y \in X$ with $T \in E(x, y)$ there is $i \in T$ with $u_i\left(f^l(x), F\right) \geq u_i(y, F)$.

External Stability, E. For all $x \in X$ and for all $l = 1, \ldots, k(x) - 1$ it holds that $u_i\left(f^l(x), F\right) > u_i\left(f^{l+1}(x), F\right)$ for all $i \in S^l(x)$.

Maximality, M. For all $x \in X$ and for all $l = 1, \ldots, k(x) - 1$ it holds that if there is $y \neq f^l(x)$ such that $S^l(x) \in E(x, y)$ then there is $i \in S^l(x)$ with $u_i\left(f^l(x), F\right) \geq u_i(y, F)$.

Internal stability is a requirement on all coalitions T that do not want to move out of a state x: there must be a move by some coalition $S^l(x)$, such that whatever

change T could implement, at least one member would not agree to that change. External stability takes into consideration our previous discussion. A coalition $S^l(x)$ that moves out of x compares the final outcome this move entails not to the payoff vector obtained in x, but to the payoff vector that would be reached if the next coalition, $S^{l+1}(x)$, got to move. Maximality requires that any moving coalition moves to an optimal state, in the sense that there is no other state they could move to which would make all members strictly better off.

An extended expectation function satisfying **I**, **E**, and **M** is called *rational*. A *stationary point* of an extended expectation function F is a state x such that $f^1(x) = x$. The set of stationary points of a rational extended expectation function F, denoted by $\mathcal{S}(F)$, is said to be an *equilibrium stable set* (ESS). A justification for the name will be given in Section 5, where we shall provide a non-cooperative foundation for rational extended expectation functions.

Note that **E** forbids two coalitions who move right after one another to implement the same terminal node. In this case the first coalition would simply not move at all and instead leave the floor to the second one. However, different coalitions can make the same move if there are other coalitions moving between them.

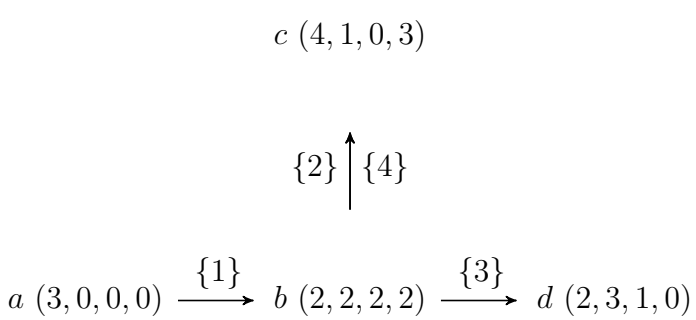
Example 3.3. Consider the game depicted in Figure 2. Consider F with $F^1(a) = (b, \{1\})$, $F^1(b) = (c, \{2\})$, $F^2(b) = (d, \{3\})$, and $F^3(b) = (c, \{4\})$. This extended expectation function[4] satisfies **E**, although both $\{1\}$ and $\{4\}$ (would) make the same move out of b. Thus, $\{c, d\}$ is an ESS. □

We first show that our axioms are an appropriate adaption of those proposed by Dutta and Vohra (2017).

Lemma 3.4. *Let F be an absorbing extended expectation function. If F satisfies* **I** *then F^1 satisfies* **I'***; if F satisfies* **M** *then F^1 satisfies* **M'***; if F satisfies* **E** *and*

[4] F is uniquely defined as the effectivity correspondence E does not allow any more moves. Throughout the paper we shall specify F only at those states $x \in X$ where $E(x, y) \neq \emptyset$ for some $y \neq x$, i.e. at those states for which non-trivial moves are allowed.

Figure 2: External stability



$k(x) \leq 2$ *for all* $x \in X$ *then* F^1 *satisfies* **E'**.

Proof. Since F is absorbing, $P_F(x)$ has a terminal node for each $x \in X$. Suppose F satisfies **I**, and let $x \in X$ be such that $x = f^1(x)$. If $y \neq x$ and $T \in E(x, y)$ then $u_i(x, F) = u_i(f^1(x), F) \geq u_i(y, F)$ for some $i \in T$ by **I**. So F^1 satisfies **I'**.

Let F satisfy **M**, let $x \in X$, and let $y \neq f^1(x)$ be such that $S^1(x) \in E(x, y)$. Then there is $i \in S^1(x)$ with $u_i(f^1(x), F^1) \geq u_i(y, F^1)$ by **M**. So, F^1 satisfies **M'**.

Let F satisfy **E** and let $x \in X$. If $k(x) = 1$, there is nothing to show. If $k(x) = 2$ then $f^2(x) = x$ and $u_i(f^1(x), F^1) > u_i(f^2(x), F^1) = U_i(x)$. So, F^1 satisfies **E'**. ■

So, while an extended expectation function specifies much more than an expectation function, the first layer, namely F^1, is an expectation function; and this function inherits the properties **I'** and **M'** if F satisfies **I** and **M**. On the other hand, **E'** will, in general, not be satisfied, but this is exactly what we wanted. We close this section by briefly illustrating that the issues raised in Examples 3.1 and 3.2 can be resolved with this approach.

Example 3.5. Recall Example 3.1 and Figure 1a. An extended expectation function that satisfies **I** must prescribe a move of $\{2\}$ to c, irrespective of $\{1\}$'s move at a. So, we either have $F(a) = ((b, \{1\}), (c, \{2\}), (a, \emptyset))$ or $F(a) = ((c, \{2\}), (a, \emptyset))$. An

extended expectation function with $F(a) = ((c, \{2\}), (b, \{1\}), (a, \emptyset))$, however, would violate **E**, as in this case $\{1\}$ could maintain her payoff in a by not moving (if she ever had the chance to move).

Recall Example 3.2 and Figure 1b. An extended expectation function that satisfies all three axioms could have $F(a) = ((b, \{1, 2\}), (c, \{1\}), (d, \{2, 3\}))$. **E** is not violated since any attempt of $\{2\}$ to boycott a move from a to b would ultimately result in c. Yet, $F(a) = ((b, \{1, 2\}), (d, \{2, 3\}), (c, \{1\}))$ violates **E** as in this case 2 could beneficially veto b and achieve outcome d. □

The foregoing examples illustrate the crucial effect of the order in which coalitions can move at a given state. We will shed more light on this order in Section 5 where we provide a non-cooperative foundation. At this moment we only note that it is intrinsic to the extended expectation function F, and our axioms are independent of it.

4 Examples

In this section we will briefly illustrate how the ESS compares to other solutions that have been proposed in the literature and that are based on farsighted dominance relations. Let $x, y \in X$. State y *farsightedly dominates* x under the effectivity correspondence E, $y \succ^f x$, if there is a sequence $y^0, (y^1, S^1), \ldots, (y^m, S^m)$, with $y^0 = x$ and $y^m = y$, such that for all $k = 1, \ldots, m$

$$S^k \in E(y^{k-1}, y^k) \qquad \text{and} \qquad U_i(y) > U_i(y^{k-1}) \text{ for all } i \in S^k.$$

Chwe (1994) adapted the stability axioms of von Neumann and Morgenstern (1944) by using this farsighted dominance relation:

Farsighted Internal Stability. A set $Y \subseteq X$ is *farsightedly internally stable* if there do not exist $x, y \in K$ such that $y \succ^f x$,

Farsighted External Stability. A set $Y \subseteq X$ is *farsightedly externally stable* if for every $x \in X \setminus Y$ there is $y \in Y$ such that $y \succ^f x$.

A set $Y \subseteq X$ is a *farsighted stable set* if it is both farsightedly internally stable and farsightedly externally stable. A set $Y \subseteq X$ is *consistent* if for all $y \in X$ and all $S \subseteq N$ with $S \in E(x, y)$ there is $z \in Y$ such that $z = y$ or $z \succ^f y$, and $u_i(z) \leq u_i(x)$ for some $i \in S$. Chwe (1994) shows that for any abstract game there is (with respect to set inclusion) a unique *largest consistent set* (LCS). We shall use the remainder of this section to show that the equilibrium stable set we propose in this article is different from the farsighted stable set, the largest consistent set, and the (strongly) rational expectation farsighted set.

Example 4.1. Consider the game given in Figure 3.[5] The farsighted stable set contains states c and d as these are terminal nodes of the game from which no coalition can deviate. States a and b are not contained in the farsighted stable set as both are farsightedly dominated by c. The only rational (extended) expectation function is as follows: at state a coalition $\{1\}$ is expected to stay in a while in state b coalition $\{2\}$ is expected to move to c. Thus, ESS, SREFS, and LCS are given by $\{a, c, d\}$. This set, however, violates farsighted internal stability. □

Example 4.2. Consider the game depicted in Figure 4. We construct two different ESS's. First let $F^1(b) = (a, \{1\})$, $F^2(b) = (c, \{2\})$, $F^1(e) = (f, \{2\})$, and $F^2(e) = (d, \{1\})$. Then $F^1(c) = (e, \{2\})$ and $F^2(d) = (b, \{1\})$ by **I**. Although $\{1\}$'s move at b leads to a lower payoff than 1 would obtain in b, the expectation of $\{2\}$'s hypothetical future behavior makes this move beneficial. So, F is a rational extended expectation function with $\mathcal{S}(F) = \{a, f\}$. This ESS, however, is neither a REFS nor externally stable as neither b nor d are farsightedly dominated.

Let now $F^1(b) = (c, \{2\})$, $F^2(b) = (b, \emptyset)$, $F^1(e) = (d, \{1\})$, and $F^2(e) = (e, \emptyset)$. Then F with $F^1(x) = (x, \emptyset)$ for all $x \neq b, e$ is another rational extended expectation

[5]The example is taken from Dutta and Vohra (2017).

Figure 3: Violation of internal farsighted stability.

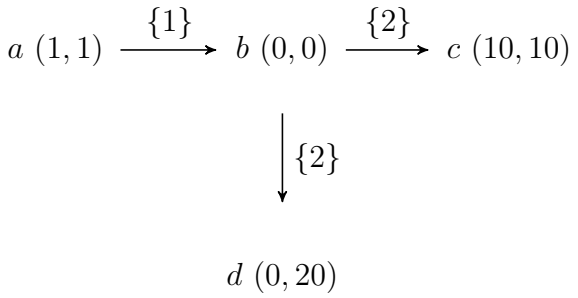
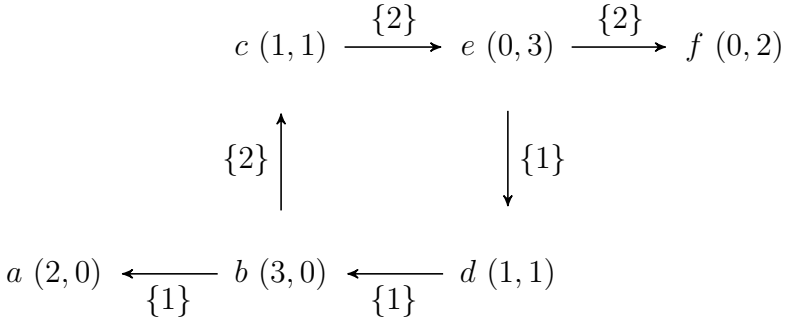


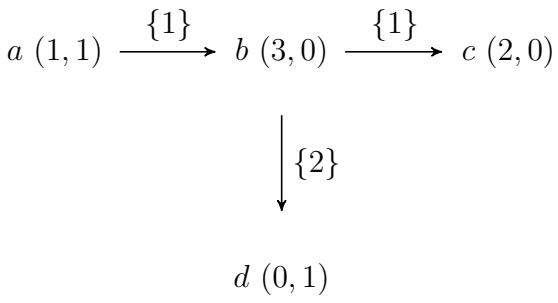
Figure 4: Violation of farsighted external stability.



function with $\mathcal{S}(F) = \{a, c, d, f\}$. As in this extended expectation function there is at most one coalition that moves in every node, the set of terminal nodes is also a SREFS by Lemma 3.4. □

Example 4.3. Consider the game depicted in Figure 5. By **I** coalition $\{2\}$ must move from b to d. Hence, one rational extended expectation function is given by $F^1(b) = (\{1\}, c)$ and $F^2(\{2\}, d)$. In this case $\mathcal{S}(F) = \{c, d\}$. However, $\{1\}$'s move at b violates **E'**, and **I'** requires $\{2\}$ to move from b to d. So, at state a coalition $\{1\}$ is expected to stay in a by **E'**. Therefore, the SREFS is $\{a, c, d\}$. This is identical to the farsighted stable set as well as the LCS, and could be supported as an ESS if $\{2\}$ was allowed to move before $\{1\}$ at b. □

Figure 5: ESS vs. REFS (SREFS).



We have seen that our solution is different from those proposed in the literature thus far. So, what do we gain from another solution? The answer comes with the next section where we show that rational extended expectation functions are exactly those that emerge from a strategy profile in the original game which is stable with respect to coalition deviations.

5 Non-Cooperative Foundation

Let G be an abstract game and let $\emptyset \neq S \subseteq N$. A *strategy* of coalition S is a map $\sigma_S : X \to X$ with $S \in E(S, \sigma_S(x))$ for all $x \in X$. Denote the set of S's strategies by Σ_S. Let $\sigma^0 : X \to X$ be defined as $\sigma^0(x) = x$ for all $x \in X$ and observe that $\sigma^0 \in \Sigma_S$ for all $S \subseteq N$. To avoid trivialities and technical issues let $\Sigma_\emptyset = \{\sigma^0\}$. A *strategy profile* is a vector $(\sigma_S)_{S \subseteq N} \in \times_{S \subseteq N} \Sigma_S$ of strategies, and the set of all strategy profiles is denoted by Σ.

5.1 Strategies and Extended Expectation Functions

A strategy profile $\sigma \in \Sigma$ induces some expectations among players, namely that each coalition moves according to their strategy. However, this expectation does not uniquely define an extended expectation function as at any state several coalitions

might wish to move. In order to define an extended expectation function F_σ that corresponds to σ, we assume that at each state x coalitions are allowed to move according to a linear order $\succeq^x$ such that $S \succeq^x \emptyset$ for all $S \subseteq N$, where $S \succeq^x T$ denotes that S moves before T at x. We then define $F_\sigma(x) = \left(F_\sigma^1(x), \ldots, F_\sigma^{k(x)}(x)\right)$ such that

1. $T \in \left\{S^1(x), \ldots, S^{k(x)-1}(x)\right\}$ if and only if $\sigma_T(x) \neq x$,
2. $S^l(x) \succeq^x S^{l+1}(x)$ for all $l = 1, \ldots, k(x) - 1$,
3. $f^l(x) = \sigma_{S^l(x)}(x)$ for all $l = 1, \ldots, k(x) - 1$,
4. $S^{k(x)}(x) = \emptyset$ and $f^{k(x)}(x) = x$.

Note that this construction is unique for any $\sigma \in \Sigma$ and any order profile $\succeq = (\succeq^x)_{x \in X}$. In order to keep notation simple we shall write F_σ, bearing in mind that F_σ depends on $\succeq$ as well. The order profile $\succeq$ shall be common knowledge. It is clear that F_σ is an extended expectation function, but F_σ need neither be absorbing nor satisfy any of our axioms.

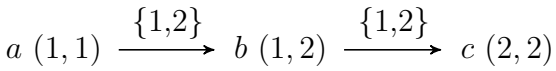
5.2 Better Responses

We shall now focus on strategy profiles. In order to keep notation simple, we shall write $u_i(\sigma, x)$ for $u_i(x, F_\sigma)$. Let $\emptyset \neq S \subseteq N$, let $\sigma_S, \sigma'_S \in \Sigma_S$ and $\sigma_{-S} \in \Sigma_{-S} = \times_{T \neq S} \Sigma_T$. Strategy σ_S is a *better response than* σ'_S *against* σ_{-S} *at* x if

$$u_i((\sigma_S, \sigma_{-S}), x) > u_i((\sigma'_S, \sigma_{-S}), x) \tag{2}$$

for all $i \in S$. Say that σ_S is *a better response than* σ'_S *against* σ_{-S} if it is a better response at all $x \in X$ with $\sigma_S(x) \neq \sigma'_S(x)$. The latter requirement is similar to the definition of *coalition deviations* in Kimya (2015). While it seems strong at first sight, the following example illustrates its importance.

Figure 6: Better responses



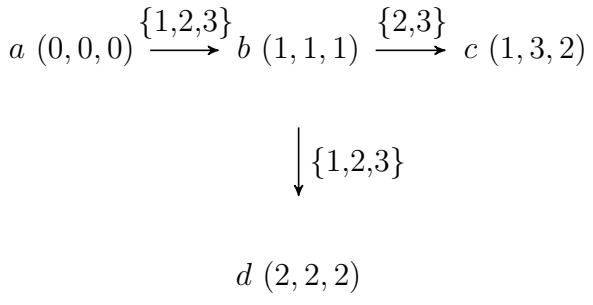
Example 5.1. Consider the game depicted in Figure 6 and strategy profile σ with $\sigma_S = \sigma^0$ for all $S \subseteq N$. At a coalition $\{1, 2\}$ could strictly improve by switching to σ' with $\sigma'_{\{1,2\}}(a) = b$ and $\sigma'_{\{1,2\}}(b) = c$. However, if we consider only that part of the game that starts at b, we would not claim that the coalition's switching from b to c is a profitable deviation, as 2 does not profit from it. So, σ' is not a better response than σ. □

One could argue that in above example there is no reason for 2 not to join 1 in implementing σ'. And one could define better responses slightly differently in order to avoid such an argument, namely by requiring that for all $x \in X$ the weak version of Inequality (2) is satisfied for all $i \in S$ and its strict version is satisfied for some $i \in S$. But we want to make our solution comparable to other notions of dominance, and they are typically defined in terms of strict inequalities for all members of the deviating coalition. We therefore want to stress that the proof of the next lemma would also hold for the alternative definition.

Lemma 5.2 (**One-shot deviation principle**). *Let $\emptyset \neq S \subseteq N$, let $\sigma_S, \sigma'_S \in \Sigma_S$ and $\sigma_{-S} \in \Sigma_{-S}$. If σ_S is a better response against σ_{-S} than σ'_S then there are $x^* \in X$ and $\sigma^*_S \in \Sigma_S$ such that σ^*_S is a better response against σ'_S as well, and $\sigma^*_S(x) = \sigma'_S(x)$ for all $x \neq x^*$.*

Proof. Let $\sigma_S, \sigma'_S \in \Sigma_S$ be as defined and let $x \in X$ be such that $u_i\left((\sigma_S, \sigma_{-S}), x\right) > u_i\left((\sigma'_S, \sigma_{-S}), x\right)$ for all $i \in S$. Such $x \in X$ exists by the definition of σ_S, and $u_i\left((\sigma_S, \sigma_{-S}), x\right) > -\infty$ for all $i \in S$. Let $(x^1, \ldots, x^m)$ be the path that is implemented by (σ_S, σ_{-S}) starting at x; that is, $x^1 = x$, $x^m = t\left(P_{F_{(\sigma_S,\sigma_{-S})}}(x)\right)$, with $F_{(\sigma_S,\sigma_{-S})}$ as

Figure 7: Better responses and objected strategies



defined in Subsection 5.1. This path is finite as $u_i\left(\left(\sigma_S, \sigma_{-S}\right), x\right) > -\infty$ for all $i \in S$. Let $h \in \{1, \ldots, m\}$ be such that $\sigma_S\left(x^h\right) \neq \sigma'_S\left(x^h\right)$ and $\sigma_S\left(x^l\right) = \sigma'_S\left(x^l\right)$ for $l > h$. Let $x^* = x^h$. By construction σ_S is a better response at x^* than σ'_S. Define now σ^*_S by $\sigma^*_S(x) = \sigma'_S(x)$ for all $x \neq x^*$ and $\sigma^*_S(x^*) = \sigma_S(x^*)$. Then σ^*_S is a better response than σ'_S. ■

Example 5.3. Let $N = \{1, 2, 3\}$, $X = \{a, b, c, d\}$, and for all $x \in X$ let $\succeq^x$ be such that $S \succeq^x T$ whenever $T \subseteq S$. Figure 7 depicts an effectivity correspondence and payoffs for all players. Coalition $\{1, 2, 3\}$ is always the first coalition to move, and irrespective of the strategy of coalition $\{2, 3\}$ there is no better response than their moving from a to b and from b to d. [6] □

5.3 Objections

We have carefully avoided the term "best response" thus far, as the strategy in the previous example has a severe caveat: player 2 has no reason to support the joint

[6] Given the effectivity correspondence in Figure 7, $\{1, 2, 3\}$ has no choices in c and d, so the behavior at a and b already defines a full strategy. In order to keep notation simple, we will leave out all trivial parts of strategies in the remainder of the paper.

move from b to d. By vetoing such a decision,[7] the order $\succeq$ would guarantee that $\{2,3\}$ would make their move next, achieving c with payoff 3 for 2. In order to formalize this idea, we need a little more notation. For $\emptyset \neq S \subseteq N$, $\sigma_S \in \Sigma_S$, and $x \in X$ define $\sigma_S^{0,x}$ by

$$\sigma_S^{0,x}(y) = \begin{cases} \sigma_S(y) & \text{if } y \neq x, \\ x & \text{if } y = x. \end{cases}$$

Note that $\sigma_S^{0,x} \in \Sigma_S$ for all $\sigma_S \in \Sigma_S$ as $S \in E(x,x)$ for all $x \in X$. The only (potential) difference between $\sigma_S^{0,x}$ and σ_S is that the former will map x on itself. We say that agent $i \in S$ *objects* σ_S *at* x *given* σ_{-S} if $\sigma_S(x) \neq x$, $\sigma_T(x) = x$ for all $T \succeq^x S$, and

$$u_i\left(\left(\sigma_S^{0,x}, \sigma_{-S}\right), x\right) \geq u_i\left(\left(\sigma_S, \sigma_{-S}\right), x\right).$$

The idea behind objections is that a player will not join S in implementing σ_S if she cannot profit by doing so. We say that agent $i \in S$ *objects* σ_S if there is some $x \in X$ such that i objects σ_S at x. Say σ_S is *objected* if it is objected by some $i \in S$.

Example 5.4. Given the order in Example 5.3 the strategy of coalition $\{1,2,3\}$ is objected at b by player 2. □

In the definition of an objected strategy σ_S against σ_{-S} at x we have imposed the condition that $\sigma_T(x) = x$ for all $T \succeq^x S$. The reason is that if a coalition S were allowed to move at x only after some other coalition T has already moved out of x, then S's behavior at x is (trivially) payoff-irrelevant. Nevertheless, S might strictly benefit from σ_S at x if they had a chance to actually implement it. We have to carefully distinguish between strategies that are not profitable at x (compared to σ_S^0) because their implementation would not be profitable, and those that cannot be

[7] We make the implicit assumption here that a coalition S can implement a move from x to $y \neq x$ only if all members unanimously agree on it. This is a standard assumption in the literature on dominance relations, and the consequence is that any member of S can veto any move out of x.

implemented because of the order $\succeq^x$. The former strategies are objected, the latter will be investigated more closely in Subsection 5.4.

Similarly to Lemma 5.2 we next formulate a one-shot deviation principle for unobjected better responses.

Lemma 5.5. *Let $\emptyset \neq S \subseteq N$, let $\sigma_S, \sigma'_S \in \Sigma_S$ and $\sigma_{-S} \in \Sigma_{-S}$ such that σ'_S is unobjected against σ_{-S}. If σ_S is an unobjected better response against σ_{-S} than σ'_S then there are $x^* \in X$ and $\sigma^*_S \in \Sigma_S$ such that σ^*_S is an unobjected better response against σ_{-S} as well, and $\sigma^*_S(x) = \sigma'_S(x)$ for all $x \neq x^*$.*

Proof. Construct x^* and σ^*_S as in the proof of Lemma 5.2. For each $x \in X$ it holds that either $u\left(\left(\sigma^*_S, \sigma_{-S}\right), x\right) = u\left(\left(\sigma'_S, \sigma_{-S}\right), x\right)$ or $u\left(\left(\sigma^*_S, \sigma_{-S}\right), x\right) = u\left(\left(\sigma_S, \sigma_{-S}\right), x\right)$. In the former case σ^*_S is unobjected at x as σ'_S is; in the latter case σ^*_S is unobjected at x as σ_S is. ■

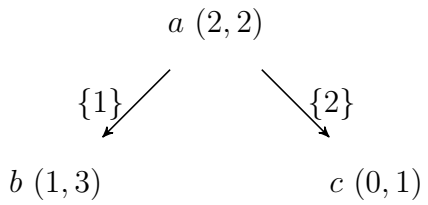
Note that the construction in Lemma 5.5 relies on σ'_S's being unobjected. In particular, it is not true that for all σ'_S for which an unobjected better response exists, there is an unobjected better response that deviates from σ'_S at only one state $x^* \in X$.

5.4 Incredible Threats

As there is a fixed order in which coalitions are allowed to move, there might be a state x and a coalition S, such that S will not have the opportunity to move at x given strategy profile σ_{-S}. Coalition S's action at x will thus never be implemented and does not affect payoffs. The following example illustrates the resulting issues.

Example 5.6. Let $N = \{1, 2\}$, $X = \{a, b, c\}$ and $\{1\} \succeq^x \{2\}$ for all $x \in X$, and consider the effectivity correspondence and payoffs depicted in Figure 8. Let $\sigma_{\{2\}}(a) = c$. In this case, $\{1\}$ should play $\sigma_{\{1\}}(a) = b$ to secure at least some positive payoff. And if she does so, $\{2\}$'s action at a is payoff-irrelevant, as she will never move anyway. So, there is no better response for $\{2\}$ against $\sigma_{-\{2\}}$. □

Figure 8: What should 1 do if 2 threatens to move to c?



The reader will have realized that this example is closely related to the problem of subgame perfect equilibria. The issue, of course, is that in the setup of an abstract game there is no initial node, there is no order in which the game is being played, and cycles are possible. So, in order to capture the idea of a subgame in this context, we have to work a little. For $x \in X$ and $\emptyset \neq S \subseteq N$ let

$$E^{S,x}(x', y) = \begin{cases} \{T \in E(x', y) : S \succeq^x T\} & \text{if } x' = x \\ E(x', y) & \text{otherwise.} \end{cases}$$

The effectivity correspondences E and $E^{S,x}$ are almost identical, the only difference being that according to $E^{S,x}$ those coalitions that are allowed to move prior to S at x are not effective for any move out of x. As they have to stay in x, S can actually implement their move. The corresponding abstract game $G^{S,x}$ is defined as $G^{S,x} = \left(N, X, E^{S,x}, (U_i(.))_{i \in N}\right)$. As the effectivity correspondence $E^{S,x}$ is different, the strategy space in $G^{S,x}$ is different as well: denote by $\Sigma^{S,x}$ the strategy space for this abstract game, and observe that $\Sigma_S^{S,x} = \Sigma_S$, that is S has the same strategy space as before. For $\sigma \in \Sigma$ let the strategy profile $\sigma^{S,x} \in \Sigma^{S,x}$ be defined by

$$\sigma_T^{S,x}(y) = \begin{cases} y & \text{if } y = x \text{ and } T \succeq^x S \\ \sigma_T(y) & \text{otherwise.} \end{cases}$$

That is, the only change between σ_T and $\sigma_T^{S,x}$ lies in the behavior of T at x if T is allowed to move before S at x: in this case T does not move out of x. Again, it is

easy to see that $\sigma_S^{S,x} = \sigma_S$ for all $S \subseteq N$ and all $x \in X$.

In some sense we can interpret $G^{S,x}$ as a subgame of the abstract game G. The game $G^{S,x}$ allows S's move at x to be implemented, as S is the first coalition to move. This construction allows us to prescribe "sensible" behavior even in states where a coalition does, originally, not move. In particular, we will later impose the condition that σ_S is not only a "good" strategy against σ_{-S}, but that also against $\sigma_{-S}^{S,x}$ at all x. This roughly corresponds to the idea of subgame perfection for coalitions.

It is worth mentioning, though, that state x might be reached more than once as a strategy profile σ might induce cycles – a phenomenon that can not appear in a standard dynamic game. The definition above ensures that even if x is reached a second time in $G^{S,x}$, it is again S who has to move first. That means that S can not throw the ball back to a coalition T that would, in the original game G, move prior to S at x.

Example 5.7. Recall Example 5.6 and Figure 8. In the game $G^{\{2\},a}$ any strategy of $\{2\}$ with $\sigma_{\{2\}}(a) = a$ is a better response than any strategy with $\sigma_{\{2\}}(a) = c$. □

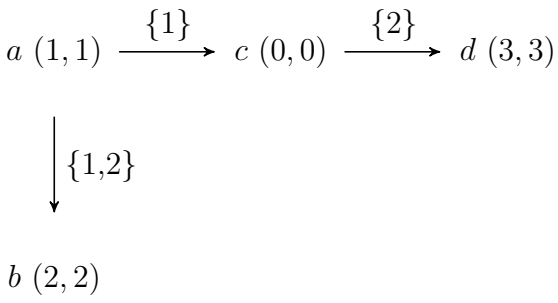
5.5 Best Responses

Defining best responses for coalitions brings the difficulty that various subcoalitions might implement certain actions in different states. The next example illustrates this point

Example 5.8. Let $N = \{1, 2\}$, $X = \{a, b, c, d\}$, for all $x \in X$ let $\succeq^x$ be given by $S \succeq^x T$ if $T \subseteq S$, and consider the effectivity correspondence and payoffs depicted in Figure 9. Then for coalition $\{1, 2\}$ it is optimal to remain in a such that $\{1\}$ can move to c and $\{2\}$ can move to d. □

We are now ready to define best responses. During the preceding discussions we have identified four requirements that one would expect: a best response for a coalition S must be a profile of strategies, one for each subcoalition of S (see Example 5.8);

Figure 9: What should $\{1,2\}$ do?



the strategy must be unobjected, even in those states where S will never be allowed to actually execute their strategy (see Examples 5.6 and 5.7); there must not be any (unobjected) better response (see Examples 5.3); and it must be optimal for each subcoalition of S as well (see Example 5.4). The following definition puts these requirements together.

Definition 5.9. Let $\emptyset \neq S \subseteq N$ and let the order profile $\succeq$ be given. A strategy profile $\sigma_{P(S)}$ is a best response against σ_{-2^S} if

1. for each $x \in X$ strategy σ_S is unobjected against $\sigma_{-S}^{S,x}$ at x in the game $G^{S,x}$,
2. for each $x \in X$ and each $\tau_S \in \Sigma_S$ that is a better response than σ_S against $\sigma_{-S}^{S,x}$ at x in the game $G^{S,x}$ there is $i \in S$ who objects τ_S,
3. σ_T is a best response for all $\emptyset \neq T \subsetneq S$.

The following example illustrates how best responses might depend on $\succeq$.

Example 5.10. Recall Example 5.8 and Figure 9. If $\{1\}$ moves before $\{1,2\}$ at a then $\sigma_{\{1,2\}}(a) = b$ a tribute to Condition 3 in Definition 5.9. □

5.6 Equilibrium

Definition 5.11. Let the order profile $\succeq$ be given. A strategy profile σ is an *equilibrium* (with respect to $\succeq$) if for all nonempty coalitions S the profile $\sigma_{P(S)}$ is a best response against σ_{-2^S}.

Recall that the definition of best responses is recursive: a coalition's best response contains best responses for each subcoalition. So, we obtain the following characterization of equilibria which is mainly stated for later reference and does not require a proof.

Proposition 5.12. *The following are equivalent:*

1. *σ is an equilibrium.*
2. *σ is a best response for N.*
3. *for each $x \in X$ and each $\emptyset \neq S \subseteq N$ strategy σ_S is unobjected against $\sigma_{-S}^{S,x}$ at x in the game $G^{S,x}$, and for each $\tau_S \in \Sigma_S$ that is a better response against $\sigma_{-S}^{S,x}$ at x in the game $G^{S,x}$ there is $i \in S$ who objects τ_S.*

As best responses depend on $\succeq$, so do equilibria and the outcome of the game. We now come to our main result that connects equilibria and rational expectation functions.

Theorem 1. *An extended expectation function F is rational if and only if there are an order profile $\succeq$ and an equilibrium σ with respect to $\succeq$ such that $F = F_\sigma$.*

Proof. Let $\succeq$ be an order profile, let σ be an equilibrium with respect to $\succeq$ and let $F = F_\sigma$. We prove that F satisfies **I**. Assume this were not the case. Then there are $x \in X$ and $T \in 2^N \setminus \left\{S^1(x), \ldots, S^{k(x)}(x)\right\}$ such that for each $l = 1, \ldots, k(x)$ there is $y \in X$ with $T \in E(x, y)$ and $u_i(y, F) > u_i\left(f^l(x), F\right)$ for all $i \in T$. Let T and x as described, let $l \in \{1, \ldots, k(x)\}$ be the minimal number with $T \succeq^x S^l(x)$ (this is well-defined as $T \succeq^x \emptyset = S^{k(x)}$), and let y be such that $u_i(y, F) > u_i\left(f^l(x), F\right)$ for

all $i \in T$. Consider the game $G^{T,x}$, recall that $\Sigma_T^{T,x} = \Sigma_T$, and let $\sigma'_T \in \Sigma_T$ be defined by

$$\sigma'_T(x') = \begin{cases} \sigma_T(x) & \text{if } x' \neq x, \\ y & \text{otherwise.} \end{cases}$$

Clearly, σ'_T is a better response against $\sigma_{-T}^{T,x}$ at x than σ_T. As σ is an equilibrium, σ'_T must be objected (otherwise $\sigma_{P(T)}$ would not be a best response). As σ'_T and σ_T coincide everywhere except x, and since σ_T is not objected, σ'_T can only be objected in some $z \in X$ with $x \in P_F(z)$. For such z it holds that $u_i\left(\left(\sigma'_T, \sigma_{-T}^{T,x}\right), z\right) = u_i(y, F)$. Moreover, $u_i\left(\left(\sigma_T, \sigma_{-T}^{T,x}\right), z\right) = u_i\left(\left(\sigma_T, \sigma_{-T}^{T,x}\right), x\right) = u_i\left(f^l(x), F\right)$ by construction. Since σ'_T is objected at z and σ_T is not, there is $i \in T$ with

$$\begin{aligned} u_i\left(\left(\sigma_T, \sigma_{-T}^{T,x}\right), x\right) &= u_i\left(f^l(x), F\right) < u_i(y, F) \\ &= u_i\left(\left(\sigma_T, \sigma_{-T}^{T,x}\right), y\right) = u_i\left(\left(\sigma'_T, \sigma_{-T}^{T,x}\right), x\right) \\ &= u_i\left(\left(\sigma'_T, \sigma_{-T}^{T,x}\right), z\right) \leq u_i\left(\left(\sigma_T^{0,z}, \sigma_{-T}^{T,x}\right), z\right) \\ &< u_i\left(\left(\sigma_T, \sigma_{-T}^{T,x}\right), z\right) = u_i\left(\left(\sigma_T, \sigma_{-T}^{T,x}\right), x\right), \end{aligned}$$

where the first two (in)equalities have been elaborated before, and the remaining ones (in that order) hold because of the definition of F, the construction of σ'_T, the definition of z, σ'_T's being objected at z, σ_T's being unobjected at z, and again the definition of z. But the overall inequality is impossible. Hence, such y cannot not exist. That means that F satisfies **I**.

We next prove that F satisfies **E**. Assume this were not the case. Then there are $x \in X$ and $l \leq k(x) - 1$ such that $u_i\left(f^{l+1}(x), F\right) \geq u_i\left(f^l(x), F\right)$ for some $i \in S^l(x)$. By definition, $u_i\left(f^l(x), F\right) = u_i\left(\left(\sigma_{S^l(x)}, \sigma_{-S^l(x)}^{S^l(x),x}\right), x\right)$ and $u_i\left(f^{l+1}(x), F\right) = u_i\left(\left(\sigma_{S^l(x)}^{0,x}, \sigma_{-S^l(x)}^{S^l(x),x}\right), x\right)$. Hence,

$$\begin{aligned} u_i\left(\left(\sigma_{S^l(x)}^{0,x}, \sigma_{-S^l(x)}^{S^l(x),x}\right), x\right) &= u_i\left(f^{l+1}(x), F\right) \\ &\geq u_i\left(f^l(x), F\right) = u_i\left(\left(\sigma_{S^l(x)}, \sigma_{-S^l(x)}^{S^l(x),x}\right), x\right), \end{aligned}$$

which means that $\sigma_{S^l(x)}$ is objected at x by i. This contradicts σ's being an equilibrium. So, F satisfies **E**.

We prove that F satisfies **M**. Assume this were not the case. Then there are $x \in X$, $l \in \{1, \ldots, k(x)-1\}$, and $y \in X$ such that $S^l(x) \in E(x,y)$, $f^l(x) \neq y$, and $u_i(y,F) > u_i\left(f^l(x), F\right)$ for all $i \in S^l(x)$. Define $\sigma'_{S^l(x)}$ by

$$\sigma'_{S^l(x)}(x') = \begin{cases} \sigma_{S^l(x)}(x') & \text{if } x' \neq x \\ y & \text{if } x' = x. \end{cases}$$

Clearly, $\sigma'_{S^l(x)}$ is a better response than $\sigma'_{S^l(x)}$ against $\sigma_{S^l(x)}^{S^l(x),x}$. Hence, as σ is an equilibrium, $\sigma'_{S^l(x)}$ must be objected. As $\sigma'_{S^l(x)}$ and $\sigma_{S^l(x)}$ coincide everywhere except x, and since $\sigma_{S^l(x)}$ is not objected, $\sigma'_{S^l(x)}$ can only be objected at some $z \in X$ with $x \in P_F(z)$. In particular, $u_i\left(\left(\sigma_{S^l(x)}, \sigma_{-S^l(x)}^{S^l(x),x}\right), z\right) = u_i\left(\left(\sigma_{S^l(x)}, \sigma_{-S^l(x)}^{S^l(x),x}\right), x\right) = u_i\left(f^l(x), F\right)$ and $u_i\left(\left(\sigma'_{S^l(x)}, \sigma_{-S^l(x)}^{S^l(x),x}\right), z\right) = u_i(y,F)$. Since $\sigma'_{S^l(x)}$ is objected at z and $\sigma_{S^l(x)}$ is not, there is $i \in S^l(x)$ with

$$\begin{aligned} u_i\left(\left(\sigma_{S^l(x)}, \sigma_{-S^l(x)}^{S^l(x),x}\right), x\right) &= u_i\left(f^l(x), F\right) < u_i(y,F) \\ &= u_i\left(\left(\sigma'_{S^l(x)}, \sigma_{-S^l(x)}^{S^l(x),x}\right), z\right) \leq u_i\left(\left(\sigma_{S^l(x)}^{0,z}, \sigma_{-S^l(x)}^{S^l(x),x}\right), z\right) \\ &< u_i\left(\left(\sigma_{S^l(x)}, \sigma_{-S^l(x)}^{S^l(x),x}\right), z\right) = u_i\left(\left(\sigma_{S^l(x)}, \sigma_{-S^l(x)}^{S^l(x),x}\right), x\right), \end{aligned}$$

where the weak inequality follows from $\sigma'_{S^l(x)}$'s being objected at z, and the last strict inequality follows from $\sigma_{S^l(x)}$'s not being objected. But the overall inequality is impossible. This proves that F satisfies **I**, **E**, and **M**. So, F is rational.

Let now F be a rational extended expectation function. Define σ by

$$\sigma_S(x) = \begin{cases} f^l(x) & \text{if } S = S^l(x) \text{ for some } l = 1, \ldots, k(x) \\ x & \text{otherwise.} \end{cases}$$

For $x \in X$ define $\succeq^x$ such that

1. $S \succeq^x \emptyset$ for all $S \subseteq N$,

2. $S^l(x) \succeq^x S^{l+1}(x)$ for all $l = 1, \ldots, k(x) - 1$,

3. for $S \in 2^N \setminus \left(S^1(x), \ldots, S^k(x)\right)$ it holds that $S^{l-1}(x) \succeq^x S \succeq^x S^l(x)$ where l is the minimal number such that for all $y \in X$ there is $i \in S$ with $u_i\left(f^l(x), F\right) \geq u_i(y, F)$. (If $l = 1$ let $S \succeq S^1(x)$.)

Since F satisfies **I**, such $\succeq^x$ exists for all $x \in X$. We show that σ is an equilibrium with respect to $\succeq$, i.e. that $\sigma_{P(S)}$ is a best response against σ_{-2^S} for all $S \subseteq N$. Let $\emptyset \neq S \subseteq N$. By Proposition 5.12 it is sufficient to show that σ_S is unobjected against $\sigma_{-S}^{S,x}$ in the game $G^{S,x}$ for all $x \in X$, and that each $\tau_S \in \Sigma_S$ that is a better response against $\sigma_{-S}^{S,x}$ in the game $G^{S,x}$ is objected. We first show that σ_S is unobjected for all $x \in X$. If x is such that $S \neq S^l(x)$ for all $l = 1, \ldots, k(x)$, there is nothing to show as $\sigma_S(x) = x = \sigma_S^0(x)$. If $S = S^l(x)$ for some l then $\sigma_S(x) = f^l(x)$, and by **E** it holds that

$$u_i(\sigma, x) = u_i\left(f^l(x), F\right) > u_i\left(f^{l+1}(x), F\right) = u_i\left(\left(\sigma_S^{0,x}, \sigma_{-S}\right), x\right)$$

for all $i \in S$. Hence, σ_S is not objected at any $x \in X$.

Next we show that there is no unobjected better response. Assume on the contrary that there are $x \in X$, $S \subseteq N$, and $\sigma'_S \in \Sigma_S$ such that σ'_S is an unobjected better response against $\sigma_{-S}^{S,x}$ in $G^{S,x}$. As σ_S is unobjected, by Lemma 5.5 we can assume without loss of generality that there is $x^* \in X$ such that $\sigma'_S(x) = \sigma_S(x)$ for all $x \neq x^*$, and $u_i\left(\left(\sigma'_S, \sigma_{-S}\right), x^*\right) > u_i\left(\left(\sigma_S, \sigma_{-S}\right), x^*\right)$ for all $i \in S$. Let $y = \sigma'_S(x^*) \neq \sigma_S(x^*)$ and note that $S \in E(x^*, y)$. Assume first that $S = S^l(x^*)$ for some $l = 1, \ldots, k(x^*) - 1$. Then $f^l(x^*) = \sigma_S(x^*) \neq y$, so by **M** there is $i \in S$ with

$$\begin{aligned} u_i\left(\left(\sigma_S, \sigma_{-S}\right), x^*\right) &= u_i\left(f^l(x^*), F\right) \geq u_i(y, F) \\ &= u_i\left(\left(\sigma'_S, \sigma_{-S}\right), x^*\right) > u_i\left(\left(\sigma_S, \sigma_{-S}\right), x^*\right) \end{aligned}$$

which is impossible. So, assume that $S \neq S^l(x^*)$ for all $l = 1, \ldots, k(x^*)$. By the definition of $\succeq^{x^*}$ there are $l \leq k(x^*)$ and $i \in S$ such that $u_i\left(\left(\sigma_S^{0,x^*}, \sigma_{-S}^{S,x^*}\right), x^*\right) =$

$\succeq^x$ for $x \in \{b, e\}$	ESS
$\{1\} \succeq^b \{2\}$, $\{1\} \succeq^e \{2\}$	$\{a, f\}$
$\{1\} \succeq^b \{2\}$, $\{2\} \succeq^e \{1\}$	$\{a, f\}$
$\{2\} \succeq^b \{1\}$, $\{2\} \succeq^e \{1\}$	$\{a, f\}$
$\{2\} \succeq^b \{1\}$, $\{1\} \succeq^e \{2\}$	$\{a, c, d, f\}$

Table 1: Order dependent equilibrium stable sets.

$u_i\left(f^l\left(x^*\right), F\right)$ and $u_i\left(f^l\left(x^*\right), F\right) \geq u_i\left(y, F\right)$. Hence,

$$u_i\left(\left(\sigma_S^{0,x^*}, \sigma_{-S}^{S,x^*}\right), x^*\right) \geq u_i\left(y, F\right) = u_i\left(\left(\sigma'_S, \sigma_{-S}^{S,x^*}\right), x^*\right) > u_i\left(\left(\sigma_S^{0,x^*}, \sigma_{-S}^{S,x^*}\right), x^*\right),$$

where the last inequality holds as σ'_S is unobjected against σ_{-S}^{S,x^*} at x^*. But this is impossible. Thus, σ is an equilibrium. ■

Example 5.13. Recall Example 4.2 and Figure 4. Depending on the order of moves in b and d we derive different equilibria and, hence, different rational extended expectation functions. Table 1 contains all four different order profiles and the corresponding ESS's. □

6 Applications

6.1 Absorbing Extended Expectation Functions

Players' utilities from an extended expectation function with non-terminal paths are lower than in any potential state. So, it is never in a coalition's interest to close a cycle or (if X is infinite) play a strategy that leads to an infinite path. We can use this observation to prove the following Corollary.

Corollary 6.1. *Every rational extended expectation function is absorbing.*

Proof. Let F be rational. Then there is an equilibrium such that $F = F_\sigma$ by Theorem 1. Assume that there is $x \in X$ such that $P_F(x)$ is non-terminal. Define strategy profiles $\sigma^1, \ldots, \sigma^{k(x)-1}$ by

$$\sigma_T^l(y) = \begin{cases} y & \text{if } y = x \text{ and } T = S^{l'}(x) \text{ for some } l' = 1, \ldots, l \\ \sigma_T(y) & \text{otherwise.} \end{cases}$$

Let l be the minimal integer such that the path $P_{F_{\sigma^l}}(x)$ is terminal. Such l exists as $P_{F_{\sigma^{k(x)-1}}}(x) = (x)$. Let $S = S^l(x)$ and note that $\sigma^l_{-S} = \sigma^{S,x}_{-S}$ and $\sigma^l_S = \sigma^{0,x}_S$. By construction

$$u_i\left(\left(\sigma_S^{0,x}, \sigma_{-S}^{S,x}\right), x\right) = u_i\left(\sigma^l, x\right) = U_i\left(t\left(P_{F_{\sigma^l}}(x)\right)\right) > -\infty = u_i\left(\left(\sigma_S, \sigma_{-S}^{S,x}\right), x\right).$$

But this means that σ_S is objected against σ_{-S}, in contradiction to σ's being an equilibrium. ■

6.2 Single-Payoff Equilibrium Stable Sets

A set $Y \subseteq X$ is a *single-payoff set* if $U_i(y) = U_i(y')$ for all $y \in Y$ and all $i \in N$. Dutta and Vohra (2017) provide a detailed analysis of single-payoff SREFS's. We show that they actually coincide with single-payoff ESS's.

Theorem 2. *Let $Y \subseteq X$ be a single-payoff set. Then Y is an* ESS *if and only if Y is a* SREFS.

Proof. Let Y be a single-payoff ESS, and let F be a rational extended expectation function with $\mathcal{S}(F) = Y$. As each player's payoffs are identical across all elements of Y, and as F satisfies **E**, at each $x \in X$ there is at most one non-empty coalition that deviates from x to some $z \in X \setminus \{x\}$. Hence, $k(x) \leq 2$ for all $x \in X$. By Lemma 3.4 the expectation function F^1 satisfies **I'**, **E'**, and **M'**. As the stationary points of F and F^1 are identical, Y is a REFS. By Theorem 1 in Dutta and Vohra (2017), Y is a SREFS.

Suppose now that Y is a SREFS. Let F' be a rational expectation function with Y as the set of stationary points, and let F be the extended expectation function with $F^1(x) = F'(x)$ for all $x \in X$ and $F^2(x) = (x, \emptyset)$ for all $x \in X \setminus Y$. Clearly, $\mathcal{S}(F) = Y$. We show that F satisfies **I**, **E**, and **M**. Assume first that F does not satisfy **I**. Then there are $x, x' \in X$ and $T \in E(x, x')$ such that $u_i(x', F) > u_i\left(f^l(x), F\right)$ for $l = 1, 2$ and all $i \in T$. Let $y = t\left(P_F\left(f^l(x)\right)\right)$ and $y' = t\left(P_F(x')\right)$ and note that $y, y' \in Y$. (The former exists as F' is absorbing and $k(x) \leq 2$ for all $x \in X$.) Then $U_i(y') = u_i(x', F) > u_i(x, F) = U_i(y) = U_i(y')$. This, however, is impossible. So, F satisfies **I**. As $k(x) \leq 2$ for all $x \in X$, conditions **M** and **E** of F are equivalent to **M'** and **E'** of F'. ■

6.3 Characteristic Function Form Games

A *game in characteristic function form* (with player set N) is a map V that maps each nonempty $S \subseteq N$ to a comprehensive, closed, and convex set $V(S) \subseteq \mathbb{R}^S$. Ray and Vohra (2015) associate with such a game V an abstract game $\left(N, X, E, (U_i(.))_{i \in N}\right)$: a state $x \in X$ is a pair $(v(x), \pi(x))$ of a partition $\pi(x)$ of the player set, and a utility vector $v(x)$ with $v_S(x) \in V(S)$ for all $S \in \pi(x)$. The utility function U_i is simply defined as $U_i(x) = v_i(x)$. The effectivity correspondence E satisfies the following two conditions:

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

(ii) For each $x \in X$, $T \subseteq N$, and each weakly Pareto efficient $v \in V(T)$ there is $y \in X$ such that $T \in E(x, y)$, $T \in \pi(y)$, and $v_T(y) = v$.

Condition (i) ensures that a deviating coalition T cannot affect the partition or the payoffs in coalitions it does not intersect with at state x. Condition (ii) ensures that coalition T can, from any state x, deviate to a state with T as a member of the partition and v as payoff vector for T, as long as v is feasible and weakly Pareto efficient for T.

Corollary 6.2. *Let V be a game in characteristic function form, and let G be an abstract game that satisfies Conditions (i) and (ii). If Y is a single-payoff* ESS *with payoff vector $v \in \mathbb{R}^N$ then v lies in the coalition structure core of V.*

Proof. By Theorem 2 any single-payoff ESS is a single-payoff SREFS. By Theorem 2 in Ray and Vohra (2015) this is the case if and only if v is separable, so v must lie in the coalitions structure core of V. ■

6.4 Matching

We consider two-sided one-to-one matching problems. Let W be a set of women, M be a set of men, and $N = W \cup M$. For each $i \in W$ let $\succeq_i'$ be a strict preference order over $M \cup \{i\}$, and for $j \in M$ let $\succeq_j'$ be a strict preference order over $M \cup \{j\}$. A matching is a map $\mu : N \to N$ such that $\mu(i) \in M \cup \{i\}$ for all $i \in W$, $\mu(j) \in W \cup \{j\}$ for all $j \in M$, and $\mu(\mu(i)) = i$ for all $i \in N$. If $\mu(i) = i$ then i is *single*, otherwise i is *matched to* $\mu(i)$. A matching μ is *individually rational* if $\mu(i) \succeq_i' i$ for all $i \in N$. A pair $(i, j) \in W \times M$ *blocks* a matching μ if $j \succ_i' \mu(i)$ and $i \succ_j' \mu(j)$. A matching μ is *stable* if it is individually rational and not blocked by any pair.

Let $\mathcal{M}$ be the set of all matchings, and define each player i's preference $\succeq_i$ over $\mathcal{M}$ by $\mu \succeq_i \mu'$ if and only if $\mu(i) \succeq_i' \mu'(i)$. Define an effectivity correspondence $E : \mathcal{M} \times \mathcal{M} \to 2^N$ by

$$E(\mu, \mu') = \left\{ S \subseteq N : \begin{array}{l} \{i, \mu'(i)\} \subseteq S \text{ whenever } \mu'(i) \notin \{i, \mu(i)\}, \text{ and} \\ \{i, \mu(i)\} \cap S \neq \emptyset \text{ whenever } \mu'(i) = i \neq \mu(i) \end{array} \right\}.$$

That is, $E(\mu, \mu')$ contains all those coalitions S whose members can transform μ into μ' via deleting old and forming new links. The tupel $\left(N, \mathcal{M}, E, (\succeq_i)_{i \in N}\right)$ forms an abstract game.[8]

Corollary 6.3. *Let $\mu \in \mathcal{M}$. Then μ is a stable matching if and only if $\{\mu\}$ is an* ESS *of the abstract game.*

[8] As $\mathcal{M}$ is finite $\succeq_i$ is equivalent to an ordinal utility function.

Proof. Theorems 1 and 2 in Mauleon et al. (2011) show that μ is stable if and only if $\{\mu\}$ is a farsighted stable set. Theorem 1 of Dutta and Vohra (2017) shows that any single-payoff farsighted stable set is a SREFS, and any single-payoff SREFS is a farsighted set. Hence, our Theorem 2 completes the proof. ■

References

Bloch, F., van den Nouweland, A., 2017. Farsighted stability with heterogeneous expectations. Working Paper.

Chwe, M.S.Y., 1994. Farsighted coalitional stability. Journal of Economic Theory 63, 299–325.

Diamantoudi, E., Xue, L., 2003. Farsighted stability in hedonic games. Social Choice and Welfare 21, 39–61.

Dutta, B., Vohra, R., 2017. Rational expectations and farsighted stability. Theoretical Economics 12, 1191–1227.

Gillies, D., 1959. Solutions to general non-zero games, in: Kuhn, A., Luce, R. (Eds.), Contributions to the Theory of Games. Princeton University Press, pp. 47–85.

Harsanyi, J.C., 1974. An equilibrium-point interpretation of stable sets and a proposed alternative definition. Management Science 20, 1472–1495.

Herings, P.J.J., Mauleon, A., Vannetelbosch, V.J., 2009. Farsightedly stable networks. Games and Economic Behavior 67, 526–541.

Jordan, J.S., 2006. Pillage and property. Journal of Economic Theory 131, 26–44.

Kimya, M., 2015. Equilibrium coalitional behavior. Working Paper.

Mauleon, A., Vannetelbosch, V.J., Vergote, W., 2011. Von Neumann-Morgenstern farsightedly stable sets in two-sided matching. Theoretical Economics 6, 499–521.

von Neumann, J., Morgenstern, O., 1944. Theory of Games and Economic Behaviour. Princeton University Press.

Ray, D., 2007. A Game-Theoretic Perspective on Coalition Formation. Oxford University Press.

Ray, D., Vohra, R., 2015. The farsighted stable set. Econometrica 83, 977–1011.

Ray, D., Vohra, R., 2017. Maximality in the farsighted stable set. Working Paper.

Xue, L., 1998. Coalitional stability under perfect foresight. Economic Theory 11, 603–627.