Farsightedness in Coalition Formation

Marco Mariotti and Licun Xue

1 Introduction

The main theme of this chapter is the formalization of the concept of 'farsightedness' in the behaviour of coalitions. Loosely speaking, a coalition is farsighted when it looks ahead at the ultimate consequences of its actions. If a subcoalition breaks up from a larger coalition, will this cause further breakups? And how will the members who are excluded from the deviating coalitions react? This is the sort of questions that are addressed in the literature we survey. As an example, the most popular cooperative solution concept, the Core, seems to embody a very limited amount of farsightedness: a coalition will deviate from a given status quo if it perceives the possibility of an immediate improvement, disregarding the possible further developments of play. In fact, things are not so simple, for even the Core can be shown to embody some amount of farsightedness (Ray 1989). One could analyse in turn the behavioural assumptions implicit in other classical solution concepts, such as for instance the Bargaining Set for cooperative games or the Strong Nash Equilibrium for games in strategic form. But, rather than proceeding case by case, we study here some recent attempts at dealing with the issue of farsightedness in a general way.

The lion's share in the chapter is taken by Greenberg's (1990) Theory of Social Situations (TOSS in short), for two reasons. Firstly, it is the quintessential general framework, which provides a unified language to treat the issues of interest here (and indeed most other game theoretic situations). In fact, TOSS is a valuable tool to pinpoint the (often implicit) behavioural assumptions of several existing solution concepts. Secondly, TOSS has been rather influential, spawning a literature which aimed to extend, modify and apply the original theory. Although the resulting contributions sometimes end up looking rather different from Greenberg's formulation, their conceptual relatedness is obvious. Besides, they can often be formally reconducted to TOSS. We discuss TOSS and some subsequent results obtained by Greenberg and others in section 2.

Because of space limitations, our survey beyond TOSS is far from being comprehensive. In fact, we focus on just a few of the concepts which have been proposed in the last decade. We think, however, that this will suffice to give the reader a flavour of the style and direction of the analysis involved.

In section 3 we present Chwe's (1994) work. He considers a class of social environments which, although less general than those allowed by TOSS, is sufficiently flexible to integrate the presentation of some negotiation processes associated with cooperative, strategic, and extensive form games. Chwe's main objective is to offer a notion of 'consistency' in the vein of the von Neumann and Morgenstern Stable Set, but which both captures foresight and is likely to be nonempty. Chwe's inclusive solution, the Largest Consistent Set, can be derived by using Greenberg's framework.

By inspecting the situation underlying Chwe's notion, Xue (1998) shows that Chwe's solution captures only a limited aspect of foresight. He then proposes a solution concept that captures 'perfect foresight' using Greenberg's framework. Xue's approach, centers around the graph-theoretic notion of a 'path' and attempts to embody in a more complex environment the logic of subgame perfection used in extensive form games. This approach is considered in section 4.

In section 5 we discuss yet another approach proposed in Mariotti (1997), which attempts to integrate some of Greenberg's ideas with the standard notion of a 'strategy' in order to capture farsightedness. As in Xue's work here, too, one tries to transfer the logic of subgame perfection to the more complex environment. The original work was formulated in terms of Greenberg's Coalitional Contingent Threat Situation for games in strategic form. In this chapter we provide a version of the same concepts which applies at a similar level of generality of Chwe's and Xue's approaches.

Section 6 concludes.

2 The Theory of Social Situations

When von Neumann and Morgenstern (1944) laid out the foundations of game theory, they proposed the 'Stable Set' as the main solution concept for situations in which the participants can form coalitions by freely communicating and making binding agreements. The idea of a Stable Set fell relatively out favour in the subsequent developments of game theory. However, Greenberg' TOSS revived some of the original motivations and even terminology of von Neumann and Morgenstern. TOSS is a much more general theory of social behaviour than that contained in *Theory of Games and*

Economic Behaviour (TGEB), but it revolves around the same central idea of 'stable standards of behaviour' as regulatory of strategic behaviour.

In TGEB a standard of behaviour amounts simply to a set of utility imputations. According to von Neumann and Morgenstern, a standard of behaviour, viewed as a solution to the game, must satisfy two requirements: (i) it is "free from inner contradictions" (internal stability) and (ii) it can "be used to discredit any non-conforming procedure" (external stability) (TGEB, p. 41). Greenberg introduced a far more encompassing notion of standard of behaviour to which these two requirements, expressing the features of stability a 'social norm' should possess, can be applied.

TOSS aims to offer a unified approach to the study of strategic interactions. This approach supersedes the classical distinction between "noncooperative" and "cooperative" (or other) descriptions of social environments. A social environment is represented as a "situation" which specifies *all* relevant information: for example, the beliefs of the players, the institutional setting such as the availability of binding agreements and social and legal restrictions on the formation of certain coalitions, and the details of negotiation process. In TOSS a unified solution concept is used for all such situations, by employing the stability criteria mentioned above. It is this separation between the solution concept and the description of the social environment that so often makes TOSS a useful language to address the issue of farsightedness. Behavioural assumptions additional to those implicit in stability are brought automatically to the fore once a given solution concept is expressed in TOSS language. In this sense, stability (a property that social norms are assumed to possess) becomes a benchmark against which to 'measure' other behavioural assumptions.

These observations can provide a response to the criticism that TOSS, being able to incorporate most existing solution concepts, is "too" general, to the point of being empty as a solution concept itself. According to this criticism, "anything" can be obtained as solution in TOSS, provided the appropriate social environment (situation) is specified. In our view, on the contrary, it precisely *because* TOSS is so general that it is useful. The pertinent criticism against TOSS should rather concern the issue whether stability (in the technical sense) is a necessary characteristic of social norms.

2.1 Definitions

2.1.1 Situation

We start by introducing the first main ingredient of TOSS, namely the notion of a "situation". This notion in turn involves two elements: a set of "positions" and an "inducement correspondence". A *position* describes the "current state of affairs". In particular, it specifies a set of individuals, the set of all possible outcomes, and the preferences of the individuals over this set of outcomes. Formally,

Definition 1 *A position, G, is a triple $G = (N(G), X(G), \{u_i(G)\}_{i \in N(G)})$, where $N(G)$ is the set of players, $X(G)$ is the set of all feasible outcomes, and $u_i(G)$ is the utility function of player i in position G over the outcomes*[1]*, that is, $u_i(G) : X(G) \to \Re$. Thus, for all $x, y \in X(G)$, and for all $i \in N(G)$, $u_i(G)(x) > u_i(G)(y)$ if and only if i prefers, in position G, the outcome x over the outcome y.*

In the theory, given a position, any feasible alternative can be "proposed". Although "proposed" is the term used by Greenberg, a more indicative term could be "under consideration". The crucial point is that in a position a feasible alternative becomes -through some procedure which may or may not be one of negotiation- the current 'status quo'. All the individuals active in the position can consider whether to accept or not the proposed status quo. An *inducement correspondence* specifies what it is feasible for any coalition (a set of players) to do if it decides to reject the proposed alternative. Consider a position G and suppose that an outcome $x^* \in X(G)$ is proposed. For a coalition $S \subset N(G)$ to be able to decide whether to accept or reject the proposed outcome, it must know (at least) the courses of action that are available. To be more specific, S has to know the set of positions it can "induce" if it rejects x^*.

Definition 2 *Let Γ be a set of positions and let*

$$\Omega = \{(G, x, S) \mid G \in \Gamma, x \in X(G), S \subset N(G)\}$$

An inducement correspondence is a mapping

$$\gamma : \Omega \to \Gamma$$

[1] Here we keep with Greenberg's formulation by using a utility function, although a preference relation is all that is needed to define the primitives of TOSS.

that specifies, for each $G \in \Gamma$*,* $x \in X(G)$*, and* $S \subset N(G)$*, a subset* $\gamma(S, G, x) \subset \Gamma$*.*

A **situation** is a pair (γ, Γ), where Γ is a set of positions, and γ is the inducement correspondence. A situation provides a *complete* description of a social environment. In particular, it specifies who can do what at every stage or the detailed "rules of the game".

The requirement that Γ is closed under γ guarantees that "the rules of the game" specify what can happen from any admissible (possibly itself an induced) position $G \in \Gamma$, when a feasible outcome $x \in X(G)$ is proposed. One additional requirement imposed on the inducement correspondence, γ, is that the set of players of each position that a coalition S can induce, includes the players in S. Formally, it is assumed:

Let (γ, Γ) be a situation. For all $G \in \Gamma$, $S \subset N(G)$, and $x \in X(G)$, if $H \in \gamma(S, G, x)$, then $S \subset N(H)$.

Note that γ may still assign the empty set to some (or even all) coalitions – some (or all) coalitions may be unable to induce any position at all. Think, for example, of the terminal nodes of an extensive form game.

2.1.2 Standards of Behavior

A situation specifies what it is feasible. We now move to the description of *behaviour*, namely a choice out of the feasible set of outcomes. This is done through the concept of a 'standard of behaviour'.

Given the description of the social environment as a situation, a standard of behavior specifies the 'solution' to every position in Γ. More precisely, consider a situation (γ, Γ) and a position $G \in \Gamma$. A natural question is: What is the set of outcomes that it is "reasonable" to expect at G? That is, what is the set that consists of all the feasible outcomes in $X(G)$ from which no coalition in $N(G)$ will choose to induce another position? This subset of $X(G)$ (whatever it may be) is called a "solution for G" and denoted by $\sigma(G)$.

Equipped with the concept of a solution for a given G, it is clear that a formal object linking *any* position to its solution must be introduced. This is because, given any position G, the acceptability of an outcome x^* in $X(G)$ depends on the outcomes that are believed to be accepted in the positions that can be induced from G when x^* is proposed. Consider for example

$H \in \gamma(S, G, x^*)$. For members of S to be able to decide whether to reject x^* and induce H, they must have information about the outcomes that are expected to result once position H is induced. This discussion motivates the following definition.

Definition 3 *Let Γ be a collection of positions. A standard of behavior (SB) for Γ is a mapping σ that assigns to every position $G \in \Gamma$ a solution $\sigma(G) \subset X(G)$.*

An SB σ for Γ is any arbitrary mapping which specifies a solution for every position. Greenberg views an SB as a particular "theory in the social sciences", namely a theory of social behaviour, given a social environment specified by a situation.

The next step is to impose some restrictions on σ. This is where the notion of stability comes in. As von Neumann and Morgenstern required stability of their solution, so it is required in TOSS that σ be "free of inner contradictions" and that it can "be used to discredit any non-conforming procedure". As we saw above, these two properties are termed "internal stability" and "external stability", which we shall define formally.

2.1.3 Stability

Consider a situation (γ, Γ), and let σ be an SB for Γ. The first fundamental assumption is that rational behaviour entails internal stability of σ as explained below.

Imagine that a way of comparing *sets* of outcomes with outcomes has been defined (recall that a solution is a subset and generally not a singleton). If all players adopt the recommendations made by σ, any group of players, S, will reject an outcome, $x \in X(G)$, if it can induce a position $H \in \gamma(S, G, x)$, whose solution $\sigma(H)$ all members of S prefer to x.

So, this principle provides a justification for rejecting outcomes. Formally,

Definition 4 *Let σ be a SB for situation (γ, Γ). σ is internally stable for (γ, Γ) if for all $G \in \Gamma$, whenever there exists a coalition $S \subset N(G)$ and a position $H \in \gamma(S, G, x)$ such that the members of S prefer $\sigma(H)$ to x, it is the case that $x \notin \sigma(G)$.*

The next assumption is that rational behavior entails external stability: the rejection of x can be justified *only* by an objection in terms of the "solution" outcomes in $\sigma(H)$, and not of any other outcome in the set of feasible ones, $X(H)$. The underlying logic is that outcomes belonging to $X(H)\backslash\sigma(H)$ have no validity as objections because, not being part of the solution set, they cannot form the basis for an agreement between the members of a rejecting coalition.

This prevents the exclusion of certain outcomes from σ from being arbitrary. This principle limits the justifications for rejecting outcomes to those expressed by internal stability. Formally,

Definition 5 *Let σ be a SB for situation (γ, Γ). σ is externally stable for (γ, Γ) if for all $G \in \Gamma$, whenever there exist no coalition $S \subset N(G)$ and a position $H \in \gamma(S, G, x)$, such that the members of S prefer $\sigma(H)$ to x, it is the case that $x \in \sigma(G)$.*

Definition 6 *Let σ be a SB for situation (γ, Γ). σ is stable if it is both internally and externally stable.*

Stability is the only requirement imposed on an SB. The above two definitions are incomplete in that it has not yet been specified how the members of a coalition S compare $\sigma(H)$ and x. This difficulty is resolved in the original formulation of TOSS by considering only two extreme behavioral assumptions (formally defined at the end of this subsection): optimistic and conservative behavior. Though these assumptions yield several interesting results, the investigation of alternative behavioral assumptions may prove to be valuable. For example, Greenberg's (1994) definition of stable value for cooperative games with transferable utilities, a uniform probability distribution is used to evaluate the solution of a position. In general, it is clear that any particular specification of a method, or set of methods, to compare sets of outcomes with outcomes is not an integral part of the TOSS, but rather a part of the description of the social environment. Indeed, one might even conceive of a separate such method of evaluation for each player. This is to be contrasted with Chwe's approach in the next section, where a particular assumption of "conservative" behaviour is built in the formal definition of his solution concept.

The two extreme behavioral assumptions introduced by Greenberg (1990) are as follows.

Optimistic Behavior A coalition S prefers $\sigma(H)$ to $x \in X(G)$ if *there exists* an outcome $y \in \sigma(H)$ that all members of S prefer to x, i.e., $u_i(H)(y) > u_i(G)(x)$ for all $i \in S$.

Conservative Behavior A coalition S prefers $\sigma(H)$ to $x \in X(G)$ if *all* outcomes in $\sigma(H)$ make all members of S better off, that is, for all $y \in \sigma(h)$, $u_i(H)(y) > u_i(G)(x)$ for all $i \in S$.

If a SB is stable under the assumption of optimistic behavior, the SB is called **Optimistic Stable Standard of Behavior (OSSB)** and if a SB is stable under the assumption of optimistic behavior, the SB is called **Optimistic Stable Standard of Behavior (OSSB)**. In some important situations, such as the Coalitional Contingent Threat Situation defined below, these two notions may coincide.

2.2 Some Results

2.2.1 OSSB and von Neumann and Morgenstern Stability

As we have explained before, the terms "standard of behavior" and "stability" have been borrowed from von Neumann and Morgenstern's seminal work. In fact, the OSSB can be formally derived from von Neumann and Morgenstern's (vN-M) **Abstract Stable Set** for some "associated abstract system". Recall that an **abstract system** is a pair $(D, \prec)$, where D is an arbitrary nonempty set, and $\prec$ is a binary relation on D, called the dominance relation. For $a, b \in D$, $b \prec a$ is interpreted to mean that a dominates b.

Definition 7 *Let $(D, \prec)$ be an abstract system and $A \subset D$. A is vN-M internally stable if $x \in A$ implies that there does not exist $y \in A$ such that $x \not\prec y$; A is vN-M externally stable if $x \in D \setminus A$ implies that there exists $y \in A$ such that $x \prec y$; A is vN-M stable if it is both vN-M internally and externally stable.*

We can formally define an abstract system $(D, \prec)$ associated with a situation (γ, Γ) as follows:

$$D \equiv \{(G, x) \mid G \in \Gamma \text{ and } x \in X(G)\}$$

and, for (G,x) and (H,y) in D, $(G,x) \prec (H,y)$ if and only if there exists an $S \subset N(G)$ such that $H \in \gamma(S,G,x)$, and for all $i \in S$, $u_i(H)(y) > u_i(G)(x)$. Also, let us define the graph C of an SB σ by:

$$C \equiv \{(G,x) \mid G \in \Gamma \text{ and } x \in \sigma(G)\}.$$

We can now state the following fundamental result, due to Shitovitz [see Greenberg (1990)].

Theorem 1 *Let (γ,Γ) be a situation. The mapping σ is an OSSB for (γ,Γ) if and only if its graph is a vN-M abstract Stable Set for the system $(D,\prec)$ that is associated with (γ,Γ).*[2]

Theorem 1 draws a formal relationship between OSSB[3] and the abstract Stable Set. This is of great practical use in applications, because it allows one to apply known results from works on abstract Stable Sets to the theory of social situations. We mention in particular the following theorem due to von Neumann and Morgenstern (1947). To state it we need an additional definition.

Let $(D,\prec)$ be an abstract system. The dominance relation, $\prec$, is called **acyclic** if there is no **finite** sequence, $a_j, j = 1,2,...,J$, of elements in D, such that for all $j = 1,2,...,J$, $a_j \prec a_{j+1}$, where $a_{J+1} = a_1$. The dominance relation, $\prec$, is called **strictly acyclic** if there exists no **infinite** sequence, a_j, of elements in D, such that $a_j \prec a_{j+1}$ for all $j, j = 1,2,\ldots$.[4]

Theorem 2 *Let $(D,\prec)$ be an abstract system, where the dominance relation $\prec$ is strictly acyclic. Then there exists a unique vN-M (abstract) Stable Set for the system $(D,\prec)$.*

This result can be proved by using transfinite induction.

Form Theorems 1 and 2, the following basic existence result, also proved by Shitovitz [see Greenberg (1990)], follows:

[2] The comparison between OSSB and vN-M abstract stable sets has to involve the graph of the SB, since the stability of an SB is imposed on the *mapping*, σ, whereas in vN-M, the stability is required of a *set*.

[3] The notion of CSSB, on the other hand, cannot be formally related to a vN-M abstract stable set. However, using the notion of "F-stability", Rothblum obtained a partial analog of this theorem for CSSB (see, Greenberg, 1990, section 11.8.)

[4] Observe that strict acyclicity implies acyclicity.

Theorem 3 *Let (γ, Γ) be a situation such that Γ contains a finite number of positions, and each position $G \in \Gamma$ contains a finite number of outcomes. Assume, in addition, that the inducement correspondence is such that for all $G, H \in \Gamma$, $x \in X(G)$ and $S \subset N(G)$, if $H \in \gamma(S, G, x)$ then $N(H) = S$. Then, (γ, Γ) admits a unique OSSB.*

It is important to note that a stable SB under behavioral assumption other than optimism cannot, in general, be formally derived from a vN-M abstract Stable Set. This is but one feature illustrating the greater generality of TOSS. This kind of results prove to be very useful in comparing different notions in coalitional analysis, as while most notions involve *implicit* assumptions about the negotiation process and behavior, TOSS does not address foresight directly.

2.2.2 Some Properties of OSSB and CSSB

We proceed to state some properties of OSSB and CSSB. Observe that by external stability, an OSSB or a CSSB cannot be identically empty-valued. That is, if σ is an OSSB or CSSB for a situation (γ, Γ), then there exists at least one position $G \in \Gamma$ such that $\sigma(G) \neq \emptyset$. σ is said to be *nonempty-valued* if $\sigma(G) \neq \emptyset$ for all $G \in \Gamma$. Let σ and σ' be two SB's for a situation (γ, Γ). Say σ *includes* σ' if $\sigma(G) \supset \sigma'(G)$ for all $G \in \Gamma$. Greenberg, Monderer, and Shitovitz (1996) proved the following result.

Theorem 4 *Let (γ, Γ) be a situation and let Σ be the set of all conservative internally stable non-empty valued SB's. Then Σ admits a largest element, σ^L, with respect to the inclusion order defined above. Moreover, σ^L is the largest nonempty-valued CSSB.*

Note that if an optimistic internally stable SB is nonempty-valued, then it is also conservative internally stable. Hence we have the following corollary.

Corollary 5 *If σ is a nonempty-valued OSSB, then there exists a CSSB that includes it. In particular, every nonempty-valued OSSB is included in the largest nonempty-valued CSSB.*

2.3 Applications

TOSS can be used to "derive", namely express in a new language, existing solution concepts. This is accomplished by identifying the situations underlying those solution concepts. Since a standard of behaviour is only constrained by the requirement of stability, and a situation entails a complete specification of all other aspects of the negotiation and reasoning process, one can often uncover subtle or hidden assumptions imposed by an existing solution concept. This can, in turn, point to possible desirable modifications of this solution concept and suggest new concepts. Such an approach will be apparent in the next two sections.

In what follows, we shall use two examples, useful for later reference, to illustrate applications of TOSS.

2.3.1 The von Neuman and Morgenstern (vN-M) situation

Let (N, v) be a finite game in characteristic function form, with $v(S)$ a compact nonempty subset of (the Euclidean space) $\Re^S$.

Definition 8 *A* Stable Set*, or a* vN-M solution*, for the cooperative game (N, v) is a set A which is a subset of $v(N)$ and satisfies the following: $y \in A$ if, and only if, there exist no coalition $S \subset N$ and a payoff $x \in A$ such that $x_S \in v(S)$ and $x_S \gg y_S$.*

The negotiation process underlying vN-M solution is one where coalitions make "tender threats" which other coalitions can, in turn, counter by other tender threats. That is, every offer in $v(N)$ that is being made can be objected to by some coalition which offers a new payoff in $v(N)$, which, in turn, can be objected to by any other coalition, and so on.

The situation $(\hat{\gamma}, \hat{\Gamma})$ that describes this negotiation process for the cooperative game (N, v) is constructed by defining, for each $S \subset N$, the position $\hat{G^S} = (N, \hat{v}(S), \{u_i(\hat{G^S})\}_{i \in N})$ as follows. The set of players in position $\hat{G^S}$ is the set N. The set of outcomes, $\hat{v}(S)$, in that position contains all those payoffs that are feasible for the grand coalition and whose projection on S yields a payoff that is feasible for coalition S, that is, $\hat{v}(S) = \{x \in v(N) \mid x_S \in v(S)\}$. Finally, for all $i \in N$ and $x \in \hat{v}(S)$, $u_i(\hat{G^S})(x) = x_i$. Next, define the situation $(\hat{\gamma}, \hat{\Gamma})$ by $\hat{\Gamma} = \{\hat{G^S} \mid S \subset N\}$ and, for all $\hat{G^S} \in \hat{\Gamma}$, $x \in X(\hat{G^S})$, and $T \subset N$, $\hat{\gamma}(T, \hat{G^S}, x) = \{\hat{G^T}\}$. This says that a coalition T can counter-propose the payoff y, $y \in \hat{v}(T)$ independently of the currently proposed payoff, x and of which coalition proposes it.

The crucial relationship (credited to P. De Marzo, 1986 in Greenberg (1990)) between an OSSB for the vN-M situation $(\hat{\gamma}, \hat{\Gamma})$, and a vN-M solution for the game (N, v), is stated below.

Theorem 6 *Let (N, v) be a cooperative game The SB σ is an OSSB for $(\hat{\gamma}, \hat{\Gamma})$ if, and only if, $\sigma(\hat{G^N}) \equiv \sigma(G^N) \equiv A$, where A is a vN-M solution for the game (N, v), and for all $\hat{G^S} \in \hat{\Gamma}, : \sigma(\hat{G^S}) \equiv A \cap \hat{v}(S)$.*

This results holds, with the appropriate reformulation of the definitions, for games with an infinite number of players.

2.3.2 The Coalitional Contingent Threat Situation

A game in strategic form is a triple $(N, \{Z_i\}_{i \in N}, \{u_i\}_{i \in N})$, where N is the set of players, Z_i is the nonempty strategy set of player i, and $u_i : Z_N \rightarrow \Re$ is player i's payoff function, where, for $S \subset N$, Z_S denotes the Cartesian product of Z_i over $i \in S$.

Consider the following negotiation process. An n-tuple of strategies $\zeta \in Z_N$, is proposed to the players. If all individuals *openly* agree to play ζ, then ζ will be played. If a coalition $S \subset N$ objects to the specified choice of ζ_S, he has to declare that if all other players will stick to the specified n-tuple of strategies ζ (hence the term "contingent threats"), then it will employ the strategy $\rho_S \in Z_S$ instead of ζ_S. The new revised proposal then becomes $\xi = (\zeta_{N \setminus S}, \rho_S)$, from which another coalition, T, can, in turn, openly threaten to deviate, and the process continues in this manner.

The foregoing process is described by the following coalitional contingent threats situation (γ, Γ).[5] For $\zeta \in Z_N$, let G_ζ denote the position where the set of players is N, the set of outcomes consists of the single N-tuple of strategies ζ, and the utility level player $i \in N$ derives from ζ is $u_i(\zeta)$. That is,

$$G_\zeta \equiv (N, \{\zeta\}, \{u_i\}_{i \in N}).$$

The set of positions in the coalitional contingent threats situation (γ, Γ) is

$$\Gamma \equiv \{G^N\} \cup \{G_\zeta \mid \zeta \in Z_N\},$$

[5] If only individuals can make contingent threat, then resulting situation is called the "individual contingent threat situation" [see Greenberg (1990)].

and the inducement correspondence γ is given by: For all $G \in \Gamma$, $\zeta \in Z_N$, and $S \subset N(G)$,

$$\gamma(S \mid G, \zeta) \equiv \{G_\xi \mid \xi = (\xi_S, \zeta_{N \setminus S}), \xi_S \in Z_S\}.$$

Since the set of outcomes at each position is a singleton, a standard of behavior in this situation is an OSSB if and only if it is a CSSB. Moreover, an OSSB (CSSB) contains the set of strong Nash equilibria. In the rest of this chapter, the coalitional contingent threat situation will be variously modified to capture foresight of the players.

3 The Largest Consistent Set

3.1 Definitions and some results

Chwe (1994) uses an alternative, simpler description of a social environment than the one considered so far. Consider a set of individuals, N, who are faced with a feasible set of alternatives Z. Each individual $i \in N$ has a strict preference relation $\prec_i$ on Z. What coalitions can do if and when they form is specified by $\{\stackrel{S}{\longrightarrow}\}_{S \subset N}$, where $\{\stackrel{S}{\longrightarrow}\}$, $S \subset N$, is an "effectiveness relation" on Z and $a \stackrel{S}{\longrightarrow} b$, where $a, b \in Z$, means that coalition $S \subset N$ can replace a by b.[6]. Restrictions on coalition formation, if there are any, are also reflected in $\{\stackrel{S}{\longrightarrow}\}_{S \subset N}$ (which is a collection of partial relations). Thus, a social environment [see also Rosenthal (1972)] is represented by $\mathcal{G} = (N, Z, \{\prec_i\}_{i \in N}, \{\stackrel{S}{\longrightarrow}\}_{S \subset N})$. Here are some examples to illustrate the flexibility and generality of $\mathcal{G}$.

Example 1 *The coalitional contingent threat situation associated with a strategic form game can be represented by $\mathcal{G}$. Let $Z = Z_N$. Thus, for all $a, b \in Z$, $a \stackrel{S}{\longrightarrow} b$ if and only if $a_{-S} = b_{-S}$. If coalitions cannot form, then for every $i \in N$, and $a, b \in Z$, $a \stackrel{\{i\}}{\longrightarrow} b$ if and only if $a_{-i} = b_{-i}$, and $a \prec_i b$ if and only if $u_i(a) < u_i(b)$. This corresponds to the "individual contingent threat situation".*

[6] It is reasonable, but not necessary, to assume $a \stackrel{T}{\longrightarrow} a$ where T is the union of all the coalitions that can replace a by some other alternatives.

Example 2 *One way to represent a cooperative game by $\mathcal{G}$ is as follows. Let Z be the set of imputations (efficient and individually rational payoff vectors in $\nu(N)$). For $a, b \in Z$ and $S \subset N$, $a \xrightarrow{S} b$ if and only if $b_S \in \nu(S)$. This bears a similarity with the vN-M situation.*

Although $\mathcal{G}$ describes the *primitives* of a social environment, its description of the social environment is incomplete, in comparison with a *situation*. Indeed, we can apply different solution concepts to $\mathcal{G}$, each of which embeds different "rules of the game". For example, the vN-M solution for cooperative games can be generalized to more complex social environment as studied in this paper. That is, one can apply the notion of "vN-M abstract Stable Set" to the study of social environment represented by $\mathcal{G}$. To this end, we introduce the following definition.

Let V be a (abstract) Stable Set for $(Z, \gtrdot)$. If $x \in Z$ is the status quo, the set of "predicted outcomes" is given by $\{y \in V \mid y = x \text{ or } y \gtrdot x\}$. That is, if some $x \in V$ is the status quo, it will prevail; however, if some $x \in Z \setminus V$ is the status quo, then some $y \in V$ such that $y \gtrdot x$ will prevail.

The following dominance relation on Z is similar to the one used in the definition of vN-M solution for cooperative games.

Definition 9 *For $a, b \in Z$, b is said to dominate a, or $b > a$, if there exists a coalition $S \subset N$ that can replace a by b, i.e., $a \xrightarrow{S} b$, and all members of the acting coalition S prefer b to a, i.e., $a \prec_i b$ for all $i \in S$.*

Let V be stable for $(Z, >)$. If $\mathcal{G}$ represents a cooperative game, then V is equivalent to the vN-M solution for this cooperative game. However, Harsanyi [10] criticizes the vN-M solution for its failing to incorporate foresight. Such a criticism applies also to an abstract Stable Set V for $(Z, >)$, which can be illustrated by an (extremely) simple example, where $N = \{1, 2\}$, $Z = \{a, b, c\}$, player 1 can replace a by b, i.e., $a \xrightarrow{S} b$, and player 2 can replace b by c, i.e., $b \xrightarrow{S} c$. This example is depicted by Figure 1, where the vector attached to each alternative is the payoff vector derived from that alternative if it prevails.

$$a_{(1,1)} \xrightarrow{\{1\}} b_{(0,0)} \xrightarrow{\{2\}} c_{(2,2)}$$

Figure 1

The unique Stable Set for $(Z, >)$ is $V = \{a, c\}$. According to the definition of V, player 1 will not replace a by b, since b itself is not stable (and hence will be replaced). But if he is farsighted, he should and will replace a by b, precisely because player 2 (who is rational) will subsequently replace b by c. That is, farsighted players do not just look at the next step. For this reason, Harsanyi suggests (within the framework of cooperative games) to replace the (direct) dominance relation $>$ by some "indirect dominance", which captures the fact that farsighted individuals consider the *final* outcomes that their actions may lead to. Chwe (1994) formalized a version of Harsanyi's indirect dominance when applied to $\mathcal{G}$. An alternative b is said to indirectly dominate another alternative a if b can replace a in a sequence of "moves", such that at each move the active coalition prefers (the final alternative) b to the alternative it faces at that stage. Formally,

Definition 10 *For $a, b \in Z$, b indirectly dominates a, or $b \gg a$, if there exist $a_0, a_1, \ldots, a_m$ in Z, where $a_0 = a$ and $a_m = b$, and coalitions $S_0, S_1, \ldots, S_{m-1}$ such that for $j = 0, 1, \ldots, m-1$, $a_j \overset{S_j}{\longrightarrow} a_{j+1}$ and for all $i \in S_j$, $a_j \prec_i a_m$.*

Now, given the indirect dominance relation $\gg$, one can consider the (abstract) Stable Set for $(Z, \gg)$. For Figure 1, the unique Stable Set for $(Z, \gg)$ is $H = \{c\}$, which captures foresight of the individuals in this example: If a is the status quo, c is the only predicted outcome. As noted by Chwe [6], however, the Stable Set for $(Z, \gg)$ can be too "exclusive" in that its exclusion of some alternatives may not be consistent with rationality and foresight. To rectify this, Chwe suggests a new solution concept – "the Largest Consistent Set". In the definition of (the largest) consistent set, a coalition rejects or deviates from an alternative only if its deviation lead *only* to alternatives that benefit its members. (In contrast, the Stable Set for $(Z, \gg)$ entails that a coalition deviates as long as this deviation might lead to *some* alternative that benefits its members.) The Largest Consistent Set has the merits of "ruling out with confidence" and being nonempty under weak condition. It turns out, however, that the Largest Consistent Set may be too inclusive. We shall illustrate these issues by the example in Figure 2 in the next section. But, first, let us introduce the formal definition of the Largest Consistent Set.

Definition 11 *Consider a social environment $\mathcal{G}$. A subset $Y \subset Z$ is consistent if $a \in Y \iff$ for every $d \in Z$ and $S \subset N$ such that $a \overset{S}{\longrightarrow} d$, there* *exists $e \in Y$, $e = d$ or $e \gg d$, such that $a \not\prec_S e$. The Largest Consistent Set (LCS) is the unique maximal consistent set with respect to set inclusion.*

Chwe proved that if Z is countable and contains no infinite sequence $a_1, a_2, \ldots$ such that $i < j$ implies $a_i \ll a_j$, then LCS is nonempty. Xue (1997) extends this result by removing the countability condition. Chwe also applies his notion to "coalitional contingent threat situation", voting games and other contexts[7].

3.1.1 LCS and TOSS

As shown by Chwe (1994), the notion of consistent set can be cast within the framework of TOSS, thereby revealing how individuals view and use their alternatives. In particular, the negotiation/reasoning process underlying Chwe's consistent set is described by situation (γ, Γ) defined as follows. For $a \in Z$, define a position $G_a = (N, X_a, (\prec_i)_{i \in N})$ where

$$X_a = \{a\} \cup \{b \in Z \mid a \ll b\}.$$

and $\Gamma = \{G_a \mid a \in Z\}$. For $G_a \in \Gamma, x \in X_a$ and $S \subset N$, $\gamma(S \mid G_a, x) = \left\{G_y \mid x \xrightarrow{S} y\right\}$.

Chwe (1994, Proposition 4) showed that the CSSB for the "Chwe situation" is formally related to his consistent set. This formal relationship can be stated in the following proposition [see Xue (1997)].

Proposition 7 *For $Y \subset Z$, define an SB ψ for the "Chwe situation" by $\psi(X_a) = X_a \cap Y$ for all $a \in Z$. Then, ψ is a CSSB if and only if Y is consistent and ψ is nonempty-valued. In particular, ψ is the largest (nonempty-valued) CSSB if and only if Y is the LCS and ψ is nonempty-valued.*

4 A 'Path' Approach

4.1 Paths and Perfect Foresight

According to Chwe (1994), the set of "predicted outcomes", when $x \in Z$ is the status quo, is given by $\{y \in LCS \mid y = x \text{ or } y \gg x\}$. We now proceed to

[7] See Page, Wooders, and Kamat (2001) for an application of Chew's notion of farsighted stability to network formation.

illustrate that the Stable Set for $(Z, \gg)$ can be too exclusive while the LCS can be too inclusive. Consider the social environment where $N = \{1, 2\}$, $Z = \{a, b, c, d\}$, and the "effectiveness relations" as well as the payoffs are depicted in Figure 2.

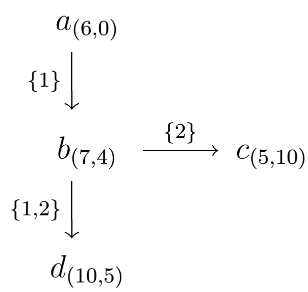

Figure 2

Assume that the status quo is a. If a prevails, the payoffs are 6 and 0 for players 1 and 2 respectively. Player 1 can replace a by b, which, if prevails, yields a payoff of 7 to player 1 and 4 to player 2. Once b becomes the (new) status quo, there are two possibilities: either player 2 can replace b by c, or players 1 and 2 together can replace b by d. Applying the definition of indirect dominance gives

$$b \gg a, d \gg b, c \gg b, \text{ and } d \gg a.$$

As a result, the unique Stable Set for $(Z, \gg)$ is $H = \{c, d\}$. a is excluded from H since $d \gg a$ and $d \in H$. Note that $c \in H$ but $c \not\gg a$. Therefore, if a is the status quo, the unique predicted outcome is d. But clearly, to reach d from a requires player 1 first to replace a by b, and once b is reached, player 1 will not join player 2 to replace b by d; instead, player 2 will replace b by c. Hence given that players are farsighted, in contemplating a deviation from a, player 1 should anticipate the final outcome (c, in this case) that will arise, and thereby will not replace a by b.

The LCS solves the exclusion of a. Indeed, $LCS = \{a, c, d\}$. Therefore, when a is the status quo, the set of predicted outcomes is $\{a, d\}$. But there remains a problem: when a is the status quo, one of the "predicted" outcomes according to the LCS is d, resulting in the same difficulty as was discussed above. The above analysis illustrates several aspects of perfect foresight.

- A farsighted player considers only the *final* outcomes that might result when making choices. Indeed, player 1, in contemplating a deviation from a, does not make his decision by comparing a with b.

- Even though, as stated in (1), it is only the final outcomes that matter, a player with perfect foresight considers also how, if at all, these final outcomes can be reached. In our example, it is feasible to reach d from a, but rational players would not follow the "path" (a, b, d). (Were b reached, player 2 would deviate and implement c.) To capture perfect foresight, we must, therefore, consider deviations "along the way" to the final outcomes.

- The exclusiveness of the Stable Set for $(Z, \gg)$ and inclusiveness of the LCS are not isolated events. They both stem from the fact that indirect dominance defined on Z fails to capture perfect foresight since it ignores the possible deviations along the way from one alternative (e.g., a) to another (e.g., d).

On the basis of these considerations Xue (1998) suggests that, in order to model perfect foresight, one needs to consider the "graph" of a social environment and use "paths" as the building blocks in the formalization of foresight. The social environment depicted in Figure 2 has been represented intentionally in a graph form to stress this point. Viewing a social environment as a graph has been overlooked in the literature on foresight, since its necessity is not obvious, particularly when $\mathcal{G}$ represents a normal form game, a cooperative game, or a social environment of more complex structure.

We now formalize perfect foresight by considering the "graph" of $\mathcal{G}$. To this end, we introduce the following definition.

Definition 12 *The directed graph generated by $\mathcal{G}$, denoted $\phi(\mathcal{G})$, consists of the set of vertices (nodes) Z and a collection of arcs where for every $a, b \in Z$, ab is an arc if and only if there exists $S \subset N$ such that $a \xrightarrow{S} b$. If ab is an arc, b is said to be adjacent from a and a adjacent to b. A path is a sequence of vertices $(v_1, v_2, \ldots, v_k)$, where for all $j = 1, 2, \ldots, k-1$, $v_j v_{j+1}$ is an arc, that is, there exists a coalition $S_j \subset N$ such that $v_j \xrightarrow{S_j} v_{j+1}$. The length of this path is $k-1$. $\phi(\mathcal{G})$ is said to be acyclic if every path consists of distinct vertices. $\phi(\mathcal{G})$ is said to be bounded if there exists a finite integer J such that every path has a length that does not exceed J.*

The following notations are introduced to facilitate the subsequent analysis. If $a \in Z$ is a vertex that lies on the path α, we shall write $a \in \alpha$. For a path α, let $\alpha|_b$, where $b \in \alpha$, denote its continuation from b (and including b)

and let $t(\alpha)$ denote its terminal node (i.e., the last node that lies on α). Also, let Π be the set of all paths and for $a \in Z$, let Π_a denote the set of paths that originate from a (including a itself). The preferences over paths in Π are the preferences over their terminal nodes, i.e., for any two paths α and β, $\alpha \prec_i \beta$ if and only if $t(\alpha) \prec_i t(\beta)$. Also, we write $\alpha \prec_S \beta$ if $t(\alpha) \prec_S t(\beta)$, i.e., if $t(\alpha) \prec_i t(\beta)$ for all $i \in S$.

For every $a \in Z$, Π_a specifies the set of "feasible outcomes" when a is the status quo (or under consideration). Note that $\mathcal{G}$ is far more general than an extensive form game: At every node, more than one coalitions may act, and $\phi(\mathcal{G})$, the graph of $\mathcal{G}$, need not be acyclic (Such is the case, for example, when $\mathcal{G}$ represents a cooperative game or a normal form game).

We shall retain the assumptions of Chwe (1994) that actions are public and payoffs are derived at a status quo only if no coalition wishes to replace it. Consider, for example, a duopoly pricing model. Then the assumption of public actions amounts to the requirement that price changes are *instantly* detected while payoffs are derived only if prices are "stable" in that no firm will wish to change its price. Public actions can also represent objections and counter-objections in an open negotiation, in which case, our analysis identifies those agreements that are acceptable to rational (and farsighted) individuals.

In the analysis that follows, perfect foresight is captured *explicitly* by the following situation, which we shall henceforth refer to as "the situation with perfect foresight": Assume that alternative $a \in Z$ is the status quo. Then the associated position is $G_a = \{N, \Pi_a, (\succ_i)_{i \in N}\}$. Thus, the set of positions is given by

$$\Gamma^* = \{G_a \mid a \in Z\}.$$

Consider a path $\alpha \in \Pi_a$ and some node $b \in \alpha$ and assume that a coalition $S \subset N$ can replace b by some alternative c that does not lie on α, i.e., $b \xrightarrow{S} c$ and $c \notin \alpha$. In doing so, S is aware of that the set of feasible paths from c is Π_c. Thus, the inducement correspondence γ^* is given by: For all $G_a \in \Gamma^*$, $\alpha \in \Pi_a$, and $S \subset N$,

$$\gamma^*(S, G_a, \alpha) = \left\{G_c \mid b \in \alpha \text{ and } b \xrightarrow{S} c\right\}.$$

A stable standard of behavior σ for (Γ^*, γ^*) specified, for every $a \in Z$, the set of paths in Π_a that might be followed by rational and farsighted individuals.

4.1.1 Properties

It is easy to verify that for the social environment depicted in Figure 2, the situation with perfect foresight admits a unique OSSB which coincides with the unique CSSB. Denoting this SB by σ, we have that $\sigma(G_b) = \{(b, c)\}$[8] and $\sigma(G_a) = \{a\}$. Hence, coalition $\{1, 2\}$ will never form. Moreover, if a is the status quo, a (and only a) will prevail. Thus, the unique (optimistic or conservative) stable SB gives rise to the outcome conforming to perfect foresight.

For an arbitrary social environment $\mathcal{G}$, the notion of (optimistic or conservative) stable SB is used in the same fashion. In particular, a stable SB enables us to answer the following questions.

- Which outcomes in Z are "stable" in that they will prevail. That is, which outcomes, if happen to be the status quo, will not be replaced by farsighted rational individuals.
- How stable outcomes are reached from "non-stable" outcomes.
- Which coalitions might form in the process of replacing a non-stable outcome with a stable one.

Before we answer these questions, we shall establish a few important properties of a stable SB. The first lemma shows that predictions by a stable SB are consistent, i.e., a "stable path" satisfies a "truncation property": the continuation of a "stable path" is stable at any stage along the way. The second lemma guarantees that the existence of a stable SB implies the existence of stable outcomes in Z. Formally,

Lemma 8 *Assume that σ is a stable SB and that $\alpha \in \sigma(G_a)$. Then, for all $b \in \alpha$, $\alpha|_b \in \sigma(G_b)$.*

Lemma 9 *If σ is a stable SB, then there exists at least one $a \in Z$ such that $a \in \sigma(G_a)$.*[9]

[8] Recall that (b, c) is the path that originates from b and terminates at c.

[9] Recall that for the social environment depicted in Figure 2, there exists a unique OSSB σ which coincides with the unique CSSB and $a \in \sigma(G_a)$.

Let σ be an OSSB/CSSB for the situation with perfect foresight. Then the set of *stable outcomes under optimism/conservatism* is given by $E_\sigma = \{a \in Z \mid a \in \sigma(G_a)\}$. Each alternative $a \in E_\sigma$ is stable under optimism/conservatism in the sense that it will prevail if it is the status quo. Put differently, no coalition (who behaves optimistically/conservatively) with the power to replace a by another alternative would (eventually) benefit by doing so. Moreover, every outcome that belongs to $Z \setminus E_\sigma$ is an *unstable outcome*. Whenever such an outcome is the status quo, there is at least one coalition that can and will (eventually) benefit from replacing it.

The above assertion relies on the existence of a stable SB. Thus, it is important to investigate the existence of a stable SB. One simple sufficient condition is as follows (For more general results, see Xue (1998)).

Proposition 10 *Assume that $\phi(\mathcal{G})$ is a bounded acyclic graph.*[10] *Then there exists a unique OSSB and a unique CSSB.*

In our analysis, foresight is formalized by considering the graph of the social environment. There is an intimate connection between our notion and subgame perfection in extensive form games of perfect information – a special class of directed graphs. Indeed, if $\mathcal{G}$ represents an extensive form game of perfect information, there is an equivalence between the CSSB for the situation with perfect foresight and the notion of subgame perfection. Consequently, our analysis extends the notion of subgame perfection to social environments with diverse coalitional interactions.

Proposition 12 provides a formal result on the relationship between the largest CSSB for the situation with perfect foresight and the largest CSSB (hence the LCS) for the Harsanyi-Chwe situation.

Proposition 11 *Let $\mathcal{G}$ be a social environment. Let ψ^ℓ be the largest CSSB for the Harsanyi-Chwe situation and σ be a nonempty-valued CSSB for the situation with perfect foresight such that for every $a \in Z$, $\alpha \in \sigma(a)$, where $t(\alpha) \neq a$, implies $t(\alpha) \gg a$. Then for all $a \in Z$, $\alpha \in \sigma(a)$ implies $t(\alpha) \in \psi^\ell(a)$, i.e., σ "refines" ψ^ℓ.*

The notion of perfect foresight can also be applied to normal form games. If $\mathcal{G}$ represents the coalitional contingent threat situation, we have the following results.

[10] Figures 1 and 2, for example, are both bounded acyclic graphs.

Proposition 12 *Let a be a strict Strong Nash Equilibrium for a finite normal form game G and assume that $\mathcal{G}$ represents the coalitional contingent threat situation Then a belongs to the set of stable outcomes under conservatism as defined by the largest CSSB for the situation with perfect foresight.*

The set of stable outcomes however, may also include strategy profiles that are not Nash equilibria. By modifying the approach presented here, Xue (2000) defines the notion of "negotiation-proof Nash equilibrium", which rectifies myopia and nestedness embedded in the notion of coalition-proof Nash equilibrium (Berheim, Peleg, and Whinston, 1987).

5 A Strategic Approach

5.1 General Definitions

When discussing the stability of outcomes with respect to coalitional deviations we normally use expressions such as 'coalition S would deviate' without specifying how exactly the *individual* members of the (prospective) coalition plan and implement the deviation. In other words, we treat each possible coalition, rather than each of the individual players, as the fundamental decision unit. Importantly, this is true even when considering games described in strategic form, in which coalitions are not primitives of the analysis (think of the concepts of a Strong Nash Equilibrium or the Coalition Proof Nash Equilibrium). Attributing a will and decision power directly to coalitions suggests that one could explicitly define a noncooperative game with coalitions *as players*. The advantage of doing this is that one can then adapt the standard tools and equilibrium concepts used in noncooperative game theory to capture the ideas of farsightedness and consistency (think of subgame perfection for example). In particular, one can define *strategies* for coalitions. Consistency and farsightedness will be expressed by a *strategic equilibrium* notion. The dual nature of noncooperative strategies as both prescription and expectation of behaviour also allows to throw some light on the problem pointed out by Luce and Raiffa (1957).

> "Presumably, an adequate descriptive theory will incorporate the expectations of each potential coalition as to the reactions of the remaining players if that coalition should actually form. How

> these expectations should be calculated is far from clear." (p. 191).

At a strategic equilibrium expectations are endogenously determined as equilibrium expectations.

This line of enquiry has been pursued in Mariotti (1997) with the definition of a "Coalitional Equilibrium". The original paper proposed this concept in the context of an underlying noncooperative game described in strategic form. Here, we first generalize the basic ideas to make them coherent within the context of this chapter. Later we will move to strategic form games.

Given are: a finite set of outcomes Z; a set of players N; a set of feasible coalitions Σ, with $S \subseteq N$ for all $S \in \Sigma$; $|N|$ preference relations $\succeq_i$, one for each player i, on subsets of Z; an (possibly empty) effectiveness correspondence $r : Z \times \Sigma \to Z$. In this last definition, $b \in r(a, S)$ denotes the fact that coalition S can move the status quo from a to b. Therefore, this is simply a notationally different way of expressing a set of effectiveness relations.

We assume that preferences satisfy the following: $B \succ_i \emptyset$ for all nonempty $B \subseteq Z$; $B \succeq_i C$ if $\{b\} \succeq_i \{c\}$ for all $b \in B$, $c \in C$. The same observations made for Greenberg's TOSS apply here: any further specification of a preference relation on sets of outcomes is not part of the solution concept, but rather part of the description of the social environment under consideration.

Imagine the following procedure, inspired by Greenberg's (1990) *coalitional contingent threat situation* defined in Section 2. At each stage an outcome a is under proposal. Any coalition S may potentially induce a new outcome b provided that $b \in r(a, S)$. Once at b, any other coalition T may induce $c \in r(b, T)$, and so on. The process continues in this way until there is an outcome from which nobody wishes to deviate.

A *history* is a finite sequence of pairs $\{(S^n, a^n)\}_{n=1,..,K}$ such that for all n: (i) $a^n \in Z$ and $S^n \in \Sigma$; (ii) $a^{n+1} \in r(a^n, S^n)$. A composite history obtained by adding to h an additional element (S, a) is written $h \oplus (S, a)$

An initial history at a status quo a is represented by $\{\emptyset, a\}$. Let $\mathcal{H}$ be the set of all histories. Given $h \in \mathcal{H}$ with $h = \{(S^n, a^n)\}_{n=1,...,K}$, let $\lambda(h) = a^K$. An $S-$*strategy* is a map $\alpha_S : \mathcal{H} \to Z$ with the property that, for all $h \in \mathcal{H}$, $\alpha_S(h) \in r(\lambda(h), S)$. An S-strategy defines what each coalition would do, after each history, *if* it could move. Let Ω_S be the set of all S-strategies, and let $\Omega = \prod \Omega_S$ (assume an ordering to list coalitions).

Given $h \in \mathcal{H}$ and $\alpha \in \Omega$, we say that $b \in Z$ is *reachable* from h via α if there exists a history $h' = \{(T^n, c^n)\}_{n=1,...,K'}\}$, called a *reaching history* for b, with the following properties: (i) $c^1 = \alpha_{T^1}(h)$; (ii) for all n, $c^n = \alpha\left(h \oplus (T^1, c^1) \oplus ... \oplus (T^{n-1}, c^{n-1})\right)$; (iii) $b = c^{K'}$. That is, there must be a sequence of coalitions whose $S-$strategies lead, starting from history h, to the outcome b.

Given $h \in \mathcal{H}$ and $\alpha \in \Omega$, we say that $b \in Z$ is *terminal* from h via α if (i) b is reachable from h via α; (ii) for all reaching histories h' for b, for all $S \in \Sigma$, $\alpha_S(h \oplus h') = b$. That is, after reaching b no coalition wants to move. Let $\tau(h, \alpha) = \{b \in Z | b \text{ is terminal from } h \text{ via } \alpha\}$. Clearly, $\tau(h, \alpha)$ may be empty. By our assumption on preferences, this case is at the bottom of the preference ranking for all players. This means that infinite histories are worse than any other set of outcomes. Players only care about the final outcome and do not mind how many moves it takes to get there.

We can now define our equilibrium notion[11]. A *Coalitional Equilibrium (CE)* is a pair $(a^*, \alpha^*) \in Z \times \Omega$ such that for all $h \in \mathcal{H}$: (i) if $\lambda(h) = a^*$, a^* is terminal from h via α^*; (ii) for all $S \in \Sigma$ there is no $\alpha_S \in \Omega_S$ such that $\tau\left(h, \left(\alpha_S, \alpha^*_{-S}\right)\right) \succeq_i \tau(h, \alpha^*)$ for all $i \in S$; (iii) if for some $S \in \Sigma$ $\alpha^*_S(h) \neq \lambda(h)$ then $\tau(h, \alpha^*) \succeq_i \{\lambda(h)\}$ for all $i \in S$. If (a^*, α^*) is a CE, say that a^* is an *equilibrium point supported by* α^*.

In applications, it is often much easier to consider the stationary version of the concept, in which $S-$strategies are only conditional on the current status quo, independently of the history that led to it. A stationary $S-$strategy is thus a map σ_S from the set of outcomes to itself. A *stationary Coalitional Equilibrium* is a CE in which coalitions are restricted to play only stationary strategies.

Part (i) of the definition of a CE simply states the stationarity of the outcome a^* when the $S-$strategy profile α^* is played. Part (ii) is analogous to the definition of a subgame perfect equilibrium in an extensive form game: after no history can a coalition deviate from the equilibrium $S-$strategy and improve. However, the definition is adapted to take into account the fact that an $S-$strategy profile may not lead to a unique outcome. A further difference with a standard extensive form is that here we are not dealing with a tree. There may combination of $S-$strategies that generate infinite cycles. Condition (iii) implies that an equilibrium $S-$strategy profile always leads to a terminal outcome. The interpretation of this condition is that, at any history, the coalition that has the power to move has the option of

[11] The terminology used here differs slightly form that in Mariotti (1997). A Coalitional Equilibrium was called there a Full Coalitional Equilibrium.

implementing the status quo.

Some remarks on the correct interpretation of a strategy and an equilibrium are in now order. It does not seem plausible to assume that all coalitions (most of which may not even ever actually form!) *plan* behaviour before playing the game. The proper interpretation of a profile of $S-$strategies is as a set of common expectations regarding contingent coalitional behaviour: an $S-$strategy expresses what it is commonly expected that coalition S would do, after any history, if it formed and thus obtained the power to move. Given such expectations, the idea is that any coalition is assumed to exploit an improvement opportunity, at any history, with a (one-step, as shown below) deviation. So, for expectations to be consistent with optimality of behaviour, we need that strategies form an equilibrium. An outcome is an equilibrium outcome if it is consistently expected that no coalition will want to depart from it. This addresses the problem raised in the previous quotation by Luce and Raiffa.

5.2 Some Basic Results and Observations

We first observe informally that the 'one-deviation property' applies in this context. If a coalition can improve on a given $S-$strategy α_S, then it can do so by using a strategy that differs from α_S only in the actions prescribed at one history. The reason for this result is that infinite histories have less value than any other history; then, standard arguments can be applied.

Next, we turn to the issue of stationarity.

Proposition 13 *Any stationary CE is also a CE.*

This again follows from standard arguments, presented for example in Harsanyi (1974). Although it is reassuring to know that the restriction to stationarity does not create new equilibria, the only justification for this restriction is simplicity of computation.

Nonstationary equilibria sometimes must be invoked to guarantee existence, as the following simple example shows.

Example 3 $N = \{1, 2, 3\}$. $Z = \{a, b, c\}$, $\{b\} = r(\{1\}, a)$, $\{c\} = r(\{2\}, b)$, $\{a\} = r(\{3\}, c)$ *and* $r(S, x)$ *is empty otherwise.* $\{b\} \succ_1 \{a\} \succ_1 \{c\}$, $\{c\} \succ_2 \{b\} \succ_2 \{a\}$, $\{a\} \succ_3 \{c\} \succ_3 \{b\}$. *There is no need to specify preferences on other subsets.*

In this example each player i can move to the next outcome in the cycle, which is his favourite. But this is threatened by the next player having the power to move to *his* best outcome, which is i's worst outcome. There is no stationary CE. To see this informally, suppose for instance that a was an equilibrium point. Then b cannot be an equilibrium point: otherwise 1 would move to it. But in a CE 3 must move from c to a, given that a is an equilibrium point. So 2 would not move from b to c and b would also be an equilibrium point, a contradiction. An analogous argument shows that b or c cannot be equilibrium points either.

But consider nonstationary strategies constructed as follows. A player stays at the status quo at any initial history (that is, of the type $(x, \emptyset)$). At non-initial histories, the initial mover is 'punished' by the other two players by implementing his worst outcome. For example, a history with outcome sequence ab must have 1 as the first mover. Then at this history 2 moves to c and 3 stays at c. If some player deviates from this punishment pattern, he is punished in the same way if possible (in the case above if 2 fails to move to c he cannot be punished -but observe that 2's deviation would clearly be non-optimal). And so on. So, for example, suppose that the history with sequence of outcomes abc (player 1 to be punished) is followed by a (player 3 has moved to a instead of punishing 1 by staying at c), so that we are at the history with action sequence $abca$ and it is 1's turn to move. According to the strategies described, what happens is that 1 moves to b and 2 stays at b, implementing 3's worst outcome.

It is not difficult to see that the strategy profile sketched above constitutes a CE. Play does not move from a, b and c when starting at those outcomes.

This type of 3-cycle example is notorious for generating existence problems. For instance, there is no von Neumann Morgenstern Stable Set in the associated abstract system. Using a non-stationary CE allows one to glean some insight in the complex social behaviour and expectations which may occur in this situation. For comparison, the LCS includes the same set of outcomes, but little light is shed on the expectations sustaining those outcomes.

In general, existence of a CE is hard to establish, although there are no examples of non-existence. Existence results can be obtained for some limited case. For instance, existence of a (stationary) CE is guaranteed whenever the effectiveness correspondence is such that all possible histories are acyclic, in the sense that they cannot reach the same outcome twice (but possibly involve the same coalition moving more than once). To see this, note that this form of acyclicity implies that there exist outcomes from which no coalition can move. The equilibrium $S-$strategies can then be constructed by backward induction, starting at these outcomes. Another existence result is given in the next subsection.

To give the flavor of the characterisation results possible in this framework, we sketch one on Pareto optimality, which generalizes Proposition 4.2 in Mariotti (1997).

Proposition 14 *Suppose that* $r(\{N\}, a) = Z$ *for all* $a \in Z$*. Let* (a^*, α^*) *be a stationary CE such that* a^* *is the unique equilibrium point supported by* α^**,in the sense that if* (a^{**}, α^*) *is also a CE, then* $a^* = a^{**}$*. Then* a^* *is Pareto optimal in* Z*, in the sense that there is no other outcome* b *such that all players prefer* $\{b\}$ *to* $\{a\}$*.*

The first assumption in the statement says that the grand coalition can move to any outcome from any outcome. To see that the result holds, suppose that $a^*$was Pareto dominated, in the sense described in the statement, by another outcome b. Then, for it to be optimal for the grand coalition not to move to b from a^* it must be the case that α^* prescribes some coalition S to move from b. Condition (iii) in the definition of a CE implies that this leads to a terminal point c preferred by S to $\{b\}$. But such a terminal point c is clearly an equilibrium point, so by uniqueness $c = a^*$. This yields the contradiction that the members of S prefer $\{a^*\}$ to itself.

A similar argument shows that in unanimity games (where all players agree on what is the best outcome), the unique equilibrium point is the unanimous best outcome.

5.3 Strategic Form Games

More precise results can be obtained in specific contexts. The original version of a CE was proposed in the context where the underlying strategic situation is described by means of a strategic form game $F = (N, \{Z_i\}_{i \in N}, \{u_i\}_{i \in N})$, where Z_i is the strategy space of player i and u_i his utility function. Therefore, the outcome space Z is the Cartesian product of the Z_i. The preference relation $\succeq_i$ was assumed to be as follows: $B \succeq_i C$ iff there exists $b \in B$ such that $u_i(b) \geq u(c)$ for all $c \in C$. Then, one can prove:

Proposition 15 *Any stationary Coalitional equilibrium point yields each player more than his maximin value.*

Proposition 16 *Let F be a two-player game such that: (i) the Pareto frontier is strictly decreasing and (ii) there exists a Pareto efficient outcome a^* that dominates the pair of maximin values. Then a stationary Coalitional Equilibrium exists, with a^* as an equilibrium point.*

These two results apply directly to the simple but interesting example of the Prisoner's Dilemma game. Proposition 15 tells us that {Cooperate, Defect} and {Defect, Cooperate}, which yield one of the player his lowest, and individually irrational, payoff, cannot be equilibrium points. Proposition 16 tells us that {Cooperate, Cooperate} can be supported as an equilibrium point. Proposition 14 tells us that {Defect, Defect} cannot be the unique equilibrium point. Finally, it is not difficult to see that {Cooperate, Cooperate} and {Defect, Defect} cannot coexist as equilibrium points. We can conclude that {Cooperate, Cooperate} is the unique equilibrium point. The reason why it is claimed that this game is not trivial is that other solution concepts yield different results. For example, the Strong Equilibrium Set, the Optimistically Stable Standard of Behavior of the coalitional contingent threat situation and the von Neumann Morgenstern Stable Set of the associated abstract situation are all empty. The Coalition Proof Nash Equilibrium is {Defect, Defect}. The Largest Consistent Set is {Cooperate, Cooperate}. One concludes that , even in such simple games, what outcome or set of outcomes prevails is very sensitive to exactly how farsighted players are assumed to be and to the details of the negotiation procedure.

Another interesting example is the following[12].

	L	R	L	R
T	$2,2,2$	$0,0,0$	$0,0,0$	$4,4,1$
B	$0,0,0$	$0,0,0$	$0,0,0$	$3,3,3$
		L	R	

Here there are two stationary Coalitional Equilibria, one with $\{(B, R, R)\}$ as the unique equilibrium point, the other with $\{(T, R, R)\}$ as the unique equilibrium point. For comparison, the Largest Consistent Set is

$$\{(T, L, L), (B, R, R), (T, R, R)\}$$

Let us try to understand these different results in terms of the expectations the player must have. In particular, observe that though (T, L, L) is included

[12] Taken from Mariotti (1997).

in the LCS it cannot be a Coalitional equilibrium point. Why? Because if it was, this would imply that (B, R, R) is not an equilibrium point (deviation by the grand coalition). In turn, the only explanation for this is that it is a common expectation that Row would deviate to (T, R, R), and so that (T, R, R) is also an equilibrium point. But this is not a consistent expectation, because there is nothing to stop Box to deviate to (T, R, L). In particular, this deviation cannot be deterred by the fear or returning to (T, R, R) (via (B, R, R)): for in this case Column would upset the stability of (T, L, L) by moving to (T, R, L).

It is clear from this analysis that the inclusion of (T, L, L) in the LCS can only be supported by expectations which are inconsistent across players about what happens at (T, R, L): on the one hand, for Column to stay at (T, L, L) he must think that play would lead back from (T, R, L) to (T, L, L) after his deviation; but, on the other hand, for Box to stay at (T, R, R) he must think that play would lead back from (T, R, L) to (T, R, R) after his deviation.

6 Concluding Remarks

We have presented a selective survey of what we believe is an important issue, the formalisation of farsightedness in coalition formation. This issue has not played a dominant role in the literature so far. By contrast, consider extensive form games, where the aspect of farsightedness of individual players is recognised as crucial, and is taken into account when using concepts such as subgame perfection, sequentiality, and so on[13].

As we have discussed, in the literature considered in this chapter there isn't yet a single equilibrium concept on which consensus has been reached. We have argued, however, that Greenberg's Theory of Social Situations provides a useful unifying framework to study the issue, and we have examined in depth some subsequent contributions which in one way or the other are related to that theory. Other approaches and insights are offered, for example, by Arce M. (1996), Chakravorti (1999), Diamantoudi (2000), Konishi and Ray (2000), Li (1992) and Suzuki and Muto (2000). Applications to the study of International Environmental Agreements are provided, for example,

[13] Similarly, Brams' (1994) Non Myopic Equilibrium is motivated by the desire to endow with farsightedness players acting in situtations which are described as strategic form games.

in Ecchia and Mariotti (1997, 1998).

What has been offered so far is an increased understanding of (the implicit assumptions underlying) existing solution concepts, as well as some proposals for improvement. But this is doubtless an exciting area of research in which much work is ongoing and much still remains to be done.

7 References

Arce M., D.G., On the Policing Properties of Coalition-Proof versus Farsighted Strong Equilibria, mimeo, 1996

Bernheim, B.D., B. Peleg, and M.D. Whinston, Coalition-Proof Nash Equilibria. I. Concepts, *Journal of Economic Theory,* **42**, 1, 1–12, 1987

Brams, S., "Theory of Moves", Cambridge University Press, Cambridge, 1994

Chakravorti, B., Far-sightedness and the Voting Paradox, *Journal of Economic Theory,* **84**, 2, 216–226, 1999

Chwe, M. S.-Y. Farsighted Coalitional Stability, *Journal of Economic Theory,* **63**,2, 299–325, 1994.

Diamantoudi, E., Cartel Stability Revised, mimeo, University of Aarhus, 2000

Ecchia, G. and M. Mariotti, Coalition Formation in International Environmental Agreements and the Role of Institutions, *European Economic Review* **42**, 3, 573-82, 1998.

Ecchia, G. and M. Mariotti, The Stability of International Environmental Coaltions with Farsighted Countries: Some Theoretical Observations, in C. Carraro (Ed.) "International Environmental Agreements: Strategic Policy Issues", Edward Elgar, Cheltenham, 1997.

Greenberg, J., "The Theory of Social Situations", Cambridge University Press Cambridge 1990

Greenberg, J., Coalition Structures *in* "Handbook of Game Theory", (R. Aumann and S. Hart, Eds.), North Holland, Amsterdam, 1994

Greenberg, J., D. Monderer, and B. Shitovitz, Multistage Situations, Econometrica, **64**, 6, 1415-37, 1996

Harsanyi, J.C., An Equilibrium-Point Interpretation of Stable Sets and a Proposed Alternative Definition, *Management Science,* **20**, 4, 1472–1495, 1974

Konish, H. and D. Ray, Coalition Formation as a Dynamic Process, mimeo, Boston College, 2000

Li, S., Far-sighted Strong Equlibrium and Oligopoly, *Economic Letters*, **40**, 1, 39-44, 1992

Luce, R.D. and H. Raiffa, “Games and Decisions”, John Wiley and Sons, New York, 1957

Mariotti, M., A Theory of Agreement in Strategic Form Games, *Journal of Economic Theory,* **74**, 1, 196-217, 1997

Page, F., M.H. Wooders, and S. Kamat, Networks and Farsighted Stability, mimeo, 2001

Ray, D., Credible Coalitions and the Core, *International Journal of Game Theory,* **18**, 2, 185-87, 1989

Rosenthal, R.W., Cooperative Games in Effectiveness Form, *Journal of Economic Theory,* **5**, 1, 88–101, 1972

Suzuki, A. and S. Muto, Farsighted Stability in an n-Person Prisoner's Dilemma, mimeo, Yamagata University, 2000

von Neumann, J. and O. Morgenstern “Theory of Games and Economic Behavior”, Princeton University Press first edition, 1944, second edition, 1947, Princeton, NJ

Xue, L., Nonemptiness of the Largest Consistent Set, *Journal of Economic Theory,* **73**, 2, 453–59, 1997

Xue, L., Coalitional Stability under Perfect Foresight, *Economic Theory,* **11**, 3, 603–627,1998

Xue, L., Negotiation-Proof Nash Equilibrium, *International Journal of Game Theory,* **29**, 3, 339-357, 2000