National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

Coalitional Stability in Strategic Situations

LICUN XUE

Department of Economics

McGill University, Montreal

May 1996

A THESIS SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF

THE DEGREE OF

DOCTOR OF PHILOSOPHY

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

ISBN 0-612-12513-0

Canada

Contents

Abstract

In many (social, economic, and political) strategic situations, conflict and cooperation coexist and group (or coalitional) behavior is as important as individual behavior. This dissertation studies several issues in such situations.

Chapter 1 provides an overview of the theoretical background and motivates the analysis undertaken.

Chapter 2 analyzes strategic situations with diverse coalitional interactions to ascertain the "stable" outcomes that will not be replaced by any rational (hence farsighted) coalition of individuals, and the coalitions that are likely to form. The analysis takes into full account the perfect foresight of rational individuals, which has been overlooked in the literature.

Chapter 3 defines "negotiation-proof Nash equilibrium", a notion that applies to environments where players can negotiate openly and directly prior to the play of a noncooperative game. The merit of the notion of negotiation-proof Nash equilibrium is twofold: (1) It resolves the nestedness and myopia embedded in the notion of *coalition-proof Nash equilibrium*. (2) The negotiation process, which is formalized by a "graph", serves as a natural alternative to the approach that models pre-play communication by an extensive form game.

Chapter 4 examines the notion of "renegotiation-proofness" in infinitely repeated games. It is shown that imposing renegotiation in all contingencies creates both conceptual and technical difficulties. A notion of self-enforcing agreements is offered: an agreement is self-enforcing if it is immune to any deviation by any coalition which cannot (confidently) count on renegotiation.

Résumé

Dans beaucoup des situations stratégiques (sociales, économiques, ou politiques), le conflit et la coopération coexistent, et le comportement du groupe (ou coalitional) est aussi important que le comportement individuel. Cette thèse étudie plusiers situations de cette nature.

Chapitre 1 fournit un survol du antécédent théorique, et motiver de l' analyse entreprise.

Chapitre 2 analyse les situations stratégiques avec intéractions diverse coalitionalles pour découvrir les résultats stables qui ne sont pas replacés par aucune coalition des individus rationnels (et par conséquent clairvoyant), et les coalitions qui sont probable de se former. L' analyse tient compte de la prévoyance parfaite des individus rationnels, qui n' est pas remarquée par la littérature.

Chapitre 3 examine les accords autoforcés que les joueurs peuvent atteindre dans une négociation publique avant de jouer des jeux non-cooperatifs. Une notion des accords autoforcés est offerte, qui fait mieux que les notions précédentes, en demandant que les accords soient autoforcés contre toute déviation, et en captant la prévoyance parfaite des individus rationnels.

Chapitre 4 examine la notion de l' épreuve de renégociation dans les jeux à répétition infinie. On a montre que l'imposition de la renégociation dans toutes éventualités crée des difficultés conceptuelles et techniques. Une notion des accords autoforcés est offrie: Un accord est autoforcé si il n'est immunisé contre chaque déviation par aucune coalition, qui ne peut pas (avec confiance) compter sur renégociations.

Acknowledgements

Those who are interested enough or obliged to read this thesis may well know that to complete a Ph.D. program is no easy task, which would not have been possible without the help of many people. One person, to whom I am indebted the most, is my thesis supervisor, Professor Joseph Greenberg (Yossi). He provided me with excellent supervision through his guidance, encouragement, and inspiration. He introduced me to a wide range of literature in economic and game theory; most of all, he lead me into the "wonderful world" of academic research: He taught me not only to be formal and precise and but also to go beyond the technicalities and concentrate on fundamental issues. The attention and support I received from Yossi were more than most graduate students could expect. Not only did he spend a lot of time with me on research, but also he made sure I was well exposed: he arranged me to meet many of our visitors; he encouraged and supported me to attend many conferences; .... But, to me, Yossi has not been merely an academic supervisor. I also consider him as a good friend. He helped me to get through many difficult (academic and personal) times.

I am grateful to Professor Venki Bala, member of my thesis committee, who has provided me with many insightful suggestions on all sort of issues in academia throughout my graduate studies. I appreciate very much his insights and friendship. My gratitude also goes to Professor Vicky Zinde-Walsh, who supported me tremendously in the early part of my graduate study. I also thank Vicky for her understanding when I switched to economic and game theory and for her continuous help after that.

I am grateful to Professor Binya Shitovitz for guidance and suggestions, especially when I was working on my very first formal proof (of an existence theorem). I am in debt to Professor G.B. Asheim, who read the entire manuscript on various occasions and provided me with many insightful comments. His acknowledgement

of my work means much to me, being at a very early stage of my academic career.

I would also like to thank Professors Dov Monderer, Michael Chwe and Rob Gilles, and Dr. X. Luo for very helpful comments on various parts of this thesis. I would like to thank Professors Daniel G. Arce M. and Pierpaolo Battigalli for helpful discussions.

Professor N.V. Long, who taught me the first graduate microeconomics, has helped me in many aspects. He was also the one who introduced me to Professor Joseph Greenberg, at a time when my knowledge of game theory or game theorists was minimal.

I thank Professor John W. Galbraith, who did not only teach me a lot of Econometrics, but also gave me a lot of help and advices during my study. Many thanks go to Professors Chris Green for his encouragement and continual interest in what I have been doing. I would like to thank Professor Tom Velk for his interest in and comments on a part of this thesis. I also take this opportunity to thank Professors Bart Hamilton, S. Hogan, and C. Ragan for their help and advices when I took their courses.

I gratefully acknowledge the financial support from the *Social Sciences and Humanities Research Council of Canada* (SSHRC) through a doctoral fellowship, from the McGill economics department, and from SSHRC and FCAR, Quebec through Professor Joseph Greenberg's research grants. I also would like to thank the McGill graduate faculty and FCAR for awarding me fellowships.

Finally, I would like to thank my wife Faye for sharing with me a life that has been so far quite stressful. Her love and support have helped to get through many difficult days which could have been unbearable.

Coalitional Stability in Strategic Situations

Chapter 1
Introduction

This dissertation analyzes several issues in strategic situations where group behavior is as important as individual behavior and there are interactions among individuals as well as among groups. In particular, it investigates the outcomes that are likely to prevail in situations with diverse interactions among individuals and groups of perfect foresight (Chapter 2); it also addresses the issues of negotiation (Chapter 3) and renegotiation (Chapter 4) in strategic situations. The interactive nature of the problems under investigation implies that individuals and groups, who are assumed to be rational, behave *strategically* in that they have to consider their knowledge and expectations of the behavior of the others. While the problems under investigation are of obvious empirical importance, their theoretical importance cannot be fully exposed without discussing game *theory* itself.

Game theory studies conflict and cooperation in situations where decision makers interact and their decisions affect each other's welfare. The publication of "Theory of Games and Economic Behavior" by von Neumann and Morgenstern in 1944 marked the foundation of game theory, and, since then, game theory has advanced considerably. The significant impact of game theory on social sciences especially on economics is evidenced by the fact that 1994 Nobel prize in economics was awarded to three game theorists. Game theory is now part of almost every economist's "tool-kit". Despite the progress that has been made, game theory is still in the early stage of its development.

Following von Neumann and Morgenstern (1944), game theory distinguishes between two approaches: the "cooperative approach" and the "noncooperative approach". The cooperative approach was the main subject of investigation in 1950s and 1960s. According to such an approach, a social environment is described

by a cooperative game that associates each group or coalition of players with a set of payoffs it can achieve for its members independent of the rest of the players, and the players are assumed to be able to communicate, to coordinate their actions, to transmit threats, and to reach binding agreements. That is, players are able to negotiate outside the formal structure of the game rules. Different negotiation processes lead to various different solution concepts: the core, the stable set, the bargaining set, to name a few. The core, for example, is the set of outcomes (or distributions of welfare) that are immune to any conceivable coalitional deviation. The description of a social environment as a cooperative game, however, does not capture the externalities of the actions of one coalition upon the remaining players. This; together with the involvement of binding agreements, greatly limit the applicability of the cooperative approach.

During the last two decades, the emphasis of game theory has shifted to the noncooperative approach, which concentrates on the individual and on what strategy a selfish individual should/would use. The noncooperative approach represents a social environment as either a normal form game or an extensive form game. A normal form game is static: players choose strategies independently and simultaneously and payoffs are derived once each player submits his choice of a strategy. On the other hand, an extensive form is dynamic: players act sequentially in a specific order. The ruling solution concepts for noncooperative games are Nash equilibrium and its refinements, most of which make no attempt to account for coalition formation or any mode of collusion among players.

The assumption that players reveal their strategies simultaneously and cannot communicate their choice is a strong assumption. Such an assumption is only plausible when a game is a two-person strictly competitive game, where what is good for one player is bad for the other player, hence communication serves no purpose. In most games, as Ordeshook (1986, p. 302) wrote, "it is probably rare, though, that communication among people, however imperfect, remains impossi-

ble". In fact, individuals are often compelled to communicate in order to achieve outcomes that are mutually beneficial. The cooperative approach models the situations where communication not only is possible but also stands as a central feature of human interactions; this approach abstracts away the details of the procedure of communication and cooperation, and concentrates, instead, on the possibility of agreements. However, the cooperative approach may also abstract away the externalities inherent to noncooperative games.

The noncooperative approach, following Nash (1951), maintains that noncooperative games are more fundamental than cooperative games and that cooperative games can and should be subsumed under the noncooperative approach by making communication and bargaining formal moves in a noncooperative extensive form game. The resulting game would have an enlarged domain of strategies, and the payoff functions could be extended in the natural way. Then one can analyze the consequences of communication and cooperative behavior by applying Nash equilibrium or its refinements to the "transformed game". Such an approach is complex and unnatural. McKinsey (1952, p. 359) pointed out,

> It is extremely difficult in practice to introduce into the cooperative games the moves corresponding to negotiations in a way which will reflect all the infinite variety permissible in the cooperative game, and to do this without giving one player an artificial advantage (because of his having the first chance to make an offer, let us say.

Moreover, modeling communication and bargaining as formal moves in an extensive form game not only is restrictive but also may bury some of the most important aspects of communication. As Aumann (1987, p. 463) wrote,

> ... problems of negotiation are usually more amorphous; it is difficult to pin down just what the procedures are. More fundamentally, there is a feeling that procedures are not really that relevant; that it is the possibilities for coalition forming, promising and threatening that are decisive, rather than whose turn it is to speak.

The division between the cooperative and noncooperative approach is unfortu-

nate, since in most social environments, conflict and cooperation coexist. Selfishness does not preclude individuals from cooperating or coordinating their actions in a mutual beneficial fashion. Moreover, our society is organized in such a way that many of our social, political, and economic activities can only be conducted by groups of individuals. Given that most social environments involve interactions among individuals as well as among groups or coalitions, game theory need to recognize that both coalitional (or group) behavior and individual behavior are equally important. Shubik (1984, p.7) wrote,

> A theory of games is , among other things, a theory of organization. It deals not so much with feasibility as with negotiation and enforceability – with the power of individuals or groups to influence the distribution of goods and welfare, whether by threats and collusion or by unilateral action.

Instead of modeling procedural details of communication, we can concentrate on what communication can achieve. Communication admits coalition formation, which enable players to coordinate their actions, through binding or nonbinding agreements. This view can apply directly to noncooperative games. This approach is a hybrid of both the noncooperative and cooperative approaches, and it can preserve the noncooperative ingredients as well as the externalities inherent to noncooperative games. Such an approach, in my view, is not less fundamental than the noncooperative approach. If the noncooperative approach is mainly motivated by the selfishness of the individuals, then it is not necessary adhere to such an approach that makes no attempt to account for coalition formation or any other mood of cooperation. After all, cooperation does not necessarily contradict selfishness. Among the first to account for coalition formation in noncooperative games, the notion of "strong Nash equilibrium" (Aumann 1959) allows any coalition to coordinate the choices of strategies of its members, and a Nash equilibrium is strong if it is immune to any conceivable coalitional deviation. However, this notion involves, at least implicitly, binding agreements (among the members of deviating coalitions), since without binding agreements, a coalitional devia-

tion may be subject to further deviations. Furthermore, if binding agreements are available, then there is no need to restrict our attention to Nash equilibria. Coalitions can and do form in the absence of binding agreements. The notion of "coalition-proof Nash equilibrium" (Bernheim et al. 1987) attempts to capture the notion of self-enforcing agreements for environment with unlimited, nonbinding preplay communication. But as we shall discuss later in this chapter and more formally in Chapter 3, this notion involves myopia and agreements that are not fully self-enforcing.

Communication can also be introduced in an environment when a normal form game is repeated finitely or infinitely many times. The notion of "renegotiation-proofness" (see Chapter 4) answers the following question: among the abundance of subgame perfect equilibria, particularly in infinitely repeated games, which equilibria are proof against renegotiation in *every contingency*? Thus, renegotiation-proofness entails that players have the opportunity to renegotiate *after every history of play*, and this fact is common knowledge. Consequently, each player, in contemplating a deviation, is confident that his deviation is followed by a renegotiation; the grand coalition renegotiates regardless of whether the renegotiated agreement will be honored. Therefore, given its demanding assumption, renegotiation-proofness should not be the only way to introduce cooperative behavior and select equilibria into repeated games.

Despite of the progress in the coalitional analysis, there are many questions that remain to answered. As Roth (1988) pointed out, coalition formation "remains one of the most difficult and important problems". Not only cooperative behavior in the current paradigm of noncooperative games need much further study, but also it is necessary to end the division between the cooperative and noncooperative approaches and seek a general framework.

> The task of formalizing the communication process, especially the prohibitions against communication among the players, is far from trivial. It appears that to include it in a generalization of game theory will be an ma-

> jor theoretical step. Lacking such a generalization, several tacks have been taken, each of which in unhappily special and arbitrary (Luce and Raiffa 1957, p.p. 164-165).

Moreover, as discussed earlier, it is not natural to study coalition formation within the framework of extensive form games where the order of moves is important: extensive form games "depend very strongly on the precise form of procedures, on the order of making offers and counter-offers, and so on" (Aumann 1987, p.463). In general, a framework that deals explicitly with coalitional dynamics is lacking.

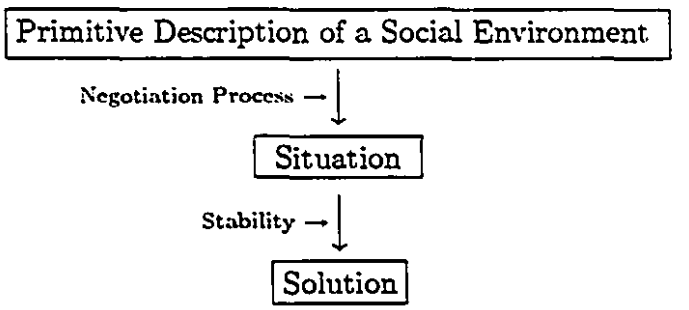
> It is a sad fact that we still lack a general theory of cooperative games in extensive form. The standard solution theories tell us next to nothing about coalitional dynamics ... (Shubik 1984, p. 68)

In fact, that a general and unified framework need to emerge was perceived by von Neumann and Morgenstern when they formally founded game theory. von Neumann and Morgenstern (1947, p. 608) raised objections to the two distinct theories they were forced to employ, and suggested that when the theory is more mature it may be unified.

"The Theory of Social Situations" (Greenberg 1990) offers an integrative approach to the study of strategic interactions. First, this approach unifies the description of "noncooperative" and "cooperative" social environments, thereby allowing for formulations of diverse interactions. Second, this approach represents a social environment as a "situation" which forces the specification of 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 exact negotiation process (e.g., how player make use of their opportunities). Third, the theory of social situations offers a unified solution concept by insisting on the single *stability* criterion.

Such a stability criterion has an appealing interpretation in the context of social norms and organizations [see von Neumann and Morgenstern (1947) and Green-

berg (1990)]. The procedure of applying this approach is as follows. Starting with a primitive description or the raw data of a social environment, the negotiation process and all institutional and behavioral assumptions can be formalized as a *situation*. Then the notion of stability is applied to derive the "solution".



This approach provides a suitable framework for the study of conflict and cooperation. This approach, together with the notion of von Neumann and Morgenstern abstract stable set, are the primary tools for the analysis in this dissertation.

The starting point of Chapter 2, "Coalitional Stability under Perfect Foresight" is the observation that many social environments comprise both "cooperative" and "noncooperative" ingredients and the diversity of coalitional interactions cannot be captured by either cooperative or noncooperative games. The primitive description of the social environment follows that of Chwe (1994), which is similar to Rosenthal's (1972) game in effectiveness form. The social environment specifies a set of alternatives (or status quo's if temporarily under consideration or outcomes if implemented) and a set of players who can rank (at least partially) the alternatives. Furthermore, it specifies that if one alternative is the "status quo", which coalition is endowed with the power to make which alternatives the new status quo's. Note that in such an environment the actions of each coalition may have externalities on the welfare of the remaining players. The question is which outcomes will prevail and which coalitions are likely to form without binding agreements in such a social environment if the actions are public, or alternatively, what (possibly binding) agreements can be reached if the players can

openly negotiate.

The noncooperative approach is not suitable here since it requires the specification of strategies for each player. Treating the social environment as an "abstract game", the notions of "abstract core" and "von Neumann Morgenstern abstract stable set" (the extensions of core and stable set to abstract games) can be applied. These notions, however, may embody myopia on the part of the players. According to the definition of abstract core, a coalition deviates in spite of the fact that further deviations may lead to outcomes that do not benefit its members (this is also related to the issue of credibility of a deviation); and according to the definition of abstract stable set, a coalition will not deviate if further deviations may occur, even though these further deviations may ultimately lead to outcomes that benefit its members. In the context of a cooperative game (which is special case of the social environment under consideration), Harsanyi (1974) argued that this lack of foresight is due to the fact that the "dominance relation" used in the definition of stable set is a myopic one, and the problem of myopia can be solved by replacing the direct dominance with an "indirect dominance" which allows the players to look arbitrarily many steps ahead. Thus, following Harsanyi (1974) one need only to apply abstract stable set with indirect dominance to the social environment under investigation. As is shown in Chapter 2, however, this does not solve the problem of myopia, because indirect dominance capture only partial foresight. This also emerges in Chwe's (1994) largest consistent set that is also based on indirect dominance. Such a discovery is made possible by analyzing the *situation* that represents the negotiation process underlying the notions built on indirect dominance. Such a discovery also motivated us to formalize perfect foresight. The idea, roughly speaking, is to view the social environment as a "graph" (which may not be acyclic). The formalization of perfect foresight as a situation and the stability criterion allows us to derive the outcomes that will not be replaced by farsighted players and the coalitions that are likely to form among

farsighted players.

Chapter 3 returns to the issue of "cooperation in noncooperative games". The notion of "negotiation-proof Nash equilibrium" is defined to answer the following question: "if players can engage in pre-play negotiation prior to the play of a non-cooperative game, what self-enforcing agreements are negotiation-proof ?" The pre-play negotiation is modeled from the viewpoint that communication admits coalition formation, but each coalitional agreement is nonbinding. It is *a priori* that every coalition can form to make joint objection to any strategy profile under consideration. Rationality (and perfect foresight) of the self-interested players dictates which coalitions might actually form in the open negotiation: rational players form a coalition only if it is in the best interest of each member to join this coalition. Thus, my approach to communication is consistent with selfishness of the players. This approach offers a natural alternative to the one that models communication as an extensive form game [see, e.g., Farrell (1987, 1988), Rabin (1994)]. Furthermore, the notion of "negotiation-proof Nash equilibrium" resolves the nestedness restriction (after a coalition deviates, only its subsets can further deviate) and myopia embedded in the definition of *coalition-proof Nash equilibrium* (Bernheim et al. 1987). A Nash equilibrium is negotiation-proof if and only if no coalition can deviate in such a way that its deviation will *ultimately* lead to another negotiation-proof Nash equilibrium that benefits all its members. The notion of negotiation-proof Nash equilibrium is also extended to extensive form games (including finitely repeated games), with emphasis on the difference between negotiation-proofness and renegotiation-proofness.

Chapter 4 examines the issue of "renegotiation-proofness" in the context of infinitely repeated games. In particular, it questions the study of cooperative behavior and equilibrium selection through renegotiation-proofness and argue that imposing (or introducing, almost mechanically,) renegotiation in every contingency is at least very demanding. This chapter defines the notion of "stable (self-

enforcing) agreements" in infinitely repeated games where players can coordinate their actions but cannot make binding contracts, thereby offering an alternative to the study of cooperation through the notion of renegotiation-proofness. It differs from *renegotiation-proofness* in that it allows for any coalition to deviate, and moreover, a deviating coalition does not count on renegotiating with nonmembers. In addition to its intuitive appeal, stable agreements can resolve the conflict between efficiency and renegotiation: the set of stable agreements is nonempty and efficient (within the set of subgame perfect equilibrium outcomes) for a large class of games including all two-player games and all games for which every efficient subgame perfect equilibrium path is stationary.

I hope to show, through this dissertation, that the study of cooperation, coalitions, and agreements not only is essential but also can be fruitful.

Chapter 2

Coalitional Stability under Perfect Foresight

Consider a social environment with diverse coalitional interactions. What outcomes are "stable" in that they will not be replaced by any coalition of rational (hence farsighted) players? What coalitions are likely to form? This chapter addresses these issues. The analysis undertaken focuses on the perfect foresight of rational players that has been overlooked by the notions suggested in the literature for similar social environment. Perfect foresight is formalized by the means of a "situation" (Greenberg 1990) which specifies explicitly how farsighted players view and use their available alternatives, and the notion of stability [von Neumann and Morgenstern (1947) and Greenberg (1990)] is used to derive the "stable outcomes" and the coalitions that are likely to form to bring about these outcomes.

2.1 Introduction

This chapter defines a solution concept for strategic social environments with diverse coalitional interactions. It improves on previous solution concepts for similar social environments in that it captures the perfect foresight of the individuals. The primitive description of a social environment follows that of Chwe (1994), which is sufficiently flexible to integrate the representation of a cooperative game, an extensive form with perfect information, and a normal form game played in such a fashion that there are moves and counter moves. Moreover, the description can accommodate social environments of more complex structure. Particularly, it allows for cooperation within a coalition and at the same time (noncooperative) interactions among coalitions (the action taken by one coalition may impose externalities upon the payoffs of the other coalitions).

In most economic and game theoretic models, individuals or agents are presumed to be rational and intelligent. In a non-strategic setting, perfect foresight

as implied by rationality and intelligence is captured by dynamic consistent plans (policies) derived from dynamic programming. In strategic social environments, however, the interactive nature of the decision making poses more challenges. Myopia in the Cournot model was criticized by a number of scholars [see, for example, Chamberlin (1933)], for the reason that each firm ignores the reactions of its rivals. The notion of coalition proof Nash equilibrium (Bernheim et al. 1987) may also be subject to the criticism of myopia [see Chwe (1994)]. In the context of cooperative games, Harsanyi (1974) criticizes the von Neumann and Morgenstern solution for its failing to incorporate foresight: in order to deter deviations, it is not sufficient that further deviations will take place; what deters farsighted individuals from deviation is that the resulting (final) payoffs would make them worse off (see Appendix 2). Hence the von Neumann and Morgenstern solution based on "direct dominance" may be subject to the "destabilizing effect of indirect dominance". "Indirect dominance" captures the fact that farsighted individuals look ahead and it is the *final* payoffs that individuals care about. This very idea can be extended to more complex social environments such as the ones studied by Chwe (1994) and this chapter. Based on Harsanyi's (1974) "indirect dominance" and motivated by the fact that the von Neumann and Morgenstern abstract stable set[1] with indirect dominance may be too exclusive in that it can rule out "arbitrarily", Chwe (1994) defines the *largest consistent set*, a weaker notion that is "not so good at picking out, but ruling out with confidence" (Chwe 1994, p. 239). It turns out, however, that the largest consistent set may be too inclusive in many situations. I shall show that the inclusiveness of the largest consistent set and the exclusiveness of the abstract stable set with indirect dominance are not isolated phenomena. They both stem from the fact that indirect dominance as defined in the literature does not capture perfect foresight: Individuals consider only the final payoffs but not how, or if at all, these payoffs can be reached; that

[1]This is a generalization of the more familiar notion of the von Neumann and Morgenstern solution (stable set) for cooperative games.

is, deviations "on the way" to the final outcomes are ignored.

The purpose of this chapter is to develop a solution concept that captures perfect foresight. By examining the negotiation/reasoning process underlying the notions based on indirect dominance, I show that indirect dominance overlooks the "graph" (formally defined in Section 3) of the social environment. The formalization of perfect foresight in this chapter recognizes the "graph structure" of the social environments and uses "paths" as the building blocks. In doing so, all deviations "on the way" to the final outcomes are considered. The necessity for "paths" is not obvious for a complex social environment that does not possess a tree structure (for example, a social environment represents a cooperative game or a normal form game). The solution concept I shall develop is derived by applying the notion of stability that is introduced by von Neumann and Morgenstern (1944) and generalized and further developed by Greenberg (1947).

The organization of the rest of this chapter is as follows: In Section 2.2, I formally define the social environment to be analyzed. Then I introduce the solution concepts suggested in the literature for such a social environment. By analyzing these solution concepts, I show why they do not capture perfect foresight and identify the underlying problems. In Section 2.3, I formalize perfect foresight by considering the graph of the social environment and using "paths" as the building blocks. Applying the notion of stability, I derive the "stable outcomes" and the coalitions that are likely to form to bring about these stable outcomes. In Section 2.4, I re-examine the literature by comparing, both formally and through examples, the negotiation and reasoning process underlying and the implications of the solution concepts in the literature with the solution concept I introduce. Section 2.5 concludes the chapter by a brief comment on the methodologies that are relevant to this chapter and points out several issues for further research. All proofs are relegated to Appendix 2.1. Appendix 2.2 provides a simple cooperative game that illustrates Harsanyi's criticism of the von Neumann and Morgenstern

solution. A theorem that generalizes Chwe's (1994) result on the nonemptiness of the largest consistent set is given in Appendix 2.3. Appendix 2.4 gives some formal discussion on modeling foresight.

2.2 Foresight and Stability in the Literature

Consider a social environment with a set of individuals, N, who face a set of alternatives Z. Each individual $i \in N$ has a strict preference relation $\prec_i$ on Z. Coalitions[2] may be endowed with the power to replace one alternative by some other alternatives. If coalition $S \subset N$ is endowed with the power to replace $a \in Z$ by some $b \in Z$, we write $a \xrightarrow{S} b$. Using Chwe's (1994) notation, a social environment is represented by $\mathcal{G} = (N, Z, \{\prec_i\}_{i \in N}, \{\xrightarrow{S}\}_{S \subset N, S \neq \emptyset})$.

In this section, I shall introduce and analyze the solution concepts in the literature for social environments represented by $\mathcal{G}$, and identify the lack of foresight in these notions. Before I proceed, I shall use some examples to illustrate the flexibility of $\mathcal{G}$, thereby facilitating the understanding of the social environments depicted by $\mathcal{G}$.

Normal Form Games. Assume that a normal form game is played in such a fashion that there are moves and counter-moves. Study of normal form game played in such a fashion can be found, for example, in Greenberg (1990), Brams (1994), and Chwe (1994). A normal form game is a triple $G = (N, \{Z^i\}_{i \in N}, \{u^i\}_{i \in N})$, where N is the set of players and for $i \in N$, Z^i is the nonempty set of strategies of player i and u^i is player i's payoff function, $u^i : Z^N \to \mathcal{R}$, where for $S \subset N$, Z^S denotes the Cartesian product of Z^i over $i \in S$, i.e., $Z^S = \prod_{i \in S} Z^i$. To represent a normal form game by $\mathcal{G}$, let $Z = Z^N$. If coalitions cannot form, then for every $i \in N$, and $a, b \in Z$, $a \xrightarrow{\{i\}} b$ if and only if $a^{-i} = b^{-i}$, and $a \prec_i b$ if and only if $u^i(a) < u^i(b)$. If coalitions can form, for all $a, b \in Z$, $a \xrightarrow{S} b$ if and only if $a^{-S} = b^{-S}$.

[2] A coalition is a nonempty subset of N.

Cooperative Games. A cooperative game is a pair (N, ν), where N is the nonempty set of players and ν is the characteristic function which assigns to every coalition $S \subset N$ a nonempty subset of $\mathcal{R}^S$ denoted $\nu(S)$. To represent this game by $\mathcal{G}$, 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 $b^S \in \nu(S)$.

Obviously, the von Neumann and Morgenstern (vN-M) solution (for cooperative games) can be generalized to more complex social environment as studied in this chapter. That is, one can apply the notion of vN-M abstract stable set to the study of social environment $\mathcal{G}$. For this purpose the following definitions are introduced.

Definition 2.1. Let $\succ$ be a dominance relation defined on Z. Then pair $(Z, \succ)$ is called an abstract system. The set, $V \subset Z$, is

(1) a *vN-M internally stable set* if V is free of inner contradictions, i.e., there do not exist $x, y \in V$ such that $y \succ x$,

(2) a *vN-M externally stable set* if V accounts for every alternative it excludes, i.e., if $x \notin V$, it must be the case that there exists $y \in V$ such that $y \succ x$, and

(3) a *vN-M (abstract) stable set* if it is both vN-M internally and externally stable.

Let V be a (abstract) stable set for $(Z, \succ)$. If $x \in Z$ is the status quo, the set of "predicted outcomes" is given by $\{y \in V \mid y = x \text{ or } y \succ x\}$. 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 $y \succ x$ 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 2.2. For $a, b \in Z$, b is said to dominate a, or $b > a$, if

(1) there exists a coalition $S \subset N$ that can replace a by b, i.e., $a \xrightarrow{S} b$, and

(2) 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 (1974) criticizes the vN-M solution for its failing to incorporate foresight. Such a criticism can also apply to an abstract stable set V for $(Z, >)$, which can be illustrated through the following (extremely) simple example[3], where $N = \{1, 2\}$, $Z = \{a, b, c\}$, player 1 can replace a by b, and player 2 can replace b by c. 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 2.1

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. But if he is farsighted, he should and will replace a by b, knowing that player 2 (who is rational) will subsequently replace b by c. That is, farsighted players do not just look one step ahead. For this reason, Harsanyi suggests to replace the dominance relation in the definition of V by some "indirect dominance" (as opposed to the "direct dominance" relation defined in Definition 2.2), which captures the fact that farsighted individuals consider the *final* outcomes that their actions may lead to. 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.[4] Formally,

[3] Appendix 2.2 offers a simple cooperative game for which the vN-M solution is subject to Harsanyi's criticism. I thank Professor Ron Holzman for pointing out this example.

[4] In the main text of his paper, Harsanyi (1974) considers an indirect dominance entailing that individuals consider also the intermediate outcomes. The indirect dominance in Definition 2.3 was mentioned by Harsanyi informally and formalized by Chwe (1994).

Definition 2.3. 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 $j = 0, 1, \ldots, m-1$, $a_j \xrightarrow{S_j} a_{j+1}$ 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 $(Z, \gg)$. Consider, again, Figure 2.1. 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 (1994), 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. I shall illustrate this issue by the example in Figure 2.2. But, first, I shall introduce the formal definition of the largest consistent set.

Definition 2.4 (Chwe). Consider a social environment $\mathcal{G}$. A subset $Y \subset Z$ is *consistent* if $a \in Y \iff$ for every d such that $a \xrightarrow{S} d$, there exists $e \in Y$, $d = e$ or $d \ll e$, such that $a \not\to_S e$. The *largest consistent set* (LCS) is the unique maximal consistent set with respect to set inclusion.

According to Chwe (1994), the set of "predicted outcomes", when $x \in Z$ is the status quo, is given by $\{y \in \text{LCS} \mid y = x \text{ or } y \gg x\}$. To illustrate that the stable set for $(Z, \gg)$ can be too exclusive while the LCS can be too inclusive, consider the example depicted in Figure 2, where $N = \{1, 2\}$ and $Z = \{a, b, c, d\}$. 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

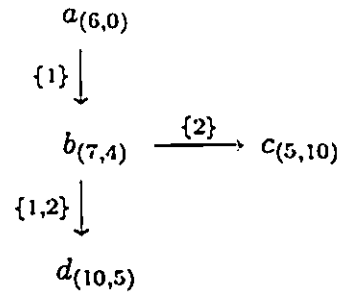


FIGURE 2.2

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 if 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 is d, resulting in the same difficulty as was discussed above. In Section 2.4, I shall show that the inclusiveness of the LCS is by no means accidental.

The above analysis illustrates the following aspects of perfect foresight.

(1) 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.

(2) Even though, as stated in (1), it is only the final outcomes that matter, a

player of 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.

(3) 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).

Therefore, to model perfect foresight, one need 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 2.2 has been represented *purposely* in a graph form to stress this point. The "graph" structure of the social environment 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.

2.3 Stability under Perfect Foresight

2.3.1 Formalization of Perfect Foresight.

In this section, I shall formalize perfect foresight by considering the "graph" of $\mathcal{G}$. To this end, I introduce the following definition.

Definition 2.5. A *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 analysis that proceeds. If $a \in Z$ is a vertex that lies on the path α, I shall write $a \in \alpha$. For a path α, let $\alpha|_b$, where $b \in \alpha$, denote its continuation from 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). The objective of this section is to determine which paths in Π_a might be followed by rational and farsighted individuals. Note that in general $\mathcal{G}$ does not represent 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 (e.g., when $\mathcal{G}$ represents a cooperative game or a normal form game; see Figures 2.5 and 2.6). Given the complex nature of $\mathcal{G}$, I shall employ the more general framework of "the theory of social situations" (Greenberg 1990). The theory of social situations unifies the representation of cooperative and noncooperative social environments; moreover, it insists upon the explicit specification of the negotiation/reasoning process (by the means of a "situation") and extends the notion of stability developed by von Neumann and Morgenstern (1947).

I shall retain the assumptions of Chwe (1994) that actions are public, binding agreements are not permissible, and payoffs are derived at a status quo only if no coalition wishes to replace it. In the spirit of the theory of social situations, perfect foresight is formalized *explicitly* by the following "situation[5]", which I

[5]A "situation" specifies how individuals view and use their alternatives; in particular, a situation specifies the "feasible outcomes" at every state or status quo and the opportunities available to the individuals (i.e., what individuals can do at every status quo and what the

shall henceforth refer to as "the situation with perfect foresight": Assume that alternative $a \in Z$ is the status quo. 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. In contemplating such a deviation from α, however, members of S do not base their decision on comparing α with Π_c. Rather, they consider the paths that might be followed by rational and farsighted individuals at c. Let $\sigma(\Pi_c) \subset \Pi_c$ denote this set of paths. In determining whether some path $\beta \in \Pi_c$ belongs to $\sigma(\Pi_c)$, each deviating coalition applies the same reasoning. Thus, the following definition is needed.

Definition 2.6. A standard of behavior (SB) σ for the situation with perfect foresight is a mapping that assigns to every $a \in Z$ a subset of Π_a, called the *solution* at a.

Obviously, in order for $\sigma(\Pi_a)$ to contain the set of paths (originating from a) that will be followed by rational and farsighted players, σ cannot be an arbitrary mapping. Following Greenberg (1990), we shall require that σ be stable. That is, σ must be free of inner contradictions and at the same time accounts for every path it excludes.

Definition 2.7. An SB σ for the situation with perfect foresight is

(1) *internally stable* if for all $a \in Z$, $\alpha \in \sigma(\Pi_a)$ implies that there do not exist $b \in \alpha$, a coalition $S \subset N$, and $c \in Z$ such that $b \xrightarrow{S} c$ and S "prefers" $\sigma(\Pi_c)$ to α,

(2) *externally stable* if for all $a \in Z$, $\alpha \in \Pi_a \setminus \sigma(\Pi_a)$ implies that there exist $b \in \alpha$, a coalition $S \subset N$, and $c \in Z$ such that $b \xrightarrow{S} c$ and S "prefers" $\sigma(\Pi_c)$ to α, and

(3) *stable* if it is both internally and externally stable.

---

consequences of their actions are).

That is, an SB σ is stable if for every $a \in Z$, $\sigma(\Pi_a)$ contains *those and only those* paths that are not rejected by any coalition, whose members are aware of and believe in the specification of the SB σ.

Note that the notion of stability requires that a *single* path α be compared with *a set of paths* $\sigma(\Pi_c)$. The way such a comparison is made depends on the players' attitude towards (Knightian) uncertainty. Following most of the application of the theory of social situations, I shall concentrate on the following two extreme behavioral assumptions.

(1) Optimism – players always hope for the best, i.e., S "prefers" $\sigma(\Pi_c)$ to α if for *some* $\beta \in \sigma(\Pi_c)$, $\alpha \prec_S \beta$, and

(2) Conservatism – players always fear the worst, i.e., S "prefers" $\sigma(\Pi_c)$ to α if for *all* $\beta \in \sigma(\Pi_c)$, $\alpha \prec_S \beta$.

If an SB σ is stable under optimism, it is called an "optimistic stable standard of behavior" (OSSB), and if σ is stable under conservatism, it is called a "conservative stable standard of behavior" (CSSB). Formally,

Definition 2.8. Let σ be an SB for the situation with perfect foresight. Then,

(1) σ is an OSSB if for all $a \in Z$, $\alpha \in \Pi_a \setminus \sigma(\Pi_a) \iff$ there exist $S \subset N$, $b \in \alpha$, and $c \in Z$ such that $b \xrightarrow{S} c$ and $\alpha \prec_S \beta$ for *some* $\beta \in \sigma(\Pi_c)$.

(2) σ is a CSSB if for all $a \in Z$, $\alpha \in \Pi_a \setminus \sigma(\Pi_a) \iff$ there exist $S \subset N$, $b \in \alpha$, and $c \in Z$ such that $b \xrightarrow{S} c$, $\sigma(\Pi_c) \neq \emptyset$, and $\alpha \prec_S \beta$ for *all* $\beta \in \sigma(\Pi_c)$.

2.3.2 The Significance of a Stable SB.

It is easy to verify that for the social environment depicted in Figure 2.2, the situation with perfect foresight admits a unique OSSB which coincides with the unique CSSB. Denoting this SB by σ, we have that $\sigma(\Pi_b) = \{(b, c)\}$[6] and $\sigma(\Pi_a) = \{a\}$. Hence, coalition $\{1, 2\}$ will never form. Moreover, if a is the status

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

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.

(Q1) 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.

(Q2) How stable outcomes are reached from "non-stable" outcomes.

(Q3) Which coalitions might form in the process of replacing a non-stable outcome with a stable one.

Before I answer these questions, I 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 "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 2.9. *Assume that σ is a stable SB and that $\alpha \in \sigma(\Pi_a)$. Then, for all $b \in \alpha$, $\alpha|_b \in \sigma(\Pi_b)$.*

Lemma 2.10. *If σ is a stable SB, then there exists at least one $a \in Z$ such that $a \in \sigma(\Pi_a)$.*

The set of *stable outcomes* is, therefore, given by $E = \{a \in Z \mid a \in \sigma(\Pi_a)\}$. Each alternative $a \in E$ is stable in the sense that it will prevail if it is the status quo. Put differently, no coalition with the power to replace a by another alternative would (eventually) benefit by doing so. Moreover, every outcome that belongs to $Z \setminus E$ 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.

As we examine paths rather than the elements in Z, we can predict not only the stable outcomes but also the coalitions that might form when an unstable outcome is the status quo. More specifically, if $a \in Z \setminus A$ is under consideration and if $\sigma(\Pi_a) \neq \emptyset$, the predicted outcomes are the terminal nodes of those stable paths that belong to $\sigma(\Pi_a)$, and the coalitions that might form are those that implement the paths in $\sigma(\Pi_a)$. Figure 2.3 serves as another example to illustrate this point.

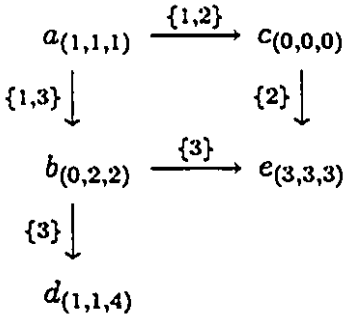


FIGURE 2.3

The situation with perfect foresight for the social environment depicted in Figure 2.3 admits a unique OSSB which coincides with the unique CSSB. When a is the status quo, the unique stable path is (a, c, e), implying that coalition $\{1, 2\}$ will form. Moreover, coalition $\{1, 3\}$ will not form: Player 1, being farsighted, realizes that, in the absence of binding agreements, were he to join player 3 and replace a by b, player 3 would then replace b by d.

A stable SB σ fully answers (Q1)-(Q3) whenever the status quo is some $a \in Z$ for which $\sigma(\Pi_a) \neq \emptyset$. If the status quo $a \in Z$ is such that $\sigma(\Pi_a) = \emptyset$ then the SB σ tells us that a cannot remain as a status quo (since $a \notin \sigma(\Pi_a)$), but σ is silent about which paths are likely to be followed, and which outcome in Z might result. It is, therefore, important to investigate those situations whose stable SBs are *nonempty-valued*, i.e., $\sigma(\Pi_a) \neq \emptyset$ for every[7] $a \in Z$. A nonempty-valued stable

[7]Observe that if σ is (externally) stable, then it must be the case that there exists at least one $a \in Z$ for which $\sigma(\Pi_a) \neq \emptyset$. If σ is nonempty-valued, this condition holds for *every* $a \in Z$.

SB provides complete answers to (Q1)-(Q3), since it has the property that no matter what the status quo is, there will always exist paths which farsighted and rational players would agree to follow. Furthermore, due to the interdependence among the solutions at different status quo's in Z, perfect foresight may not emerge in a stable SB if it is not nonempty-valued. I shall return to this issue in Section 2.5 and Appendix 2.4.

Proposition 2.12 below provides a sufficient condition for a stable SB to be nonempty-valued. To this end, we first need to define some dominance relations on Π (the set of all paths).

Definition 2.11. For $\alpha, \beta \in \Pi$, $a \in Z$, and $S \subset N$, we write $\alpha \angle_S^a \beta$ if $a \in \alpha$ and there exists $b \in \beta$ such that $a \xrightarrow{S} b$ and $\alpha \prec_S \beta$. We also write $\alpha \angle^a \beta$ if $\alpha \angle_S^a \beta$ for some $S \subset N$, $\alpha \angle_S \beta$ if $\alpha \angle_S^a \beta$ for some $a \in Z$, and $\alpha \angle \beta$ if $\alpha \angle_S^a \beta$ for some $a \in Z$ and some $S \subset N$.

That is, a path β dominates path α if α possesses a vertex a that some coalition, S, can replace with vertex b that lies on the path β, and every member of S prefers the terminal node of β over the terminal node of α.

Proposition 2.12. *Assume that Π does not admit an infinite sequence of (not necessarily distinct) paths $\alpha_1, \alpha_2, \ldots$ such that $\alpha_i \angle^a \alpha_{i+1}$ for some $a \in Z$. Then, if σ is either an OSSB or a CSSB, σ is nonempty-valued.*

The following example illustrates that the condition in the above proposition is not unnecessarily strong, even when the set of alternatives, Z, is a finite set.

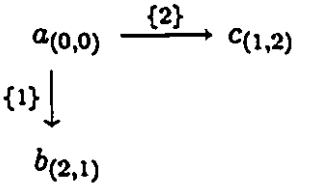


FIGURE 2.4

The social environment depicted in Figure 2.4 violates the condition in Propo-

sition 2.12 since $(a,b)\ \angle^a\ (a,c)\ \angle^a\ (a,b)$. Moreover, there exists a unique OSSB, σ, which coincides with the unique CSSB which satisfies $\sigma(\Pi_a) = \emptyset$. That is, a is rejected by both players, but players 1 and 2 cannot agree on which alternative, b or c, should replace a.

2.3.3 Existence of a Stable SB.

I now turn to the existence of (nonempty-valued) OSSB and CSSB. Consider first, the OSSB. There are several conditions that guarantee the existence of OSSB for the situation with perfect foresight. One such condition is the *strict acyclicity* of $\angle$. The dominance relation $\angle$ is said to be *strictly acyclic* if there do not exist an infinite sequence of (not necessarily distinct) paths $\alpha_1, \alpha_2, \ldots$ in Π such that $\alpha_i \angle \alpha_{i+1}$ for all $i = 1, 2, \ldots$. Using Proposition 2.12, we have

Proposition 2.13. *Assume that $\angle$ is strictly acyclic. Then there exists a unique nonempty-valued OSSB.*

It turns out that strict acyclicity of $\angle$ is also sufficient for the existence of a nonempty-valued CSSB (which need not be unique). This result follows from Proposition 2.13 together with the following proposition.

Proposition 2.14. *If there exists a nonempty-valued OSSB, then there exists a nonempty-valued CSSB and a largest*[8] *nonempty-valued CSSB σ^ℓ. Moreover, for every nonempty-valued conservative or optimistic stable SB σ, we have $\sigma(\Pi_a) \subset \sigma^\ell(\Pi_a)$ for every $a \in Z$.*

The existence of CSSB require less demanding conditions than that of OSSB. The example of "Condorcet paradox" in Figure 2.5 illustrates that CSSB exists when OSSB fails to exist. Indeed, there exists a nonempty-valued CSSB σ such that $\sigma(\Pi_x) = \Pi_x$ for all $x \in \{a,b,c\}$. Therefore, the CSSB predicts that each $x \in \{a,b,c\}$ might arise.

[8] For two SB's, σ and σ', $\sigma \geq \sigma'$ if $\sigma(\Pi_a) \supset \sigma'(\Pi_a)$ for all $a \in Z$.

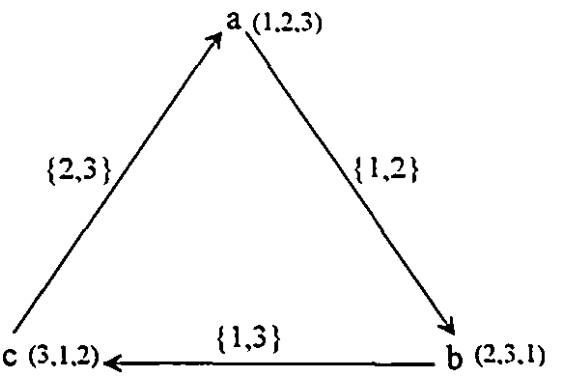


FIGURE 2.5

A weak sufficient condition for the existence of a nonempty-valued CSSB is given in Theorem 3.11. For any two paths $\alpha, \beta \in \Pi$, I shall write $\alpha \subset \beta$, if there exists $b \in \beta$ such that $\beta|_b = \alpha$.

Theorem 2.15. *Assume that there does not exist an infinite sequence of paths* $\alpha_1, \alpha_2, \ldots$ *in* Π *such that for all* $i, j = 1, 2, \ldots, \alpha_i \angle \alpha_{i+1}$ *and* $i < j$ *implies* $\alpha_i \not\subset \alpha_j$. *Then the situation with perfect foresight admits a nonempty valued CSSB.*

Note that the condition in Theorem 2.15 does not hold for the social environment depicted in Figure 2.4. It holds, however, for the social environment depicted in Figure 2.5. The sketch of the proof is as follows: First, note that an empty-valued SB is conservative internally stable while an (nonempty-valued) SB σ such that $\sigma(\Pi_x) = \Pi_x$ for all $x \in Z$ is conservative externally stable. The central idea is to show, by Zorn's Lemma, that under the given condition, there exists a minimum nonempty-valued conservative externally stable SB, which is also conservative internally stable.

2.4 Re-examination of the Related Literature

In this section, I shall compare, both formally and through examples, the stable SBs for the situation with perfect foresight with the notions discussed in Section 2.2. Such comparison is made by examining the negotiation/reasoning

process (which is formalized by means of a situation) underlying these notions. As shown by Chwe (1994), the notion of consistent set can be cast within the framework of the theory of social situations, thereby revealing how individuals view and use their alternatives. In particular, the *situation* that describes the negotiation/reasoning process underlying Chwe's consistent set is as follows: For every $a \in Z$, the set of feasible outcomes when a is the status quo is

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

For $b \in X_a$, if there is a coalition $S \subset N$ such that $b \xrightarrow{S} c$ for some $c \in Z$, then the set of feasible outcomes at c is given by X_c.

For this (Chwe) situation, again we can apply the notion of stability. Some $b \in X_a$ is likely to arise or stable if no coalition wishes to replace b by some $c \in Z$, by considering the set of likely (stable) outcomes at c (which is a subset of X_c). Let ψ be an SB that assigns to every $a \in Z$ a subset of X_a. The following definition is parallel to Definition 2.8.

Definition 2.16. Let ψ be an SB for the "Chwe situation". Then,

(1) ψ is an OSSB if for all $a \in Z$, $b \in X_a \setminus \psi(X_a) \iff$ there exist $S \subset N$ and $c \in Z$ such that $b \xrightarrow{S} c$ and $b \prec_S d$ for *some* $d \in \psi(X_c)$.

(2) ψ is a CSSB if for all $a \in Z$, $b \in X_a \setminus \psi(X_a) \iff$ there exist $S \subset N$ and $c \in Z$ such that $b \xrightarrow{S} c$, $\psi(X_c) \neq \emptyset$, and $b \prec_S d$ for *all* $d \in \psi(X_c)$.

Chwe (1994) shows that the CSSB for the "Chwe situation" is formally related to his consistent set. Proposition 2.17 states this formal relationship. This proposition is slightly stronger than that of Chwe.

Proposition 2.17. *For $Y \subset Z$, define an SB ψ by $\psi(X_a) = X_a \cap Y$ for all $a \in Z$. Then, ψ is a CSSB for the "Chwe situation" 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.*

Therefore, the following assumptions are embedded in the definition of the LCS. The first two assumptions signify the difference between the LCS and the notion proposed in this chapter.

(1) For every $a \in Z$ the set of feasible outcomes is given by $X_a = \{a\} \cup \{b \in Z \mid b \gg a\} \subset Z$. For example, for the social environment depicted by Figure 2.2, $X_a = \{a, b, d\}$ and $X_b = \{b, c, d\}$. This is in sharp contrast to the situation with perfect foresight where the set of feasible outcomes at a or b is the set of paths originating from a or b.

(2) $b \in Z$ is likely to arise or stable if $b \in X_a$ (i.e., feasible at a) and no coalition wishes to deviate from b. Therefore, deviations in the process of reaching b from a are ignored. Indeed, for the social environment in Figure 2.2, since d belongs to both X_a and X_b and d is the "end of the play", d is included in both $\psi^c(X_a)$ (the solution when a is the status quo) and $\psi^c(X_b)$ (the solution when b is the status quo), where ψ^c is the unique CSSB for the "Chwe situation". The situation with perfect foresight employs paths as the building blocks and all deviations along every path are considered. In particular, the path (a, b, d) does not belong to $\sigma(\Pi_a)$, where σ is the unique CSSB (also the unique OSSB) for the situation with perfect foresight, since once b is reached, player 2 will deviate and implement c.

(3) Individuals are conservative: A deviation occurs only if all resulting outcomes benefit the deviating coalition.

A very interesting result is that the OSSB for the "Chwe situation" is formally related to the stable set for $(Z, \gg)$. Such a result is derived by a theorem due to Shitovitz [Theorem 4.5 (Greenberg 1990)].

Claim 2.18. *σ is an OSSB for the "Chwe situation" if and only if $Y = \bigcup_{a \in Z} \sigma(X_a)$ is a vN-M abstract stable set for $(Z, \ll)$.*

For this reason, I shall refer to the "Chwe situation" as Harsanyi-Chwe sit-

uation. Proposition 2.17 and Claim 2.18 imply that the negotiation/reasoning process underlying Chwe's LCS is exactly the same as the one underlying the abstract stable set for $(Z, \gg)$, and the difference between these notions lies in the different behavioral assumptions embedded in them. Therefore, the exclusiveness of the stable set for $(Z, \ll)$ and the inclusiveness of the LCS are not isolated phenomena. For Figure 2.2, the unique OSSB ψ^o and the unique CSSB ψ^c for the Harsanyi-Chwe situation are such that $\psi^o(X_b) = \psi^c(X_b) = \{c, d\}$; hence either c or d might arise were b the status quo. Thus, if player 1 is optimistic, he will reject a, hoping that d might arise; if player 1 is conservative, he will not rule a out, fearing that c might arise. (In contrast, the unique stable SB for the situation with perfect foresight entails that were player 1 to replace a by b he would *necessarily* end up with c.) Furthermore, d is included in both $\psi^o(X_a)$ and $\psi^c(X_a)$, implying that d might arise if a is the status quo. Therefore, I shall argue that both the vN-M stable set for $(Z, \gg)$ and the LCS do not capture perfect foresight for the reason that they ignore the deviations on the way of replacing an alternative by another one. Now I am going to provide more examples to illustrate the importance of paths in the formalization of perfect foresight.

First consider the following "coordination" game played in such a fashion that there are moves and counter moves and assume that coalition $\{1, 2\}$ cannot form.

	ℓ	r
u	1,1	0,0
d	0,0	2,2

FIGURE 2.6

The LCS contains both (u, ℓ) and (d, r). (u, ℓ) is included, since, for example, player 1's deviation from u to d is deterred by his own further deviation back to u. This is the consequence of ignoring the graph structure of the social environment. The situation with perfect foresight consider the graph of the game. In contemplating a deviation from (u, ℓ) to (d, ℓ), player 1 realizes that the only sta-

ble path from (d, ℓ) (prescribed by OSSB and CSSB for the situation with perfect foresight) leads to (d, r). Therefore a player with prefect foresight will deviate from (u, ℓ).

The social environment depicted in Figure 2.3 is another example illustrating the lack of foresight or rationality in the abstract stable set for $(Z, \gg)$ and the LCS. Both the LCS and the unique stable set for $(Z, \gg)$ predict that coalition $\{1, 3\}$ might form, since e is included in both notions and c indirectly dominates a via both b and c. As discussed in Section 2.3, however, the unique stable path for the situation with perfect foresight is (a, c, e); hence only coalition $\{1, 2\}$ will form.

Theorem 2.19 provides a formal result on the relationship between the CSSBs (hence the stable outcomes), in particular, the largest CSSB, for the situation with perfect foresight and the largest CSSB (hence the LCS) for the Harsanyi-Chwe situation.

Theorem 2.19. *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 $a \in Z$, $\alpha \in \sigma(\Pi_a)$, where $t(\alpha) \neq a$, implies $t(\alpha) \gg a$. Then for all $a \in Z$, $\alpha \in \sigma(\Pi_a)$ implies $t(\alpha) \in \psi^\ell(X_a)$, i.e., σ "refines" ψ^ℓ.*

Recall that $\psi^\ell(X_a) \subset X_a = \{a\} \cup \{b \in Z \mid b \gg a\}$. Therefore, the condition that for every $a \in Z$, $\alpha \in \sigma(\Pi_a)$, where $t(\alpha) \neq a$, implies $t(\alpha) \gg a$ enables the CSSB for the situation with perfect foresight to be formally compared with the CSSB for the Harsanyi-Chwe situation. This condition holds, for example, for the social environments depicted in Figures 2.2 and 2.3. Furthermore, the implication of Theorem 2.19 is most compelling when $\mathcal{G}$ represents an extensive form game with perfect information.

Extensive Form Games. An extensive form game with perfect information can be represented by $\mathcal{G}$ in the following way: Let Z be the set of nodes and partition

Z into $Z_0, Z_1, Z_2, \ldots, Z_n$, where Z_i, $i \in N$, is the set of nodes that belong to player i and Z_0 is the set of terminal nodes. Then an extensive form game can be represented by $\mathcal{G}$: For every $a \in Z \setminus Z_0$ and every $i \in N$, let $a \prec_i b$ for all $b \in Z_0$, and $a \xrightarrow{\{i\}} b$ if $a \in Z_i$ and b is adjacent from a.

Consider an extensive form game such that the graph $\phi(\mathcal{G})$ (or the game tree in this case) is bounded. The LCS is nonempty and coincides with the unique stable set for $(Z, \gg)$. At the "root" of the game tree, both notions predict that the set of outcomes that are likely to prevail is Z_0, the set of *all* terminal nodes. However, if the (unique) CSSB for the situation with perfect foresight is nonempty valued, then it is formally related to the notion of subgame perfection; if, in addition, the situation with perfect foresight admits a (unique) nonempty valued OSSB, then it refines the CSSB. Formally,

Claim 2.20. *Let $\mathcal{G}$ represent an extensive form game with perfect information and $\phi(\mathcal{G})$ be bounded. Assume that the unique CSSB σ^c for the situation with perfect foresight be nonempty valued. Then for every $a \in Z$, $\sigma^c(\Pi_a)$ coincides with the set of subgame perfect equilibrium paths for the subgame originating from a. If, in addition, there exists a (unique) nonempty valued OSSB σ^o, then $\sigma^o(\Pi_a) \subset \sigma^c(\Pi_a)$ for every $a \in Z$.*

In his concluding remarks, Chwe (2.20) recognizes several issues that the notion of the LCS fails to address, yet no constructive solution was offered. The notion suggested in this chapter resolves most if not all of these issues. Now, I shall use the following example to illustrate that the issue of "preemption" that is not well addressed by the LCS (and the stable set for $(Z, \gg)$) can be analyzed by the stable SB for the situation with perfect foresight.

When a is the status quo, both stable set for $(Z, \gg)$ and the LCS predict that b be the unique outcome. If players are rational and farsighted, however, player 1 will "preempt" player 2's move, and player 2 will "wait" and let player 1 move. Indeed, the unique OSSB (which is also the unique CSSB) for the situation with

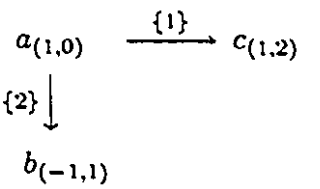


FIGURE 2.7

perfect foresight, where paths are the building blocks, predicts that when a is the status quo, the unique stable path is (a, c) (which Pareto dominates b).

I shall conclude this section by formally relating the OSSB for the situation with perfect foresight to the vN-M stable set for a abstract system. Such a result is, again, the special case of Shitovitz's [Theorem 4.5 (Greenberg 1990)] result on the formal relationship between OSSB for any situation and the stable set for the corresponding abstract system.

Claim 2.21. *σ is an OSSB for the situation with perfect foresight if and only if the set $Y = \bigcup_{a \in Z} \sigma(\Pi_a)$ is a vN-M stable set for $(\Pi, \angle)$.*

Thus, Harsanyi's criticism should not be viewed as a criticism to the notion of vN-M stability in general; rather, it is a criticism that can apply only to certain abstract systems such as $(Z, >)$ given in Section 2.2. This point has not been clarified in the literature.

2.5 Concluding Remarks

The analysis of coalitional stability and foresight has demonstrated that to model perfect foresight one need to consider the graph of the social environment even if such a social environment does not represent an extensive form game. Also, the analysis clarifies that it is not the notion of stability that is farsighted or myopic; it is the abstract system or the negotiation/reasoning process (which can be formalized by a situation). The notion of stability in the theory of social situations (Greenberg 1990) resembles that of von Neumann and Morgenstern (1947). One of the advantages of the framework of the theory of social situations, however, lies in the explicit specification of the negotiation/reasoning process as well as the

individuals' attitude towards strategic (Knightian) uncertainty, which enables us to examine the assumptions that might be otherwise implicit (or hidden) in existing notions. For example, as was shown in the previous section, the definitions of the stable set with indirect dominance and the LCS embed several assumptions that can be revealed by analyzing the corresponding (Harsanyi-Chwe) situation. This is exactly where our formalization of perfect foresight was motivated and initiated. Moreover, the theory of social situations allows for difference behavioral assumptions, while the vN-M stable set implicitly assumes optimistic behavior (as the vN-M stable set is formally related to OSSB).

One of the implications of this chapter is that the representation and analysis of a cooperative and a noncooperative environment can be bridged, and the notion of stability can be applied regardless of the social environment's cooperative or noncooperative nature. In particular, the notion of stability is not necessarily linked to cooperative games, especially in view of the fact that many noncooperative solution concepts can be derived by using the notion of stability [see Greenberg (1990)] and the social environment studied in this chapter is by no means a pure cooperative one. In a pure non-cooperative dynamic environment, the concept of subgame perfection (and its variants) captures perfect foresight. In view of Claim 2.20, the theory of situation and the notion of stability enable the extension of the concept of subgame perfect to social environments of more complex structure (although this is not the motivation of this chapter).

The idea that farsighted individuals look arbitrary steps ahead is analogous to the consideration of consistency in Dutta et al 's (1989) definition of "consistent bargaining set" for cooperative games. Recall that the core rules out a payoff vector if there is an objection to this payoff vector; hence the core does not assess the "credibility" of an objection. The bargaining set [Aumann and Maschler (1964) and Mas-Colell (1989)] goes one step further by considering only "justified objections", i.e., those objections that do not have counter-objections. The

"credibility" of counter-objections is, however, left unattended. The consistent bargaining set of Dutta et al. (1989) entails that every objection in a "chain" of objections is tested in precisely the same way as its predecessor. Note that the formalization as well as the intuition of the notion of consistent bargaining set are different from those of perfect foresight. In the definition of consistent bargaining set, it is the "credibility" of an objection that matters. In our formalization of foresight, however, it is the final ("credible") outcomes resulting from an objection (which itself does not have to be "credible") that matter. That is, if players are farsighted, a coalition may object to a payoff vector as long as such an objection (which itself may not be "credible") will ultimately lead to ("credible") outcomes that benefit its members.

Now, I shall point out several questions for future research. First, in our analysis, individuals are assumed to be patient. This implies that for example, in Figure 2.5, if individuals are optimistic, OSSB does not exist since each coalition always hopes that its favorite alternative might arise and thereby always rejects the status quo. Introducing discounting into this model may help resolve this issue. Moreover, with discounting, we may be able to evaluate paths of infinite length. Secondly, the stable SB for the situation with perfect foresight may fail to be nonempty-valued. For the social environment in Figure 2.4, when a is the status quo, the stable SB is silent on which path will arise. That is, the solution for a is empty. Now, suppose there is another alternative x adjacent to a and x gives a payoff of -1 to each player.

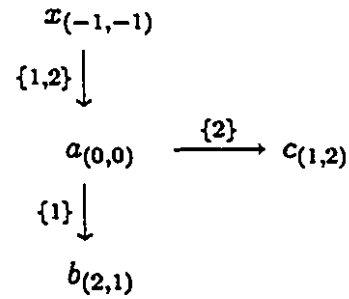


FIGURE 2.8

In this case, the unique CSSB (OSSB) predicts that x will prevail if it is the status quo, thereby failing to capture the perfect foresight of the players. To resolve this issue, we need to "break down" the emptiness of the solution for a. Section 3.3 in Chapter 3 offers, at least implicitly, a way of doing so. Thirdly, this chapter assumes a simple information structure. It might be interesting to investigate the consequences of incomplete information and imperfect information.

Appendix 2.1: Proofs

Proof of Lemma 2.9. Let σ be an OSSB. Assume in negation that $\exists a \in Z$ and $\alpha \in \Pi_a$ such that $\alpha \in \sigma(\Pi_a)$ but $\alpha|_b \notin \sigma(\Pi_b)$ for some $b \in \alpha$. By external stability, $\exists c \in \alpha|_b$ (hence $c \in \alpha$) , $d \in Z$, and $S \subset N$ such that $\alpha|_b \prec_S \beta$ (hence $\alpha \prec_S \beta$) for some $\beta \in \sigma(\Pi_d)$. This contradicts the internal stability of σ.

Similarly, we can show that Lemma 2.9 holds if σ is a CSSB. ∎

Proof of Lemma 2.10. By external stability, σ cannot be identically empty. Therefore, $\sigma(\Pi_a) \neq \emptyset$ for some $a \in Z$. By Lemma 2.9, every terminal node of a path in $\sigma(\Pi_a)$ satisfies Lemma 3.6. ∎

Proof of Proposition 2.12. Let σ be an OSSB or CSSB. Assume in negation that there exists $a \in Z$ such that $\sigma(\Pi_a) = \emptyset$. Then, $a \notin \sigma(\Pi_a)$. By external stability, $\exists S \subset N$ and $b \in Z$ such that $a \xrightarrow{S} b, \sigma(\Pi_b) \neq \emptyset$ and $a \prec_S \beta$ for every $\beta \in \sigma(\Pi_b)$. Let $W_a = \{b \in Z \mid \exists S$ such that $a \xrightarrow{S} b$ and $\sigma(\Pi_b) \neq \emptyset\}$; then $W_a \neq \emptyset$. Let $\Phi \subset \Pi_a$ such that for every $b \in W_a$, there exists $\alpha \in \Phi$ such that $a|_b \in \sigma(\Pi_b)$. By assumption, Φ admits a maximal element with respect to $\angle^a$. Let ζ be such a maximal element. Then by external stability, $\zeta \in \sigma(\Pi_a)$, contradicting that $\sigma(\Pi_a) = \emptyset$. ∎

Proof of Proposition 2.13. The sufficiency of strict acyclicity is due to a theorem due to von Neumann and Morgenstern (1947) on the existence of abstract stable sets, and a theorem due to Shitovitz [Theorem 4.5 (Greenberg 1990)] that establishes a formal relationship between the graph of an OSSB and a von Neumann

and Morgenstern abstract stable set. Moreover, it follows from Proposition 2.12 that this OSSB is also nonempty-valued. ∎

Proof of Proposition 2.14. Obviously, a nonempty-valued OSSB is also a nonempty-valued conservative internally stable SB. By a theorem of Greenberg, Monderer, and Shitovitz (1995), there exists a largest nonempty-valued conservative internally stable SB with respect to set inclusion that is also the largest nonempty-valued CSSB. ∎

Proof of Theorem 2.15.

Let Σ be the set of conservative externally stable SB's such that $\sigma \in \Sigma$ implies

(C1) σ satisfies "truncation property", i.e., for every $a \in Z, \alpha \in \sigma(\Pi_a)$ implies $\alpha|_b \in \sigma(\Pi_b)$ for all $b \in \alpha$, and

(C2) for every $a \in Z, \alpha \in \Pi_a \setminus \sigma(\Pi_a)$ and $\alpha|_b \in \sigma(\Pi_b)$, where b is adjacent from a, imply that there exists $S \subset N$ and $c \in Z$ such that $a \rightarrow_S c$ and $\alpha \prec_S \beta$ for all $\beta \in \sigma(\Pi_c)$.

Obviously, let σ^0 be such that $\sigma(\Pi_a) = \Pi_a$ for all $a \in Z$; then $\sigma^0 \in \Sigma$. For $a \in Z$ and $\sigma \in \Sigma$, define

$$\mathrm{CDOM}(\sigma, a) = \{\alpha \in \Pi \mid a \in \alpha \text{ and } \exists b \in Z \text{ and } S \subset N \text{ such that } a \rightarrow_S b,$$
$$\sigma(\Pi_b) \neq \emptyset, \text{ and } \alpha \prec_S \beta \text{ for all } \beta \in \sigma(\Pi_b)\}.$$

Claim A1: Let $\sigma \in \Sigma$ and $a \in Z$. Then $\alpha_1 \in \mathrm{CDOM}(\sigma, a)$ implies that there exist $J_1 < \infty$ and $\alpha_2, \ldots, \alpha_{J_1}$ in Π_a such that $\alpha_1 \angle \alpha_2 \angle \cdots \angle \alpha_{J_1}$, $i < j < J_1$ implies $\alpha_i \not\subset \alpha_j$ and $\alpha_{J_1} \in \sigma(\Pi_a)$. Moreover, let $b \in \alpha_{J_1}$ be adjacent from a; then $\zeta \in \Pi_a$ and $\zeta|_b \in \sigma(\Pi_b)$ imply $\zeta \notin \mathrm{CDOM}(\sigma, a)$ and therefore $\zeta \in \sigma(\Pi_a)$.

$\alpha_1 \in \mathrm{CDOM}(\sigma, a)$ implies that there exists $S_2 \subset N$ and $b_2 \in Z$ such that $a \rightarrow_{S_2} b_2$ and $\alpha_1 \prec_{S_2} \beta$ for all $\beta \in \sigma(\Pi_{b_2})$. If $\zeta \in \Pi_a$ and $\zeta|_{b_2} \in \sigma(\Pi_{b_2})$ imply that $\zeta \notin \mathrm{CDOM}(\sigma, a)$ and hence by (C2) $\zeta \in \sigma(\Pi_a)$, then we are done. Otherwise, let $\alpha_2 \in \Pi_a$ be such that $\alpha_2|_{b_2} \in \sigma(\Pi_{b_2})$ and $\alpha_2 \in \mathrm{CDOM}(\sigma, a)$. That is, there exist a coalition $S_3 \subset N$ and $b_3 \in Z$ such that $a \rightarrow_{S_3} b_3$ and $\alpha_2 \prec_{S_3} \beta$ for

all $\beta \in \sigma(\Pi_{b_3})$. If $\zeta \in \Pi_a$ and $\zeta|_{b_3} \in \sigma(\Pi_{b_3})$ imply $\zeta \notin \text{CDOM}(\sigma, a)$ and hence $\zeta \in \sigma(\Pi_a)$, then we are done. Otherwise, let $\alpha_3 \in \Pi_a$ be such that $\alpha_3|_{b_3} \in \sigma(\Pi_{b_3})$ and $\alpha_3 \in \text{CDOM}(\sigma, a)$. Continuing in this fashion, there exist a sequence of paths $\alpha_1, \alpha_2, \ldots$ such that for all $i = 1, 2, \ldots, \alpha_i \angle \alpha_{i+1}$. Moreover, by (C1), $i < j$ implies $\alpha_i \not\subset \alpha_j$. By assumption, such a sequence is finite.

Claim A2: Every $\sigma \in \Sigma$ is nonempty valued.

Assume otherwise there exists $a \in Z$ such that $\sigma(\Pi_a) = \emptyset$. Then $a \notin \sigma(\Pi_a)$. By external stability, $a \in \text{CDOM}(\sigma, a)$. Then Claim A2 follows from Claim A1.

Now, define a partial ordering $\geq$ on Σ such that for every $\sigma, \sigma' \in \Sigma$, $\sigma' \geq \sigma$ if and only if $\sigma'(\Pi_a) \subset \sigma(\Pi_a)$ for all $a \in Z$. Let Θ be a chain in Σ, i.e., every two elements in $\Theta \subset \Sigma$ are comparable.

Claim A3: Let $\sigma \in \Sigma$ and $a \in Z$. Then $\alpha_1 \in \text{CDOM}(\sigma, a)$ implies that there exist $J_1 < \infty$ and $\alpha_2, \ldots, \alpha_{J_1}$ in Π_a such that $\alpha_1 \angle \alpha_2 \angle \cdots \angle \alpha_{J_1}$ and $i < j < J_1$ implies $\alpha_i \not\subset \alpha_j$. Moreover, there exists $\sigma' \in \Theta$ such that $\sigma' \geq \sigma$ and if $b \in \alpha_{J_1}$ is adjacent from a, then, $\zeta \in \Pi_a$ and $\zeta|_b \in \sigma'(\Pi_b)$ imply $\zeta \notin \text{CDOM}(\sigma'', a)$ for all $\sigma'' \in \Theta$ and therefore $\zeta \in \sigma'(\Pi_a)$. (In particular, $\alpha_{J_1} \in \sigma'(\Pi_a)$.)

This follows from (repeatedly applying) Claim A1.

Claim A4: Σ has a maximal element with respect to $\geq$.

By Zorn's lemma, it suffices to show that every chain in Σ has an upper bound in Σ. Let Θ be a chain. Define η by $\eta(\Pi_a) \equiv \bigcap_{\sigma \in \Theta} \sigma(\Pi_a)$ for all $a \in Z$.

To show that η belongs to Σ, it suffices to show that η is nonempty valued. Assume otherwise, that there exists $a \in Z$ such that $\eta(\Pi_a) = \emptyset$. Hence, $a \notin \eta(\Pi_a)$, implies that there exists $\sigma^1 \in \Theta$ such that $a \notin \sigma^1(\Pi_a)$. By Claim A3, there exist $J_1 < \infty$ and $\alpha_1, \alpha_2, \ldots, \alpha_{J_1}$ in Π_a, where $\alpha_1 = a$, such that $\alpha_1 \angle \alpha_2 \angle \cdots \angle \alpha_{J_1}$ and $i < j < J_1$ implies $\alpha_i \not\subset \alpha_j$. Moreover, there exists $\sigma_2 \in \Theta$ such that if $b \in \alpha_{J_1}$ is adjacent from a then, $\zeta \in \Pi_a$ and $\zeta|_b \in \sigma_2(\Pi_b)$ imply $\zeta \notin \text{CDOM}(\sigma', a)$ for all $\sigma' \in \Theta$ and therefore $\zeta \in \sigma_2(\Pi_a)$. In particular, $\alpha_{J_1} \in \sigma_2(\Pi_a)$. If there do not exits $c \in \alpha_{J_1}$ and σ' such that $\alpha_{J_1} \in \text{CDOM}(\sigma', c)$, then $\alpha_{J_1} \in \sigma(\Pi_a)$ for all $\sigma \in \Theta$,

contradicting $\eta(\Pi_a) = \emptyset$. Let $c \in \alpha_{J_1}$ be the closest node to a such that for some $\sigma_3 \in \Theta$ such that $\sigma_3 \geq \sigma_2$, $\alpha_{J_1} \in \text{CDOM}(\sigma_3, c)$. Applying Claim A3, there exists $J_2 < \infty$ and $\alpha_{J_1}, \alpha_{J_1+1}, \ldots, \alpha_{J_1+J_2}$ in Π_a such that $\alpha_{J_1} \angle \alpha_{J_1+1} \angle \cdots \angle \alpha_{J_1+J_2}$ and $J_1 < i < j < J_1 + J_2$ implies $\alpha_i \not\subset \alpha_j$. Also, there exists $\sigma_4 \in \Theta$ such that if $d \in \alpha_{J_1+J_2}$ is adjacent from c then, $\zeta \in \Pi_a$ and $\zeta|_d \in \sigma_2(\Pi_b)$ imply $\zeta|_c \notin \text{CDOM}(\sigma_4, c)$ for all $\sigma' \in \Theta$ and therefore $\zeta \in \sigma_4(\Pi_c)$ and $\zeta \in \sigma_4(\Pi_a)$. In particular, $\alpha_{J_1+J_2} \in \sigma_4(\Pi_a)$. Moreover, since $\sigma_1 \leq \sigma_2 \leq \sigma_3 \leq \sigma_4$, $1 < i < j < J_1 + J_2$ implies $\alpha_i \not\subset \alpha_j$. Continuing in this fashion, there exists an infinite sequence $\alpha_1, \alpha_2, \ldots$ such that for all $i, j = 1, 2, \ldots, \alpha_i \angle \alpha_{i+1}$ and $i < j$ implies $\alpha_i \not\subset \alpha_j$. A contradiction.

Therefore, Σ admits a maximum element. Let η be such a maximum element. Then,

Claim A5: η is conservative internally stable.

Otherwise, there exists $a \in Z$ and $\alpha \in \eta(\Pi_a)$ and $\alpha \in \text{CDOM}(\eta, a)$. Define η' as

$$\eta'(\Pi_b) = \begin{cases} \eta(\Pi_b) \setminus \{(\theta, \alpha)\} & \text{if } \theta \in \Pi_b \text{ and } t(\theta) = a \\ \eta(\Pi_b) & \text{otherwise} \end{cases}$$

where (θ, α) denotes the path combining θ and α. By Claim A3, for all $b \in Z$ such that there exists $\beta \in \eta(\Pi_b)$ with $\beta|_a = \alpha$, there exists $\xi \in \eta(\Pi_b)$ such that $\xi \neq \beta$. Therefore η' is nonempty valued and conservative externally stable. Obviously, η' satisfies (C1) and (C2) and hence belongs to Σ. Since $\eta' > \eta$, contradicting that η is the maximum element in Σ.

η is a CSSB since it is both conservative internally and externally stable. ∎

Proof of Proposition 2.17.

Let $Y \subset Z$ be a consistent set. Then, $a \in Y \iff \forall d$ such that $a \xrightarrow{S} d, \exists c \in Y \in Y \cap X_d$ such that $a \not\succ_S e$. Since $\sigma(X_a) = X_a \cap Y$, $a \in Y \iff a \in \sigma(X_a) \iff \forall d$ such that $a \xrightarrow{S} d$, $\exists e \in \sigma(X_d)$ such that $a \not\succ_S e$. Obviously, if Y is consistent and σ is nonempty-valued, then σ is a conservative stable SB. To complete the proof, we need only to show that if σ is conservatively stable then

it is nonempty-valued. Indeed, otherwise, $\exists a \in Z$ such that $\sigma(X_a) = \emptyset$, implying $a \notin \sigma(X_a)$. By external stability, $\exists b \in Z, S \subset N$ such that $a \xrightarrow{S} b$, $\sigma(X_b) \neq \emptyset$, and $a \prec_S c \ \forall c \in \sigma(X_b)$. But $c \in \sigma(X_b)$ and $a \prec_S c$ imply $a \ll c$ and hence $c \in X_a$. Therefore $c \in \sigma(X_a)$. A contradiction.

The second part of the proposition follows from the theorem of Greenberg et. al. (1995) that if a situation admits a nonempty-valued CSSB, then it admits a largest nonempty-valued CSSB with respect to set inclusion. ∎

Proof of Claim 2.18. Again, this claim can be derived as a special case of Shitovitz's theorem [Theorem 4.5 (Greenberg 1990)]. ∎

Proof of Theorem 2.19. Let σ be a nonempty-valued CSSB for the situation with perfect foresight. In view of the theorem of Greenberg, Monderer, and Shitovitz (1995) that was used in the proof of Proposition 2.14, it suffices to show that $\hat{\sigma}$ defined by

$$\hat{\sigma}(X_a) = \cup\{t(\alpha) \mid \alpha \in \sigma(\Pi_a)\}$$

is a nonempty-valued conservative internally stable SB for the Harsanyi-Chwe situation. Indeed, let $b \in \hat{\sigma}(X_a)$; then $b \in \hat{\sigma}(X_b)$ and hence $b \in \sigma(\Pi_b)$. Therefore, there does not exist $c \in Z$ and $S \subset N$ such that $b \xrightarrow{S} c$, $\sigma(\Pi_c) \neq \emptyset$, and $b \prec_S \eta$ for all $\eta \in \sigma(\Pi_c)$. Since $\eta \in \sigma(\Pi_c)$ implies $t(\eta) \in \hat{\sigma}(X_c)$, there does not exist $c \in Z$ and $S \subset N$ such that $b \xrightarrow{S} c$, $\hat{\sigma}(X_c) \neq \emptyset$, and $b \prec_S d$ for all $d \in \hat{\sigma}(X_c)$. Hence $\hat{\sigma}$ is conservative internally stable. Since σ is nonempty-valued, $\hat{\sigma}$ is also nonempty-valued. ∎

Proof of Claim 2.20. It is easy to verify that the situation with perfect foresight is equivalent to the "tree situation" in Greenberg (1990). Then Claim 4.5 follows from Theorem 8.2.2 in Greenberg (1990) and Proposition 2.14 in this chapter. ∎

Proof of Claim 2.21. This is a special case of Shitovitz's theorem [Theorem 4.5 (Greenberg 1990)]. The proof in our context follows easily from the definition of OSSB and that of abstract stable set. ∎

Appendix 2.2: A Simple Cooperative Game

Professor Ron Holzman (in 1994) pointed out to me the following four-person transferable utility (TU) game that can be used to illustrate Harsanyi's criticism.

$$\begin{aligned} v(S) =& 1, \text{ if } |S| \geq 3 \text{ or } S = \{3,4\}, \\ & \frac{1}{2}, \text{ if } S = \{1,3\},\{1,4\},\{2,3\},\{2,4\}, \\ & 0, \text{ otherwise} \end{aligned}$$

where $v(S)$ denotes the value of coalition S.

This TU game has a finite stable set that consists of the following 7 points:

$$\begin{aligned} K =& \{(0,0,\tfrac{1}{2},\tfrac{1}{2}),(\tfrac{1}{4},\tfrac{1}{4},\tfrac{1}{2},0),(\tfrac{1}{4},0,\tfrac{1}{2},\tfrac{1}{4}), \\ & (0,\tfrac{1}{4},\tfrac{1}{2},\tfrac{1}{4}),(\tfrac{1}{4},\tfrac{1}{4},0,\tfrac{1}{2}),(\tfrac{1}{4},0,\tfrac{1}{4},\tfrac{1}{2}),(0,\tfrac{1}{4},\tfrac{1}{4},\tfrac{1}{2})\} \end{aligned}$$

Consider $x = (\frac{1}{4},0,\frac{1}{2},\frac{1}{4}) \in K$. $x < y = (0,\frac{1}{8},\frac{1}{2},\frac{3}{8})$, since $x \xrightarrow{\{2,4\}} y$ and $x \prec_{\{2,4\}} y$. The only imputation in K that dominates y is $z = (\frac{1}{4},\frac{1}{4},0,\frac{1}{2})$ since

$$y \xrightarrow{\{1,2,4\}} z \text{ and } y \prec_{\{1,2,4\}} z.$$

If players are farsighted, coalition $\{2,4\}$ will replace x by y, knowing that coalition $\{1,2,4\}$ will replace y by z. Note that no coalition will replace z, since no imputation will make either player 1, 2, or 4 better off.

Appendix 2.3 Nonemptiness of the LCS

Casting the LCS within the framework of the theory of social situations provides additional benefit: it extends Chwe's (1994) result on the nonemptiness of the LCS.

Theorem 2.22. *If there is no infinite sequence $a_1, a_2, \ldots$ in Z such that $i < j$ implies $a_i \ll a_j$, then there exists a nonempty valued CSSB.*

Hence by Proposition 2.17, the largest consistent set is nonempty. Moreover, the nonempty-valuedness of the CSSB implies that for all $a \in Z \setminus \text{LCS}$, there is

$b \in$ LCS such that $a \ll b$. Note that in contrast to Chwe (1994, Proposition 2), Theorem 2.22 does not require Z to be countable.

The sketch of the proof is as follows. First, note that an SB σ such that $\sigma(X_a) = X_a$ for all $a \in Z$ is nonempty-valued conservative externally stable. I use Zorn's Lemma to show that there exists a minimal nonempty-valued conservative externally stable SB and then show this SB is also conservative internally stable. This is dual to a theorem of Greenberg, Shitovitz and Monderer (1995) that the maximal nonempty valued conservative stable SB is conservative externally stable.

Proof of Theorem 2.22. For $a \in Z$ and an SB σ, define

$$\begin{aligned}\text{CDOM}(\sigma, X_a) =&\{b \in X_a \mid \exists S \subset N \text{ and } c \in Z \text{ such that } b \rightarrow_S c, \sigma(X_c) \neq \emptyset \\ &\text{and } \forall d \in \sigma(X_c),\ b \prec_S d\}\end{aligned}$$

Let $\mathcal{K}$ be the set of SBs σ with the following properties:

(A.1) $\forall a, b, c \in Z,\ c \in \sigma(X_a) \cap X_b \implies c \in \sigma(X_b)$.

(A.2) $\forall a \in Z,\ \exists b \in Z$ such that $b \in \sigma(X_b)$ and $\sigma(X_b) \subset \sigma(X_a)$.

(A.3) σ is conservative externally stable, i.e., $\sigma(X_a) \supset X_a \backslash \text{CDOM}(\sigma, X_a)$, $\forall a \in Z$.

$\mathcal{K} \neq \emptyset$ since $\sigma^0 \in \mathcal{K}$ where $\sigma^0(X_a) \equiv X_a$ for every $a \in Z$. Define a partial ordering "$\leq$" on $\mathcal{K}$ such that for every $\sigma, \sigma' \in \mathcal{K}, \sigma \leq \sigma'$ if and only if $\sigma(X_a) \supset \sigma'(X_a)$ for every $a \in Z$.

Claim 1. $\mathcal{K}$ has a maximal element (with respect to "$\leq$").

By Zorn's Lemma, it suffices to show that every chain in $\mathcal{K}$ has an upper bound in $\mathcal{K}$. Let $\mathcal{C}$ be a chain. It suffices to show that $\eta \in \mathcal{K}$ where $\eta(X_a) \equiv \bigcap_{\sigma \in \mathcal{C}} \sigma(X_a)$ for every $a \in Z$.

We first prove that η satisfies (A.1). By the definition of η, for every $a, b, c \in Z$, $c \in \eta(X_a) \cap X_b$ implies $c \in \sigma(X_a) \cap X_b$ for every $\sigma \in \mathcal{C}$. Since every σ in $\mathcal{C}$ satisfies (A.1), $c \in \sigma(X_b)$ for every $\sigma \in \mathcal{C}$. Thus $c \in \eta(X_b)$.

To prove that η satisfies (A.2), we need the following property of η:

(C.1) Let $\sigma \in \mathcal{C}$ and $a, b \in Z$. Then, $\sigma(X_a) \supset \sigma(X_b)$ implies $\eta(X_a) \supset \eta(X_b)$.

Indeed, $\eta(X_b) \subset \sigma(X_b) \subset \sigma(X_a) \subset X_a$. Since η satisfies (A.1), $c \in \eta(X_b) = \eta(X_b) \cap X_a \implies c \in \eta(X_a)$, i.e., $\eta(X_a) \supset \eta(X_b)$.

We now proceed to prove that η satisfies (A.2). If for every $a \in Z, a \in \eta(X_a)$, then we are done. Otherwise, there exists $a_0 \in Z$ such that $a_0 \notin \bigcap_{\sigma \in \mathcal{C}} \sigma(X_{a_0}) \equiv \eta(X_{a_0})$, implying that there exists $\sigma_1 \in \mathcal{C}$ such that $a_0 \notin \sigma_1(X_{a_0})$. Since $\sigma_1 \in \mathcal{K}$, (A.2) implies that there exists $a_1 \in Z$ such that $a_1 \in \sigma_1(X_{a_1}) \subset \sigma_1(X_{a_0})$. By (C.1), $\eta(X_{a_0}) \supset \eta(X_{a_1})$. If $a_1 \in \eta(X_{a_1})$, we are done. Otherwise, there exists $\sigma_2 \in \mathcal{C}$ such that $a_1 \notin \sigma_2(X_{a_1})$. Since $a_1 \in \sigma_1(X_{a_1})$ and $\mathcal{C}$ is a chain, $\sigma_1 < \sigma_2$. In particular, $\sigma_2(X_{a_1}) \subset \sigma_1(X_{a_1})$. Applying (A.2) again, there exists $a_2 \in Z$ such that $a_2 \in \sigma_2(X_{a_2}) \subset \sigma_2(X_{a_1})$. Thus $a_2 \in \sigma_2(X_{a_2}) \subset \sigma_2(X_{a_1}) \subset \sigma_1(X_{a_1}) \subset \sigma_1(X_{a_0})$. By (C.1), $\eta(X_{a_0}) \supset \eta(X_{a_1}) \supset \eta(X_{a_2})$. If $a_2 \in \eta(X_{a_2})$, then we are done. Otherwise, there exists $\sigma_3 \in \mathcal{C}$ such that $a_2 \notin \sigma_3(X_{a_2})$. Continuing this inductively, there exists $\sigma_1(X_{a_0}) \supset \sigma_1(X_{a_1}) \supset \sigma_2(X_{a_1}) \supset \sigma_2(X_{a_2}) \supset \cdots \supset \sigma_k(X_{a_{k-1}}) \supset \sigma_k(X_{a_k}) \supset \cdots$ such that for $i \geq 1, a_i \in \sigma_i(X_{a_i}) \setminus \sigma_{i+1}(X_{a_i})$ and $\eta(X_{a_0}) \supset \eta(X_{a_1}) \supset \cdots \supset \eta(X_{a_k}) \supset \cdots$. If there exists $J < \infty$ such that $a_J \in \eta(X_{a_J})$, then η satisfies (A.2) and we are done. Otherwise, since $1 \leq i < j$ implies $a_i \neq a_j$ and $a_j \in \sigma_i(X_{a_i}) \subset X_{a_i}$, there exists a infinite sequence $a_1, a_2, \ldots, a_k, \ldots$ such that $a_i \ll a_j$ if $i < j$, contradicting the assumption of the theorem.

Finally, we show that η also satisfies (A.3). Otherwise, there exists $a \notin \eta(X_a) \cup$ CDOM(η, X_a). Since by (A.2) η is nonempty valued, CDOM$(\sigma, X_a) \subset$ CDOM(η, X_a) for every $\sigma \in \mathcal{C}$. Hence $a \notin$ CDOM(σ, X_a) for every $\sigma \in \mathcal{C}$. But σ belongs to $\mathcal{K}$ and hence is conservative externally stable. Thus $a \in \sigma(X_a)$ for every $\sigma \in \mathcal{C}$, implying $a \in \eta(X_a)$, a contradiction.

Since η satisfies (A.1), (A.2), and (A.3), $\eta \in \mathcal{K}$. Thus, every chain in $\mathcal{K}$ has an upper bound in $\mathcal{K}$, and therefore by Zorn's Lemma, $\mathcal{K}$ has a maximal element.

Claim 2. Let σ^* be the maximal element in $\mathcal{K}$. Then σ^* is also conservative internally stable.

Assume in negation that there exists $b \in Z$ such that $b \in \sigma^*(X_a)$ and $b \in \text{CDOM}(\sigma^*, X_a)$. Define $\sigma'(X_x) = \sigma^*(X_x) \setminus \{b\}$ for every $x \in Z$. Since $\sigma^* < \sigma'$, to reach the desired contradiction, it suffices to show that $\sigma' \in \mathcal{K}$. Since σ^* satisfies (A.1), so does σ'. To see that σ' satisfies (A.2), let $x \in Z$. Since $\sigma^* \in \mathcal{K}$, by (A.2) there exists $c \in Z$ such that $\sigma^*(X_x) \supset \sigma^*(X_c) \ni c$. If $c \neq b$, then $\sigma'(X_x) \supset \sigma'(X_c) \ni c$ and we are done. Otherwise, $c = b$; hence, $\sigma^*(X_x) \supset \sigma^*(X_b) \ni b$. But $b \in \text{CDOM}(\sigma^*, X_a)$ implies that there exists $c \in Z$ and S such that $b \rightarrow_S c$, $\sigma^*(X_c) \neq \emptyset$ and for every $d \in \sigma^*(X_c), b \prec_S d$, i.e., $b \Rightarrow d$. Thus, $b \notin \sigma^*(X_c)$ and by (A.1), $\sigma^*(X_b) \supset \sigma^*(X_c)$. Again, since $\sigma^* \in \mathcal{K}$, (A.2) implies that there exists $h \in Z$ such that $h \in \sigma^*(X_h) \subset \sigma^*(X_c)$. Also, $b \notin \sigma^*(X_c)$ implies $h \neq b$. Thus $\sigma'(X_x) \supset \sigma'(X_h) \ni h$; hence σ' satisfies (A.2) and is nonempty valued. Consequently, $\text{CDOM}(\sigma', X_x) \supset \text{CDOM}(\sigma^*, X_x)$ for every $x \in Z$, which implies σ' is conservative externally stable, i.e., σ' satisfies (A.3). So $\sigma' \in \mathcal{K}$, a contradiction.

Since σ^* is both internally and externally stable, it is a CSSB. ∎

Appendix 2.4 Foresight, Feasible Outcomes, and Nonempty-valuedness of a Stable SB

Given a social environment $\mathcal{G} = (N, Z, \{\prec_i\}_{i \in N}, \{\overset{S}{\rightarrow}\}_{S \subset N, S \neq \emptyset})$, let F_a denote the set of "outcomes" individuals in N regard "feasible" at every $a \in Z$. $\{F_a\}_{a \in Z}$ together with $\{\rightarrow_S\}_{S \subset N}$ can be called a "situation" in the sense of Greenberg (1990). Such a situation describes how individuals view their available alternatives. Foresight or myopia on the part of the individuals is reflected, particularly, in the specification of $\{F_a\}_{a \in Z}$.

If $F_a \equiv \{a\}$ for all $a \in Z$, then the situation entails myopia of the individuals. Consider, for example, Figure 2.1. The unique OSSB and CSSB σ is such that

$$\sigma(F_c) = \{c\}, \sigma(F_b) = \emptyset, \text{ and } \sigma(F_a) = \{a\}.$$

The unique vN-M abstract stable set (or the set of stable outcomes) is given by $V = \sigma(F_a) \cup \sigma(F_b) \cup \sigma(F_c) = \{a, c\}$. That $\sigma(F_b)$ is empty can be interpreted as follows: b, the only feasible outcome at b is ruled out from $\sigma(F_b)$, since player 2 can "induce" c and he prefers $\sigma(F_c) = \{c\}$ to b; yet this reasoning is not reflected in $\sigma(F_b)$, which asserts that player 2 will not stay at b but does not specify *what he will do*. A problem arises immediately since $\sigma(F_b)$ play a key role in determining $\sigma(F_a)$, the solution at a. That $\sigma(F_b)$ fails to specify *what player 2 will do* is due to that c is not a feasible outcome at b.

A remedy is to consider c as a feasible outcome when b is the status quo, since c can replace a. Generally, let $F_a \equiv X_a$ where

$$X_a = \{a\} \cup \{b \in Z \mid b \gg a\} \subset Z$$

as discussed in Section 2.4. In this case, the OSSB is formally related to the vN-M abstract stable set for $(Z, \gg)$, and CSSB to the LCS. In doing so, both OSSB and CSSB are free of myopia. This, however, is not the case for a slightly more complex example depicted in Figure 2.2, as discussed in Section 2.2. According to the specification of $\{F_a\}_{a \in Z}$, d belongs to $F_a \equiv X_a$. But such a specification does not address how d is reached from a; consequently, deviations "along the way" from a to d are ignored.

In the example depicted in Figure 2.2, the perfect foresight of player 1 should enable him to assess if d can be reached when a is the status quo. This cannot be achieved if d is simply considered as a feasible at a. The path that leads from a to d should be considered instead. Generally, let $F_a \equiv \Pi_a$ for all $a \in Z$, where a is the set of paths originate from a. Then this "path" situation captures perfect foresight in that every coalition, in choosing an (joint) action, considers that another coalition might react, a third coalition might in turn react, as so on. Rationality determines which paths will be followed hence the coalitions that will form to implement such a path. The perfect foresight will emerge in the OSSB or CSSB if it is non-empty valued. Otherwise, as discussed in Section 2.5, an stable

SB may display myopia. Consider, again, the example depicted in Figure 2.8. An empty solution at a represents the fact that both players cannot agree on any path to follow; each player wish to induce his favorite outcome. Consequently, either path (a, b) or (a, c) *might actually arise*; this fact should be used to determine the solution at x, given that an empty solution at a tells nothing but that players cannot agree upon either path. In this case, we need to modify the definition of a stable SB to account for *what might actually arise* whenever players cannot reach an agreement as what path to follow.

Chapter 3
Negotiation-Proof Nash Equilibrium

This chapter defines "negotiation-proof Nash equilibrium", a notion that applies to environments where players can negotiate openly and directly prior to the play of a noncooperative game. It recognizes the possibility that a group of self-interested players may choose to coordinate, nonbindingly and voluntarily, their choice of strategies and make a joint objection, and it takes the perfect foresight of rational players fully into account. The merit of the notion of negotiation-proof Nash equilibrium is twofold: (1) It resolves the nestedness and myopia embedded in the notion of *coalition-proof Nash equilibrium*. (2) The negotiation process, which is formalized by a "graph", serves as a natural extension to approach that models pre-play communication by an extensive form game.

3.1 Introduction

The most fundamental solution concept for noncooperative games is that of Nash equilibrium. One common interpretation of Nash equilibrium is as a self-enforcing agreement. That is, if players communicate and agree on a certain profile of strategies without a binding agreement, then these strategies must constitute a Nash equilibrium. But communication may achieve better outcomes for the players since it creates the opportunity for negotiation and coordination. In this paper I analyze the consequence of open negotiation prior to the play of a noncooperative game. I defined the notion of "negotiation-proof Nash equilibrium", which recognizes the possibility that a group of self-interested players may choose to coordinate, nonbindingly and voluntarily, their choice of strategies, and takes the perfect foresight of rational players fully into account.

The are several approaches to communication. The notion of *correlated equilibrium* (Aumann 1974) considers mediated communication: a mediator (or a correlation device) helps the players communicate and share information. Mediated

communication can achieve payoffs that are not possible in any Nash equilibrium, and it does so by extending the set of equilibria. Alternatively, one can consider direct unmediated communication prior to the play of a noncooperative game, exploring the coordination role of communication. One approach to direct communication is to explicitly model the procedure of communication as a dynamic game, which specifies how messages are interchanged, the order of offers and counter-offers, and etc. [see, e.g., Farrell (1987, 1988) and Rabin (1994)]. The result, however, may be sensitive to the exact procedure employed and strong restrictions often have to be made to isolate the desired result. Also, one may argue that modeling communication as a noncooperative game may not fully capture the coordination role of communication, since the communication game itself may in turn call for coordination. Another approach to direct communication focuses on the possibility that players can coordinate their choice of strategies via self-enforcing agreements that are mutually beneficial, leaving the details of communication unmodeled [see, e.g., Bernheim et al.'s (1987)]. I shall first motivate my analysis by examining such an approach to direct communication, and then discuss the relation of my analysis to the first approach.

Bernheim et al.'s (1987) notion of *coalition-proof Nash equilibrium* (CPNE) "is designed to capture the notion of an efficient self-enforcing agreement for environments with unlimited but nonbinding, pre-play communication" ($p.3$). One motivation is that the notion of *strong Nash equilibrium* (SNE) fails to capture the fact that a coalitional deviation may be subject to further deviations in the absence of binding agreements. An agreement is coalition-proof if it is efficient within the class of "self-enforcing" agreements. In turn, an agreement is "self-enforcing" if and only if no *proper subset* of players, taking the strategies of its complement as fixed, can deviate in such a way that benefits all its members. Therefore, in the definition of CPNE, self-enforceability of agreements is restricted to an important aspect: only subsets of a deviating coalition can further deviate. While such a

(*nestedness*) restriction enables CPNE to be defined recursively, it also implies that the definition of CPNE may involve agreements that are open to further deviations. Consider the 3-player game in Table 3.1, where player 1 chooses rows, player 2 chooses columns, and player 3 chooses matrices.

TABLE 3.1

	L	R
U	2,2,1	1,0,0
D	0,0,0	3,3,0

A

	L	R
U	0,0,0	0,2,0
D	0,0,0	1,4,1

B

The game in Table 3.1 has two Nash equilibria (in pure strategies): (U, L, A) and (D, R, B). However, (U, L, A) is not coalition-proof by the following argument: Players 1 and 2 can jointly deviate to (D, R, A) which renders both players 1 and 2 higher payoffs. Such a deviation is "self-enforcing" because, according to the nestedness restriction in the definition of CPNE, only subsets of $\{1, 2\}$ can further deviate. Without the nestedness restriction, the self-enforceability of the deviation to (D, R, A) is evidently in doubt. Players 2 and 3 have incentive to further deviate from (D, R, A) to (D, R, B) in a self-enforcing way, thereby upsetting its self-enforceability.

TABLE 3.2

	L	C	R
U	1,1,1	0,0,0	0,0,0
M	0,0,0	0,0,0	0,0,0
D	0,0,0	0,0,0	0,0,0

A

	L	C	R
U	0,5,0	0,0,0	4,4,0
M	0,0,0	2,2,2	0,0,0
D	0,0,0	0,0,0	3,3,0

B

Aside from the critique of the nestedness restriction, the definition of CPNE also fails to account for the foresight of rational players, as noticed by Chwe (1994).

The myopia embedded in the definition of CPNE can be illustrated by the example in Table 3.2 taken from Chwe (1994). For this game, the unique CPNE is (M, C, B). Although (D, R, B) renders both players 1 and 2 higher payoffs than (M, C, B) does, players 1 and 2 will not jointly deviate to (D, R, B). According to the definition of CPNE, such a joint deviation is not self-enforcing, the reason being that player 1 can subsequently deviate to (U, R, B), a "self-enforcing agreement" under the nestedness assumption. But this implies, evidently, that players 1 and 2 are myopic: were they farsighted, their joint deviation to (D, R, B) should be encouraged, not discouraged, by player 1's further deviation to (U, R, B).

This chapter offers a model of pre-play communication that overcomes the difficulties of CPNE as illustrated through the examples in Tables 3.1 and 3.2. The suggested notion, "negotiation-proof Nash equilibrium", exploits open nonbinding negotiation that takes place prior to the play of an one-shot noncooperative game. The pre-play negotiation is conducted as follows: Suppose a strategy profile is considered by all the players. A group or coalition of players *can* make a joint objection by announcing *openly*, "if the rest of you stick with your strategies, we shall adopt new strategies so-and-so". This objection is simply a declaration of joint intention or a joint "contingent threat" that comprises no binding power. Given the new, revised strategy profile, another coalition, not necessarily a subset of the original objecting coalition, *can* make a further objection by announcing *openly* the new strategies its members will adopt contingent on the strategies of nonmembers. The process continues in this manner, until no coalition has an incentive to make any further objection. Since players are rational (and hence farsighted) and binding agreements are not possible, a coalition, in contemplating an objection, has to consider the *ultimate* consequences of its objection; and a self-interested player joins a coalition only if it is in his best interest to do so.

The above pre-play negotiation process takes after the "coalitional contingent threat situation" (Greenberg 1990) but for the following two distinct features:

(i) The pre-play negotiation is proceeded by an *one-shot noncooperative game*, hence a meaningful agreement must be self-enforcing, i.e., must be a Nash equilibrium;

(ii) In the pre-play negotiation, players are farsighted in that each coalition[9], in making an objection, considers that another coalition may make counter-objections, a third coalition may make further objections, and etc. What matters to farsighted players is the final agreements that their objections will lead to; hence they may strategically "deviate" to an agreement, which is not necessarily a Nash equilibrium, in order to induce a final agreement (necessarily a Nash equilibrium) that benefits all its members.

Loosely speaking, a Nash equilibrium is negotiation-proof if and only if no coalition can make an objection to it in such a way that its objection will *ultimately* lead to another negotiation-proof Nash equilibrium that benefits all its members. Such a definition is intrinsically "circular",[10] and is achieved by employing von Neumann and Morgenstern's (1947) "abstract stable set".

In the above pre-play negotiation, it is *feasible* that any coalition can form and object to any strategy profile. However, a rational and self-interested player is not bounded to join any coalition. The formation of any coalition is purely voluntary and is driven by each member's pursuing his own interest, and a group of *rational* players forms a coalition only if it is in the best interest of each member not to quit this coalition. Thus, our negotiation process captures the intrinsic noncooperative behavior of the players. In Table 3.1, for example, its is *feasible* for players 1 and 2 to form a coalition and jointly "deviate" from (U, L, A) to (D, R, A). But, player 1, being *rational* and hence *farsighted*, will not join player 2 to make such a deviation, knowing player 3 (or players 2 and 3) will further deviate to (D, R, B). It is not essential who can make a proposal that players 1 and 2 form a coalition to jointly deviate from (U, L, A) to (D, R, A): player 1 will neither initiate nor

[9] A single player is a singleton coalition.

[10] Recall that the nestedness restriction enables CPNE to be defined recursively.

accept such a proposal.

Therefore, it is *a priori* that any coalition can form, but rationality dictates which coalitions (would) actually form. Here, I take the view that "it is the possibilities for coalition forming, promising and threatening that are decisive, rather than whose turn it is to speak" (Aumann 1987). The negotiation allows the players to negotiate openly and directly, and to exercise their "bargaining power" embedded in the structure of the game, in particular, the intrinsic properties of payoffs. As discussed earlier, some models of pre-play communication impose procedures that can be represented by extensive form games. For example, in Rabin's (1994) [see also Farrell (1987, 1988)] model of pre-play communication (for two-player games), players make repeated simultaneous proposals of equilibria; if the players propose the same equilibrium, they have an agreement to play that equilibrium.[11] The pre-play negotiation process postulated in this paper may be viewed as a natural extension to these models.

In the next section, the pre-play negotiation process among rational (and far-sighted) players is formalized as a "(directed) graph". Such a graph need not be acyclic and does not stipulate that each "node" should belong to a single player.[12] This is in contrast to an extensive form game, which is a acyclic graph and requires each node to belong to a single player. The graph captures the dynamics as well as the diverse coalitional interactions in the negotiation process. Although, such a complex negotiation process cannot be represented by an extensive form game in discrete time, it might be possible to accommodate such a process in an extensive form game in continuous time, such as the framework used by Perry and Reny (1994).

The rest of this chapter is organized as follows. In the next section, following

[11] Such a procedure may be "at variance with common procedure" [see Rabin (1994, *p.*389)].

[12] In fact, when a strategy profile x is under consideration, it is feasible that any coalition can make an objection. Rationality dictates which coalitions would actually form, and x is not negotiation-proof as long as there exists one coalition of rational players who will ultimately benefit by objecting to x. Therefore, it is not necessary to restrict that only a particular coalition or player can object to x.

the formalization of the negotiation process, a formal definition of "negotiation-proof Nash equilibrium" is offered using von Neumann and Morgenstern abstract stable set. Section 3.3 provides a way to improve the notion of negotiation-proof Nash equilibrium. Section 3.4 extends negotiation-proofness to dynamic games. Section 3.5 offers a brief discussion of several attempts in the literature to relax the nestedness restriction of CPNE and to capture the foresight of players in strategic settings. It also briefly discuss the possibility of allowing for correlated strategies.

3.2 Negotiation-Proof Nash Equilibrium

Consider a strategic form game $\mathcal{G} \equiv (N, \{Z_i\}_{i\in N}, \{u_i\}_{i\in N})$, where N is the set of players and for every $i \in N$, Z_i is the set of strategies of player i, and u_i is the payoff function of player i, $u_i\colon Z \to \Re$, where $Z \equiv \prod_{i\in N} Z_i$. For $S \subset N$, let $Z_S \equiv \prod_{i\in S} Z_i$, and for all $x, y \in Z$, I write $x \prec_S y$ if $u_i(x) < u_i(y)$ for all $i \in S$.

Assume that $x \in Z$ is under consideration. As discussed in the introduction, it is a priori that any coalition *can* form to jointly object to x. If a coalition $S \subset N$ objects to x by choosing $y_S \in Z_S$ contingent on $x_{N\setminus S}$, then the resulting new strategy profile is $y = (y_S, x_{N\setminus S})$; in this case, I shall write $x \to_S y$ to denote "S objects y to x". Thus, for all $S \subset N$, $\to_S$ is a (binary) relation on Z that specifies what S can do *if and when* it forms. Given that players are rational and hence farsighted, if a coalition S forms and objects y to x, it must be the case that

(C1) such an objection leads to a final agreement z that benefits all members of S.

Recall the example in Table 3.1. Coalition $\{1, 2\}$ does not form for the exact reason that (C1) is violated. The example Table 3.3 illustrate that a coalition forms when (C1) holds.

Both (U, L, A) and (D, R, B) are CPNE's. In fact, they are also SNE's. I shall argue, however, that players 1 and 2, being farsighted (as implied by rationality), will jointly deviate from (U, L, A) to (D, R, A), because player 3, for his own

TABLE 3.3

		L	R
U		2,2,2	0,0,1
D		0,0,1	2,2,0

A

		L	R
U		0,0,0	1,1,1
D		0,0,0	3,4,1

B

interest, will subsequently deviate to (D, R, B), which renders both players 1 and 2 higher payoffs than (U, L, A). Note that players 1 and 2 do not (strictly) prefer (D, R, A) to (U, L, A). Thus, in contemplating a deviation, a coalition of farsighted players considers the *final* agreement that its deviation leads to, as asserted by (C1). Note that the joint deviation of players 1 and 2 from (U, L, A) to (D, R, A) is self-enforcing: neither player 1 nor player 2 has an incentive to object to such an agreement, knowing that the joint deviation leads to (D, R, B). Again, it is not essential that who proposes this agreement; it is the existence of such an agreement that invalidates (U, L, A).

TABLE 3.4

		L	R
U		2,2,2	0,0,1
M		2,0,1	4,0,2
D		0,0,1	2,2,0

A

		L	R
U		0,0,0	1,1,1
M		0,0,2	0,1,0
D		0,0,0	3,4,1

B

As condition (C1) asserts, it is not sufficient for coalition $\{1, 2\}$ to form and make a joint objection in such a way that this objection can *feasibly* lead to a final agreement that makes both players 1 and 2 better off. Further deviations along the way to the final agreement that players 1 and 2 hope to reach, may leads to an agreement that make player 1 or 2 worse off. Consider the example in Table

3.4, which is a modification of Table 3.3. It is still *feasible* that players 1 and 2's joint deviation from (U, L, A) to (D, R, A) leads to (D, R, B). But, will player 2 join player 1 to deviate from (U, L, A) to (D, R, A) in the hope that player 3 will subsequently deviate to (D, R, B)? The answer is no. Once (D, R, A) is reached, it is inevitable that player 1's further deviation to (M, R, A) would prevail. Thus, although it is feasible for players 1 and 2 to jointly deviate from (U, L, A) to (D, R, A), player 2, being self-interested and farsighted, will not form a coalition with player 1 to make such a deviation.

The above two examples illustrate that rational and farsighted players, in contemplating their deviations, consider all further deviations and recognize the other players are also rational and farsighted; they "strategically" deviate from an agreement *if and only if* such a deviation will *ultimately* lead to final agreements that make them better off. That is, a coalition forms to make a joint objection if and only if such an objection can lead to a final agreement that benefits all its members, and no coalition has an incentive to prevent this final agreement from being reached by deviating along the way to this final agreement. The perfect foresight is captured by considering, as a whole, the succession or "path" of deviations that lead to a final agreement. Consider, in the examples in both Tables 3.3 and 3.4, the path that players 1 and 2 deviate from (U, L, A) to (D, R, A) and player 3 subsequently deviates to (D, R, B). This path "prevails" in the example depicted by Table 3.3 because it survives all rational deviations of farsighted coalitions (or players); that is, players 1 and 2's joint deviation from (U, L, A) to (D, R, A) will lead to (D, R, B). The same path, however, does not prevail in the example depicted by Table 3.4.

Thus, we can represent the pre-play negotiation process by a (*directed*) *graph* that 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 $a \rightarrow_S b$. Assume that some $y \in Z$ can replace x through a succession of deviations and,

at every "stage", the deviating coalition prefers y to the agreement from which it deviates.[13] This succession of deviations that replace x with y a called a "path" (of deviations) from x to y. Formally,

Definition 3.1. A path from $x \in Z$ is a sequence of strategy profiles $(x^0, x^1, \ldots, x^m)$ in Z, where $x^0 = x$, such that there exist coalitions $S^0, S^1, \ldots, S^{m-1}$ and $x^j \rightarrow_{S^j} x^{j+1}$ and $x^j \prec_{S^j} x^m$, for all $j = 0, 1, \ldots, m-1$.

For a game $\mathcal{G}$, let Π denote the set of all paths, including all "degenerate" paths, i.e., all elements in Z. For $\alpha \in \Pi$, let $f(\alpha)$ denote the final "node" (strategy profile) that lies on path α, i.e., $f(\alpha) = x^m$, and if x is strategy profile that lies on α, I shall write $x \in \alpha$. For $\alpha, \beta \in \Pi$, if $f(\alpha) \prec_S f(\beta)$ for some $S \subset N$, I shall write $\alpha \prec_S \beta$.

As discussed in the introduction, the open negotiation is proceeded by a non-cooperative game; hence a meaningful agreements must belong to the set of Nash equilibria (self-enforcing agreements) of $\mathcal{G}$. Therefore, only those paths that lead to Nash equilibria are of interest. Let NE denote the set of Nash equilibria of $\mathcal{G}$ and let $\Pi_{NE} \equiv \{\alpha \mid f(\alpha) \in NE\}$. In order to determine whether a path $\alpha \in \Pi_{NE}$ will prevail, deviations along α have to be considered. For any $x \in \alpha$, if a coalition can initiate another path β that makes its members better off than α, then α is said to be "dominated" by β. That is,

Definition 3.2. For $\alpha, \beta \in \Pi_{NE}$, α is dominated by β, or $\alpha < \beta$, if there exists $x \in \alpha$ and $y \in \beta$ such that $x \rightarrow_S y$ and $\alpha \prec_S \beta$.[14]

β itself may be dominated by another path, say, γ. Thus whether α will prevail depends whether β will prevail; whether β will prevail depends, in turn, on whether γ will prevail; and so on. We wish to identify a set of paths $\Sigma \subset \Pi_{NE}$

[13] The latter condition implies that we restrict our attention to those "paths" that can possibly be followed by rational players.

[14] For the example in Table 3, let $\alpha \equiv ((U, L, A), (D, R, A), (D, R, B))$ (i.e., path α consists of players 1 and 2's deviation from (U, L, A) to (D, R, A) and player 3's further deviation to (D, R, B)) and $\beta \equiv ((D, R, A), (M, R, A))$; then, $\alpha < \beta$.

such that it contains *those and only those paths* that are not objected by any coalition, whose members are aware of and believe in the specification of such Σ. That is, Σ is both consistent and self-policing: moreover, it "justifies" every path it excludes. This is precisely the intuition behind the von Neumann and Morgenstern abstract stable set. Recall,

Definition 3.3. Let D be an arbitrary nonempty set and $\angle$ be a binary relation on D, called the *dominance relation*[15]. The pair $(D, \angle)$ is called an *abstract system*. $K \subset D$ is

(1) *internally stable* if K is free of inner contradiction, i.e., there do not exist $a, b \in K$, such that $a \angle b$;

(2) *externally stable* if K accounts for every element it excludes, i.e., if $a \in D \setminus K$, then there exists $b \in K$ such that $a \angle b$.

(3) an *abstract stable set* if K is both internally and externally stable.

Let Σ be an abstract stable set for $(\Pi_{NE}, <)$, then it contains those paths that are to prevail in the pre-play negotiation, once Σ becomes common knowledge. Note that the abstract stable set for $(\Pi_{NE}, <)$ takes noncooperative behavior of self-interested players fully into consideration. If a path α in Σ involves any coalition, it implies that members of this coalition recognize the interdependence of their welfares and choose to coordinate their choice of strategies. Should some player find it not in his best interest to joint such a coalition, α would have been ruled out from Σ. Consider, again, the path that players 1 and 2 deviate from (U, L, A) to (D, R, A) and player 3 subsequently deviates to (D, R, B) in the examples in both Tables 3.3 and 3.4. This path belongs to the unique abstract stable set for $(\Pi_{NE}, <)$ associated with example in Table 3.3, implying that players 1 and 2 will form a coalition if (U, L, A) is under consideration. The same path, however, does not belong to the unique abstract stable set for $(\Pi_{NE}, <)$ associated with Table 3.4, because player 2 knows that once (D, R, A) is reached, players 1

[15] $a \angle b$ means that a is dominated by b.

and 3 will deviate to (M, R, A) (which belongs to the abstract stable set): hence players 1 and 2 will not form a coalition when (U, L, A) is under consideration.

If Σ is an abstract stable set for $(\Pi_{NE}, <)$ then it is "dynamically consistent". That is, every "stable path" in Σ satisfies "truncation property": the continuation of a "stable path" is stable at any stage along the way. Formally,

Lemma 3.1. *Assume that Σ is an abstract stable set for $(\Pi_{NE}, <)$ and that $\alpha \in \Sigma$. Then, $\alpha|_x \in \Sigma$ for all $x \in \alpha$, where $\alpha|_x$ is the continuation of α from x.*

If a Nash equilibrium x is not objected by any coalition who believes the specification of Σ (that is stable), then x is said to be "negotiation-proof".

Definition 3.4. Let Σ be an abstract stable set for the abstract system $(\Pi_{NE}, <)$; then the set of Negotiation-Proof Nash Equilibria (NPNE's) of $\mathcal{G}$ is given by

$$Q_\Sigma \equiv \{x \in NE \mid (x) \in \Sigma\} \equiv \{x \in Z \mid \exists \alpha \in \Sigma \text{ such that } x = f(\alpha)\}.$$

If Σ is an abstract stable set for $(\Pi_{NE}, <)$, then Q_Σ is nonempty (by the external stability of Σ). Q_Σ contains *those and only those* self-enforcing agreements from which no coalition can initiate such a deviation that will ultimately lead to some self-enforcing agreement in Q_Σ that benefits all its members. That is, on one hand, if x belongs to Q_Σ, then *no* coalition (or player) *will* ultimately benefit by objecting to x; on the other hand, if x does not belong to Q_Σ, then it must be the case that *at least one* coalition (or player) *will* ultimately benefit by objecting to x. Again, what matters is the existence of a coalition that ultimately benefits from its objection; it is not essential who turn it is to "move" when a strategy profile is under consideration.

For the game in Table 3.1, both (U, L, A) and (D, R, B) are negotiation-proof. For the game in Table 3.2, although the unique NPNE is (M, C, B), which coincides with the unique CPNE, the underlying logic is very different: In CPNE, players 1 and 2 will not deviate to (D, R, B) because of both the nestedness restriction and myopia embedded in the definition of CPNE as discussed in the

introduction. According to the definition of NPNE, however, players 1 and 2 will not deviate to (D, R, B) because such a deviation cannot eventually benefit them. For the example in Table 3.3, (D, R, B) is the unique NPNE, which "refines" CPNE and SNE. The set of NPNE's of the game in Table 4 comprises (U, L, A), (M, R, A), and (D, R, B).

Following von Neumann and Morgenstern (1947), the dominance relation $<$ on Π_{NE} is said to be *strictly acyclic* if there does not exist an infinite sequence of paths $\alpha^1, \alpha^2, \ldots$ in Π_{NE} such that $\alpha^j < \alpha^{j+1}$ for all $j = 1, 2 \ldots$.

Proposition 3.2. *If $<$ is strictly acyclic, then, the set of NPNE's of $\mathcal{G}$ is uniquely defined and nonempty.*

The examples in Tables 3.1 through 3.4 all satisfy the condition in Proposition 3.2.

Corollary 3.3. *Let $\mathcal{G}$ be a game such that NE is finite and all Nash equilibria can be weakly Pareto-ranked. Then, the set of NPNE's of $\mathcal{G}$ is uniquely defined and nonempty. Moreover, if $\mathcal{G}$ has a unique Pareto efficient Nash equilibrium (within NE), then it is the unique NPNE.*

TABLE 3.5

A

	L	R
U	2,2,2	0,0,0
D	0,0,0	3,3,0

B

	L	R
U	2,2,0	0,0,0
D	0,0,0	1,1,1

Thus, for games with common interests and coordination games, pre-play negotiation achieves full efficiency; and if a game has a unique Nash equilibrium (for example, the Cournot oligopoly model), then it is also the unique NPNE. The property of NPNE in Corollary 3.3 is not shared by CPNE. It is easy to verify that the game in Table 3.5 does not admit a CPNE or an SNE. But the

unique NPNE is (U, L, A).

NPNE may differ from CPNE or SNE even for two-player games.

TABLE 3.6

	L	R
U	2,2	0,0
D	0,0	1,2

Both (U, L) and (D, R) are CPNE's and SNE's. However, the unique NPNE is (U, L): player 1, being farsighted, will objects (U, R) to (D, R).

3.3 Weakly Negotiation-Proof Nash Equilibrium

The game in Table 3.6 illustrates that the foresight of rational players enables NPNE to provide "sharp prediction". However, the dominance relation $<$ may endow a deviating coalition (or player) with too much "power". To illustrate this, consider a slight modification of Table 3.6, which gives rise to the familiar "battle of the sexes" game in Table 3.7.

TABLE 3.7

	L	R
U	2,1	0,0
D	0,0	1,2

In this case, paths $\alpha \equiv ((U, R), (D, R))$ and $\beta \equiv ((U, R), (U, L))$ dominate each other. Therefore, for $\Sigma \subset \Pi_{NE}$ to be (internally) stable, either α or β must be excluded from Σ. Indeed, $(\Pi_{NE}, <)$ admits a stable set Σ^1 that rules out α and another Σ^2 that rules out β. Consequently, (U, L) is an NPNE according to Σ^1 and (D, R) is an NPNE according to Σ^2.[16] The exclusion of one path, say α from Σ^1, is attributed to that β belongs to Σ^1 and α is dominated by β . Note,

[16]The Nash equilibrium in mixed strategies yields payoffs of $(\frac{2}{3}, \frac{2}{3})$, is not an NPNE according to either Σ^1 or Σ^2, because it is Pareto dominated by both (U, L) and (D, R).

however, that β itself is also dominated by α. Therefore, it does not seem sound to rule out one path based on another if these two paths dominate each other. For this reason, I define a "stronger" dominance relation $\ll$ based on $<$: $\alpha \ll \beta$ if $\alpha < \beta$ and $\beta \not< \alpha$. Since farsighted players look arbitrarily many steps ahead, the dominance relation $\ll$ relation can be generalized as follows.

Definition 3.5. For $\alpha, \beta \in \Pi$, $\alpha \ll \beta$ if

(1) $\alpha < \beta$, and

(2) there do not exist $\beta^0, \beta^1, \ldots, \beta^m$ in Π_{NE}, where $\beta^0 = \beta$ and $\beta^m = \alpha$, such that for $j = 0, 1, \ldots, m-1$, $\beta^j < \beta^{j+1}$.

Using $\ll$ we can define the notion of "Weakly Negotiation-Proof Nash Equilibrium (WNPNE)" as follows.

Definition 3.6. Let Σ be an abstract stable set for the abstract system $(\Pi_{NE}, \ll)$; then the set of WNPNE's of $\mathcal{G}$ is given by

$$W_\Sigma \equiv \{x \in NE \mid (x) \in \Sigma\} \equiv \{x \in Z \mid \exists \alpha \in \Sigma \text{ such that } x = f(\alpha)\}.$$

The notion of WNPNE coincides with the notion of NPNE for the games in Tables 3.1 to 3.6. However, for the game in Table 3.7, $(\Pi_{NE}, \ll)$ admits a *unique* abstract stable set that includes both α and β; hence the set of WNPNE's $W_\Sigma = \{(U, L), (D, R)\}$ is uniquely defined.[17] WNPNE may exist when NPNE, CPNE, or SNE fails to exist. Consider the following example.

TABLE 3.8

	L	R
U	1,2,3	0,0,0
D	0,0,0	2,3,1

A

	L	R
U	0,0,0	3,1,2
D	0,0,0	0,0,0

B

[17] The Nash equilibrium in mixed strategies is not weakly negotiation-proof.

The game does not admits a CPNE, an NPNE, or an SNE; however, there exists a unique stable set for $(\Pi_{NE}, \ll)$, giving rise to three WNPNE's: $(U, L, A,)$, (D, R, A), and (U, R, B). The implication of the examples in Tables 3.7 and 3.8 is that pre-play negotiation cannot pin down the exact equilibrium to be played in these games.[18]

Proposition 3.4. *Let $\mathcal{G}$ be a finite game. Then, the set of WNPNE's is nonempty and uniquely defined.*

3.4 Extensive Form Games

Although the primary concern of this paper is normal form games, the notions of NPNE and WNPNE can also be extended to dynamic games. (We shall focus on WNPNE in this section.) In doing so, we have to be explicit about whether there is on-going open negotiation as the game unfolds. In the absence of on-going negotiation, players negotiate openly only before they engage in an extensive form game and will not have the opportunity to meet again once the game starts. In this case, we need only to consider negotiation-proof agreements. If on-going negotiation is exercised, then players negotiate prior to the start of every subgame; that is, players *renegotiate after every history of play.* In this case, agreements has to be "renegotiation-proof". Such a distinction is important particularly from the view point of a single player. Renegotiation-proofness entails that every player, in contemplating a deviation, is *certain* that all players *will* meet and renegotiate after his deviation; in fact, he believes negotiation will occur after any deviation by any player (or coalition) in any future period. If for whatever reason a player is uncertain whether renegotiation will take place after a unilateral deviation and is averse to such a uncertainty, then negotiation-proofness may well be relevant.[19]

Negotiation-proofness for extensive form games can be defined in the same

[18]Such is the case whenever a game has multiple (weakly) negotiation-proof equilibria. See Subsection 3.5.4.

[19]In the context of repeated games, the consideration that renegotiation might not take place after every history appears, for example, in Pearce (1987), Bergin and MacLeod (1993), and Chapter 4 of this dissertation.

fashion as negotiation-proofness for normal form games except that for extensive form games only "meaningful" agreements have to be subgame perfect equilibria. Let SPE denote the set of subgame perfect equilibria for $\mathcal{G}$ and let $\Pi_{SPE} \equiv \{\alpha \mid f(\alpha) \in SPE\}$.

Definition 3.7. Let Σ be an abstract stable set for the abstract system $(\Pi_{SPE}, \ll)$; then the set of Weakly Negotiation-Proof Nash Equilibria (WNPNE's) of $\mathcal{G}$ is given by

$$W_\Sigma \equiv \{x \in SPE \mid \exists \alpha \in \Sigma \text{ such that } x = f(\alpha)\}.$$

Consider the following game [from Bernheim et al. (1987)] repeated twice without discounting.

TABLE 3.9

	L	C	R
U	5,5	0,6	0,0
M	6,0	4,4	0,0
D	0,0	0,0	2,2

There exists a unique WNPNE: In the first period, players choose (U, L); the second period play is (D, R) if any player deviates in the first period and (M, C) otherwise. The equilibrium payoffs are $(9, 9)$.

Renegotiation-proofness entails that renegotiation precedes *every* subgame. For extensive form games with finite number of stages, we can use a simple recursive definition as in, for example, Bernheim and Ray (1989) and Ferreira (1996).

Definition 3.8.

(1) For a single stage game $\mathcal{G}$, $x \in Z$ is renegotiation-proof if and only if it is a WNPNE.

(2) Let $t > 1$. Assume that renegotiation-proof equilibrium has been defined for all games with less than t stages. Then for any game $\mathcal{G}$ with

t stages, $x \in Z$ is renegotiation-proof if and only if x is a WNPNE for $\bar{\mathcal{G}} \equiv (N, \bar{Z}, \{u_i\}_{i \in N})$, where

$$\bar{Z} = \{x \in Z \mid \text{the restriction of } x \text{ to any proper subgame of } \mathcal{G} \text{ constitutes a WNPNE for that subgame}\}.$$

For the example in Table 3.9, the unique WNPNE is not renegotiation-proof: player 1, say, will deviate in the first period by playing M, being certain that in the second period player 2 will join him to renegotiate and abandon the punishment equilibrium (D, R) for (M, C).[20] The unique renegotiation-proof equilibrium is to repeat (M, C), which yield payoffs of $(8, 8)$.

Renegotiation-proof equilibrium exists for finite games.

Proposition 3.5. *Let $\mathcal{G}$ be a finite game in extensive form. Then there exists a renegotiation-proof equilibrium. Moreover, every negotiation-proof equilibrium is subgame perfect.*

3.5 Discussion

3.5.1 CPNE and the Nestedness Restriction

One of the motivations to define NPNE and WNPNE is to resolve the nestedness restriction and the myopia embedded in the definition of CPNE. I first discuss briefly several notions in the literature that attempt to relax the nestedness restriction.

Recall, first, the following definition of CPNE using von Neumann and Morgenstern abstract stable set (Greenberg 1989 and 1990). For a game $\mathcal{G}$, let

$$D \equiv \{(S, x) \mid S \subset N \text{ and } x \in Z\},$$

and for (S, x) and (T, y) in D,

$$(S, x) \angle (T, y) \iff T \subset S, x_{N \setminus T} = y_{N \setminus T}, \text{ and } x \prec_S y.$$

[20] If the game is repeated more than twice, each player is certain that players 1 and 2 will renegotiate *every time* he or his opponent deviates.

Theorem 3.5 (Greenberg 1989). *Let K be an abstract stable set for $(D, \angle)$. Then the set of CPNE's is given by $\{x \mid (N, x) \in K\}$.*

The nestedness restriction is evident in the above definition. This (nestedness) restriction can be relaxed in several ways, depending on whether the agreements of a deviating coalition are common knowledge [see Greenberg (1994)]. In the "coalition contingent threat situation" (Greenberg 1990), each deviation is made publicly (and is hence common knowledge) and further deviations are not restricted to subcoalitions. This negotiation process is delineated by a dominance relation on Z[21] defined as follows.

$$x \angle' y \iff \exists S \subset N, \text{ such that } x_{N\setminus S} = y_{N\setminus S}, \text{ and } x \prec_S y.$$

The abstract stable set for the abstract system $(Z, \angle')$ consists of those and only those agreements that players, who may be myopic, can reach in open pre-play negotiation. Such an abstract stable set may contain strategy profiles that are not Nash equilibria [see Greenberg (1990)], in which case, it is necessary to enforce, via binding contracts, these agreements, or to assume that $\mathcal{G}$ is not played as an one-shot noncooperative game.

Arce M. (1994) argued that "coalition building" often occurs in political situations; that is, new members are added efficiently to an existing coalition so that the final outcome benefits all members of the new coalition. Therefore, the nestedness restriction of CPNE is "inverted". This implies that cooperation becomes possible in prisoner's dilemma, since once a coalition forms, it will never break.

The nestedness assumption can also be relaxed under the assumption that the agreements of a deviating coalition are not common knowledge. Loosely speaking, the negotiation process underlying the definition of CPNE can be viewed as follows: A deviating coalition S, upon reaching an agreement among its members, leaves the scene of negotiation and members of S will never approach nonmembers. In Chakravorti and Kahn's (1993) definition of "universal coalition-proof

[21] Without the nestedness assumption it is sufficient to define the dominance relation on Z.

equilibrium", a subset of S, say T, is allowed to approach and attract some members that are not in S, say some $Q \subset N \setminus S$, in contemplating further deviations. Since Q is not aware of the previous agreement of S , Chakravorti and Kahn postulated that Q joins T only if any actions of $T \cup Q$ that hurt some member of Q will also hurt some member of T. Moreover, in defining their notion, Chakravorti and Kahn employed semi-stable set (Roth 1974) rather than (abstract) stable set used in this paper.

3.5.2 Agreements among Farsighted Players

Study of agreements among farsighted players in strategic environments can be found, for example, in Chwe (1994) and Xue (1995), where a strategic form game is a special case of the model [introduced in Chwe (1994)] they analyzed. Chwe (1994) formalized Harsanyi's (1974) "indirect dominance" in an attempt to capture foresight. For a normal form game, a strategy profile y is said to indirectly dominate another strategy profile x if y can be reached from x through a succession of deviations, and at each "stage", the deviating coalition prefers y to the agreement from which it deviates. Thus, this indirect dominance is defined on the set of strategy profiles. Based on such an indirect dominance, Chwe (1994) defined "the largest consistent set (LCS)" and applied it to the negotiation processes underlying the "coalitional contingent threat situation" and CPNE. In both cases, LCS may involve agreements that are not Nash equilibria. Moreover, the implicit behavior assumption [See Xue (1995)] underlying the LCS is different from the one embedded in the notion of abstract stable set that has been used to define NPNE and WNPNE. Chwe (1995) also applied LCS to open pre-play negotiation but assumed that players only consider Nash equilibrium strategies in the negotiation; while in this paper a coalition may deviate to an agreement that is not necessarily a Nash equilibrium, as long as such a deviation will eventually lead to some final agreement (necessarily a Nash equilibrium) that benefit all its members. Furthermore, Xue (1995) showed that indirect dominance captures only

partial foresight in that it ignores deviations on the way to the final agreements. Xue (1995) offered a formalization of *perfect foresight* by considering the "paths" of deviations. The notions of NPNE and WNPNE are built on this formalization of perfect foresight.

3.5.3 Correlated Strategies

CPNE can be extended to allow for correlated strategies. In Moreno and Wooders (1994), for example, a correlation device (or mediator) is available every time a coalition forms, and a coalitional deviation is carried out through such a correlation device. In their notion of "coalition-proof correlated equilibrium", self-enforceability of a deviation resembles that of CPNE. Correlated strategies can also be introduced to the pre-play negotiation analyzed in this paper, and then players bargain to determine which correlated equilibria are negotiation-proof. If a correlated equilibrium is negotiation-proof, then this equilibrium is implemented by the corresponding correlation device that makes a private recommendation to each player.

3.5.4 Concluding Remarks

To model pre-play communication is no doubt a task of great difficulty; this difficulty is magnified only by the restrictive framework of dynamic games. Instead of modeling how messages are interchanged among the players, this paper offers a model of pre-play communication in which players negotiate openly and directly. I assume that communication admits the possibility of *coalition formation* in that any group of players can coordinate their choice of strategies, thereby making joint objections in the negotiation. I set aside the details of communication that lead to the formation of a coalition; instead, I assume that every coalition can form and exploit rationality of the self-interested players to ascertain which coalitions will actually form (or "survive"), thereby fully capturing noncooperative behavior intrinsic to a noncooperative game. Moreover, a strategy profile x is not negotiation-proof as long as there exists one coalition of rational and self-

interested players who will ultimately benefit by making an objection to x. Thus, it is not necessary to stipulate that only a particular coalition or player can object to x.[22]

As the analysis in this paper indicates, pre-play negotiation does not necessarily pin down the exact equilibrium to be played. If a game has multiple negotiation-proof equilibria, one solution might be to randomize (through the use of a correlation device or mediator) with equal probability among these equilibria, on the ground that pre-play negotiation has exhausted all the "bargaining power" embedded in the structure of the game, and hence all negotiation-proof equilibria are equally "plausible". This is similar to the idea that a exogenous rule is employed to break a tie. Such a solution is obviously more prescriptive than descriptive. Alternatively, one may argue that communication does not offer a compelling justification for equilibrium analysis when multiple negotiation-proof equilibria arise [see also Rabin (1994)], and resort to weak solution concept like *rationalizability*. If one does insist on equilibrium analysis, it might be necessary to consider a equilibrium selection procedure such as the one proposed by Harsanyi and Selten (1988), who used an evolutionary process to identify a unique equilibrium.

APPENDIX

Proof of Lemma 3.1. Assume otherwise that $\alpha \in \Sigma(\Pi_\alpha)$, but $\alpha|_x \notin \Sigma$ for some $x \in \alpha$. By external stability of Σ, there exists $\beta \in \Sigma$ such that t $\alpha|_x < \beta$. By Definition 3.2, $\alpha < \beta$, contradicting the internal stability of Σ. ∎

Proof of Proposition 3.2. Since $<$ is acyclic, by a theorem of von Neumann and Morgenstern (1947), $(\Pi_{NE}, <)$ admits a unique abstract stable set Σ. By external stability, $\Sigma \neq \emptyset$. Therefore, the set of NPNE's is nonempty and uniquely defined. ∎

[22]This is one of the important features that distinguish the graph from an extensive from game.

Proof of Corollary 3.3. Since NE is finite and the set of Nash equilibria can be weakly Pareto ranked, $<$ is strictly acyclic. Then, it follows from Proposition 3.2 that the set of NPNE's is uniquely defined and nonempty.

Let x be the Pareto efficient Nash equilibrium within NE and $\Sigma \equiv \{\alpha \in \Pi_{NE} \mid f(\alpha) = x\}$. Then, $\beta \in \Pi_{NE} \setminus \Sigma$ if and only if $f(\beta) \prec_N x$. Since $f(\beta) \rightarrow_N x$, $\beta < x$. Therefore, $\beta \in \Pi_{NE} \setminus \Sigma$ if and only if $\beta < x$. But $x \in \Sigma$. Hence Σ is stable for $(\Pi_{NE}, <)$. Uniqueness follows from the fact that Σ must be contained in any stable set, since for all $\alpha \in \Sigma$, there does not exists $\beta \in \Pi_{NE}$ such that $\alpha < \beta$. ∎

Proof of Proposition 3.4. We need only to show that $(\Pi_{NE}, \ll)$ admits a unique abstract stable set. By a theorem of von Neumann and Morgenstern (1947), it suffices to show that $\ll$ is strictly acyclic. That is, there does not exist an infinite sequence of paths $\alpha^1, \alpha^2, \ldots$ in Π_{NE} such that $\alpha^j \ll \alpha^{j+1}$ for all $j = 1, 2 \ldots$. Indeed, let $\alpha^1, \alpha^2, \ldots$ be a sequence of paths in Π_{NE} such that $\alpha^j \ll \alpha^{j+1}$ for all $j = 1, 2 \ldots$. I claim that $i < j$ implies that $\alpha^i \neq \alpha^j$. Otherwise, $\alpha^i \ll \alpha^{i+1} \ll \cdots \ll \alpha^j = \alpha^i$; hence $\alpha^i < \alpha^{i+1} < \cdots < \alpha^j = \alpha^i$. Then, by Definition 5, $\alpha^i \not\ll \alpha^{i+1}$. A contradiction. Π_{NE} is finite since NE is finite. Thus, $\alpha^1, \alpha^2, \ldots$ must be a finite sequence and hence $\ll$ is strictly acyclic. ∎

Proof of Proposition 3.5. Since $\mathcal{G}$ is finite, recursively applying Proposition 3.4 yields the existence of renegotiation-proof equilibrium. The second assertion follows from Definition 3.8 and the "one-stage deviation principle". ∎

Chapter 4

Self-enforcing Agreements in Infinitely Repeated Games

This chapter defines the notion of "stable (self-enforcing) agreements" in infinitely repeated games where players can coordinate their actions but cannot make binding contracts. It differs from *renegotiation proofness* in that it allows for any coalition to deviate, and moreover, a deviating coalition does not count on renegotiating with nonmembers. In addition to its intuitive appeal, stable agreements can resolve the conflict between efficiency and renegotiation: the set of stable agreements is nonempty and efficient (within the set of subgame perfect equilibrium outcomes) for a large class of games including all two-player games and all games for which every efficient subgame perfect equilibrium path is stationary.

4.1 Introduction

The theory of repeated games has succeeded in explaining cooperation through long-term interactions: a cooperative outcome can be supported by a subgame perfect equilibrium of an infinitely repeated game. Thus, cooperation can be achieved through self-enforcing agreements, provided that only unilateral deviations are considered. However, this very "folk theorem" asserts that, in general, any feasible and individually rational payoff vector can be supported by a subgame perfect equilibrium [see, e.g., Fudenberg and Maskin (1986)]. In particular, many Pareto inferior payoffs can be supported by subgame perfect equilibria. Thus, repetition allows for, but by no means singles out, cooperative outcomes.

The literature on renegotiation-proofness in infinitely repeated games [see, e.g., Bernheim and Ray (1989), Farrell and Maskin (1989), Asheim (1991)] attempts to refine the set of subgame perfect equilibria by assuming that the grand coalition (and only the grand coalition) has the opportunity to negotiate anew (out of a "bad" equilibrium) after *every* history. An immediate question that arises is why

are deviations restricted to single players or else the grand coalition? In addition, why is it that only the grand coalition can renegotiate? While both individual rationality and Pareto optimality are important, coalitional rationality should also be considered. There is another, and more subtle reason to object to the notion of renegotiation-proofness. As I shall shortly illustrate, it entails that the grand coalition *must* renegotiate after *every* history. This is a very demanding assumption: Each player, in contemplating a deviation, is certain that the grand coalition will *necessarily* form to renegotiate. In particular, the grand coalition might renegotiate toward the very (cooperative) outcome from which deviations will occur, precisely because of the imposition of renegotiation. This is illustrated through the example in Table 4.1 taken from Asheim (1991).

TABLE 4.1

	a_2	b_2	c_2	d_2
a_1	3,3	-5,-5	-5,-5	-5,4
b_1	-5,-5	1,2	-5,-5	-5,3
c_1	-5,-5	-5,-5	2,1	-5,2
d_1	4,-5	2,-5	3,-5	0,0

Suppose that this game is repeated infinitely many times and the discount factor is 0.5. Let π denote the path of the infinite repetition of (a_1, a_2). Note that π can only be supported by a subgame perfect equilibrium with Pareto inferior punishments: Player 1's deviation (to d_1) is punished by π_1, the path of playing (b_1, b_2) for one period and then reverting to π. Similarly, Player 2's deviation (to d_2) is punished by π_2, the path of playing (c_1, c_2) for one period and then reverting to π .[23] These punishments are subject to "renegotiation": Player 1, for example, can deviate to d_1, because he realizes that at the next period players 1 and 2 *will definitely* renegotiate in order to avoid the Pareto inferior path π_1. It follows

[23]If, in addition, player $i \in \{1,2\}$ deviates from π_j, $j \in \{1,2\}$, π_i restarts. This specifies a *simple strategy profile* in the sense of Abreu (1988).

that π is not supported by a renegotiation-proof equilibrium. In fact, the only subgame perfect equilibrium that is not subject to renegotiation (see Section 3) is the infinite repetition of the Nash equilibrium of the stage game, (d_1, d_2), which yields each player the lowest payoff within the set of subgame perfect equilibria.

As our discussion above demonstrates, the imposition of renegotiation after every history may well be implausible: After a player deviates, the grand coalition renegotiates only to find itself in the same position in the next stage. When this occurs, the deviating player can no longer count on renegotiation. The model presented in this chapter captures, among other things, this phenomenon: deviating players cannot count on renegotiating with the rest of society. More specifically, my model builds on the following three ingredients: First, I allow every coalition, not only single players and the grand coalition, to deviate. Second, a deviating coalition (or player) believes that other players will not be willing to renegotiate. This is captured formally by assuming that nonmembers of the deviating coalition will partition themselves into singletons (and thus, will not be able to correlate their actions). Thus, a coalition bases its deviations on what it can "enforce" by solely coordinating the actions of its own members.[24] The third ingredient of the analysis in this chapter is that players are assumed to be "conservative" or "uncertainty averse" in the sense that they always fear the worst outcome, from the set of "plausible" outcomes. I then define the notion of "stable (self-enforcing) agreements".

When applied to the example in Table 4.1, the only stable agreement is π. Indeed, a single player, in contemplating a deviation from the cooperative outcome π, realizes that the other player will *not* be willing to renegotiate. Therefore, a single player, by acting alone, cannot avoid any punishment (subgame perfect) equilibrium, and hence will not deviate from π. The grand coalition $\{1, 2\}$, on

[24] There are, of course, other possibilities that may well be worth pursuing. What I point out in this chapter is that the existing literature on "renegotiation proofness" should be examined more carefully, and that the results of such an examination may be encouraging.

the other hand, will deviate from any Pareto inferior outcome, in particular, the infinite repetition of the Nash equilibrium of the stage game (d_1, d_2), since its members can jointly "enforce" π, from which, as argued above, no single player will deviate.

The organization of the rest of this chapter is as follows: In Section 4.2 I formalize the notion of "stable (self-enforcing) agreements", which incorporates both coalitional rationality and dynamic consistency. I also investigate some properties of stable agreements. In Section 4.3 the notion of stable agreements is related to several notions in the literature including renegotiation-proofness, perfectly coalition proof Nash equilibrium, and the β-core. All proofs are relegated to the appendix.

4.2 Self-enforcing Agreements

In this section I formally define the notion of "stable (self-enforcing) agreements" if coalitions can form and no binding agreements can be signed.

Consider a (stage) normal form game $\mathcal{G} = (N, \{A_i\}_{i\in N}, \{u_i\}_{i\in N})$, where N is the finite set of players, A_i is the action set of player $i \in N$, and $u_i : A \to \Re$ is the payoff function of player $i \in N$, where $A = \prod_{i\in N} A_i$. For every $i \in N$, A_i is assumed to be compact and u_i continuous. Let $\mathcal{G}^\infty$ denote the infinite repetition of $\mathcal{G}$ and let Π denote the set of paths (action profiles), i.e., $\Pi = A^\infty$. All players discount future payoffs using the same discount factor $\delta \in (0,1)$. Thus, player i's (normalized discounted) payoff from $\alpha = (\alpha^1, \alpha^2, \ldots) \in \Pi$ is

$$U_i(\alpha) = (1-\delta)\sum_{t=1}^{\infty} \delta^t u_i(\alpha^t).$$

Let $H = \cup_{\tau=0}^{\infty} A^\tau$, where $A^0 \equiv \emptyset$, be the set of all histories. A (pure) strategy for $i \in N$ is a mapping $f_i \colon H \to A_i$.

A stable agreement (for N) is a path in Π from which no coalition $S \subset N$ would wish to deviate.[25] Let PEP denote the set of perfect equilibrium paths, a

[25] All inclusions in this chapter are weak.

set that is assumed to be nonempty. For $\alpha \in PEP$ and $\tau \geq 1$, let $\alpha|_\tau$ denote the continuation of α from τ (including τ) on. Suppose a path $\alpha \in PEP$ is considered by the grand coalition N. A coalition $S \subset N$, in contemplating a deviation from α at some period $\tau \geq 1$, has to compare $\alpha|_\tau$ with "the set of paths that are likely to occur were S to deviate".[26] Denote this set by $\sigma(S \mid \alpha, \tau)$. Paths that do not belong to $\sigma(S \mid \alpha, \tau)$ are considered "implausible" continuations should S deviate from α at stage τ.

To make the analysis more tractable and in view of the fact that all continuations of the game, from any history, are isomorphic to $\mathcal{G}^\infty$, I assume that for all $\alpha, \alpha' \in PEP$ and all $\tau, \tau' \geq 1$, and for every $S \subset N$, $\sigma(S \mid \alpha, \tau) = \sigma(S \mid \alpha', \tau')$. That is, the mapping σ is assumed to be independent of histories. While this is, certainly, a restrictive assumption, it is weaker than that of stationarity in the literature of renegotiation-proofness, because the latter entails that such a mapping is also independent of S (see Section 3).

Definition 4.1. A standard of behavior (or norm) is a mapping σ that assigns to every $S \subset N$ a subset of PEP.

Following Greenberg (1990), I shall require that the standard of behavior σ be "stable": σ must be free of inner contradictions, i.e., "internally stable" and at the same time must account for every path it excludes, i.e., "externally stable". As emphasized above, I assume that when a coalition S deviates its members believe that non-members will partition themselves into singletons. Therefore, in determining $\sigma(S)$, S considers only further deviations of two forms: either by a single individual in $N \setminus S$, or, by *subsets of* S.[27] Thus, within our context, "internal stability" stipulates that for all $S \subset N$, if $\alpha \in \sigma(S)$ then there do not exist a coalition $T \subset S$ or $T = \{j\}$ for some $j \in N \setminus S$ and a stage $\tau \geq 1$ such that "T prefers $\sigma(T)$ to $\alpha|_\tau$." And, "external stability" stipulates that for all $S \subset N$,

[26] An important part of the analysis that follows concerns the nature of this set.

[27] Recall that agreements are non-binding.

if $\alpha \in PEP \setminus \sigma(S)$ then there exist coalition $T \subset S$ or $T = \{j\}$ for some $j \in N \setminus S$ and a stage $\tau \geq 1$ such that "T prefers $\sigma(T)$ to $\alpha|_\tau$." The standard of behavior σ is stable if it is both internally and externally stable.

To formally define stability, we must first be precise on the meaning of "T prefers $\sigma(T)$ to $\alpha|_\tau$", or, more generally, on the meaning of "T prefers A to α" where A is a subset of PEP and α belongs to PEP. To motivate our definition of this preference relation (between a single path and a set of paths), consider the infinite repetition of the following 3-player game,

TABLE 4.2

	ℓ	r
u	9,0,1	0,0,2
d	0,0,1	0,9,1

A

	ℓ	r
u	9,0,1	0,0,2
d	0,0,1	0,9,1

B

where player 1 chooses rows, player 2 chooses columns, and player 3 chooses matrices. I claim that the coalition $T = \{1, 2\}$ "prefers A to α", where $A = PEP$ and α is the agreement (in A) that results from "repeating (u, r, L) forever". Indeed, observe that players 1 and 2 can coordinate their actions in the following way: After any history, play (d, ℓ) if player 3 is currently playing[28] R; play (u, ℓ) if player 3 is currently playing L and if (u, ℓ) has been played no more times than (d, r); otherwise play (d, r). By using these coordinated actions, both players 1 and 2 would be better off than they are under α from *any* path that might result. It is important to note that players 1 and 2 can only coordinate their *own* actions, and such a coordination might not suffice to define a unique path. Indeed, player 3's choice is not determined. He can ("rationally") choose, at each stage, either L or R. But, *no matter how* player 3 would play, by coordinating their actions,

[28]Since α is an agreement, the actions of players 1 and 2 may depend also on the current action of player 3.

players 1 and 2 would be better-off than they are under α.[29]

It is precisely this reasoning that underlines the following definition.

Definition 4.2. Let $T \subset N$, $A \subset PEP$, and $\alpha \in PEP$. We say that T *prefers* A *to* α if there exists a set $B \subset A$ such that

(i) $\beta \in B \Longleftrightarrow \nexists \eta \in B$ such that for some $\tau \geq 1$, $\beta^t = \eta^t$ for all $t < \tau$, $\beta^\tau_{-T} = \eta^\tau_{-T}$, and $\beta^\tau_T \neq \eta^\tau_T$;

(ii) $U_i(\alpha) < U_i(\beta)$ for all $\beta \in B$ and for all $i \in T$.

Condition (i) captures the fact that members of T can coordinate their actions, and condition (ii) captures the fact that the set of paths that respect this coordination and the original set A, are preferred by each (conservative) player in T over the agreement α.

Definition 4.3. A standard of behavior σ is stable if for all $S \subset N$, $\alpha \in PEP \setminus \sigma(S) \Longleftrightarrow \exists \tau \geq 1, T \subset S$ or $T = \{j\}$ for some $j \in N \setminus T$, such that T prefers $\sigma(S)$ to $\alpha|_\tau$.

Let σ be a stable standard of behavior. The set $\sigma(N)$ is called the set of *stable* (or *self-enforcing*) *agreements*. The set of stable agreements, $\sigma(N)$, captures coalitional rationality and dynamic consistency: it contains those and only those agreements in PEP that are not rejected by any $S \subset N$ whose members are aware of and believe in σ, and realize that subsets of S or single players may pursue further deviations, and that any coalition that further deviates goes through the same reasoning. The reader is invited to verify that in the example in Table 4.1, this set consists of the unique (Pareto) efficient perfect equilibrium path (PEP), i.e., $\sigma(N) = \{\pi\}$.

It is evident that our definition of a stable standard of behavior is inspired by the theory of social situations. Indeed, (see, e.g., Greenberg (1989)) the standard

[29] This example also illustrates the importance of considering all coalitional deviations, since the grand coalition is not able to improve upon, for example, the infinite repetition of (u, r, L).

of behavior that assigns to each subgame the set PEP is "conservatively stable" in the sense that

$$\alpha \in \Pi \setminus PEP \Longleftrightarrow \exists \tau \geq 1,\ i \in N,\ \text{and } b_i \in A_i \text{ s.t. } U_i(\alpha|_\tau) < U_i((b_i, \alpha^\tau_{-i}), \beta),$$
$$\forall \beta \in PEP.$$

Our definition builds on this result and extends it to allow for coalitions to deviate. This extension complicates the analysis in two ways. When deviations are restricted to single players, (1) it is sufficient to consider only one stage deviation (because $\mathcal{G}^\infty$ is continuous at infinity), and (2) the set of actions a deviating player $i \in N$ can take is not restricted; it can be any element b_i in A_i. Neither (1) nor (2) remain valid when we extend, as we do, the analysis to allow for coalitions to deviate. Indeed, the example in Table 4.2 above illustrates that when coalitions can form, the "one-stage deviation principle" does not hold; we must allow coalitions to coordinate their actions in several (or infinite) stages. A coordinated one-stage deviation may not suffice to make every member better-off.

Now I shall illustrate, through another example, that a coalition $S \subset N$ cannot base its deviation on arbitrary actions in $A_S = \prod_{i \in S} A_i$. Consider the infinite repetition of the prisoner's dilemma in Table 4.3 and assume $\delta = 0.4$. It is easy to verify that the cooperative outcome of the infinite repetition of (d, ℓ) cannot be supported by a subgame perfect equilibrium. In fact, the unique PEP, π, is to repeat (u, r) infinitely. Let σ be a standard of behavior such that $\sigma(S) = \pi$, for all $S \subset \{1, 2\}$. Were we to consider the joint deviation of $\{1, 2\}$ from π at some period to (d, ℓ), π would be ruled out. Such a deviation, however, cannot be carried out without a binding agreement, since itself is subject to individual deviations. To insist upon the self-enforceability of all agreements, a *coalitional* deviation has to be consistent with σ. Therefore, π, the unique PEP, will be the unique stable (self-enforcing) agreement.

Now I proceed to investigate some properties of stable agreements. The following lemma indicates that individual deviations (in Definition 4.3) do not rule out

TABLE 4.3

	ℓ	r
u	3,0	1,1
d	2,2	0,3

any PEP. This assertion holds even if the deviating player can choose an arbitrary action at the period of deviation. This is in contrast to the notion of renegotiation proofness, which entails that a deviating player, believing that the grand coalition will necessarily renegotiate, may deviate from a PEP (see Sections 4.1 and 4.3).

Lemma 4.1. *Let $\alpha \in PEP$. Then there do not exist $\tau \geq 1$ and $i \in N$ such that i prefers PEP to $\alpha|_\tau$. In fact, $\alpha \in PEP$ if and only if there do not exist $\tau \geq 1$, $i \in N$, and $b_i \in A_i$ such that i prefers $\{b\} \times PEP$ to $\alpha|_\tau$, where $b = (b_i, \alpha^\tau_{-i})$.*

Therefore, to derive PEP rather than impose it, we can modify Definition 4.3 by maintaining that a deviating player $i \in N$ can choose arbitrary action in A_i at the period of deviation.

The following proposition states that the stable standard of behavior exists[30] and stable agreements are efficient within PEP if every efficient path $\alpha \in PEP$ is "stationary", i.e., $\alpha = (a, a, a, \dots)$ for some $a \in A$, as was the case for the example in Table 4.1.

Proposition 4.2. *There exists a stable standard of behavior σ such that $\sigma(\{i\}) = PEP$, for all $i \in N$. Moreover, if every efficient PEP is "stationary", then every stable agreement in $\sigma(N)$ is efficient (within PEP).*

The following lemmas provide sufficient conditions on the stage game to guarantee the nonemptiness of $\sigma(N)$. Again, the example in Table 4.1 satisfies each condition.

[30] Recall that for every $i \in N$, A_i is compact, u_i is continuous, and hence the set of PEP is compact [see Abreu (1988)].

Lemma 4.3. *If $|N| = 2$, then $\sigma(N)$ coincides with the efficient frontier of PEP.*

Lemma 4.4. *If PEP admits a unique efficient path α, then $\sigma(N) = \{\alpha\}$.*

Remark 4.1. More general sufficient condition to guarantee the nonemptiness of $\sigma(N)$ is yet to be established. I have been unable to find a counter example such that $\sigma(N) = \emptyset$. Even if for some game $\sigma(N) = \emptyset$, i.e., the grand coalition cannot reach any self-enforcing agreement, a stable standard of behavior σ can still be useful: By external stability σ cannot be empty valued (i.e., there exists $S \subset N$ such that $\sigma(S) \neq \emptyset$), therefore, σ will "predict" the coalitions that are likely to form.

4.3 Related Literature

4.3.1 Renegotiation Proofness.

The notion of *stable agreements* is motivated by the difficulties in the renegotiation proofness literature, and defined by applying the notion of *stability* and "the theory of social situations" (Greenberg 1990). Application of the notion of stability and the theory of social situations to repeated games can be found in Greenberg (1989) and Asheim (1991). In this subsection, I shall provide a brief review of several theories of renegotiation-proofness that exhibit different attempts to improve the notion of renegotiation-proofness, and show that existing difficulties cannot be resolved under the assumption of renegotiation. The notion of stability and the theory of social situation, again, provide a common framework for our discussion.

Definition 4.4. Let $\Sigma \subset$ PEP. Then Σ is[31]

(i) *internally R-stable* if for every $\alpha \in \Sigma$, (1) there does not exist $\tau \geq 1$ such

[31]Note that " N prefers Σ to $\alpha|_\tau$" is equivalent to "there exists $\beta \in \Sigma$ such that $U_i(\beta) > U_i(\alpha|_\tau)$ for all $i \in N$". To facilitate comparison with the definition of stable agreements in the previous section, we retain the same notations as in the previous section. Also, I distinguish individual deviations from the deviations of the grand coalition to allow a deviating player to choose any $b_i \in A_i$ at period τ. Such a distinction is unnecessary in the definition of stable agreements, in view of Lemma 4.1.

that N prefers Σ to $\alpha|_\tau$ and (2) there do not exist $\tau \geq 1$, $i \in N$ and $b_i \in A_i$ such that i prefers $\{b\} \times \Sigma$ to $\alpha|_\tau$, where $b = (b_i, \alpha^\tau_{-i})$;

(ii) *externally R-stable* if for every $\alpha \in PEP \setminus \Sigma$, (1) there exists $\tau \geq 1$ such that N prefers Σ to $\alpha|_\tau$ or (2) there exist $\tau \geq 1$, $i \in N$ and $b_i \in A_i$ such that i prefers $\{b\} \times \Sigma$ to $\alpha|_\tau$, where $b = (b_i, \alpha^\tau_{-i})$.

I label the stability notion in the above definition R-stability for its relation to renegotiation-proofness, which considers only deviations of single players and the grand coalition and imposes renegotiation after every history. Note the stationarity of Σ: Σ is independent of either histories or deviating coalitions (singletons or the grand coalition). Two equivalent notions of renegotiation-proofness, *weakly renegotiation-proofness* (*WRP*) (Farrell and Maskin, 1987) and *internal consistency* (*IC*) (Bernheim and Ray, 1989), can be defined as follows.

Definition 4.5. Σ is internally R-stable if and only if $\{x \in \Re^N \mid x_i = U_i(\alpha), \alpha \in \Sigma\}$ is weakly renegotiation-proof (internally consistent).

Therefore, WRP and IC test only for internal R-stability. In general, internally R-stable set need not be unique and one internally R-stable set may contain a path that is Pareto dominated by some path in another internally R-stable set. To solve this problem, Farrell and Maskin (1987) proposed *strong renegotiation-proofness* (*SRP*). Let Σ and Σ' be internally R-stable. Then Σ is "dominated" by Σ', denoted $\Sigma < \Sigma'$, if there exist $\alpha \in \Sigma$ and $\beta \in \Sigma'$ such that α is Pareto dominated by β.

Definition 4.6. If Σ is internally R-stable and there does not exist another internally R-stable set Σ' such that $\Sigma < \Sigma'$, then $\{x \in \Re^N \mid x_i = U_i(\alpha), \alpha \in \Sigma\}$ is strongly renegotiation-proof.

However, the criterion in the notion of *SRP* is too demanding of a "candidate" set Σ. As a result, a SRP set may not exist. Bernheim and Ray (1989) insisted that the "challenging" set Σ' should itself not be subject to such challenges, and

propose the notion of *consistent set*. Let Σ and Σ' be internally R-stable. Then Σ is "indirectly dominated" by Σ', denoted $\Sigma \ll \Sigma'$, if there exist a finite sequence of internally R-stable sets, $\Sigma_1, \ldots, \Sigma_k$ such that $\Sigma < \Sigma_1 < \cdots < \Sigma_k < \Sigma'$.

Definition 4.7. Let Σ be an internally R-stable set such that $\Sigma' \ll \Sigma$ for every internally R-stable set Σ' such that $\Sigma \ll \Sigma'$. Then, $\{x \in \Re^N \mid x_i = U_i(\alpha), \alpha \in \Sigma\}$ is consistent.

The notions of SRP and consistent set, however, may, on one hand, eliminate subgame perfect equilibria that are defeated only by equilibria that are themselves not viable [see, e.g., Asheim (1991), Bergin and MacLeod (1993)], and, on the other hand, may fail to account for every path they exclude. To resolve these issues, the notion of *Pareto perfect equilibrium* (*PPE*) in infinitely repeated games (Asheim, 1991) – the extension of Pareto perfection in finitely repeated games (Bernheim and Ray, 1985) – insists on both internal and external R-stability. In particular, external R-stability implies that non-viable equilibria must be defeated by viable ones. Under the stationarity assumption, Pareto perfect equilibrium is defined as follows:

Definition 4.8. α is a Pareto perfect equilibrium (PPE) path if and only if Σ is both internally and externally R-stable and $\alpha \in \Sigma$.

However, the existence of a stable Σ is problematic even in simple two-player games: For the example in Table 4.1, the infinite repetition of the Nash equilibrium of the stage game, (d_1, d_2), constitutes the unique nonempty internally R-stable set. Thus, it is the only candidate for a Pareto perfect equilibrium path. But this path cannot account for the exclusion of other perfect equilibrium paths. Consequently, Pareto perfect equilibrium fails to exist. Relaxing the assumption of stationarity, existence of Pareto perfect equilibrium is restored [see Asheim (1991)]. But in this case, there are multiple Pareto perfect equilibria: infinite repetition of (a_1, a_2), (b_1, b_2), (c_1, c_2), or (d_1, d_2) can each be supported

by a Pareto perfect equilibrium. Asheim (1991) concluded that this is due to "the inherent difficulty of imposing renegotiation-proofness" in such a case.

The multiplicity of theories of renegotiation-proofness has demonstrated the attempts to improve the notion of renegotiation-proofness. The attempts discussed above, however, maintain the assumption that renegotiation occurs after *every* history. This implies, from a deviating player's point of view, that after his deviation, the grand coalition will *necessarily* form to renegotiate. This assumption may well be implausible as illustrated by the example in Table 4.1: A player deviates from π, counting that the grand coalition negotiates back to π itself; the grand coalition never realizes that it is its renegotiation back to π that encourages a single player to deviate from π. Moreover, the imposition of renegotiation after every history results in a conflict between efficiency and renegotiation. Indeed, the infinite repetition of the Nash equilibrium of the stage game, (d_1, d_2), is the unique WRP equilibrium and the unique SRP equilibrium, and it also gives rise to the unique consistent set. This equilibrium, however, is unanimously least preferred among the set of subgame perfect equilibria (SPE).

Pearce (1987) recognized that imposition of renegotiation after every history may be too strong: Cooperation requires punishments, so any theory of renegotiation should consider how renegotiation affects the sustainability of punishments. In Pearce's notion of renegotiation-proofness, renegotiation occurs only if the proposed equilibrium is as good as the original equilibrium in every subgame. In the case of Table 4.1, infinite repetition of (a_1, a_2), (b_1, b_2) and (c_1, c_2) can each be supported by Pearce's notion of renegotiation-proofness.

This chapter tackles the problem of renegotiation-proofness by insisting upon that a coalition must base its deviation on what it can "enforce" by solely coordinating the actions of its own members. Thus, a single player, when contemplating a deviation, cannot count on renegotiating with the rest of the players and has to

consider the worst (punishment) subgame perfect equilibrium,[32] while the grand coalition can base its deviation on any single PEP from which no proper subsets wish to deviate. The notion of stable agreement, by offering a way out of the conceptual difficulty of renegotiation-proofness, also resolves the conflict between efficiency and renegotiation: For two-player games, where the only coalition that can form is the grand coalition, we have

(1) $\sigma(N)$, the set of stable agreements, contains only those paths that are Pareto optimal within PEP. For the example in Table 4.1, $\sigma(N) = \{\pi\}$. But as the same example demonstrates, WRP, SRP, and consistency may select only Pareto inferior equilibria within the set of SPE's. Moreover, a PPE may also be inefficient [see Asheim (1991)].[33]

(2) The set of stable agreements is nonempty. But SRP equilibrium may fail to exist [see Bernheim and Ray (1989)]. Stationary Pareto perfect equilibrium may also fail to exist. This is the case in Table 4.1 where Pareto optimal payoffs can only be supported by Pareto inferior payoffs.

For games with more than two players, consideration of deviations of partial coalitions is important as demonstrated by the example in Table 4.2. The notion of stable agreements takes into consideration this important aspect of "coalitional rationality", which renegotiation-proofness fails to address.

4.3.2 Perfectly Coalition-Proof Nash Equilibrium and Strong Perfect Equilibrium.

Bernheim, Peleg, and Whinston (1987) applied their coalition proof Nash equilibrium to dynamic games with finite horizon and proposed the notion of *perfectly coalition-proof Nash equilibrium* (*PCPNE*). This definition was extended by

[32]Were he to assume that the grand coalition would necessarily renegotiate, he would more likely reject a (cooperative) path, as implied by Lemma 1 and Definition 4.

[33]A sufficient condition for SRP, consistency, and Pareto perfection to select only Pareto efficient equilibria within the set of SPE's is that any efficient payoff within the set of SPE payoffs can be supported by payoffs which are themselves efficient within the set of SPE payoffs [see Asheim (1991)]. The example in Table 4.1 violates this condition.

Asheim (1988) to dynamic games with infinite horizon [see also Asilis and Kahn (1992)]. Unlike renegotiation-proofness, PCPNE considers all coalitional deviations. It assumes, however, that after every history every coalition will form to "renegotiate". In particular, for two-player games, PCPNE coincides with Pareto perfect equilibrium (Asheim, 1991). This signifies the difference between PCPNE and the notion suggested in this chapter.

Rubinstein's (1980) *strong perfect equilibrium* is more demanding[34] in that it requires an equilibrium to survive all conceivable deviations, many of which are not credible. In particular, Pareto efficiency in the space of all feasible outcomes is imposed.

4.3.3 The β-core.

The β-core (Aumann, 1959) of the repeated game is the core of its β-characteristic function. Let X_i be the set of strategies of $i \in N$, i.e., $X_i = \{x_i \mid x_i : H \to A_i\}$. The β-characteristic function $v : N^2 \to \Re^N$ is given by: for all $S \subset N$,

$$v(S) = \bigcap_{x_{-S} \in X_S} \bigcup_{x_S \in X_S} \{u \in \Re^N \mid u_j \leq U_j(x_S, x_{-S}), \forall j \in S\}.$$

The β-core is the set of payoff vectors ζ in $v(N)$ for which there does not exist $S \subset N$ such that for some $\xi \in v(S)$, $\xi_i > \zeta_i$ for all $i \in S$. The similarity between our notion and β-core is that each coalition is certain about its ability to coordinate the actions of its members but has to consider all contingencies created by nonmembers. But the notion of stable agreements differs from the β-core in the following aspects.

(1) In determining $v(S)$, S has to consider the entire range of strategies of the members in $N \setminus S$, including, for example, dominated strategies of $N \setminus S$. In the definition of stable agreements, however, even though a deviating coalition assumes that no coalition that contains nonmembers will form to renegotiate, it does require individual rationality of nonmembers.

[34] A strong perfect equilibrium is always perfectly coalition proof.

(2) The definition of β-core does not consider the credibility or the self-enforceability of an "objection".[35] In the definition of stable agreements, a coalition S considers the possible "internal" deviations and therefore the credibility of an objection is verified.

(3) The β-core does not consider dynamic consistency. In fact, β-core is a notion in a static setting. Our notion captures the dynamic consistency at both individual level (every stable agreement is in PEP) and coalitional level.

4.4 Conclusion

In this chapter, I defined a notion of stable agreement in infinitely repeated games where players can coordinate their actions but cannot make binding contract. The notion of stable agreements is *not* a new definition of renegotiation-proofness; rather it is intended to serves as an alternative to the study of cooperation and equilibrium selection in repeated games through the notion of renegotiation-proofness. While it is interesting and instructive to test whether a subgame perfect equilibrium is renegotiation-proof, renegotiation-proofness is not and should not the only way of equilibrium selection or accounting for coordination in repeated games. Imposing renegotiation after every history is a very strong assumption, particularly from the viewpoint of single players. This motivates the study of stable agreements in this chapter. My definition of stable agreement is based on a pessimistic view of a deviating coalition: A deviating coalition, which is uncertain whether renegotiation will take place, consider the worst possibility that renegotiation might not occur; a coalition deviates only if it can guarantee its members higher payoffs by solely coordinate the actions of its members. That is, in contemplating its deviation, a coalition cannot confidently count on renegotiating with other players, although it is *possible* that renegotiation might actually occur after its deviation.

[35]This lack of "credibility" can be amended when N is finite, as was shown by Ray (1983) and Greenberg (1990).

Appendix

I first introduce the following notations to facilitate the proofs: For $A \subset PEP$ and $T \subset N$, let $\Phi_T(A)$ denote the subsets of A that satisfy conditions (i) and (ii) in Definition 2. Hence T prefers A to some $\alpha \in PEP$ if and only if there exists $\phi_T \in \Phi_T(A)$ such that for all $\beta \in \phi_T$, $U_i(\alpha) < U_i(\beta)$ for all $i \in T$.

Proof of Lemma 4.1. Otherwise, $\exists i \in N$ and $\phi_{\{i\}} \in \Phi[PEP]$ such that $U_i(\alpha^*) < U_i(\beta^*)$, where $U_i(\alpha^*) = \min_{\alpha \in PEP} U_i(\alpha)$ and $U_i(\beta^*) = \min_{\beta \in \phi_{\{i\}}} U_i(\beta)$. Since U_i is continuous at infinity, $\exists \bar{t} > 0$ such that $U_i(\alpha^*) < U_i(\overline{\beta^*})$, where $\overline{\beta^*} = (\beta^{*1}, \ldots, \beta^{*\bar{t}}, \alpha^*)$. Let τ be the smallest $\bar{t}$ such that the above holds. Now, consider the period $\tau - 1$, then $U_i(\alpha^*) < U_i(\beta^{*\tau-1}, \alpha^*)$, violating the stability of PEP (see page 7).

For the second assertion, "if" is obvious, in view of the stability of PEP given on page 7. The "only if" part follows from the proof of the first assertion (with minor modification). ∎

Proof of Proposition 4.2.

The proof of existence resembles Greenberg's (1990) results on the existence of OSSB in the hierarchical situation. For each $S \subset N$, define, recursively, two subsets of PEP, $A(S)$ and $B(S)$, as follows:

For all $i \in N$,

$$B(\{i\}) = PEP, \text{ and}$$
$$A(\{i\}) = \left\{ \begin{array}{c} \alpha \in B\{i\} \mid \exists \tau \geq 1, \phi_{\{i\}} \in \Phi_T[B(\{i\})] \text{ s.t.} \\ U_i(\alpha|_\tau) < U_i(\beta), \forall \beta \in \phi_{\{i\}}. \end{array} \right\}.$$

For $S \subset N$, assume that $A(T)$ and $B(T)$ are defined for all $T \subset S$ such that $T \neq S$. Define

$$B(S) = \left\{ \begin{array}{c} \alpha \in PEP \mid \exists \tau \geq 1, T \subset S \text{ with } T \neq S \text{ and } \phi_T \in \Phi_T[B(T) \setminus A(T)] \\ \text{s.t. for all } i \in T, U_i(\alpha|_\tau) < U_i(\beta), \forall \beta \in \phi_T. \end{array} \right\}$$

and

$$A(S) = \left\{ \begin{array}{c} \alpha \in B(S) \mid \exists \tau \geq 1 \text{ and } \phi_S \in \Phi_S[B(S)] \text{ s.t. for all } i \in T, \\ U_i(\alpha|_\tau) < U_i(\beta), \forall \beta \in \phi_S. \end{array} \right\}$$

I claim that $B(S)$ is compact for all $S \subset N$. Since PEP is a compact set, $B(\{i\})$ is compact. Therefore, for $|S| > 1$, it suffices to show that $B(S)$ is closed. Let $\{\alpha_j\}$ be a sequence of paths in $\beta(S)$ with $\alpha_j \to \alpha$, we need to show that $\alpha \in \beta(S)$. Otherwise,$\exists \tau \geq 1, T \subset S$ with $T \neq S$ and $\phi_T \in \Phi_T[B(T) \setminus A(T)]$ s.t. for all $i \in T, U_i(\alpha|_\tau) < U_i(\beta), \forall \beta \in \phi_T$. Since U_i is a continuous function for all $i \in N$, there exists J such that for all $j \geq J$, for all $i \in T, U_i(\alpha_j|_\tau) < U_i(\beta), \forall \beta \in \phi_T$. Then $\alpha_j \notin B(S)$. Contradiction.

Now, define

$$A^*(S) = \left\{ \begin{array}{c} \alpha \in B(S) \mid \exists \tau \geq 1 \text{ and } \phi_S \in \Phi_S[B(S) \setminus A(S)] \text{ s.t. for all } i \in T, \\ U_i(\alpha|_\tau) < U_i(\beta), \forall \beta \in \phi_S. \end{array} \right\}$$

I claim that $A(S) = A^*(S)$. I first show that $A(S) \subset A^*(S)$. Consider $\alpha \in A(S)$. Then $\exists \tau \geq 1$ and $\phi_S \in \Phi_S[B(S)]$ s.t. for all $i \in T$, $U_i(\alpha|_\tau) < U_i(\beta), \forall \beta \in \phi_S$. Since $B(S)$ is compact, $\exists \tau \geq 1$ and $\phi_S^* \in \Phi_S[B(S)]$ s.t. for all $i \in T$, $U_i(\alpha|_\tau) < U_i(\beta), \forall \beta \in \phi_S^*$ and $\beta \in \phi_S^*$ implies $\beta \notin A(S)$. Therefore $\alpha \in A^*(S)$. To show the converse inclusion, assume in negation that $\exists \alpha \in A^*(S) \setminus A(S)$. Then $\alpha \in A^*(S)$ implies that $\exists \tau \geq 1$ and $\phi_S \in \Phi[B(S) \setminus A(S)]$ s.t. for all $i \in T$ $U_i(\alpha|_\tau) < U_i(\beta), \forall \beta \in \phi_S$. Then $\exists \phi'(S) \in \Phi_S[B(S)]$ such that $\phi_S \subset \phi'_S$ and $\phi_S \neq \phi'_S$. Since $\alpha \notin A(S)$, $\forall \phi'(S) \in \Phi_S[B(S)]$ such that $\phi_S \subset \phi'_S$ and $\phi_S \neq \phi'_S, \exists \beta \in \phi'_S$ and $i \in S$ such that $U_i(\alpha) \not< U_i(\beta)$. If $\beta \notin A(S)$, contradiction, since $\beta \in B(S) \setminus A(S)$ and yet $\beta \notin \phi_S$. Otherwise $\beta \in A(S)$. Then, $\exists \tau \geq 1$ and $\phi_S^* \in \Phi_S[B(S)]$ s.t. for all $i \in T$, $U_i(\beta|_\tau) < U_i(\eta), \forall \eta \in \phi_S^*$ and $\eta \in \phi_S^*$ implies $\eta \notin A(S)$. Again we can replace β with some $\eta \in \phi_S^*$. If $\exists i \in S$ such

that $U_i(\alpha) \not< U_i(\eta)$, then $\eta \notin \phi_S$, which implies that η need not belong ϕ'_S. Contradiction.

Now, I shall claim that the standard of behavior σ given by $\sigma(S) = B(S) \setminus A(S)$ for all $S \subset N$ is stable.[36] Indeed, $\alpha \in PEP \setminus \sigma(S) \Leftrightarrow \alpha \in [PEP \setminus B(S)] \cup A(S) \Leftrightarrow \alpha \in [PEP \setminus B(S)] \cup A^*(S)$. By the definition of $B(S)$ and $A^*(S)$, σ is stable.

To prove the second assertion, assume in negation that there is $\alpha \in \sigma(N)$ such that for some $\beta \in PEP$, $U_i(\alpha) < U_i(\beta)$ for all $i \in N$. Then $\beta \notin \sigma(N)$, since $|\phi_N| = 1, \forall \phi_N \in \Phi_N[\sigma(N)]$. By external stability, $\exists \tau \geq 1$, $T \subset S$ and $\phi_T \in \Phi_T[\sigma(T)]$ such that for all $i \in T, U_i(\beta|_\tau) < U_i(\eta)\ \forall \eta \in \phi_T$. Since β is stationary, it follows that $\exists \tau \geq 1$, $T \subset S$ and $\phi_T \in \Phi_T[\sigma(T)]$ such that for all $i \in T, U_i(\alpha|_\tau) < U_i(\eta)\ \forall \eta \in \phi_T$, violating the internal stability of σ. ∎

Proof of Lemma 4.3. From Lemma 1 and using the notations in the proof of Proposition 2, $A(\{i\}) = \emptyset$, for all $i \in N$ and $B(\{1,2\}) = PEP$. Then $A(\{1,2\})$ coincides with the Pareto frontier of the PEP. ∎

Proof of Lemma 4.4. Obvious. ∎

[36] Note that every PEP is immune to individual deviations.

Bibliography

1. ABREU, D., On the Theory of Infinitely Repeated Games with Discounting, *Econometrica* **56** (1988), 383–396.
2. ABREU, D. AND D. PEARCE, A Perspective on Renegotiation in Repeated Games, Game Equilibrium Models (R. Selten, Ed.), Vol 2., Springer-Verlag, 1991.
3. ARCE M., D.G., Stability Criteria for Social Norms with Applications to the Prisoner's Dilemma, *Journal of Conflict Resolution* **38** (1994), 749–765.
4. ASHEIM, G.B., Extending Renegotiation-Proofness to Infinite Horizon Games, *Games and Economic Behavior* **3** (1991), 278–84.
5. ASHEIM, G.B., Renegotiation-Proofness in Finite and Infinite Stage Games through the Theory of Social Situations, Discussion Paper A-137, University of Bonn (1988).
6. ASILIS, C. AND C. KAHN , Semi-stability in Game Theory: A Survey of Ugliness, Game Theory and Economic Applications: Proceedings of the International Conference held at the Indian Statistical Institute, New Delhi, India, December 18-22, 1990 (B. Dutta et al., Eds.), Springer-Verlag, 1992.
7. AUMANN, R., Acceptable Points in Cooperative General n-Person Games, "Contributions to the Theory of Games IV", Princeton University Press, Princeton, N.J., 1959.
8. AUMANN, R., Subjectivity and Correlation in Randomized Strategies, *Journal of Mathematical Economics* **1** (1974), 67–95.
9. AUMANN, R., Game Theory, *in* "The New Palgrave: A Dictionary of Economics", (J. Eatwell, M. Milgate, and P. Newman, Eds.), MacMillan Publ., 1987.
10. AUMANN, R. AND M. MASCHLER, The Bargaining Set for Cooperative Games, *in* "Advances in Game Theory", (M. Dresher, L.S. Shapley, and A.W. Tucker,

Eds.), Princeton University Press, Princeton, NJ, 1964.

11. BERGIN, J. AND B.W. MACLEOD, Efficiency and Renegotiation in Repeated Games, *Journal of Economic Theory* **61** (1993), 42–73.
12. BERNHEIM, B.D., B. PELEG, AND M.D. WHINSTON, Coalition-Proof Nash Equilibria. I. Concepts, *Journal of Economic Theory* **42** (1987), 1–12.
13. BERNHEIM, B.D. AND D. RAY, Pareto Perfect Nash Equilibrium, mimeo, Stanford University (1985).
14. BERNHEIM, B.D. AND D. RAY, Collective Dynamic Consistency in Repeated Games, *Games and Economic Behavior* **1** (1989), 295–326.
15. BRAMS, S., "Theory of Moves", Cambridge University Press, Cambridge, 1994.
16. CHAKRAVORTI, B. AND C. KAHN, Universal Coalition-Proof Equilibrium, Bellcore Discussion Paper (1993).
17. CHAMBERLIN, E.H., "The Theory of Monopolistic Competition", Harvard University Press, Cambridge, Massachusetts, 1933.
18. CHWE, M. S.-Y., Farsighted Coalitional Stability, *Journal of Economic Theory* **63** (1994), 299–325.
19. DUTTA, B., D. RAY, K. SENGUPTA, AND R. VOHRA, A Consistent Bargaining Set, *Journal of Economic Theory* **49** (1989), 93-112.
20. FARRELL, J., Cheap Talk, Coordination, and Entry, *Rand Journal of Economics* **18** (1987), 34–39.
21. FARRELL, J. AND E. MASKIN, Renegotiation in Repeated Games, *Games and Economic Behavior* **1** (1989), 327–360.
22. FARRELL, J. AND G. SALONER, Coordination through Committees and Markets, *Rand Journal of Economics* **19** (1988), 235–252.
23. FERREIRA, J.L., A Communication-Proof Equilibrium Concept, *Journal of Economic Theory* **68** (1996), 249–257.
24. FUDENBERG, D. AND E.S. MASKIN, The Folk Theorem in Repeated Games

with Discounting or with Incomplete information, *Econometrica* **54** (1986). 533–554.

25. GREENBERG, J., Deriving Strong and Coalition-Proof Nash Equilibrium from an Abstract System, *Journal of Economic Theory* **49** (1989), 195–202.
26. GREENBERG, J. , An Application of the Theory of Social Situations to Repeated Games, *Journal of Economic Theory* **49** (1989), 278–293.
27. GREENBERG, J., "The Theory of Social Situations", Cambridge University Press, Cambridge, 1990.
28. GREENBERG, J., Coalition Structures, *in* "Handbook of Game Theory", (R. Aumann and S. Hart, Eds.), Elsevier Science B.V., 1994.
29. GREENBERG, J., D. MONDERER, AND B. SHITOVITZ, Multistage Situations, to appear in *Econometrica* (1995).
30. HARSANYI, J.C., An Equilibrium-Point Interpretation of Stable Sets and a Proposed Alternative Definition, *Management Science* **20** (1974), 1472–1495.
31. HARSANYI, J.C. AND R. SELTEN, "A General Theory of Equilibrium Selection in Games", MIT Press, Cambridge, 1988.
32. LUCE, R.D. AND H. RAIFFA, "Games and Decisions", John Wiley and Sons, New York, 1957.
33. MAS-COLELL, A., An Equivalent Theorem for a Bargaining set, *Journal of Mathematical Economics* **18** (1989), 129–138.
34. MCKINSEY, J. , "Introduction to the Theory of Games", McGraw-Hill, New York, 1952.
35. MORENO , D. AND J. WOODERS, Coalition-Proof Equilibrium, Discussion Paper, Universidad Carlos III de Madrid (1994).
36. ORDESHOOK, P.C., "Game Theory and Political Theory", Cambridge University Press, Cambridge, 1986.
37. PEARCE, D., Renegotiation-Proof Equilibria: Collective Rationality and Intertemporal Cooperation, Cowles Foundation Discussion Paper No. 855, Yale

University (1987).

38. PEARCE, D., Repeated Games: Cooperation and Rationality, "Advances in Economic Theory: Sixth World Congress", Cambridge University Press, 1992.
39. PERRY, M. AND P.J. RENY, A Noncooperative View of Coalition Formation and the Core, *Econometrica* **62** (1994), 795–818.
40. RABIN, M., A Model of Pre-game Communication, *Journal of Economic Theory* **63** (1994), 370–391.
41. RAY, R., Credible Coalitions and the Core, mimeo, Stanford University (1983).
42. ROSENTHAL, R.W., Cooperative Games in Effectiveness Form, *Journal of Economic Theory* **5** (1972), 88–101.
43. ROTH, A.E. (Ed.), "The Shapley Value: Essays in Honor of Lloyd S. Shapley", Cambridge University Press, Cambridge, 1988.
44. RUBINSTEIN, A., Strong Perfect Equilibrium in Supergames, *International Journal of Game Theory* **9** (1980), 1–12.
45. SHUBIK, M., "Game Theory in the Social Sciences", MIT Press, Cambridge, Massachusetts, 1984.
46. VON NEUMANN, J. AND O. MORGENSTERN, "Theory of Games and Economic Behavior", Princeton University Press, Princeton, NJ, first edition, 1944, second edition, 1947.