DEPARTMENT OF ECONOMICS

Working Paper

STABLE AGREEMENTS
IN INFINITELY REPEATED GAMES

Licun Xue

Working Paper No. 2000-13
Centre for Dynamic Modelling in Economics

ISSN 1396-2426

UNIVERSITY OF AARHUS • DENMARK

CENTRE FOR DYNAMIC MODELLING IN ECONOMICS
DEPARTMENT OF ECONOMICS - UNIVERSITY OF AARHUS - DK - 8000 AARHUS C - DENMARK
☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

WORKING PAPER

STABLE AGREEMENTS IN INFINITELY REPEATED GAMES

Licun Xue

Working Paper No. 2000-13

DEPARTMENT OF ECONOMICS
SCHOOL OF ECONOMICS AND MANAGEMENT - UNIVERSITY OF AARHUS - BUILDING 350
8000 AARHUS C - DENMARK ☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

Stable Agreements in Infinitely Repeated Games

Licun Xue *

Department of Economics, University of Aarhus

DK-8000 Aarhus C, Denmark

E-mail: LXue@econ.au.dk

Revised, November 2000

Abstract

This paper studies infinitely repeated games where players can form coalitions to coordinate their actions via self-enforcing agreements. The proposed notion of "stable agreements" extends a characterization of the set of subgame perfect equilibrium paths by Greenberg (1989, 1990) to account for self-enforcing coalitional deviations. An agreement is stable if no coalition can deviate in such a way that by solely coordinating the actions of its own members, it guarantees a higher payoff for each member. Existence of the proposed notion is established and its relation to other notions is investigated.

Journal of Economic Literature Classification Numbers: C70, C72

Keywords: repeated games, renegotiation, self-enforcing agreements, coalitions

* I would like to thank Joseph Greenberg, Geir B. Asheim, an associate editor and a referee for their valuable comments and suggestions.

1 Introduction

The theory of repeated games has succeeded in explaining the possibility of cooperation among self-interested individuals 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". 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 makes it possible to achieve, but by no means singles out, cooperative outcomes.

As a noncooperative notion, subgame perfect equilibrium captures the dynamic consistency of *individual* behavior. Indeed, a profile of strategies is subgame perfect if after no history of play should a *single* player have incentive to deviate *unilaterally* from his strategy; in particular, punishment for every individual deviation must be credible in that if players were called upon to carry it out, no single player would have an incentive to back out unilaterally. Thus, as a notion of implicit self-enforcing agreement, subgame perfect equilibrium is too weak since it does not account for behavior of groups or coalitions.

The notion of renegotiation-proofness[1] in infinitely repeated games introduced by Bernheim and Ray (1989) and Farrell and Maskin (1989) is one of the attempts in the literature to capture dynamic consistency at the collective level. At the heart of this notion is the assumption that the grand coalition (and only the grand coalition) has the opportunity to negotiate anew (out of a "bad" equilibrium) after *every* history. In particular, renegotiation-proofness entails that after an unilateral deviation from an equilibrium, the grand coalition (i.e., the coalition of all players) *will* renegotiate and abandon the prescribed punishment *as long as* there exits a Pareto superior equilibrium (pos-

[1] See Pearce (1992) and Bergin and MacLeod (1993) for excellent surveys.

sibly the original equilibrium). Such a notion of renegotiation-proofness can be criticized for ignoring the credibility of renegotiated equilibrium [see, e.g., Pearce (1987) and Abreu and Pearce (1991)] because it does not require the "renegotiated equilibrium" to be renegotiation-proof. Moreover, a player, in contemplating a deviation, anticipates a renegotiated equilibrium that is most favorable to him; thus, renegotiation-proofness embeds over-optimism on the part of a deviating player. This is illustrated through the example in Table 1 taken from Asheim (1991).

Table 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 the above 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.[2] These punishments are subject to "renegotiation" according to Bernheim and Ray (1989) and Farrell and Maskin (1989): Player 1, for example, *will* deviate by playing d_1, because he believes that in the next period players 1 and 2 *will* renegotiate in order to avoid the Pareto inferior path π_1. It follows that π is not supported by a renegotiation-proof equilibrium. In fact, the only subgame perfect equilibrium that is not subject to such a renegotiation is the one in which the Nash equilibrium of the stage game, (d_1, d_2), is played after every

[2] 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).

history; this equilibrium yields each player the lowest payoff within the set of subgame perfect equilibria.

To escape from the difficulties associated with renegotiation-proofness, we offer an alternative notion of self-enforcing agreements – "stable agreements" – to capture collective dynamic consistency. For an agreement to be "stable", it must be immune to deviations of all coalitions, not just single players and the grand coalition. Our notion aims to be inclusive but it rules out with confidence: Each deviating coalition is cautious or averse to strategic uncertainty in that it will not deviate from a given agreement unless doing so "guarantees" a higher payoff for each of its members. To be more specific, an agreement is stable if no coalition can deviate and achieve a higher payoff for each of its members by solely coordinating the choice of strategies of its members in a self-enforcing fashion. (Thus, while it is *feasible* for any coalition to form and make a joint deviation from a given agreement, rationality dictates which coalitions might *actually* form.) Our notion can be viewed as the weakest notion accounting for collective consistency; it builds on a characterization of the set of subgame perfect equilibria by Greenberg (1989, 1990): if players are cautious then the set of equilibrium paths satisfies internal and external stability akin to that of von Neumann and Morgenstern (1944).

The organization of the rest of the paper is as follows: Section 2 formalizes the notion of "stable agreements". Properties of stable agreements, including existence, are studied. In Section 3 the notion of stable agreements is related to several notions in the literature including notions of renegotiation-proofness, perfectly coalition proof Nash equilibrium, and the β-core. All proofs are relegated to the appendix.

2 Stable Agreements

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 Π the set of paths, i.e., $\Pi = A^\infty$. For $\alpha \in \Pi$ and a stage τ, let $\alpha|_\tau$ denote the continuation of α from τ (including τ) on. All players are assumed to discount future payoffs using the same discount factor $\delta \in (0,1)$. Thus, the (normalized discounted) payoff of player $i \in N$ from $\alpha = (\alpha^1, \alpha^2, \ldots) \in \Pi$ is

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

Let $H = \cup_{\tau=0}^{\infty} A^\tau$, where $A^0 \equiv \emptyset$, be the set of all histories. H can be ordered by $\geq$: For $h, h' \in H$, $h' \geq h$ implies that h is a sub-history of h'. A (pure) strategy for $i \in N$ is a mapping $f_i : H \to A_i$. A coalition S is a nonempty subset of N, denoted by $S \subset N$.[3] For $S \subset N$, let $A_S \equiv \prod_{i \in S} A_i$. Let $-S$ denote the complement of S, i.e., $N \setminus S$. Let PEP denote the set of subgame-perfect-equilibrium outcome paths, a set that is assumed to be nonempty.

The objective of this section is to formalize a notion of "stable agreements". We shall define our notion in the space of outcome paths, although a definition in payoff space is also possible [see, e.g. Pearce (1992)]. A stable agreement (for N) is a path in Π from which no coalition $S \subset N$ can deviate in such a way that by solely coordinating the strategies of its members, it can *guarantee* each of its members a higher payoff. Our definition builds on a characterization of the set of subgame perfect equilibria by Greenberg (1989, 1990). We first state this characterization and then introduce group or coalitional behavior. Following Greenberg (1989, 1990), let Ξ be a mapping that assigns to every history $h \in H$ a subset of outcome paths. Ξ is called a *standard of behavior* and $\Xi(h)$ is the "solution" or the set of "plausible continuations" once h transpires. Each player $i \in N$, in contemplating a deviation from a path $\alpha \in \Pi$ at period τ by choosing some $b_i \in A_i$, has to compare $\alpha|_\tau$ with the set of paths starting with (b_i, α^τ_{-i}) and followed by

[3] All inclusions in this paper are weak.

$\Xi(h)$, where $h = (\alpha^\tau, (b_i, \alpha^\tau_{-i}))$. A player is said to be uncertainty averse or "conservative" if he does not engage the above deviation unless he prefers every path in $\{(b_i, \alpha^\tau_{-i})\} \times \Xi(h)$ to $\alpha|_\tau$.

In view of the fact that all subgames are isomorphic to $\mathcal{G}^\infty$, it seems reasonable to assume that $\Xi(h) = \Xi(h') = \Theta$ for all $h, h' \in H$; that is the "solution set" is the same across all histories. Following von Neumann and Morgenstern (1944), Greenberg (1989, 1990) requires that Θ be free of internal contradiction or be "internally stable" and account for every outcome it excludes or be "externally stable"[4]. Formally,

Definition 1 *(Greenberg, 1989, 1990) Let Θ be a nonempty subset of Π.*

- *Θ is internally stable if for every $\alpha \in \Theta$, there does not exist $\tau \geq 1$, $i \in N$, and $b_i \in A_i$ such that $U_i(\alpha|_\tau) < U_i((b_i, \alpha^\tau_{-i}), \beta)$[5] for all $\beta \in \Theta$.*
- *Θ is externally stable if for every $\alpha \in \Pi \setminus \Theta$, there exist $\tau \geq 1$, $i \in N$, and $b_i \in A_i$ such that $U_i(\alpha|_\tau) < U_i((b_i, \alpha^\tau_{-i}), \beta)$ for all $\beta \in \Theta$.*
- *Θ is stable if it is both internally and externally stable, that is, $\alpha \in \Pi \setminus \Theta$ if and only if there exist $\tau \geq 1$, $i \in N$, and $b_i \in A_i$ such that $U_i(\alpha|_\tau) < U_i((b_i, \alpha^\tau_{-i}), \beta)$ for all $\beta \in \Theta$.*

Using Abreu's (1988) characterization of PEP, Greenberg (1989) showed that $\Theta = PEP$ is the unique maximal (conservative) stable standard of behavior with respect to set inclusion. Thus, PEP is a set-valued notion that is internally consistent and justifies each path it excludes under the assumption that players are conservative or cautious in evaluating the likely outcomes following their deviations (conservativeness or cautiousness is reflected in that a player $i \in N$ deviates from a path α if and only if there exists $b_i \in A_i$ such that $U_i(\alpha|_\tau) < U_i((b_i, \alpha^\tau_{-i}), \beta)$ *for all* $\beta \in \Xi$). Note that each player needs only to consider one-stage deviations.

[4] The literature on renegotiation-proofness also exhibits various attempts of imposing stability (consistency). See Section 3.1.

[5] $((b_i, \alpha^\tau_{-i}), \beta)$ is the path starting with action profile (b_i, α^τ_{-i}) and continued with β.

Now we proceed to define a notion of stable agreements for environments where coalitions can form to coordinate the actions of their members in a self-enforcing fashion. PEP can be regarded as the set of "individually stable agreements" in that no individual has an incentive to deviate from any path in PEP, knowing that he cannot dictate the choice of the rest of the players and is averse to such a strategic uncertainty. When coalitions can form, a natural extension is that each deviating coalition recognizes its ability to coordinate the actions of its own members in a self-enforcing manner but it cannot dictate the choices of nonmembers and is averse to such a strategic uncertainty. When coalitions can form, the conservatism of the players has two aspects: As in Definition 1, each coalition is conservative in evaluating the set of "likely outcomes" induced by its deviation; Moreover, each coalition cannot dictate the choices of nonmembers and therefore, in determining the set of "likely outcomes", considers only what it can achieve by solely coordinating the choices of its members in a self-enforcing fashion. Thus, each coalition is averse to the strategic uncertainty that cannot be resolved by solely coordinating the actions of its members.

For a path $\alpha \in PEP$[6] to be stable for N, it must be the case that no coalition $S \subset N$ can benefit from a self-enforcing deviation. Since coalitions differ in what the coordination of their members can achieve, it is natural to postulate a solution concept (standard of behavior) as a mapping Σ that specifies for every $S \subset N$ a subset $\Sigma(S) \subset PEP$. For every $S \subset N, \Sigma(S)$ is the set of self-enforcing agreements for S, based on which S contemplates its deviations from any $\alpha \in PEP$; that is, if $S \subset N$ deviates from $\alpha \in PEP$ at stage τ by choosing $b_S \in A_S$, it realizes that the set of continuations following its deviation is $\Sigma(S)$. Since each $\beta \in \Sigma(S)$ is self-enforcing for S, β must be immune not only to self-enforcing deviations of S but also to those of any $T \subsetneq S$; moreover, each $T \subset S$, in determining $\Sigma(T)$, needs to apply

[6] For simplicity of exposition, here we start with PEP and introduce coalitional deviations. An earlier version of this paper starts with Π and therefore has to incorporate Definition 1 in the formulation of "stable agreements".

the same reasoning[7]. Thus, for all $S \subset N$, $\alpha \in \Sigma(S)$ implies no $T \subset S$ can benefit from self-enforcing deviations. Also, like Definition 1, we would like to require $\Sigma(S)$ not to rule out arbitrarily; in particular, $\alpha \in PEP \setminus \Sigma(S)$ implies some $T \subset S$ will benefit from a self-enforcing deviation.

When a coalition $S \subset N$ deviates from $\alpha \in PEP$ at stage τ by choosing some $b_S \in A_S$, it must realize not only that the set of continuations following its deviation is $\Sigma(S)$ but also that it cannot choose $b_S \in A_S$ arbitrarily, in the absence of a binding agreement. Thus, to complete the formulation of self-enforcing deviations, we need to put restrictions on the action combination the deviating coalition S chooses. For the afore-mentioned deviation of S to be self-enforcing, each path in $\{(b_S, \alpha^\tau_{-S})\} \times \Sigma(S)$ has to be self-enforcing for S, that is,

$$\emptyset \neq \{(b_S, \alpha^\tau_{-S})\} \times \Sigma(S) \subset \Sigma(S). \tag{C}$$

To illustrate the above condition, consider the infinite repetition of the prisoner's dilemma in Table 2 and assume $\delta = 0.4$. It is easy to verify that the

Table 2

	ℓ	r
u	3,0	1,1
d	2,2	0,3

cooperative outcome of the infinite repetition of (d, ℓ) cannot be supported by a subgame perfect equilibrium. In fact, the unique path in 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 to (d, ℓ) at some stage, π would be ruled out. Such a deviation, however, cannot be carried out in the absence of a binding agreement, since each individual has an incentive not to take its part in this joint deviation[8]. Maintaining the assumption that all agreements must be self-enforcing, deviations of a

[7] Thus, those paths in $\Sigma(S)$ are considered "unavoidable" by S.

[8] This is in contrast to strong perfect equilibrium (see Section 3.2) that uses this deviation to rule out π.

non-singleton coalition have to respect condition (C).

The following definition extends Definition 1 to account for self-enforcing coalitional deviations. As in Definition 1, the "if" part reflects internal stability while "only if" part external stability.

Definition 2 *A standard of behavior* Σ *is stable if for all* $S \subset N$, $\alpha \in PEP \setminus \Sigma(S)$ *if and only if there exist a coalition* $T \subset S$*, a stage* $\tau \geq 1$*, and* $b_T \in A_T$ *such that* $\emptyset \neq \{(b_T, \alpha^{\tau}_{-T})\} \times \Sigma(T) \subset \Sigma(T)$ *and* T *"prefers"* $\{(b_T, \alpha^{\tau}_{-T})\} \times \Sigma(T)$ *to* $\alpha|_{\tau}$*.*

To complete the above definition, we need to be precise about the phrase "T 'prefers' $\{(b_T, \alpha^{\tau}_{-T})\} \times \Sigma(T)$ to $\alpha|_{\tau}$". Unlike Definition 1, we need to take into account the fact that the one-stage deviation principle no longer holds when coalitional deviations are considered. To illustrate this point, consider the example in Table 3

Table 3

	ℓ	r
u	9,0,1	0,0,2
d	0,0,1	0,9,1
	L	

	ℓ	r
u	0,0,1	0,0,2
d	3,3,1	0,0,1
	R	

where player 1 chooses rows, player 2 chooses columns, and player 3 chooses matrices. Let α be the path that results from "repeating (u, r, L) forever", and let $\Sigma(T) = PEP$, where $T = \{1, 2\}$. Evidently, a coordinated one-stage deviation from α by T cannot suffice to guarantee both its members higher payoffs, given that the set of plausible continuations is specified by $\Sigma(T)$; that is, there does not exist $b_T \in A_T$ such that (C1) is satisfied and $U_i(\alpha) < U_i(\beta)$ for all $\beta \in \{b\} \times \Sigma(T)$ and for $i = 1, 2$. Coalition T, however, can coordinate the actions of its members in more than one (or even infinite) stages in a way that respects $\Sigma(T)$. Suppose T deviates from α at $\tau = 1$ by choosing (u, ℓ) and continues to coordinate the actions of its members in the following way: After any history, play (d, ℓ) if player 3 is currently playing[9] 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, players 1 and 2 would be better-off than they are under α.[10]

It is precisely this reasoning that underlines the following definition.

Definition 3 T prefers $\{(b_T, \alpha^{\tau}_{-T})\} \times \Sigma(T)$ to $\alpha|_{\tau}$ (in Definition 2) *if there exists a set* $B \subset \{(b_T, \alpha^{\tau}_{-T})\} \times \Sigma(T)$ *such that*

(C1) $\beta \in B \Longleftrightarrow \nexists \eta \in B$ *such that for some* $\kappa \geq 1$, $\beta^t = \eta^t$ *for all* $t < \kappa$, $\beta^{\kappa}_{-T} = \eta^{\kappa}_{-T}$, *and* $\beta^{\kappa}_T \neq \eta^{\kappa}_T$, *and*

(C2) $U_i(\alpha) < U_i(\beta)$ *for all* $\beta \in B$ *and for all* $i \in T$.

Condition (C1) captures the fact that T can engage in a multi-stage coordination in response to a given path α while it cannot dictate the choices of nonmembers; condition (C2) captures the fact that T is averse to the strategic uncertainty (in B) that it cannot resolve by solely coordinating the actions of its members; conditions (C1) and (C2) extend the assumption of

[9] Note that the actions of players 1 and 2 may depend also on the current action of player 3. This reflects that each coalition can only coordinate the action of its members and cannot dictate the choice of nonmembers. This feature is embedded in any equilibrium analysis.

[10] 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).

conservatism on the behavior of single players to coalitions while incorporating necessary multi-stage coordinations of coalitions.[11]

For a stable standard of behavior Σ, the set $\Sigma(N)$ is called the set of *stable agreements*. Σ captures coalitional rationality and dynamic consistency; $\Sigma(N)$ contains those and only those paths that are not rejected by any $S \subset N$ whose members can coordinate their actions in a self-enforcing way without dictating the choices of the rest of the players. It is easy to verify that in the example in Table 1, this set consists of the unique (Pareto) efficient perfect equilibrium path (PEP), i.e., $\Sigma(N) = \{\pi\}$.

Now we proceed to investigate some properties of stable agreements. The following proposition states that the stable standard of behavior exists and no stable agreement is Pareto dominated by another stable agreement.

Proposition 1 *There exists a stable standard of behavior* Σ *such that* $\Sigma(S) \subset PEP$ *for all* $S \subset N$*. Moreover, if* $\alpha \in \Sigma(N)$*, then there does not exist* $\beta \in \Sigma(N)$ *such that* $U_i(\alpha) < U_i(\beta)$ *for all* $i \in N$*.*

The following propositions characterize the set of stable agreements in two special cases. The example in Table 1 serves as an example where both propositions apply.

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

Proposition 3 *If* PEP *admits a unique efficient path* α*, then* $\Sigma(N) = \{\alpha\}$*.*

Thus, as is the case for two-player games, the set of stable agreements can still be large. Such an indeterminacy may be resolved by mediation, social convention, or bargaining. A mediator, for example, may simply recommend each stable agreement with equal probability (or a stable agreement with the same expected payoffs) on the ground of "fairness". This is, however, beyond the scope of this paper.

[11] Alternatively, one could consider one-stage correlated deviation. But correlation complicates the formalization of self-enforcing deviations.

Remark 1 *Existence of* Σ *does not necessarily imply that* $\Sigma(N)$*, the set of stable agreements, is nonempty. If for some game* $\Sigma(N) = \emptyset$*, i.e., the grand coalition cannot reach any self-enforcing agreements, then a stable standard of behavior* Σ *can be used to "predict" the coalitions that are likely to form. Note that by external stability,* Σ *cannot be empty-valued (i.e., there exists* $S \subset N$ *such that* $\Sigma(S) \neq \emptyset$*); hence,* Σ *is never silent as a notion of self-enforcing agreements.*

3 Related Literature

3.1 Stability and Renegotiation Proofness

Although our notion is not one of renegotiation-proofness, its connection to various theories of renegotiation-proofness is evident. First, both our theory and the notions of renegotiation proofness allow for coalitional deviations, although renegotiation proofness restricts coalitional deviations to those of the grand coalition. Secondly, the notion of *stable agreements* is defined by applying the notion of *stability* that was originated by von Neumann and Morgenstern (1944) and extended by Greenberg (1990); the theories of renegotiation-proofness exhibit various attempts to apply the notion of stability. As Rubinstein (1992) wrote "... the renegotiation literature (as well as the new approach suggested by Greenberg (1990)) is returning to the internal and external consistency ideas suggested by von Neumann and Morgenstern (1944). For example, (weak) renegotiation proofness of Bernheim and Ray (1989) and Farrell and Maskin (1989) imposes a version of internal stability (stronger than ours) while Asheim's (1991) Pareto perfect equilibrium imposes also external stability in addition to the same internal stability as Bernheim and Ray (1989) and Farrell and Maskin (1989).

However, the notions of renegotiation proofness can be criticized for taking Pareto criterion too far as discussed in the introduction. They stipulate that the grand coalition will renegotiate and abandon a punishment whenever there is a Pareto dominating equilibrium available even though the later equilibrium may rely on punishments that are as severe. Implicitly, a deviating player counts too heavily on renegotiation. Our notion explores a natural extension of the uncertainty aversion on the part of players embedded in the notion of subgame perfection. Our notion can be viewed as the weakest notion that accounts for coalitional deviations. For two-player games, it is easy to see that an efficient (weakly) renegotiation-proof[12] equilibrium also belongs to the set of stable agreements for N. However, we do recognize the importance and relevance of renegotiation in formalizing notions stronger than ours. In a future project, we shall extend our analysis to account for credible renegotiation.

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 Asheim (1988) to dynamic games with infinite horizon. Unlike renegotiation proofness, PCPNE considers all coalitional deviations. But, like renegotiation proofness, PCPNE stipulates that renegotiation occurs after *every* history. In particular, for two-player games, PCPNE coincides with Pareto perfect equilibrium (Asheim, 1991), which "refines" renegotiation proofness. Existence of PCPNE is not guaranteed. Indeed, the example depicted by Table 1 does not admits a PCPNE or Pareto perfect equilibrium.

Rubinstein's (1980) *strong perfect equilibrium* is more demanding than PCPNE in that it requires an equilibrium to survive all conceivable deviations, many of which are not credible[13]. In particular, Pareto efficiency in the space of all feasible outcomes is imposed. It is easy to see that in two-player games, a strong perfect equilibrium is a PCPNE (thus also a Pareto perfect

[12] Recall that a subgame perfect equlibrium is (weakly) renegotiation-proof if no two of its continuation equilibria can be strictly Pareto ranked.

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

equilibrium) and a PCPNE is (weakly) renegotiation-proof and also a stable agreement for N.

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 $S \subset N$,

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

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, members of $N \setminus S$ are assumed to be individually rational.

(2) The definition of β-core does not consider the credibility or the self-enforceability of an "objection". 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, as a notion for static settings, does not consider dynamic consistency. Our notion captures the dynamic consistency at both coalitional and individual levels (every stable agreement belongs to PEP).

4 Appendix

We first introduce the following notations to facilitate the proofs: For $\Omega \subset \mathit{\Pi}$ and $T \subset N$, let $\Phi_T(\Omega)$ denote the subsets of Ω that satisfy conditions (C1) and (C2) in Definition 3. Hence T prefers Ω to some $\alpha \in \mathit{\Pi}$ if and only if there exists $\phi_T \in \Phi_T(\Omega)$ such that for all $\beta \in \phi_T$, $U_i(\alpha) < U_i(\beta)$ for all $i \in T$.

Proof of Proposition 1.

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

For all $i \in N$,

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

By Definition 1 (in particular, internal stability of PEP) and one-stage deviation principle for individuals, $A(\{i\}) = \emptyset$ for all $i \in N$. For $S \subset N$, assume that $A(T)$ and $B(T)$ are defined for all $T \subsetneq S$. Define

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

We shall show that the standard of behavior Σ given by $\Sigma(S) = B(S) \setminus A(S)$ for all $S \subset N$ is stable in the sense of Definition 2. This is accomplished in the following three steps.

STEP 1. We first show that $B(S)$ is compact for all $S \subset N$. Since PEP is a compact set[14], $B(\{i\})$ is compact. Therefore, for $|S| > 1$, it suffices to show

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

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 $\phi_T \in \Phi_T[B(T) \setminus A(T)]$ such that. 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

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

STEP 2. We then show that $A(S) = A^*(S)$. We first show that $A(S) \subset A^*(S)$. Consider $\alpha \in A(S)$. Then $\exists \tau \geq 1$ and $\phi_S \in \Phi_S[B(S)]$ such that 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 $\phi_S^* \in \Phi_S[B(S)]$ such that. 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 $\phi_S \in \Phi[B(S) \setminus A(S)]$ such that. 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 $\phi_S^* \in \Phi_S[B(S)]$ such that. 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.

STEP 3. Finally, We show that the standard of behavior Σ given by $\Sigma(S) = B(S) \setminus A(S)$ for all $S \subset N$ is stable. Indeed, by the definition of $\Sigma(S)$, $B(S)$, and $A(S)$, we have that $\alpha \in PEP \setminus \Sigma(S)$ if and only if $\alpha \in [PEP \setminus B(S)] \cup A(S)$. Since $A(S) = A^*(S), \alpha \in PEP \setminus \Sigma(S)$ if and only if $\alpha \in [PEP \setminus B(S)] \cup A^*(S)$. Then by the definition of $B(S)$ and $A^*(S), \alpha \in PEP \setminus \Sigma(S)$ if and only if there do not exist $\tau \geq 1, T \subset S$ and $\phi_T \in \Phi_T[\Sigma(T)]$ such that for all $i \in T, U_i(\alpha|_\tau) < U_i(\beta), \forall \beta \in \phi_T$. Hence Σ is stable in the sense of Definition 2.

The second part of Proposition 1 follows from the fact that $|\phi_N| = 1, \forall \phi_N \in \Phi_N[\Sigma(N)]$. ■

Proof of Proposition 2. Follows from the fact that $\Sigma(\{i\}) = PEP$ for all $i = 1, 2$, and $\left|\phi_{\{1,2\}}\right| = 1$. ■

Proof of Proposition 3. Let $\beta \in PEP \setminus \{\alpha\}$. Given that α is the only efficient path in PEP and PEP is compact, $U_i(\alpha) > U_i(\beta)$ for all $i \in N$. It is easy to see that by external stability, $\alpha \in \Sigma(N)$. By internally stability, $\beta \notin \Sigma(N)$ (otherwise N can benefit from a self-enforcing deviation). ■

References

[1] Abreu, D. (1988), "On the Theory of Infinitely Repeated Games with Discounting", *Econometrica* **56**, 383-396.

[2] Abreu, D. and D. Pearce (1991), "A Perspective on Renegotiation in Repeated Games" in *Game Equilibrium Models* (R. Selten, Ed.), Vol 2. Springer-Verlag.

[3] Asheim, G. B. (1991), "Extending Renegotiation-Proofness to Infinite Horizon Games", *Games and Economic Behavior* **3**, 278-284.

[4] Asheim, G. B. (1988), "Renegotiation-Proofness in Finite and Infinite Stage Games through the Theory of Social Situations", Discussion Paper A-137, University of Bonn.

[5] Aumann, R. J. (1959), "Acceptable Points in General Cooperative n-Person Games" in *Contributions to the Theory of Games IV* (H.W. Kuhn and R.D. Luce, Eds.). Princeton University Press.

[6] Bergin, J. and W. B. MacLeod (1993), "Efficiency and Renegotiation in Repeated Games", *Journal of Economic Theory* **61**, 42-73.

[7] Bernheim, B. D. and D. Ray (1989), "Collective Dynamic Consistency in Repeated Games", *Games and Economic Behavior* **1**, 295-326.

[8] Bernheim, B. D., B. Peleg, and M. Whinston (1987), "Coalition-Proof Nash Equilibrium 1. Concepts", *Journal of Economic Theory* **42**, 1-42.

[9] Farrell, J. and E. Maskin (1989), "Renegotiation in Repeated Games", *Games and Economic Behavior* **1**, 327-360.

[10] Fudenberg, D. and E. Maskin (1986), "The Folk Theorem in Repeated Games with Discounting or with Incomplete information", *Econometrica* **54**, 533-554.

[11] Greenberg, J. (1990), *The Theory of Social Situations: An Alternative Game-Theoretic Approach*. Cambridge University Press.

[12] Greenberg, J. (1989), "An Application of the Theory of Social Situations to Repeated Games", *Journal of Economic Theory* **49**, 278-293.

[13] Pearce, D. (1992), "Repeated Games: Cooperation and Rationality" in *Advances in Economic Theory: Sixth World Congress*. Cambridge University Press.

[14] Pearce, D. (1987), "Renegotiation-Proof Equilibria: Collective Rationality and Intertemporal Cooperation", Cowles Foundation Discussion Paper No. 855, Yale University.

[15] Rubinstein, A. (1992), "Repeated Games: Cooperation and Rationality - Comments" in *Advances in Economic Theory: Sixth World Congress*, Cambridge University Press.

[16] Rubinstein, A. (1980), "Strong Perfect Equilibrium in Supergames", *International Journal of Game Theory* **9**, 1-12.

[17] von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton: Princeton University Press.

Working Paper

1999-29 Michael Jansson: Consistent Covariance Matrix Estimation for Linear Processes.

2000-1 Niels Haldrup and Peter Lildholdt: On the Robustness of Unit Root Tests in the Presence of Double Unit Roots.

2000-2 Niels Haldrup and Peter Lildholdt: Local Power Functions of Tests for Double Unit Roots.

2000-3 Jamsheed Shorish: Quasi-Static Macroeconomic Systems.

2000-4 Licun Xue: A Notion of Consistent Rationalizability - Between Weak and Pearce's Extensive Form Rationalizability.

2000-5 Ebbe Yndgaard: Labour, An Equivocal Concept for Economic Analyses.

2000-6 Graham Elliott and Michael Jansson: Testing for Unit Roots with Stationary Covariates.

2000-7 Nikolaj Malchow-Møller and Bo Jellesmark Thorsen: A Dynamic Agricultural Household Model with Uncertain Income and Irreversible and Indivisible Investments under Credit Constraints.

2000-8 Niels Haldrup, Antonio Montanés and Andreu Sanso: Measurement Errors and Outliers in Seasonal Unit Root Testing.

2000-9 Erik Harsaae: En kritisk vurdering af den generelle ligevægtsmodel.

2000-10 Rasmus Højbjerg Jacobsen: Why the ECB Should be Ultra-Liberal.

2000-11 Michael Rosholm and Michael Svarer: Structurally Dependent Competing Risks.

2000-12 Efforsyni Diamantoudi and Licun Xue: Farsighted Stability in Hedonic Games.

2000-13 Licun Xue: Stable Agreements in Infinitely Repeated Games.

CENTRE FOR DYNAMIC MODELLING IN ECONOMICS

DEPARTMENT OF ECONOMICS - UNIVERSITY OF AARHUS - DK - 8000 AARHUS C - DENMARK
☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

Working papers, issued by the Centre for Dynamic Modelling in Economics:

1999-11	Martin Paldam: The Big Pattern of Corruption. Economics, Culture and the Seesaw Dynamics.
1999-21	Martin Paldam: Corruption and Religion. Adding to the Economic Model?
2000-1	Niels Haldrup and Peter Lildholdt: On the Robustness of Unit Root Tests in the Presence of Double Unit Roots.
2000-2	Niels Haldrup and Peter Lildholdt: Local Power Functions of Tests for Double Unit Roots.
2000-3	Jamsheed Shorish: Quasi-Static Macroeconomic Systems.
2000-4	Licun Xue: A Notion of Consistent Rationalizability - Between Weak and Pearce's Extensive Form Rationalizability.
2000-6	Graham Elliott and Michael Jansson: Testing for Unit Roots with Stationary Covariates.
2000-8	Niels Haldrup, Antonio Montanés and Andreu Sanso: Measurement Errors and Outliers in Seasonal Unit Root Testing.
2000-12	Effrosyni Diamantoudi and Licun Xue: Farsighted Stability in Hedonic Games.
2000-13	Licun Xue: Stable Agreements in Infinitely Repeated Games.