Finite Horizon Bargaining and the Consistent Field*

Armando Gomes† Sergiu Hart‡ Andreu Mas-Colell§

April 1997

Abstract

This paper explores the relationships between noncooperative bargaining games and the consistent value for non-transferable utility (NTU) cooperative games. A dynamic approach to the consistent value for NTU games is introduced: the consistent vector field. The main contribution of the paper is to show that the consistent field is intimately related to the concept of subgame perfection for finite horizon noncooperative bargaining games, as the horizon goes to infinity and the cost of delay goes to zero. The solutions of the dynamic system associated to the consistent field characterize the subgame perfect equilibrium payoffs of the noncooperative bargaining games. We show that for transferable utility, hyperplane and pure bargaining games, the dynamics of the consistent field converge globally to the unique consistent value. However, in the general NTU case, the dynamics of the consistent field can be complex. An example is constructed where the consistent field has cyclic solutions; moreover, the finite horizon subgame perfect equilibria do not approach the consistent value.

Journal of Economic Literature Classification Numbers: C71, C72.

---

*This research was started in 1991, as part of the first author's Ph.D. thesis at Harvard University. We gratefully acknowledge financial support from: CAPES, Brazil (Gomes); the U.S.-Israel Binational Science Foundation (Hart and Mas-Colell); the Israeli Academy of Sciences and Humanities (Hart); the Spanish Ministry of Education (Mas-Colell); and the Generalitat de Catalunya (Mas-Colell).

†Department of Finance, The Wharton School, University of Pennsylvania, Philadelphia. *e-mail:* gomes@wharton.upenn.edu.

‡Department of Economics; Department of Mathematics; and Center for Rationality and Interactive Decision Theory, The Hebrew University of Jerusalem. *e-mail:* hart@math.huji.ac.il.

§Department of Economics, Universitat Pompeu Fabra, Barcelona. *e-mail:* mcolell@upf.es.

1. Introduction

This paper belongs to a general research program which studies the relationships between equilibria of n-player noncooperative games and axiomatically generated solutions for the cooperative game described in coalitional form. Here we carry out an exploration based on the theory of differentiable dynamic systems.

For n-person situations of pure bargaining (where the cooperation of all players is needed to achieve an outcome different from the threat values), the classical solution concept proposed by axiomatic cooperative game theory is the *Nash* (1950) *bargaining solution.* Interestingly, the Nash solution has also been arrived at as a limit — when the cost of delay in agreement becomes small — of the subgame perfect equilibria of models of bargaining in extensive form (in particular, of the Ståhl (1972) – Rubinstein (1982) model of alternating offers; see Binmore (1987) and the book of Osborne and Rubinstein (1990) for these and other models).

Similarly, for n-player games with transferable utility (TU), the *Shapley* (1953) *value* is a central solution concept derived by axiomatic cooperative game theory. Again, bargaining models in extensive form have been proposed, whose solutions coincide with, or converge to, the Shapley value (Harsanyi (1981), Gul (1989), Hart and Moore (1990), Winter (1994), Hart and Mas-Colell (1996b)).

The theory is less settled for the general non-transferable utility (NTU) games in coalitional form. In this paper we focus on the *consistent (NTU-) value*, an axiomatic solution proposed by Maschler and Owen (1989, 1992), which generalizes both the Nash solution for the pure bargaining case and the Shapley value for the TU case. The point of departure for our current research is Hart and Mas-Colell (1996b), which contains an analysis of an infinite horizon noncooperative bargaining game whose stationary subgame perfect equilibria are close, when the parameter that measures the cost of delay in agreement is low, to the consistent values.

The present paper starts by developing a dynamic approach to the consistent value.

It generalizes to NTU games some dynamic processes put forward by Maschler, Owen and Peleg (1988) for pure bargaining games and by Maschler and Owen (1989) for hyperplane games (an extension of TU games), and which, in these cases, globally converge to the unique consistent value. Motivated by the axiomatic concept of consistency we introduce the concept of *consistent (vector) field*. Roughly speaking, the consistent field is defined, for every payoff configuration at the Pareto frontier, as the direction to move locally along the frontier in order to reduce the "inconsistency" of the payoff. The singularities of the consistent field are the consistent values and the solutions (or flows) of the dynamic system associated with the consistent field constitute a natural way by which players starting from arbitrary payoffs could adjust.

The main contribution of the paper is to show that the consistent field is intimately related to subgame perfection for finitely horizon noncooperative bargaining games, providing thus an unexplored link between the cooperative and noncooperative theoretical formulations. The specific noncooperative bargaining game we study is the *finite horizon* version of the bargaining game introduced by Hart and Mas-Colell (1996b). Informally, this noncooperative game is a sequential game where the players have up to T stages to reach an agreement. At each stage a player is selected at random to propose a particular way to split the gains from cooperation, and will be ousted from the game with a probability of $1-\rho$ if an unanimous agreement is not reached. The subgame perfect Nash equilibrium (SPNE) of this game is easily obtained by backward induction, and the problem we address is to develop a characterization of the SPNE payoffs, denoted $w(\rho, T)$, for a low cost of delay factor $1-\rho$ and a large number T of potential rounds of negotiation.

We show that the limit of $w(\rho, T)$ depends on the relative rates at which $1-\rho$ converges to 0 and T converges to infinity. As ρ converges to 1 and T converges to infinity, in such a manner that the probability ρ^T of all players remaining at the last stage of the game converges to 1 — thus the convergence of T to infinity is much slower than the convergence of ρ to 1 — the SPNE payoffs $w(\rho, T)$ converge to r, a well defined, efficient point; we call r the *Raiffa point*. In the two-player case, this result was obtained by Sjöström (1991), the point r being the Raiffa bargaining solution (see Luce and Raiffa (1957, §6.7)).

Next, assume that the rate at which T converges to infinity increases, so that the probability of all players remaining at the last stage converges now to some $\mu < 1$ (i.e., $\rho^T \to \mu$ or $T(1-\rho) \to -\ln\mu$). We prove that in this case the SPNE payoffs $w(\rho, T)$ converge to the solution, at time $t = -\ln\mu \in [0, \infty)$, of the dynamic system associated with the consistent field and having the Raiffa point r as its initial condition at $t = 0$. We also show that if this solution trajectory starting at r converges, as t goes to infinity, to a (local) attractor a of the consistent field, then a is the limit of any sequence of SPNE payoff when ρ^T converges to zero (and, of course, $\rho \to 1$ and $T \to \infty$). Finally, we show that any point in the limit set of the trajectory of the consistent field solutions through the Raiffa point can be reached as the limit of SPNE payoffs of an appropriate sequence of finite horizon games with $\rho^T \to 0$ (and $\rho \to 1$, $T \to \infty$).

All these results indicate that we can attach significance to the dynamic properties of the consistent field both on cooperative and noncooperative theory grounds, and therefore we conclude that it is a vector field well worth analyzing in more depth. In that vein, we show that the global dynamics are convergent to the unique consistent value in the pure bargaining and in the hyperplane games cases. For the general case we analyze the local dynamics of the consistent field around a consistent value. We show that this local dynamics is composed of a "game part", which depends only on the particular consistent value, and a "geometry part", which depends only on the curvature of the Pareto frontier at the consistent value. Exploiting this relationship we can construct examples with a wide variety of local behaviors: sink, source, saddle point. We can also, using Hopf bifurcation theory, construct an NTU game where the consistent field has cyclical solutions, thus indicating that the limit of SPNE solutions of finite horizon bargaining games could well be a point which is *not* a consistent value. This is in contrast to the global convergence of the consistent field in the TU, hyperplane, and pure bargaining cases. All this confirms once more the intuition that pure bargaining games and TU (or, more generally, hyperplane) games are the most well behaved of the NTU games, and thus the easier to analyze. The game theoretic behavior of general NTU games is however considerably more complex than what one may be led to suspect from an analysis of these two cases. We refer to Hart and Mas-Colell (1996a) for further elaboration of this point.

The paper is organized as follows: Section 2 presents the basic model — the underlying cooperative game in coalitional form, and the noncooperative bargaining game — followed by a preliminary analysis of the subgame perfect equilibria of the latter. Section 3 recalls the definition of the consistent NTU-value, and introduces the consistent field and its associated dynamics. The results connecting the SPNE payoffs with the dynamics of the consistent field are stated in Section 4. A (local) analysis of the consistent field is carried out in Section 5; we then provide various examples for the behavior of the resulting dynamics (and thus, *a fortiori*, of the SPNE payoffs). Proofs are relegated to the Appendix.

2. The Model

Let (N,V) be a *non-transferable utility (NTU) n-person game in coalitional form*. The set of players is $N = \{1,2,\ldots,n\}$ and V is the coalitional (characteristic) function. For each coalition $S \subset N$, the set $V(S)$ — a subset of $\Re^S$ — is the set of all allocations that are feasible for the members of S.

We make the following standard assumptions on (N,V):

(A1) For each coalition S, the set $V(S)$ is *closed*, *convex* and *comprehensive* (i.e., if $x \in V(S)$ and $y \leq x$ then $y \in V(S)$). Moreover, $0 \in V(S)$.

(A2) For each coalition S, the boundary (or Pareto frontier) of $V(S)$, denoted by $\partial V(S)$, is C^2 (i.e., at each boundary point there is a single outward normal direction, which varies in a continuously differentiable manner with the point) and *nonlevel* (i.e., the outward normal vector at any point of $\partial V(S)$ is positive in all coordinates).

(A3) *Monotonicity*: If $Z \subset S$ then $V(Z) \times \{0^{S\setminus Z}\} \subset V(S)$ (i.e., completing a vector in $V(Z)$ with 0's for the coordinates in $S\setminus Z$ results in a vector in $V(S)$).

The noncooperative game we analyze is the finitely repeated version of the game introduced by Hart and Mas-Colell (1996b). The *n-person noncooperative bargaining game* (N,V,ρ,T), where $\rho \in [0,1]$ and T is a positive integer, is described inductively as follows:

> The game is a perfect information game consisting of at most T rounds of negotiation. In each round t there is a set $S \subset N$ of active players who can reach an agreement, starting in the first round $(t=1)$ with $S=N$. One player in S is chosen randomly, with all players in S equally likely to be selected. Say player i has been chosen. Then i proposes a feasible payoff vector in $V(S)$ to the other players in S. They can either agree or not (they are asked in some prespecified order). The game ends with the proposed payoffs if all players in S agree, or with payoffs equal to 0 if it is the last round $t=T$ and there is no agreement. Otherwise, the game moves to the next round $t+1$, where with probability ρ the set of active players does not change, and with probability $1-\rho$ it becomes $S\backslash i$. In the latter case, the payoff of the "dropped out" player i is 0.

The *subgame perfect Nash equilibrium (SPNE)* of the above finite game with perfect information can be easily obtained by backward induction. Suppose that in the last step of negotiation, T, the players in the coalition S have "survived". The equilibrium strategies for these remaining players are as follows. If player $i \in S$ is chosen to be the proposer (which occurs with probability $1/|S|$), then i's strategy is to propose that vector $e_{S,i} \in \partial V(S)$ such that $e^j_{S,i} = 0$ for all $j \in S\backslash i$ (efficiency then uniquely determines the payoff of player i). The strategy for each $j \in S\backslash i$ is to accept any $x_S \in V(S)$ such that $x^j_S \geq 0$. The (expected) equilibrium payoff vector (before the selection of the proposer) is then $e_S := (1/|S|)\sum_{i\in S} e_{S,i}$ (note that the convexity of $V(S)$ implies that $e_S \in V(S)$).

At round $T-1$ of the negotiation the SPNE strategies are as follows. Suppose that S is the set of remaining players at this stage. If player $i \in S$ is chosen to be the proposer then i's strategy is to propose that allocation $a_{S,i} \in \partial V(S)$ such that $a^j_{S,i} =$

$\rho e_S^j + (1-\rho)e_{S\backslash i}^j$ for all $j \in S\backslash i$. The strategy for each $j \in S\backslash i$ is to accept any $x_S \in V(S)$ such that $x_S^j \geq \rho e_S^j + (1-\rho)e_{S\backslash i}^j$.

The strategy profile above is the unique subgame perfect equilibrium strategy for the game starting at round $T-1$. To prove that, observe that with probability ρ the proposer i will remain for the next and last stage, and the expected payoff of the remaining S players in the continuation game is, as seen above, e_S. With probability $1-\rho$ player i will drop out of the game, and the expected payoff of the remaining players in the continuation game is $e_{S\backslash i}$. It follows that the most player j expects to get by rejecting an offer of the proposer i is $\rho e_S^j + (1-\rho)e_{S\backslash i}^j$, which implies that the strategy profile is the unique SPNE of the game.

To formalize this, define a *payoff configuration (p.c.) a* to be a collection of payoff vectors for all coalitions: $a = (a_S)_{S\subset N}$ with $a_S \in V(S)$ for all $S \subset N$. The backward induction arguments are then captured by the following function $F_\rho : V \to V$, where $V := \prod_{S\subset N} V(S)$ is the set of all payoff configurations.

Definition 2.1. *The* backward induction function $F_\rho : V \to V$ *maps each payoff configuration $a = (a_S)_{S\subset N} \in V$ to a payoff configuration $F_\rho(a) = (b_S)_{S\subset N} \in V$ given by*

(i) $b_{S,i}^j = \rho a_S^j + (1-\rho)a_{S\backslash i}^j$ *for all $S \subset N$ and all $j \in S\backslash i$;*
(ii) $b_{S,i} \in \partial V(S)$ *for all $i \in S \subset N$;*
(iii) $b_S = \frac{1}{|S|}\sum_{i\in S} b_{S,i}$ *for all $S \subset N$.*

The backward induction function provides the expected payoffs $F_\rho(a)$ at any stage of the game, given that the payoffs in the continuation game are specified by the configuration $a = (a_S)_{S\subset N}$. The function F_ρ is well-defined because of the assumptions (A1) - (A3) imposed on the game (N,V).

The *SPNE payoff configuration* of the noncooperative game (N,V,ρ,T) can be conveniently represented as the payoff configuration $w(\rho,T) = (w_S(\rho,T))_{S\subset N}$ where $w_S(\rho,T)$ is the $|S|$-dimensional vector representing the unique SPNE payoff vector of the noncooperative game $\left(S, V_{|S}, \rho, T\right)$ restricted to the coalition $S \subset N$.

We have just proven above that the continuation games of the noncooperative game (N,V,ρ,T) starting at round T and $T-1$ have SPNE payoff configurations given by $w(\rho,1)=e$ and $w(\rho,2)=F_\rho(e)$, respectively. Now[1] $e=F_\rho(0)$, where $0=(0^S)_{S\subset N}$ is the payoff configuration with all coalitional payoff vectors equal to 0, and $F_\rho(e)=F_\rho(F_\rho(0))=F_\rho^2(0)$ is the second iterate of F_ρ evaluated at 0. Proceeding inductively one obtains the strategy profile of the SPNE and its corresponding payoff configuration: $w(\rho,t)=F_\rho(w(\rho,t-1))$ for all t. Therefore

Proposition 2.2. *The SPNE payoff configuration of the noncooperative game (N,V,ρ,T) is given by $w(\rho,T)=F_\rho^T(0)$, the T-th iterate of the function F_ρ evaluated at the payoff configuration 0, where F_ρ, the backward induction function, is given by Definition 2.1 above.*

3. The Consistent Field

We now turn to the study of dynamics associated with the concept of the consistent NTU-value, which was introduced by Maschler and Owen (1989, 1992) and analyzed by Hart and Mas-Colell (1996b). In this section we develop the concept of the *consistent field*. This is a vector field defined over the Pareto frontier of the game that, informally speaking, gives the direction that reduces the "inconsistency" in the payoff configuration.

We start by recalling the definition of the *consistent value* of an NTU-game (N,V). Similarly to the Shapley value, let π be a permutation of the n players, and define recursively the vector of marginal contributions d_π (with $\pi(i)$-th coordinate $d_\pi^{\pi(i)}$) by

$$\begin{aligned} d_\pi^{\pi(1)} &= \max\left\{a^{\pi(1)} : a\in V(\{\pi(1)\})\right\}, \text{and for } i>1 \text{ by} \\ d_\pi^{\pi(i)} &= \max\left\{a^{\pi(i)} : a\in V(\{\pi(1),\ldots,\pi(i)\}) \text{ and } a^{\pi(j)}=d_\pi^{\pi(j)} \text{ for all } j<i\right\}. \end{aligned}$$

So, for a given order π, each player $\pi(i)$ gets $d_\pi^{\pi(i)}$, which is the highest possible given that all the previous players $\pi(j)$ (for $j<i$) got $d_\pi^{\pi(j)}$. Consider now the vector of expected marginal contributions $\Psi(N,V):=(1/n!)\sum_\pi d_\pi$. Since $\Psi(N,V)$ is an average

[1] We thank Vincent Feltkamp for pointing out that the induction may be conveniently started at 0.

of vectors on the boundary $\partial V(N)$ of the convex set $V(N)$, it will not in general be efficient. However, for a *hyperplane game* (where, for each $S \subset N$, the set $V(S)$ is a half-space, and so its boundary $\partial V(S)$ is a hyperplane), the expected marginal contribution vector $\Psi(N,V)$ is efficient. It is the *consistent value* of the hyperplane game[2] (N,V). Further, for each coalition $S \subset N$ we have $\Psi(S,V) \in \partial V(S)$; the efficient payoff configuration $(\Psi(S,V))_{S\subset N}$ is called the *consistent value payoff configuration* of the hyperplane game (N,V).

For a general NTU-game, the construction of the consistent value is based on the hyperplane case. For each efficient payoff configuration $a = (a_S)_{S\subset N} \in \partial V := \prod_{S\subset N} \partial V(S)$ with supporting normal vectors $\lambda_S \equiv \lambda_S(a_S) \in \Re^S_{++}$ to the boundary $\partial V(S)$ at a_S, associate the *supporting hyperplane game* (N, V'_a) defined by $V'_a(S) := \left\{c \in \Re^S : \lambda_S \cdot c \leq \lambda_S \cdot a_S\right\}$ for all $S \subset N$. Let $b \equiv b(a) := (\Psi(S, V'_a))_{S\subset N}$ be the consistent p.c. of the supporting hyperplane game (N, V'_a). If $b = a$ then a is a *consistent value payoff configuration* of the NTU-game (N,V).

Following Hart and Mas-Colell (1996b, Proposition 4), a consistent value payoff configuration $a = (a_S)_{S\subset N}$ for the game (N,V) can be characterized by

(i) $a_S \in \partial V(S)$ for all $S \subset N$;

(ii) $\lambda_S \cdot a_S = \max\{\lambda_S \cdot c : c \in V(S)\}$ for all $S \subset N$; and

(iii) $\sum_{j\in S\setminus i} \lambda^i_S (a^i_S - a^i_{S\setminus j}) = \sum_{j\in S\setminus i} \lambda^j_S (a^j_S - a^j_{S\setminus i})$ for all $S \subset N$ and each $i \in S$.

Conditions (i) and (ii) state that the payoff vector a_S is on the Pareto frontier of the coalition S and that λ_S is an outward normal vector to the boundary of $V(S)$ there. The last condition (iii) may be viewed as a "preservation of average differences" requirement: the average contribution to i from the remaining players equals the average contribution of i to the remaining players. We refer to Hart and Mas-Colell (1996b) for further details. In particular, under our assumptions, consistent value p.c.'s exist and are always non-negative.

[2]In the special that (N,V) is a TU-game, this is the Shapley value.

Maschler, Owen and Peleg (1988) and Maschler and Owen (1989) have proposed dynamic processes adapted to the consistent value for pure bargaining games[3] and for hyperplane games. We proceed to do this here for the general NTU case.

The dynamic approach to the consistent value that we analyze is an explicit procedure that, starting from an arbitrary efficient p.c., adjusts the payoffs in the direction indicated by the above characterization of the consistent value. The adjustment process can be described as follows. Given an efficient p.c. $a = (a_S)_{S \subset N} \in \partial V$, the payoffs for each coalition S are adjusted, assuming that the players in S already agree with the payoffs a_Z for the smaller coalitions $Z \subsetneqq S$. Considering the supporting hyperplane game (S, V'_a) at a, and fixing the payoffs for the subcoalitions of S, then, in order to bring about consistency for the coalition S, the payoff a_S would need to be changed to a payoff $b_S(a)$ in the hyperplane $V'_a(S)$ satisfying the preservation of average differences; i.e.,

$$\begin{aligned} \lambda_S(a_S) \cdot b_S(a) &= \lambda_S(a_S) \cdot a_S, \text{ and} \\ \sum_{j \in S \setminus i} \lambda_S^i(a_S)(b_S^i(a) - a_{S \setminus j}^i) &= \sum_{j \in S \setminus i} \lambda_S^j(a_S)(b_S^j(a) - a_{S \setminus i}^j). \end{aligned} \tag{3.1}$$

The change in the payoff a_S is equal to $C_S(a) := b_S(a) - a_S$, and the *consistent field* at a is defined to be $C(a) = (C_S(a))_{S \subset N}$. Thus, the S-coordinate C_S of the consistent field vector gives the direction in which to move locally along the efficient frontier $\partial V(S)$ so that the consistency of the payoffs for the players in coalition S is reduced, given that the payoffs for the subcoalitions of S are unchanged. The explicit expression produced by (3.1) is given in the following definition.

Definition 3.1. *The* consistent field *(or* C-field*) associated with the NTU-game (N, V) is the vector field $C(\cdot)$ over the boundary ∂V, with $C(a) = (C_S(a))_{S \subset N}$ for any $a \in \partial V$ defined by the expression*

$$C_S^i(a) = \frac{1}{|S| \lambda_S^i(a)} \sum_{j \in S \setminus i} \left(\lambda_S^j(a)(a_S^j - a_{S \setminus i}^j) - \lambda_S^i(a)(a_S^i - a_{S \setminus j}^i) \right) \tag{3.2}$$

[3] In the pure bargaining case, the consistent value coincides with the Nash bargaining solution.

for all $S \subset N$ and $i \in S$, where $\lambda_S(a)$ is the unit normal vector to the boundary $\partial V(S)$ at a_S.

It is a simple computation to verify the equivalence of (3.1) and (3.2); it is actually useful to write this expression yet another way, namely

$$C_S^i(a) = \frac{1}{|S|\lambda_S^i(a)} \left(\lambda_S(a) \cdot a_S - \sum_{j \in S \setminus i} \left(\lambda_S^j(a) a_{S\setminus i}^j - \lambda_S^i(a) a_{S\setminus j}^i \right) \right) - a_S^i. \tag{3.3}$$

Note that $\lambda_S(a) \cdot C_S(a) = 0$, thus C is indeed a vector field over the boundary ∂V. The singularities of the consistent field, i.e., the payoff configurations a such that $C(a) = 0$, are precisely the consistent value p.c.'s of the game (N,V). Finally, note that whenever $a \geq 0$ and $a_S^i = 0$ for some i and S, then $C_S^i(a) \geq 0$ (see (3.3) and use $\lambda_S(a) \cdot a_S \geq \lambda_S(a) \cdot (a_{S\setminus i}, 0)$, which holds since $(a_{S\setminus i}, 0) \in V(S)$ by the monotonicity assumption (A3)). Thus the consistent vector field "points inward" at boundary points of $\partial_+ V := \partial V \cap \{a \geq 0\}$.

Associated with the consistent field C over the boundary ∂V is the dynamic system $da/dt = C(a)$. For any a in the non-negative part of the boundary $\partial_+ V$, there is a unique function $\Lambda_t(a) : [0, \infty) \to \partial_+ V$ that satisfies

$$\begin{aligned} \frac{d\Lambda_t(a)}{dt} &= C(\Lambda_t(a)), \text{ and} \\ \Lambda_0(a) &= a. \end{aligned} \tag{3.4}$$

We will refer to $\Lambda_t(a)$ as the *solution of the consistent field* starting, at $t = 0$, from the non-negative efficient p.c. a. Note that the solutions are defined on the interval $[0, \infty)$ because, by the "pointing inwards" property of the field, every solution that starts in the non-negative part of the boundary will remain there. For $t = \infty$, we define $\Lambda_\infty(a)$ as the ω-limit set of the solution, i.e., the set of all limit points of $\Lambda_t(a)$ as $t \to \infty$.

Example 1: Pure Bargaining Games

An n-person pure bargaining game satisfies $V(S) \cap \Re_+^S = \{0\}$ for all $S \neq N$. For this particular case, the consistent field has the same dynamics as the process introduced by

Maschler, Owen and Peleg (1988). For $S \neq N$ we have $(\Lambda_t(a))_S \equiv 0$; so, the only dynamics that matters is for $S = N$. For any $a = a_N \in \partial_+ V(N)$ with supporting normal $\lambda(a) = \lambda_N(a)$, (3.2) becomes

$$C^i(a) = \frac{1}{n\lambda^i(a)} \sum_{j \in S \backslash i} \left(\lambda^j(a)a^j - \lambda^i(a)a^i\right)$$

(we have dropped the subscript N throughout for ease of reading). The dynamics of the C-field is simple: the unique singularity of the C-field, which is the unique consistent value and coincides with the Nash bargaining solution, is a global attractor for the field. (This can be verified by showing that the function $L(a) := \prod_{i \in N} a^i$ is a Lyapunov function for the C-field; see Maschler, Owen and Peleg (1988)).

Example 2: Hyperplane Games

Let the hyperplane game (N, V) be given for each coalition S by the unit normal vector λ_S and by the number ν_S such that $V(S) = \left\{a_S \in \Re^S : \lambda_S \cdot a_S \leq \nu_S\right\}$. The expression (3.3) for the C-field becomes

$$C^i_S(a) = \frac{1}{|S|\lambda^i_S}\left(\nu_S - \sum_{j \in S \backslash i} (\lambda^j_S a^j_{S \backslash i} - \lambda^i_S a^i_{S \backslash j})\right) - a^i_S.$$

First observe that the C-field is a linear function of a (since λ_S does not change with a_S) and that the unique singularity of the C-field is the unique consistent value. To characterize the dynamics of the C-field we simply need to determine the sign of the real part of the eigenvalues of $DC(a)$, the derivative of the field at a. The expression above for $C^i_S(a)$ immediately implies that the matrix $DC(a)$ is triangular and all its diagonal entries are -1 (indeed: $\partial C^i_S(a)/\partial a^j_Z = 0$ if Z is not a subset of S; moreover, $\partial C^i_S(a)/\partial a^i_S = -1$ and $\partial C^i_S(a)/\partial a^j_S = 0$ for $j \neq i$). Therefore all the eigenvalues of $DC(a)$ are equal to -1, implying that the solution of the C-field converges exponentially to the consistent value (e.g., see Palis and de Melo (1982)). Again, we conclude that the dynamics is very simple: there is only one consistent value which is a global attractor. A result similar to this has been obtained by Maschler and Owen (1989) for the "correction function" they propose.

4. From Subgame Perfect Equilibria to Consistent Field Solutions

We now address the problem of characterizing the SPNE solutions of the noncooperative game (N, V, ρ, T) as the probability of breakdown decreases to zero and the number of periods of negotiation increases. That is, we want to find the limit of the SPNE payoff configurations $w(\rho, T)$ as $\rho \to 1$ and $T \to \infty$. Recall the result of Proposition 2.2 that $w(\rho, T) = F_\rho^T(0)$.

This section will show that the solutions Λ_t of the consistent field are intimately related to the subgame perfect Nash equilibria of the finitely repeated noncooperative bargaining game. We start by highlighting a basic relationship between the consistent field C and the backward induction function F_ρ. All the proofs are in the Appendix.

Proposition 4.1. *The derivative of* $F_\rho(a)$ *with respect to* ρ *at any point* $a \in \partial V$ *satisfies*

$$\frac{dF_\rho(a)}{d\rho}_{|\rho=1} = -C(a).$$

This result will be most useful because we are interested in the limit as ρ converges to 1 of the iterates of the function F_ρ. Observing that for $\rho = 1$ we have $F_1(a) = a$ for all $a \in \partial V$, the result roughly states that, for efficient payoffs and ρ close enough to 1, $F_\rho(a)$ can be approximated by $a + (1-\rho)\,C(a)$. This suggests a natural relationship between the limit of F_ρ^T as $T \to \infty$ and $\rho \to 1$ and the solution of the dynamic system associated with the consistent field.

We first consider the case where there is no breakdown, i.e. when $\rho = 1$.

Proposition 4.2. *For any payoff configuration* $a \in V$*, the limit as* $T \to \infty$ *of* $F_1^T(a)$ *exists and is an efficient payoff configuration:* $\lim_{T\to\infty} F_1^T(a) \in \partial V$*.*

In particular, this proposition implies that the limit as $T \to \infty$ of the SPNE payoffs, i.e. $\lim_{T\to\infty} F_1^T(0)$, exists and is efficient (this was shown by Sjöström (1991) in the two-player case). We call this point the Raiffa point (see Luce and Raiffa (1957, §6.7)).

Definition 4.3. *The* Raiffa payoff configuration *of the game* (N, V)*, denoted* $r \equiv r(N, V)$*, is given by* $r := \lim_{T\to\infty} F_1^T(0)$.

Depending on the rates at which $\rho \to 1$ and $T \to \infty$, the SPNE p.c. may converge to different limits. Specifically, those turn out to depend on the limit of ρ^T, the probability that all players remain up to the last stage of the game. If $\rho \to 1$, $T \to \infty$ and $\rho^T \to 1$, meaning that T converges to infinity much slower than ρ converges to 1, then we will see that $w(\rho, T)$ converges to the Raiffa p.c. r. As the relative rate at which T converges to infinity increases, so that the probability ρ^T of all players remaining at the last stage converges to some $\mu < 1$, then we will see that $w(\rho, T)$ converges to an appropriate point on the solution path of the consistent field starting at r. Formally,

Theorem 4.4. *Let r be the Raiffa payoff configuration (given by Definition 4.3) and let Λ_t be the solution of the consistent field (given by (3.4)). Then:*

(i) If $\rho \to 1$, $T \to \infty$ and $\rho^T \to \mu \in (0,1]$, then $w(\rho, T)$ converges to $\Lambda_{-\ln \mu}(r)$.

(ii) If $\Lambda_t(r)$ converges as $t \to \infty$ to a local attractor (sink) a of the consistent field, then $w(\rho, T)$ converges to a as $\rho \to 1$, $T \to \infty$ and $\rho^T \to 0$.

(iii) Any payoff configuration in $\Lambda_\infty(r)$, the ω-limit set of the solution of the consistent field through r, can be obtained as the limit of $w(\rho_k, T_k)$ for appropriate sequences $\rho_k \to 0$ and $T_k \to \infty$ with $\rho_k^{T_k} \to 0$ (as $k \to \infty$).

Note that when $\mu = 1$ we get $\Lambda_{-\ln \mu}(r) = \Lambda_0(r) = r$, therefore $w(\rho, T)$ converges to the Raiffa point r as $\rho \to 1$, $T \to \infty$ and $\rho^T \to 1$. As for the second part of the theorem, it includes the two cases of pure bargaining and of hyperplane games, where, as we saw in Examples 1 and 2 of Section 3, the unique consistent value is a global attractor.

5. Local Analysis of the Consistent Field

The results of Theorem 4.4 indicate that the dynamic properties of the consistent field are of importance in describing the solutions of the noncooperative bargaining games. We now proceed to analyze in more depth the dynamics of the consistent field.

The dynamics of the consistent field for a general NTU game can be significantly more complicated than the dynamics for pure bargaining games and for hyperplane games. For these two particular cases, as shown in Section 3, there is a unique consistent value which is a global attractor for the C-field.

We propose to study the dynamic properties of the consistent field in a neighborhood of a consistent value. As it is well known, the linear system $DC(a)\,(x-a)$, where DC is the derivative of C, can be used as an approximation of the consistent field $C(x)$ around the consistent value a. By a standard result in dynamic system (the Hartman-Grobman Theorem, e.g., Palis and de Melo (1982)), the local dynamics of the C-field and the dynamics of the linear system are equivalent if the consistent value is a hyperbolic equilibrium (i.e., if the eigenvalues of $DC(a)$ have non-zero real part). Moreover, these dynamics are determined by the eigenvalues of $DC(a)$.

As a first step we develop an expression for the derivative of the C-field at a consistent value.

Theorem 5.1. *The derivative of the C-field $DC(a)$ at a consistent value payoff configuration $a = (a_S)_{S\subset N}$ is a block triangular matrix. For all $S \subset N$, the diagonal block matrix corresponding to S is $DC_S(a) : T_{a_S}\partial V(S) \to T_{a_S}\partial V(S)$, where $T_{a_S}\partial V(S)$ is the tangent plane to the boundary $\partial V(S)$ at a_S. Moreover, $DC_S(a)v_S$ for any $v_S \in T_{a_S}\partial V(S)$ can be expressed as*

$$DC_S(a)v_S = \underbrace{G_S(a)}_{\text{game part}}\ \underbrace{D\lambda_S(a_S)}_{\text{geometry part}}\ v_S\ -\ v_S$$

where $\lambda_S(a_S)$ is the unit length outward normal to $\partial V(S)$ at a_S and $G_S(a) = (G_S^{ij})_{i,j\in S}$ is the matrix given by

$$G_S^{ii} = -\frac{1}{|S|\lambda_S^i(a_S)} \sum_{j\in S\setminus i} (a_S^i - a_{S\setminus j}^i) \text{ and}$$

$$G_S^{ij} = \frac{1}{|S|\lambda_S^i(a_S)} \left(a_S^j - a_{S\setminus i}^j\right), \text{ for } i \neq j.$$

The derivative $DC_S(a)$ is thus naturally decomposed into a *game part* $G_S(a)$, which depends only on a and $\lambda_S(a_S)$, and a *geometry part* $D\lambda_S(a_S)$, which is the Gauss curvature map of the boundary $\partial V(S)$ at a_S. The theorem is proved in the Appendix.

We now proceed to exhibit a family of NTU games with various dynamics for the consistent field around a consistent value: repulsor (source), saddle point, and cycle. (Recall that in the cases of pure bargaining and hyperplane games there is a unique global attractor.)

Example 3: A family of consistent fields with varied local dynamics

There are 3 players: $N = \{1,2,3\}$. We begin by fixing the games in all but the grand coalition:

$$\begin{aligned} V(1) &= V(2) = V(3) = \left\{c \in \Re^1 : c \leq 0\right\} \\ V(12) &= \left\{(c_1,c_2) \in \Re^2 : \frac{c_1}{14} + \frac{c_2}{9} \leq 2\right\} \\ V(23) &= \left\{(c_2,c_3) \in \Re^2 : \frac{c_2}{14} + \frac{c_3}{9} \leq 2\right\} \\ V(13) &= \left\{(c_1,c_3) \in \Re^2 : \frac{c_3}{14} + \frac{c_1}{9} \leq 2\right\}. \end{aligned}$$

Every subgame is therefore a hyperplane game, and, for all $S \subsetneqq \{1,2,3\}$, the consistent solution $a_{S\subsetneqq}$ of the subgame (S,V) — which is unique — coincides with the Raiffa point (and the Nash bargaining solution). Specifically:

$$\begin{aligned} a_1 &= a_2 = a_3 = 0 \\ a_{12} &= (14, 9, \cdot) \\ a_{23} &= (\cdot, 14, 9) \\ a_{13} &= (9, \cdot, 14). \end{aligned} \tag{5.1}$$

We now come to the construction of $V_h(123)$, which will depend on a parameter $h = (h_1,h_2)$, where $0 \leq h_1, h_2 \leq 8$. Consider first the family of ellipsoids

$$E_h := \left\{x \in \Re^3 : z = M^T x \text{ is such that } (z_1)^2 + \sqrt{3}h_1(z_2)^2 + \sqrt{3}h_2(z_3)^2 = 3\right\},$$

where M is the orthonormal matrix

$$M := \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{pmatrix}.$$

This family is such that $a_{123} := (10,10,10)$ belongs to E_h and the unit normal to E_h at $(10,10,10)$ is $(1/\sqrt{3},1/\sqrt{3},1/\sqrt{3})$, for all h. Moreover, if $V_h(123)$ equals $E_h^- := E_h + \Re^3_-$, the comprehensive hull of E_h, then it is immediate to check that the payoff configuration a (given by (5.1) and $a_{123} = (10,10,10)$) is a consistent value.

However, it is not possible to put $V_h(123) = E_h^-$, because the monotonicity of the game would be violated. But this can be fixed. Note first that if we were to take $V(123) = W := \left\{(c_1,c_2,c_3) \in \Re^3 : c_1 + c_2 + c_3 \leq 30\right\}$ — which has the same unit normal at $(10,10,10)$ — then the game would be monotone. This implies that we can let $V_h(123)$ be a set that is close to W, it coincides with E_h^- in an appropriately small ε-neighborhood of $(10,10,10)$, and is such that all our assumptions on the game are satisfied.

Next, observe that if $V(123) = W$ then the grand coalition's component of the Raiffa point is $(10,10,10)$. Suppose we were to alter this set by modifications that affect only neighborhoods of the intersections of W with the three axes. Then $(10,10,10)$ would remain a consistent value, but the Raiffa point would move. Moreover, by suitable such small modifications we can place the Raiffa point anywhere we want in the ε-neighborhood of $(10,10,10)$ (for ε sufficiently small). By continuity, all this remains true for $V_h(123)$. Summarizing: in constructing $V_h(123)$ we can also make sure that the Raiffa point for the grand coalition is placed at a specific point — to be determined later — in the ε-neighborhood of $(10,10,10)$.

We are now ready to exhibit values of the parameter h for which the local behavior of the consistent field at a is not attracting. A straightforward computation yields that the two non-zero eigenvalues of the matrix $DC_{123}(a)$ are

$$\frac{3\sqrt{3}}{40}(h_1+h_2) - 1 \pm \frac{1}{40}\sqrt{27h_1^2 - 154h_1h_2 + 27h_2^2}.$$

Therefore:

(i) For $h_1 = h_2 = 5$, the two eigenvalues have positive real part. Hence the consistent value a is a repulsor (source).

(ii) For $h_1 = 1$ and $h_2 = 6$, the two eigenvalues are real, one is positive and the other is negative. Hence the consistent value a is a saddle point.

(iii) For $h_1 = 3$ and $h_2 = 40\sqrt{3}/9 - 3 \approx 4.698$, the eigenvalues have zero real part and non-zero imaginary part. Moreover, as h_1 moves from below 3 to above 3 (and h_2 is unchanged), the real part moves from negative to positive. Therefore, by the Hopf Bifurcation Theorem,[4] there must be values of h_1 close to 3 such that there is a cycle in the ε-neighborhood of a. In particular, the Raiffa point for the grand coalition may be located on the cycle (remember that we can place it anywhere in the ε-neighborhood, without affecting the local behavior of the consistent field in this neighborhood). In this case, by Theorem 4.4, we get an example of a sequence of finite horizon bargaining problems with $\rho \to 1, T \to \infty$ and $\rho^T \to 0$ and such that the subgame perfect Nash equilibrium payoffs converge to a point (on the cycle) that is *not* a consistent value.

A. Appendix

This appendix contains the proofs not given in text.

Proof of Proposition 4.1. Define the function $b_{S,i}(\rho, a)$ by

$$\begin{aligned} b^j_{S,i}(\rho, a) &= \rho a^j_S + (1-\rho) a^j_{S\setminus i}, \text{for all } j \in S\setminus i \\ b^i_{S,i}(\rho, a) &= h^i_S\left(\rho a^{-i}_S + (1-\rho) a_{S\setminus i}\right) \end{aligned}$$

where a^{-i}_S denotes the vector a_S without the i-th coordinate, and $h^i_S(c^{-i}_S)$ is defined as the i-th coordinate of the point on the boundary of $V(S)$ with remaining $S\setminus i$ coordinates equal to c^{-i}_S. To simplify notation we will omit the arguments (ρ, a) of the function $b_{S,i}$. By definition, $\left(F_\rho(a)\right)_S = (1/|S|)\sum_{i\in S} b_{S,i}$.

[4]Hopf (1942) Bifurcation Theorem: Let $x' = F(x, \alpha)$ be a one-parameter family of planar systems with an equilibrium $x(\alpha)$ and eigenvalues $\lambda(\alpha) = \mu(\alpha) + i\eta(\alpha)$. Suppose that, for some value α^0 of the parameter, the equilibrium $x(\alpha^0)$ is non-hyperbolic with purely imaginary eigenvalues (i.e., $\mu(\alpha^0) = 0$). Moreover, as α crosses α^0 in some direction, $\mu(\alpha)$ changes from negative to positive and $x(\alpha)$ changes from sink to source. Then, for all α on the side of α^0, and close enough to it, there is a periodic orbit surrounding the equilibrium $x(\alpha)$ with radius of magnitude $|\alpha - \alpha^0|^{1/2}$. Also, if $x(\alpha^0)$ is asymptotically stable, then the closed orbit is stable and surrounds the unstable equilibrium. Otherwise, the closed orbit is unstable and occurs for parameter values that make the equilibrium a sink.

We first obtain the derivative of $b_{S,i}$ with respect to ρ:

$$\begin{aligned}
\frac{\partial b_{S,i}^j}{\partial \rho} &= a_S^j - a_{S\setminus i}^j, \text{ for all } j \in S\setminus i, \text{ and} \\
\frac{\partial b_{S,i}^i}{\partial \rho} &= \nabla h_S^i \left(\rho a_S^{-i} + (1-\rho) a_{S\setminus i} \right) \cdot \left(a_S^{-i} - a_{S\setminus i} \right).
\end{aligned}$$

Using the fact that $(\partial h_S^i / \partial a_j) \left(a_S^{-i} \right) = -\lambda_S^j (a_S) / \lambda_S^i (a_S)$ for a point $a_S \in \partial V(S)$, we obtain that the derivative $\partial b_{S,i}/\partial \rho$ evaluated at $\rho = 1$ is

$$\begin{aligned}
\frac{\partial b_{S,i}^j}{\partial \rho}_{|\rho=1} &= a_S^j - a_{S\setminus i}^j, \text{ for all } j \in S\setminus i, \text{ and} \\
\frac{\partial b_{S,i}^i}{\partial \rho}_{|\rho=1} &= -\frac{1}{\lambda_S^i (a_S)} \sum_{j \in S\setminus i} \lambda_S^j (a_S) \left(a_S^j - a_{S\setminus i}^j \right).
\end{aligned}$$

Substituting into the expression of $\left(F_\rho(a)\right)_S$ yields:

$$\begin{aligned}
\frac{\partial \left(F_\rho(a)\right)_S^i}{\partial \rho}_{|\rho=1} &= \frac{1}{|S|} \sum_{j \in S\setminus i} \frac{\partial b_{S,j}^i}{\partial \rho}_{|\rho=1} + \frac{1}{|S|} \frac{\partial b_{S,i}^i}{\partial \rho}_{|\rho=1} \\
&= \frac{1}{|S|} \sum_{j \in S\setminus i} \left(a_S^i - a_{S\setminus j}^i \right) - \frac{1}{|S| \lambda_S^i (a_S)} \sum_{j \in S\setminus i} \lambda_S^j (a_S) \left(a_S^j - a_{S\setminus i}^j \right) \\
&= \frac{1}{|S| \lambda_S^i (a_S)} \sum_{j \in S\setminus i} \left(\lambda_S^i (a_S) \left(a_S^i - a_{S\setminus j}^i \right) - \lambda_S^j (a_S) \left(a_S^j - a_{S\setminus i}^j \right) \right) \\
&= -C_S^i (a),
\end{aligned}$$

which proves the result.

Proof of Proposition 4.2. Let $a \in V$ and put $b := F_1(a)$. For each $S \subset N$ and each $i \in S$, let $\alpha_S^i \geq 0$ be such that $(a_S^i + \alpha_S^i, a_S^{-i}) \in \partial V(S)$. Then $b_S^i = a_S^i + (1/s)\alpha_S^i$, where $s := |S|$. Similarly, let $\beta_S^i \geq 0$ be such that $(b_S^i + \beta_S^i, b_S^{-i}) \in \partial V(S)$. Because $b_S^{-i} \geq a_S^{-i}$, comprehensiveness implies that $b_S^i + \beta_S^i \leq a_S^i + \alpha_S^i$, or $\beta_S^i \leq \alpha_S^i - (1/s)\alpha_S^i = (1-1/s)\alpha_S^i$; hence $\max_{i \in S} \beta_S^i \leq (1-1/s) \max_{i \in S} \alpha_S^i$. Therefore each iteration of F_1 decreases $\max_{i \in S} \alpha_S^i$ by the fixed factor $1 - 1/s < 1$, so it converges to 0. Together with $F_1(a) \geq a$, this implies that the sequence $F_1^T(a)$ converges to a point on ∂V.

We come now to the proof of Theorem 4.4**.** This will be done by a sequence of lemmas. The idea of the proof is as follows: Assume $\rho \to 1$ and $T \to \infty$ with $\rho^T \to \mu$, or $T(1-\rho) \to -\ln\mu$. Split T into two parts, T_1 and T_2. If T_1 converges to ∞ relatively slowly — in particular, if $\rho^{T_1} \to 1$ or $T_1(1-\rho) \to 0$ — then we prove in Lemma A.1 that $a := F_\rho^{T_1}(0) \to \lim_{t\to\infty} F_1^t(0)$, which as seen in Proposition 4.2 exists and is the Raiffa point r. Therefore $F_\rho^T(0) = F_\rho^{T_2}\left(F_\rho^{T_1}(0)\right) = F_\rho^{T_2}(a)$ will be close to $F_\rho^{T_2}(r)$. Also, once a is sufficiently close to the efficient boundary ∂V (recall that r lies on ∂V), then all its F_ρ iterates $F_\rho^t(a)$ will also be close to ∂V — this is Lemma A.4. Next, Lemma A.6 shows that, for c on the boundary, $F_\rho(c)$ can be approximated by the solution $\Lambda_{1-\rho}(c)$ of the C-field $C(\cdot)$. Iterating this implies that $F_\rho^{T_2}(r)$ is close to $\Lambda_{T_2(1-\rho)}(r)$, which in turn converges to $\Lambda_{-\ln\mu}(r)$ since $T_2(1-\rho) \to -\ln\mu$. This completes the proof, once all approximations are made precise.

All constants appearing below (η, K, and so on) depend only on the game (N,V). We use the Euclidean norm in each $\Re^S$, and for payoff configurations we take $\|a\| := \max_{S\subset N}\|a_S\|$.

The first lemma deals with the case when T converges to infinity slowly relative to the convergence of ρ to 1; in this case, starting from 0 one gets the Raiffa point r. More generally,

Lemma A.1. *There exists a constant $\eta > 0$ such that, if $\rho \to 1$, $T \to \infty$ and $T \leq -\eta\ln(1-\rho)$, then $F_\rho^T(a) \to \lim_{t\to\infty} F_1^t(a)$ for all $a \in V$.*

Proof. The function $F(\rho,a) = F_\rho(a)$ is differentiable with continuous differential over the compact set $[0,1]\times V_+$. Hence

$$\left\|F_\rho(a) - F_1(b)\right\| \leq K\left((1-\rho) + \|a-b\|\right),$$

where $K > \max_{(\rho,a)\in[0,1]\times V_+} \|DF(\rho,a)\| + 1$. This implies that

$$\left\|F_\rho(a) - F_1(a)\right\| \leq K(1-\rho),$$

and also that

$$\left\|F_\rho^2(a) - F_1^2(a)\right\| = \left\|F_\rho(F_\rho(a)) - F_1(F_1(a))\right\| \leq K\left((1-\rho) + K(1-\rho)\right).$$

By induction we then get

$$\left\|F_\rho^T(a)-F_1^T(a)\right\| \leq K(1-\rho)(1+K+\cdots+K^{T-1}) \leq \frac{K(1-\rho)K^T}{K-1}. \tag{A.1}$$

Take $\eta < 1/\ln K$; then $\ln(1-\rho)+T\ln K \leq (1-\eta\ln K)\ln(1-\rho) \to -\infty$ (since $\ln(1-\rho) \to -\infty$). Therefore $(1-\rho)K^T \to 0$, implying that the right hand side of (A.1) converges to 0. Hence $F_\rho^T(a) \to \lim_{T\to\infty} F_1^T(a)$.

The next three lemmas show that the F_ρ iterates of points close to the boundary $\partial_+ V$ stay close to it, and moreover that the distance becomes of the order of $\varepsilon(1-\rho)$ after no more than $-(\eta/2)\ln(1-\rho)$ iterations. Recall the notation $b_{S,i}(a,\rho)$ used in the proof of Proposition 4.1 above.

Lemma A.2. *There is a constant K such that $\left\|b_{S,i}(a,\rho)-a_S\right\| \leq K(\operatorname{dist}(a_S,\partial V(S))+(1-\rho))$ for all $i\in S\subset N$, all a_S in $V_+(S)$ and all $\rho\in(0,1]$.*

Proof. Let $c_S \in \partial V(S)$ be a closest point to a_S on the boundary, so $d := \operatorname{dist}(a_S,\partial V(S)) = \|a_S-c_S\|$. Denote $b'_{S,i} := \rho a_S + (1-\rho)(a_{S\setminus i},0)$. We have $\left\|a_S-b'_{S,i}\right\| = (1-\rho)\left\|(a_{S\setminus i},0)-a_S\right\| \leq M(1-\rho)$, for an appropriate bound M (both $(a_{S\setminus i},0)$ and a_S are in $V_+(S)$). Thus $\operatorname{dist}(b'_{S,i},\partial V(S)) \leq M(1-\rho)+d$. Let λ_S be the unit supporting normal to $\partial V(S)$ at c_S, then the distance of $b'_{S,i}$ to $\partial V(S)$ along the i'th coordinate is at most $(M(1-\rho)+d)/\lambda_S^i$. So $\left\|b_{S,i}-b'_{S,i}\right\| \leq (M(1-\rho)+d)/\lambda_S^i$, and altogether $\left\|b_{S,i}-a_S\right\| \leq M(1-\rho)+(M(1-\rho)+d)/\lambda_S^i$. The non-levelness assumption implies that the λ_S^i are all bounded away from 0, and the result follows.

Lemma A.3. *There is a constant K such that $\operatorname{dist}(F_\rho(a),\partial V) \leq K(\operatorname{dist}(a,\partial V)+(1-\rho))^2$ for all $a\in V_+$ and all $\rho\in(0,1]$.*

Proof. The boundary $\partial V(S)$ is C^2, therefore there is K_1 such that, for any c_S and c'_S in $\partial_+ V(S)$ we have

$$0 \leq \lambda_S \cdot (c_S - c'_S) \leq K_1 \left\|c_S - c'_S\right\|^2,$$

where λ_S is the unit outward normal vector to $\partial_+ V(S)$ at c_S. Let c_S be one of the vectors $b_{S,i} \equiv b_{S,i}(a,\rho)$, and consider the hyperplane game V' with $V'(S) := \{x_S \in \Re^S : \lambda_S \cdot x_S \leq \lambda_S \cdot c_S\} \supset V(S)$. Then

$$\begin{aligned}
\operatorname{dist}((F_\rho(a))_S, \partial V(S)) &\leq \operatorname{dist}((F_\rho(a))_S, \partial V'(S)) = \operatorname{dist}(\frac{1}{|S|}\sum_{i\in S} b_{S,i}, \partial V'(S)) \\
&= \frac{1}{|S|}\sum_{i\in S}\lambda_S \cdot (c_S - b_{S,i}) \leq \frac{1}{|S|}\sum_{i\in S} K_1 \left\| c_S - b_{S,i}\right\|^2 \\
&\leq K_1 \left(2K_2 \left(\operatorname{dist}(a_S, \partial V(S)) + (1-\rho)\right)\right)^2,
\end{aligned}$$

where K_2 is the constant from Lemma A.2 (recall that c_S is one of the vectors $b_{S,i}$, implying the inequality $\left\| c_S - b_{S,i}\right\| \leq 2K_2 \left(\operatorname{dist}(a_S, \partial V(S)) + (1-\rho)\right)$ by Lemma A.2).

Lemma A.4. *There exists a constant $d_0 > 0$ such that for every $\varepsilon > 0$ there is $\rho_0(\varepsilon) \in (0,1)$ with the property that $\operatorname{dist}(F^t_\rho(a), \partial V) \leq \varepsilon(1-\rho)$ for all $\rho \in (\rho_0(\varepsilon), 1)$, all $t \geq -(\eta/2)\ln(1-\rho)$ and all $a \in V_+$ with $\operatorname{dist}(a, \partial V) \leq d_0$, where η is given by Lemma A.1.*

Proof. Let $d_0 := 1/(4Ke^{4/\eta})$ and $\rho_0 := \max\{1 - d_0/\varepsilon, 1 - d_0, 1 - 2\varepsilon/(K(\varepsilon+2)^2), 1 - \varepsilon/(2d_0)\}$,where K is given by Lemma A.3. Take a with $\operatorname{dist}(a, \partial V) \leq d_0$. For every $t \geq 0$, denote $x_t := \operatorname{dist}(F^t_\rho(a), \partial V)$. Lemma A.3 implies that $x_t \leq f(x_{t-1})$ for all $t \geq 1$, where $f(x) := K(x + (1-\rho))^2$. For every $\rho \in (\rho_0, 1)$ put $\xi := \varepsilon(1-\rho)$; we have: (i) $\xi < d_0$ (since $1-\rho < d_0/\varepsilon$); (ii) $f(d_0) < d_0$ (indeed: $1-\rho < d_0$, implying that $f(d_0) < 4Kd_0^2 \leq d_0$); and (iii) $f(\xi/2) < \xi/2$ (since $1-\rho < 2\varepsilon/(K(\varepsilon+2)^2)$). Therefore the equation $f(x) = x$ has two real roots, $x^* < x^{**}$, that satisfy $0 < x^* < \xi/2 < d_0 < x^{**}$. Now $x_t - x^* \leq f(x_{t-1}) - f(x^*) = f'(y_t)(x_{t-1} - x^*)$ for an appropriate intermediate point y_t. The sequence x_t starts at $x_0 = \operatorname{dist}(a, \partial V) \leq d_0$, thus it never leaves the interval $[0, d_0]$, and so $y_t \in [0, d_0]$ too. Hence $0 < f'(0) \leq f'(y_t) \leq f'(d_0) = 2K(d_0 + 1 - \rho) < e^{-4/\eta}$. From this it follows that $x_t - x^* \leq e^{-4/\eta}(x_{t-1} - x^*)$, implying $x_t - x^* \leq e^{-4t/\eta} d_0$. If $t \geq -(\eta/2)\ln(1-\rho)$ then $e^{-4t/\eta} d_0 \leq (1-\rho)^2 d_0 < \varepsilon(1-\rho)/2 = \xi/2$ (since $1-\rho < \varepsilon/(2d_0)$), therefore $x_t - x^* < \xi - x^*$ (recall that $x^* < \xi/2$), which finally yields $x_t < \xi = \varepsilon(1-\rho)$.

From Lemmas A.1 - A.4 we obtain the following

Corollary A.5. *For every $\rho \in (0,1)$ there is an integer $n(\rho) > 0$ with $n(\rho) \to \infty$ and $n(\rho)(1-\rho) \to 0$ as $\rho \to 1$, such that for every $\varepsilon > 0$ there is $\rho_0(\varepsilon) \in (0,1)$ satisfying*

$$\begin{aligned} \left\| F_\rho^{n(\rho)}(0) - r \right\| &\leq \varepsilon \text{ and} \\ \operatorname{dist}(F_\rho^{n(\rho)+t}(0), \partial V) &\leq \varepsilon(1-\rho) \end{aligned}$$

for all $\rho \in (\rho_0(\varepsilon), 1)$ and all $t \geq 0$.

Proof. Let η be the constant given by Lemma A.1, and let d_0 and ρ_0 be given by Lemma A.4. Define $m(\rho) := -(\eta/2)\ln(1-\rho)$ (rounded to the nearest integer) and $n(\rho) := 2m(\rho)$, then $n(\rho) \to \infty$ and $n(\rho)(1-\rho) \to 0$ as $\rho \to 1$ (since $(1-\rho)\ln(1-\rho) \to 0$). Put $a(\rho) := F_\rho^{m(\rho)}(0)$ and $b(\rho) := F_\rho^{n(\rho)}(0) = F_\rho^{m(\rho)}(a(\rho))$.

Applying Lemma A.1 with $T = m(\rho)$ and also with $T = n(\rho)$ gives $\lim_{\rho\to 1} a(\rho) = \lim_{\rho\to 1} b(\rho) = \lim_{t\to\infty} F_1^t(0) = r \in \partial V$, so in particular $\operatorname{dist}(a(\rho), \partial V) \leq d_0$ and $\|b(\rho) - r\| \leq \varepsilon$ for all ρ close enough to 1. Using Lemma A.4 (with $a_0 = a(\rho)$) then yields $\operatorname{dist}(F_\rho^{n(\rho)+t}(0), \partial V) = \operatorname{dist}(F_\rho^{m(\rho)+t}(a(\rho)), \partial V) \leq \varepsilon(1-\rho)$ for all $t \geq 0$ and all ρ close enough to 1.

The next lemma shows that, on the boundary $\partial_+ V$, applying the F_ρ function is almost like moving $1-\rho$ along the solution paths Λ_t of the C-field.

Lemma A.6. *For every $\varepsilon > 0$ there is $\rho_0(\varepsilon) \in (0,1)$ such that $\left\| F_\rho(a) - \Lambda_{1-\rho}(a) \right\| \leq \varepsilon(1-\rho)$ for all $\rho \in (\rho_0(\varepsilon), 1)$ and all $a \in \partial_+ V$.*

Proof. The function $F : [0,1] \times \partial_+ V \to \partial_+ V$ has a compact domain and is of class C^1, and $\left(dF_\rho(a)/d\rho\right)_{|\rho=1} = -C(a)$ by Proposition 4.1. Then we have that, for every $\varepsilon > 0$ there is $\rho(\varepsilon)$ with $0 < \rho(\varepsilon) < 1$ such that

$$\left\| \frac{F_\rho(a) - F_1(a)}{1-\rho} - C(a) \right\| \leq \varepsilon \text{ for all } \rho \in (\rho(\varepsilon), 1) \text{ and all } a \in \partial_+ V. \tag{A.2}$$

By the same reasoning, since $\left(d\Lambda_{1-\rho}(a)/d\rho\right)|_{\rho=1} = -C(a)$, we can take $\rho(\varepsilon)$ so that we also have

$$\left\| \frac{\Lambda_{1-\rho}(a) - \Lambda_0(a)}{1-\rho} - C(a) \right\| \leq \varepsilon \text{ for all } \rho \in (\rho(\varepsilon), 1) \text{ and all } a \in \partial_+ V. \tag{A.3}$$

But $F_1(a) = \Lambda_0(a) = a$. Therefore

$$\left\|F_\rho(a) - \Lambda_{1-\rho}(a)\right\| \leq 2\varepsilon(1-\rho) \text{ for all } \rho \in (\rho(\varepsilon), 1) \text{ and all } a \in \partial_+ V.$$

The next two lemmas are standard.

Lemma A.7 (Gronwall). *There exists a constant $K > 0$ such that*

$$\|\Lambda_t(b) - \Lambda_t(a)\| \leq e^{Kt} \|b-a\| \text{ for all } t \geq 0 \text{ and all } a, b \in \partial_+ V.$$

Proof. C is a vector field of class C^1 defined over the compact set $\partial_+ V$ (for a proof of this well-known result see Palis and de Melo (1982)).

Lemma A.8. *Assume that $a \in \partial_+ V$ is a local attractor of the consistent field. Then there exist constants $\delta > 0$ and $K > 0$, and a norm $\|\cdot\|'$ on $\prod_{S \subset N} \Re^S$ such that*

$$\|\Lambda_t(b) - a\|' \leq e^{-Kt} \|b-a\|' \text{ for all } t \geq 0 \text{ and all } b \in \partial_+ V \text{ with } \|b-a\|' < \delta.$$

Proof. See Hirsch and Smale (1974, Theorem in §9.1).

We can now proceed to

Proof of Theorem 4.4.

(i) Let $\mu \in (0,1]$, and consider a sequence (ρ_k, T_k) such that $\rho_k \to 1, T_k \to \infty$ and $\rho_k^{T_k} \to \mu$, or, equivalently, $T_k(1-\rho_k) \to -\ln\mu$. We have to prove that $\lim_{k\to\infty} F_{\rho_k}^{T_k}(0) = \Lambda_{-\ln\mu}(r)$.

Fix $\varepsilon > 0$; let $n_k := n(\rho_\kappa)$ be given by Corollary A.5 and put $a_k := F_{\rho_k}^{n_k}(0)$. Then $\|a_k - r\| \leq \varepsilon$ and $\operatorname{dist}(F_{\rho_k}^t(a_k), \partial V) \leq \varepsilon(1-\rho_k)$ for all $t \geq 0$ and all k large enough (such that $\rho_k > \rho_0(\varepsilon)$), say $k \geq k_0$. From now on assume $k \geq k_0$. Also, put $T_k' := T_k - n_k$; then $T_k'(1-\rho_k) \to -\ln\mu$.

Denote $b_k^t := F_{\rho_k}^t(a_k) \in V_+$. Then $\operatorname{dist}(b_k^t, \partial V) \leq \varepsilon(1-\rho_k)$, and so there is $c_k^t \in \partial_+ V$ such that $\left\|b_k^t - c_k^t\right\| \leq \varepsilon(1-\rho_k)$. Let $K_1 > 0$ be such that $\left\|F_\rho(a) - F_\rho(a')\right\| \leq K_1 \|a-a'\|$

for all $\rho \in [0,1]$ and all $a, a' \in V_+$, and let $K_2 > 0$ be the constant given by the Gronwall Lemma A.7. We claim that

$$\left\| c_k^t - \Lambda_{t(1-\rho_k)}(c_k^0) \right\| \leq (K_1+2)\varepsilon(1-\rho_k)\left(1 + e^{K_2(1-\rho_k)} + \ldots + e^{(t-1)K_2(1-\rho_k)}\right) \quad \text{(A.4)}$$

for all $t \geq 0$ and all $k \geq k_0$. The proof is by induction on t (for each k), using the following inequality (recall that $b_k^t = F_{\rho_k}(b_k^{t-1})$):

$$\begin{aligned}
\left\| c_k^t - \Lambda_{t(1-\rho_k)}(c_k^0) \right\| \quad \leq \quad & \left\| c_k^t - b_k^t \right\| + \left\| F_{\rho_k}(b_k^{t-1}) - F_{\rho_k}(c_k^{t-1}) \right\| \\
& + \left\| F_{\rho_k}(c_k^{t-1}) - \Lambda_{1-\rho_k}(c_k^{t-1}) \right\| \\
& + \left\| \Lambda_{1-\rho_k}(c_k^{t-1}) - \Lambda_{1-\rho_k}(\Lambda_{(t-1)(1-\rho_k)}(c_k^0)) \right\|.
\end{aligned}$$

The first term is bounded by $\varepsilon(1-\rho_k)$, the second by $K_1 \left\| c_k^{t-1} - b_k^{t-1} \right\| \leq K_1\varepsilon(1-\rho_k)$ and the third by $\varepsilon(1-\rho_k)$ (by Lemma A.6). As for the fourth term, it is at most $e^{K_2(1-\rho_k)} \left\| c_k^{t-1} - \Lambda_{(t-1)(1-\rho_k)}(c_k^0) \right\|$ (by the Gronwall Lemma A.7), which is bounded by the induction hypothesis by

$$e^{K_2(1-\rho_k)}(K_1+2)\varepsilon(1-\rho_k)\left(1 + e^{K_2(1-\rho_k)} + \ldots + e^{(t-2)K_2(1-\rho_k)}\right).$$

Adding the four terms yields the right hand side of (A.4).

As $k \to \infty$ we have

$$(1-\rho_k)\left(1 + e^{K_2(1-\rho_k)} + \ldots + e^{(T_k'-1)K_2(1-\rho_k)}\right) \leq \frac{(1-\rho_k)\, e^{T_k' K_2(1-\rho_k)}}{e^{K_2(1-\rho_k)} - 1} \to \frac{e^{-K_2 \ln\mu}}{K_2},$$

implying by (A.4) that

$$\limsup_{k\to\infty} \left\| c_k^{T_k'} - \Lambda_{T_k'(1-\rho_k)}(c_k^0) \right\| \leq \varepsilon(K_1+2)e^{-K_2\ln\mu}/K_2 =: \varepsilon K_3.$$

Now

$$F_{\rho_k}^{T_k}(0) = F_{\rho_k}^{T_k'}\left(F_{\rho_k}^{n_k}(0)\right) = F_{\rho_k}^{T_k'}(a_k) = b_k^{T_k'},$$

and we have

$$\begin{aligned}
\left\| F_{\rho_k}^{T_k}(0) - \Lambda_{-\ln\mu}(r) \right\| \quad \leq \quad & \left\| b_k^{T_k'} - c_k^{T_k'} \right\| + \left\| c_k^{T_k'} - \Lambda_{T_k'(1-\rho_k)}(c_k^0) \right\| \\
& + \left\| \Lambda_{T_k'(1-\rho_k)}(c_k^0) - \Lambda_{T_k'(1-\rho_k)}(r) \right\| \\
& + \left\| \Lambda_{T_k'(1-\rho_k)}(r) - \Lambda_{-\ln\mu}(r) \right\|.
\end{aligned}$$

As $k \to \infty$, the first term converges to 0 (it is $\leq \varepsilon(1-\rho_k)$). The second is $\leq \varepsilon K_3$ in the limit. For the third one, note that $\left\|a_k - c_k^0\right\| \leq \varepsilon(1-\rho_k)$ and $\|a_k - r\| \leq \varepsilon$, which implies $\left\|c_k^0 - r\right\| \leq 2\varepsilon$ and thus, by the Gronwall Lemma A.7,

$$\left\|\Lambda_{T_k'(1-\rho_k)}(c_k^0) - \Lambda_{T_k'(1-\rho_k)}(r)\right\| \leq 2\varepsilon e^{K_2 T_k'(1-\rho_k)} \to 2\varepsilon e^{-K_2 \ln\mu} =: \varepsilon K_4$$

Finally, the fourth term converges to 0 since $T_k'(1-\rho_k) \to -\ln\mu$. Therefore $\limsup_{k\to\infty}\left\|F_{\rho_k}^{T_k}(0) - \Lambda_{-\ln\mu}(r)\right\|$ $\varepsilon(K_3+K_4)$; since ε is arbitrary, this completes the proof that $F_{\rho_k}^{T_k}(0) \to \Lambda_{-\ln\mu}(r)$.

(ii) For the second part, assume that a is a local attractor and that $\Lambda_t(r) \to a$ as $t \to \infty$. Consider a sequence (ρ_k, T_k) such that $\rho_k \to 1$, $T_k \to \infty$ and $\rho_k^{T_k} \to 0$, or $T_k(1-\rho_k) \to \infty$, as $k \to \infty$. Fix $\varepsilon > 0$ and let $\mu > 0$ be small enough so that $\left\|\Lambda_{-\ln\mu}(r) - a\right\|' \leq \varepsilon$, where $\|\cdot\|'$ is the norm given by Lemma A.8. By the result of part (i), if we let T_k^1 be (the integer part of) $\mu/(1-\rho_k)$ and put $a_k := F_{\rho_k}^{T_k^1}(0)$, then $\|a_k - a\|' \leq 2\varepsilon$ for all k large enough, say $k \geq k_0$. Assume moreover that $\rho_k > \rho_0(\varepsilon)$ and $n(\rho_k) \leq T_k^1 \leq T_k$ for $k \geq k_0$, where $\rho_0(\varepsilon)$ and $n(\rho)$ are given by Corollary A.5. From now on assume $k \geq k_0$.

As in the proof of part (i) above, put $b_k^t := F_{\rho_k}^t(a_k)$ and let $c_k^t \in \partial_+ V$ be such that[5] $\left\|b_k^t - c_k^t\right\|' \leq \varepsilon(1-\rho_k)$. Let $K_1 > 0$ be such that $\left\|F_\rho(a) - F_\rho(a')\right\|' \leq K_1 \|a - a'\|'$ for all $\rho \in [0,1]$ and all $a, a' \in V_+$, and let $K_2 > 0$ be the constant given by Lemma A.8. Then

$$\begin{aligned}\left\|c_k^t - a\right\|' &\leq \left\|c_k^t - b_k^t\right\|' + \left\|F_{\rho_k}(b_k^{t-1}) - F_{\rho_k}(c_k^{t-1})\right\|' \\ &\quad + \left\|F_{\rho_k}(c_k^{t-1}) - \Lambda_{1-\rho_k}(c_k^{t-1})\right\|' + \left\|\Lambda_{1-\rho_k}(c_k^{t-1}) - a\right\|' \\ &= (K_1+2)\varepsilon(1-\rho_k) + \left\|\Lambda_{1-\rho_k}(c_k^{t-1}) - a\right\|',\end{aligned} \tag{A.5}$$

for all $t \geq 0$ and all $k \geq k_0$. We now use induction on t (for each k). Assume that $\left\|c_k^{t-1} - a\right\|' \leq \delta/2$, where δ is given by Lemma A.8; this holds for $t = 0$ if ε is chosen less than $\delta/4$, since $\left\|c_k^0 - a\right\|' = \|a_k - a\|' \leq 2\varepsilon$. Then $\left\|\Lambda_{1-\rho_k}(c_k^{t-1}) - a\right\|' \leq e^{-K_2(1-\rho_k)}\left\|c_k^{t-1} - a\right\|'$, so (A.5) applied inductively yields

$$\left\|c_k^t - a\right\|' \leq (K_1+2)\varepsilon(1-\rho_k)\left(1 + e^{-K_2(1-\rho_k)} + \ldots + e^{-(t-1)K_2(1-\rho_k)}\right)$$

[5] The $\|\cdot\|'$ norm is equivalent to the Euclidean norm $\|\cdot\|$, so there is $K > 0$ such that $\|x\|' \leq K\|x\|$ for all x; one needs only to replace ε by ε/K to get the same estimates for $\|\cdot\|'$.

$$
\begin{aligned}
& +e^{-tK_2(1-\rho_k)}\left\|c_k^0-a\right\|' \\
\leq \quad & (K_1+2)\varepsilon(1-\rho_k)\frac{1}{1-e^{-K_2(1-\rho_k)}}+2\varepsilon e^{-tK_2(1-\rho_k)}.
\end{aligned}
$$

Now $K_2(1-\rho_k)/(1-e^{-K_2(1-\rho_k)}) \to 1$ as $k\to\infty$, hence the first term above is bounded by $2\varepsilon(K_1+2)/K_2$ for all k large enough. Therefore $\left\|c_k^t-a\right\|' \leq \varepsilon\left(2(K_1+2)/K_2+2\right) =: \varepsilon K_3$, which in particular is $\leq \delta/2$ if ε was chosen appropriately small.

To complete the proof, note that

$$
\begin{aligned}
\left\|F_{\rho_k}^{T_k}(0)-a\right\|' &= \left\|F_{\rho_k}^{T_k-T_k^1}(a_k)-a\right\|' = \left\|b_k^{T_k-T_k^1}-a\right\|' \\
&\leq \left\|b_k^{T_k-T_k^1}-c_k^{T_k-T_k^1}\right\|' + \left\|c_k^{T_k-T_k^1}-a\right\|' \leq \varepsilon(1-\rho_k)+\varepsilon K_3.
\end{aligned}
$$

(iii) The third part of the proposition, which states that $\lim_{\substack{T\to\infty \\ \rho\to 1\ \rho^T\to 0}} w(\rho,T) \supset \Lambda_\infty(r)$, where $\Lambda_\infty(r)$ denotes the $\omega-$limit of the solution of the consistent field through r, can be proved using the result of part (i) already established.

Take any $y\in\Lambda_\infty(r)$, then there exists a sequence $\mu_k\to 0$ such that $y_k=\Lambda_{-\ln\mu_k}(r)\to y$. For each k, by the result of part (i) there is T_k large enough, say $T_k\geq k$, such that $\left\|F_{\rho_k}^{T_k}(0)-y_k\right\| \leq 1/k$, where ρ_k is given by $T_k(1-\rho_k)=\mu_k$. Hence $F_{\rho_k}^{T_k}(0)\to y$ as $k\to\infty$.

Proof of Theorem 5.1. We first note that $C_S(a)$ depends only on a_Z for $Z\subset S$ (more precisely: only for $Z=S$ and $Z=S\backslash i$ for all $i\in S$). Therefore $DC(a)$ is a block triangular matrix. The diagonals of $DC(a)$ are the matrices $DC_S : T_{a_S}\partial V(S)\to T_{a_S}\partial V(S)$. To evaluate DC_S, express $C_S(a)$ as $b_S(a)-a_S$, where

$$
b_S^i(a)=\frac{1}{|S|\lambda_S^i(a)}\left(\lambda_S(a)\cdot a_S-\sum_{j\in S\backslash i}\lambda_S^j(a)a_{S\backslash i}^j\right)+\frac{1}{|S|}\sum_{j\in S\backslash i}a_{S\backslash j}^i.
$$

We want to evaluate $DC_Sv_S=Db_Sv_S-v_S$ for any vector $v_S\in T_{a_S}\partial V(S)$. Represent, by an abuse of notation, $b_S(a)$ as $b_S(\lambda_S(a),\lambda_S(a)\cdot a_S)$, where

$$
b_S^i(\lambda_S,\zeta_S)=\frac{1}{|S|\lambda_S^i}\left(\zeta_S-\sum_{j\in S\backslash i}\lambda_S^j a_{S\backslash i}^j\right)+\frac{1}{|S|}\sum_{j\in S\backslash i}a_{S\backslash j}^i.
$$

Using the chain rule we get

$$Db_S v_S = \left(D_{\lambda_S} b_S + D_{\zeta_S} b_S a_S^T\right) D\lambda_S v_S,$$

since $D\zeta_S = D\left(a_S^T \lambda_S(a)\right) v_S = a_S^T D\lambda_S v_S + \lambda_S^T v_S = a_S^T D\lambda_S v_S$ (recall that $v_S \in T_{a_S}\partial V(S)$, and so $v_S \perp \lambda_S$). Evaluating $G = D_{\lambda_S} b_S + D_{\zeta_S} b_S a_S^T$ yields for $i \neq j$

$$G^{ij} = \frac{\partial b_S^i}{\partial \lambda_S^j} + \frac{\partial b_S^i}{\partial \zeta_S} a_S^j = \frac{1}{|S|\lambda_S^i}\left(a_S^j - a_{S\setminus i}^j\right).$$

As for $i = j$, we get

$$\begin{aligned}
G^{ii} &= \frac{\partial b_S^i}{\partial \lambda_S^i} + \frac{\partial b_S^i}{\partial \zeta_S} a_S^i \\
&= -\frac{1}{|S|\left(\lambda_S^i\right)^2}\left(\zeta_S - \sum_{j\in S\setminus i} \lambda_S^j a_{S\setminus i}^j\right) + \frac{1}{|S|\lambda_S^i} a_S^i \\
&= -\frac{1}{|S|\left(\lambda_S^i\right)^2} \sum_{j\in S\setminus i} \lambda_S^j \left(a_S^j - a_{S\setminus i}^j\right).
\end{aligned}$$

But a is a consistent value, hence $\sum_{j\in S\setminus i} \lambda_S^i (a_S^i - a_{S\setminus j}^i) = \sum_{j\in S\setminus i} \lambda_S^j (a_S^j - a_{S\setminus i}^j)$, implying

$$G^{ii} = -\frac{1}{|S|\lambda_S^i} \sum_{j\in S\setminus i} (a_S^i - a_{S\setminus j}^i),$$

and the proof is thus complete.

References

Binmore, K. (1987), "Perfect Equilibria in Bargaining Models", in *The Economics of Bargaining*, ed. by K. Binmore and P. Dasgupta. Blackwell, 77-105.

Gul, F. (1989), "Bargaining Foundations of Shapley Value," *Econometrica*, 57, 81-95.

Harsanyi, J.C. (1981), "The Shapley Value and the Risk Dominance Solution of Two Bargaining Models for Characteristic-Function Games," in *Essays in Game Theory and Mathematical Economics*, ed. by R.J. Aumann *et al*. Bibliographisches Institut Mannheim, 43-68.

Hart, O. and Moore, J. (1990), "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98, 1119-1157.

Hart, S. and Mas-Colell, A. (1989), "Potential, Value and Consistency," *Econometrica*, 57, 589-614.

Hart, S. and Mas-Colell, A. (1996a), "Harsanyi Values of Large Economies: Non-Equivalence to Competitive Equilibria," *Games and Economic Behavior,* 13, 74-99.

Hart, S. and Mas-Colell, A. (1996b), "Bargaining and Value," *Econometrica*, 64, 357-380.

Luce, R.D. and Raiffa, H. (1957), *Games and Decisions*. J. Wiley.

Maschler, M. and Owen, G. (1989), "The Consistent Shapley Value for Hyperplane Games," *International Journal of Game Theory*, 18, 389-407.

Maschler, M. and Owen, G. (1992), "The Consistent Shapley Value for Games without Side Payments," in *Rational Interaction*, ed. by R. Selten. Springer-Verlag, 5-12.

Maschler, M., Owen, G. and Peleg, B. (1988), "Paths Leading to the Nash Set," in *The Shapley Value, Essays in Honor of Lloyd Shapley*, ed. by A.E. Roth. Cambridge University Press, 321-330.

Nash, J. (1950), "The Bargaining Problem," *Econometrica*, 18, 155-162.

Osborne, M. and Rubinstein, A. (1990), *Bargaining and Markets*. Academic Press.

Palis, J. and de Melo, W. (1982), *Geometric Theory of Dynamical Systems, an Introduction*. Springer-Verlag.

Shapley, L.S.. (1953), “A Value for n-Person Games,” in *Contributions to the Theory of Games II*, ed. by H.W. Kuhn and A.W. Tucker. Princeton University Press, 307-317.

Sjöström, T. (1991), “Ståhl Bargaining Model”, *Economics Letters,* 36, 153-157.

Ståhl, I. (1972), *Bargaining Theory*, Stockholm School of Economics.

Winter, E. (1994), “The Demand Commitment Bargaining and Snowball Cooperation,” *Economic Theory*, 4, 255-273.