# A Folk Theorem for the One-dimensional Spatial Bargaining Model*

Seok-ju Cho
Department of Political Science
Yale University

John Duggan
Department of Political Science
and Department of Economics
University of Rochester

November 25, 2014

**Abstract**

We show that in the one-dimensional bargaining model based on the protocol of Baron and Ferejohn (1989), if voting is simultaneous, publicly observed, and no agent has the power to unilaterally impose a choice, then arbitrary policies can be supported by subgame perfect equilibria in stage-undominated voting strategies when agents are patient. Moreover, in the model with a bad status quo, arbitrary outcomes can be supported for *arbitrary positive discount factors.* We formulate sufficient conditions for supporting arbitrary alternatives in terms of a system of equations. The system has a straightforward solution in the model with no discounting, and after verifying non-singularity of this system, we use the implicit function theorem to derive the folk theorem when agents are sufficiently patient.

## 1 Introduction

Many real-world examples of collective decision making, and especially policy making in committees and legislatures, can be fruitfully understood as dynamic bargaining games. A common framework for such applications is the distributive model of Baron and Ferejohn (1989) and the spatial models that have grown from it. Whereas the latter authors consider the assignment of pork across legislative districts, i.e., division of a dollar, Banks and Duggan (2000, 2006) generalize the model to include policy making in spatial environments, where policies can be multidimensional and can include ideological components. A special case of interest, and one central to applied work in political economy, is the one-dimensional model, in which policies belong to an ideological spectrum and agents have single-peaked preferences over policies. Most work has assumed simultaneous voting and focussed on

*We thank Hulya Eraslan and the anonymous editor and referees for helpful comments. All errors and omissions are our own.

subgame perfect equilibria that are stationary and such that votes are stage-undominated: the agents' proposal strategies are history-independent, voting strategies depend only on the current proposal, and each individual accepts or rejects a proposal according to whether the utility from the proposal exceeds the expected discounted payoff from rejection.[1]

An anti-folk theorem for the one-dimensional bargaining model with majority voting is provided by Cho and Duggan (2009), who show that the set of policy outcomes that can be supported in subgame perfect equilibrium collapses to the median ideal alternative as agents become patient. A key feature of the model is that agents are assumed to vote sequentially and that with positive probability the order of voting starts with the median agent and then alternates from one side of the median to the other. This implies (after some argument) that when the median alternative is proposed, it passes with positive probability. As a consequence, patient voters become more willing to reject unappealing proposals in favor of a chance that the median alternative is eventually implemented, and the agents become more demanding as they become more patient. The result carries over directly to the model with simultaneous voting if we restrict attention to subgame perfect equilibria in which the agents do not condition on the past votes of particular agents; when voting is simultaneous and by secret ballot, for example, an asymptotic version of the median voter theorem holds.

In the analysis of stationary equilibria, the choice between sequential voting (as in Cho and Duggan (2009)) and simultaneous voting with the elimination of stage-dominated voting strategies is immaterial: the two approaches are essentially equivalent and generate the same equilibrium proposal strategies and voting outcomes. When voting is simultaneous, and we consider a stationary equilibrium, elimination of stage-dominated votes refines away equilibria in which implausible outcomes of voting (such as every agent voting to reject when acceptance is preferred by all) occur due to the fact that no agent may be pivotal. But when voting is simultaneous and publicly observed, and when we allow for equilibria in which agents condition on past votes of particular agents, the stage-dominance refinement loses its bite, and results become sensitive to the voting protocol.

This paper establishes that in the one-dimensional bargaining model with simultaneous and publicly observed voting such that no agent has the power to unilaterally impose a choice, arbitrary policies can be supported in subgame perfect equilibrium when agents are patient, even if we eliminate weakly dominated strategies in voting stages. In fact, in the

[1] Banks and Duggan (2000,2006) establish existence of stationary equilibrium and convergence to the median ideal alternative as agents become patient. Cho and Duggan (2003) and Cardona and Ponsati (2011) prove uniqueness of stationary equilibria in a class of one-dimensional models. See Eraslan (2002) and Eraslan and McLennan (2012) for uniqueness results in the distributive model.

model with a bad status quo (following Baron and Ferejohn (1989) and Banks and Duggan (2000)), arbitrary outcomes can be supported for *arbitrary positive discount factors.* A key to the argument is that we can "trap" agents into making proposals that are bad for themselves. An obvious deviation from such a strategy profile is for an agent to propose the median alternative, but when voting is simultaneous and publicly observed, we can induce all agents to vote against the median (or any other alternative) when proposed. Obviously, it is Nash for all to do so, because no one is pivotal. To make it stage-undominated, we specify that if an agent votes to accept the median and all others reject, then next period's proposer proposes something bad for that agent (and that proposal is accepted). So, when agents condition on past votes, we can break the stage-dominance refinement. The folk theorem result compliments the anti-folk theorem of Cho and Duggan (2009) and proves a rather complete picture of the policy outcomes that are consistent with strategic incentives when agents are patient.

The proof of the folk theorem relies on a lemma showing that every "plausible" alternative for an agent $i$ can be supported as an equilibrium outcome when $i$ proposes, where sets $Y_1, \ldots, Y_n$ of alternatives are plausible if for each agent $i$, every alternative $x \in \text{int}Y_i$ is strictly better for $i$ than the worst (discounted) lottery over plausible alternatives $y_j$ for other agents $j$ such that the probability on $y_j$ is equal to the recognition probability of agent $j$. We can induce agent $i$ to propose any alternative $x \in \text{int}Y_i$ as follows: if the agent proposes $x$, then we specify voting strategies along lines described above so that $x$ is accepted; otherwise, if the agent proposes the wrong alternative, then we specify that the proposal is rejected, and that the next proposer $j$ proposes a sufficiently bad alternative in $Y_j$ for agent $i$, which is then accepted. Thus, the only possible gain from proposing the wrong alternative is to obtain the status quo for one period, but this gain becomes negligible as the agents become patient. We formulate sufficient conditions for supporting arbitrarily large plausible sets of alternatives in terms of a system of equations. The system has a straightforward solution in the model with no discounting, and after verifying non-singularity of this system, we use the implicit function theorem to derive the folk theorem when agents are sufficiently patient.

Other folk theorem results are known in the distributive model of bargaining: Baron and Ferejohn (1989) give a result in the symmetric model with majority voting; Norman (2002) establishes a folk theorem for the finite-horizon model; and Herings, Meshalkin, and Predtetchinski (2012) prove a folk theorem in an environment that generalizes the distributive model, but they assume that payoffs are bounded between zero and one and that the set of feasible payoff vectors contains the vertices of the unit simplex in $\mathbb{R}^n$, a condition that is not satisfied in the spatial environment. In the context of dynamic elections, Duggan and Fey (2006) provide a folk theorem in a Downsian model, where parties can make short-term commitments to platforms, and Duggan (2014) establishes a folk theorem for a dynamic electoral model with privately informed candidates.

Of course, folk theorems have been proved for abstract dynamic games. Fudenberg and Maskin (1986) state the folk theorem for subgame perfect equilibria of discounted repeated games, and this is extended to repeated games with imperfect public monitoring by Fudenberg, Levine, and Maskin (1994). Because proposals and voting alternate in the bargaining model, however, it is not the simple repetition of a fixed stage game. Wen (2002) proves a folk theorem in repeated sequential games, which assumes that in each period, players play an extensive form game such that players are partitioned into groups, groups move sequentially, players within a group choose simultaneously, and their feasible action sets are independent of others choices. The timing of moves in this framework has some similarity to the bargaining model, with a proposer first choosing a proposal the agents then simultaneously casting ballots, but it differs in the respect that the identity of the proposer can change stochastically over time and, importantly, in that the proposal considered in voting stages is determined by the previous choice of a player. Horner et al. (2011) and Fudenberg and Yamamoto (2011) establish folk theorems in dynamic games with imperfect public monitoring, but they assume finite sets of states and alternatives, and for the perfect monitoring case, their results apply only to irreducible games; all of those assumptions are violated when the bargaining model is translated to their frameworks,[2] and their analyses do not incorporate weak dominance refinements on voting strategies.[3]

## 2 The Model

Let $N = \{1, \ldots, n\}$ be a number $n \geq 3$ of agents who play an infinite-horizon bargaining game over a closed interval $X = [\underline{x}, \overline{x}] \subseteq \mathbb{R}$ of alternatives. The bargaining in every period is described as follows. If no alternative has been accepted prior to period $t$, then (1) an agent $i \in N$ is recognized with probability $\rho_i$, where $\rho = (\rho_1, \ldots, \rho_n) \in \Delta$, the unit simplex in $\mathbb{R}^n$. These recognition probabilities are exogenously fixed throughout the game. (2) If recognized, then agent $i$ makes a proposal, say $x \in X$. (3) After observing $x$, all $j \in N$ simultaneously choose ballots $b_j \in \{a, r\}$ to accept or reject the proposal. Let $\mathcal{D} \subseteq 2^N \setminus \{\emptyset\}$

[2] The bargaining model assumes that proposers choose from a continuum of alternatives, and for each possible proposal, the subsequent voting stage must be identified by a corresponding state. Because the game ends once a proposal is accepted, the bargaining model admits absorbing states, violating irreducibility.

[3] Dutta (1995) shows that the set of equilibrium outcomes is monotonic in discount factor for stochastic games with a deterministic transition satisfying the condition that each player's worst equilibrium across states and equilibria is weakly decreasing in the discount factor; in addition, he implicitly assumes this worst payoff is achieved at some state and for some equilibrium.

denote an exogenously fixed *decisive* coalitions. (4) If $\{j \in N \mid b_j = a\} \in \mathcal{D}$, then proposal $x$ is chosen and bargaining ends with outcome $x$ in period $t$ and every subsequent period. Otherwise, each $i \in N$ receives the utility from the status quo $q$ in period $t$, and the above procedure (1)–(4) is repeated in period $t+1$.

We allow that the status quo $q$ may or may not be an element of $X$. We endow each agent $i$ with a von Neumann-Morgenstern utility function $u_i\colon X \cup \{q\} \to \mathbb{R}$ that, with a discounter factor $\delta_i \in [0,1)$, represents agent $i$'s preference over sequences of outcomes and lotteries over them. If alternative $x$ is chosen in period $t$, then agent $i$ receives payoff

$$(1-\delta_i^{t-1})u_i(q) + \delta_i^{t-1}u_i(x), \tag{1}$$

which reflects $i$'s discounted utility from the status quo for the first $t-1$ periods and from $x$ for every subsequent period. If no alternative is ever accepted, then agent $i$ receives $u_i(q)$. Let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n) \in (0,1)^n$ denote a vector of discount factors.

We assume $u_i$ is strictly quasi-concave and continuously differentiable in $X$.[4] These assumptions guarantee that each agent $i$ has a unique utility-maximizing alternative in $X$, or "ideal point," denoted $\hat{x}_i$. We require a minimal condition of heterogeneity of agents' preferences. *We assume throughout that there exist $i, j \in N$ such that $\rho_i > 0$, $\rho_j > 0$, $u_i(\underline{x}) < u_i(\overline{x})$, and $u_j(\underline{x}) > u_j(\overline{x})$.*[5] To ease notation, we normalize agents' utility from the status quo by assuming that $u_i(q) = 0$ for every agent. Since $u_i$ may take positive or negative values on $X$, agents' preference for delay are unrestricted. In particular, the status quo may belong to the set of alternatives, as assumed in Banks and Duggan (2006), or it may be unanimously bad, as in the distributive model of Baron and Ferejohn (1989) and the spatial model of Banks and Duggan (2000).

The voting rule, represented by $\mathcal{D}$, is arbitrary subject to the following two restrictions. First, the universal coalition is decisive, i.e., $N \in \mathcal{D}$. This condition is self-explanatory and obviously weak. Second, the voting rule is non-collegial, i.e., for all $i \in N$, we have $N \setminus \{i\} \in \mathcal{D}$ and $\{i\} \notin \mathcal{D}$. That is, no agent has the power to unilaterally pass or veto a proposal. We capture majority rule by letting $\mathcal{D}$ consist of all coalitions containing more than $n/2$ agents; we capture any quota rule with threshold $k$ by defining $\mathcal{D}$ as the collection of groups with $k$ or more members; and clearly many other rules in which agents voting rights are unequal are allowed. Since we assume three or more agents, majority rule and all quota rules with $2 \le k \le n-1$ satisfy our assumptions. Our analysis does not apply to bargaining games that operate under unanimity rule; indeed, under unanimity rule, each

[4]By continuous differentiability, we mean that $u_i$ extends to a $C^1$ function on an open superset of $X$.

[5]If the endpoints $\underline{x}$ and $\overline{x}$ do not satisfy this condition, our arguments can be adapted to any subinterval $Z = [\underline{z}, \overline{z}] \subseteq X$ with endpoints satisfying the restriction. The existence of alternatives in $X \setminus Z$ increases the number of one-shot deviations, but when $\delta$ is close to one, these will not be profitable given our construction.

agent can obtain the status quo ad infinitum by rejecting all proposals, so it is clear that the alternatives accepted in equilibrium must Pareto dominate the status quo, regardless of the agents' time preferences.[6]

To discuss strategies and preferences over them, we need to define histories of the bargaining game. A *history* is a finite or infinite sequence of actions of agents and nature consistent with the bargaining protocol defined above. A *complete history* is either the initial history, denoted $\emptyset$, or any history ending with the rejection of a proposal. A *proposer history for $i$* is any history in which agent $i$ has just been selected by nature, and so the agent must next propose an alternative. A *voting history* is any history in which a proposal has just been made and so all agents must next vote on it. A *terminal history* is any history in which a proposal is just accepted. We let $H_t^\circ$ denote the set of $t$-period complete histories, $H_t^{pi}$ the set of $t$-period proposer histories for $i$, and $H_t^v$ the set of $t$-period voting histories. Let

$$B \;=\; \left\{ \mathbf{b} \in \{a, r\}^n \;\middle|\; \{j \in N \mid b_j = a\} \notin \mathcal{D} \right\}.$$

be the set of voting profiles resulting in the rejection of a proposal, where $\mathbf{b} = (b_1, \ldots, b_n)$ denotes a profile of ballots. Technically, we specify $H_0^\circ = \{\emptyset\}$, and for $t = 1, 2, \ldots$, we define

$$\begin{aligned} H_t^{pi} &= H_{t-1}^\circ \times \{i\} \\ H_t^v &= \left( \bigcup_{i \in N} H_t^{pi} \right) \times X \\ H_t^\circ &= H_t^v \times B. \end{aligned}$$

For each $t = 1, 2, \ldots$, and each $x \in X$, let $H_t^\bullet(x)$ denote the set of *terminal histories* in which $x$ is chosen in period $t$, i.e., $H_t^\bullet(x) = H_{t-1}^\circ \times N \times \{x\} \times (\{a, r\}^n \setminus B)$. Then the sets of *$t$-period histories* and of *finite histories* are defined as

$$H_t \;=\; \left( \bigcup_{i \in N} H_t^{pi} \right) \cup H_t^v \cup H_t^\circ \cup \left( \bigcup_{x \in X} H_t^\bullet(x) \right) \quad \text{and} \quad H \;=\; \bigcup_{t=0}^{\infty} H_t,$$

respectively. Also, the sets of proposer histories for $i$, of voting histories, of complete histories, and of terminal histories are

$$H^{pi} \;=\; \bigcup_{t=1}^{\infty} H_t^{pi}, \;\; H^v \;=\; \bigcup_{t=1}^{\infty} H_t^v, \;\; H^\circ \;=\; \bigcup_{t=0}^{\infty} H_t^\circ, \;\; H^\bullet \;=\; \bigcup_{x \in X} H^\bullet(x),$$

respectively.

[6] In the model with a bad status quo, every interior alternative Pareto dominates the status quo, and the folk theorem result indeed extends to unanimity rule. In particular, the proof of Lemma 1, below, can then be modified to cover unanimity rule, in which case equilibrium strategies must include plans that punish each voter when that voter alone rejects an equilibrium proposal.

A (pure) strategy for an agent $i$ in the bargaining game must describe what proposal $i$ would make if recognized and how $i$ would respond others' proposal depending on the past play of the game. A *proposal strategy for* $i$ is a function $p_i \colon H^{pi} \to X$ where for each $h \in H^{pi}$, $p_i(h)$ is the alternative $i$ would propose at $h$. A *voting strategy for* $i$ is a function $v_i \colon H^v \to \{a, r\}$ where for each $h \in H^v$, $v_i(h)$ is $i$'s vote at $h$. We define $v \colon H^v \to \{a, r\}^n$ by $v(h) = (v_1(h), \ldots, v_n(h))$, and we denote a strategy profile by $\sigma = (p_i, v_i)_{i \in N}$.

Beginning at any finite history $h \in H$, a strategy profile $\sigma$ determines a transition probability for histories following $h$ in an obvious way.[7] For each finite $h'$ following $h$, let $\zeta^\sigma(h'|h)$ denote the transition probability from $h$ to $h'$, i.e., the probability that $h'$ is reached from $h$ given strategy profile $\sigma$. With this, we can calculate agent $i$'s expected payoff following any history $h$, denoted $U_i^\sigma(h|\boldsymbol{\delta})$. For any $h \in H$, let $\tau(h)$ be the length of history $h$ in terms of periods, i.e, $\tau(h) = t$ if and only if $h \in H^t$. Recalling that $u_i(q) = 0$ for all $i \in N$, we then have

$$U_i^\sigma(h|\boldsymbol{\delta}) \quad = \quad \sum_{x \in X} \sum_{h' \in H^\bullet(x)} \zeta^\sigma(h'|h) \delta_i^{\tau(h')-1} u_i(x),$$

where we make use of the fact that, because of our focus on pure strategies, the support of $\zeta^\sigma(\cdot|h)$ on $H^\bullet$ is countable.

As is standard, a subgame perfect equilibrium is a strategy profile such that every player's action at every history is optimal with respect to the remainder of the game. Formally, a strategy profile $\sigma = (p_i, v_i)_{i \in N}$ is a *subgame perfect equilibrium* if for every $i \in N$: for every $h \in H^{pi}$ and every $x \in X$, we have

$$U_i^\sigma(h, p_i(h)|\boldsymbol{\delta}) \quad \geq \quad U_i^\sigma(h, x|\boldsymbol{\delta});$$

and for every $h \in H^v$ and every $b_i \in \{a, r\}$, we have

$$U_i^\sigma(h, (v_i(h), v_{-i}(h))|\boldsymbol{\delta}) \quad \geq \quad U_i^\sigma(h, (b_i, v_{-i}(h))|\boldsymbol{\delta}).$$

In addition, we require that equilibrium voting strategies are not weakly dominated in the stage voting game with continuation play as defined in $\sigma$. We say a strategy profile $\sigma = (p_i, v_i)_{i \in N}$ is *in stage-undominated voting strategies* if for every $i \in N$, every $h \in H^v$, and every $b_i \in \{a, r\}$, either for every $\mathbf{b}_{-i} \in \{a, r\}^{N \setminus \{i\}}$, we have

$$U_i^\sigma(h, (v_i(h), \mathbf{b}_{-i})|\boldsymbol{\delta}) \quad \geq \quad U_i^\sigma(h, (b_i, \mathbf{b}_{-i})|\boldsymbol{\delta});$$

or for some $\mathbf{b}_{-i} \in \{a, r\}^{N \setminus \{i\}}$, the above inequality holds strictly. We let $\Sigma(\boldsymbol{\delta})$ denote the set of subgame perfect equilibria in stage-undominated voting strategies.

[7]See Cho and Duggan (2009) for the formal development for a similar game.

## 3 Results

In this section, we establish a folk theorem for the general model and corollary for the model with bad status quo. For each $i \in N$ and each strategy profile $\sigma$, we define

$$X(\boldsymbol{\delta}) \;\; = \;\; \left\{ (x_1,\dots,x_n) \in X^n \;\middle|\; \begin{array}{c} \text{there exists } \sigma \in \Sigma(\boldsymbol{\delta}) \text{ such that for all } i, \\ p_i(\emptyset, i) = x_i \text{ and } \{j \in N | v_j(\emptyset, i, x_i) = a\} \in \mathcal{D} \end{array} \right\},$$

as the set consisting of every vector $(x_1, \dots, x_n)$ of alternatives that is supportable in the strong sense that there is a subgame perfect equilibrium in stage-undominated voting strategies such that each agent $i$ proposes the alternative $x_i$ in the first period, and that proposal is accepted.

We next define a consistency condition for a profile $(Y_1, \dots, Y_n)$ of subsets of alternatives, and we show that under this condition, for all $(x_1, \dots, x_n) \in \times_{i \in N} Y_i$, there is a subgame perfect equilibrium in stage-undominated voting strategies such that each agent $i$ proposes $x_i$ in the first period, and that proposal is accepted. Let $\Omega$ denote the set of compact intervals contained in $X$, i.e.,

$$\Omega \;\; = \;\; \{Y \subseteq X \mid Y \text{ is convex and compact }\}.$$

We say $\mathbf{Y} = (Y_1, \dots, Y_n) \in \Omega^n$ is *plausible* at $\boldsymbol{\delta}$ if for every $i \in N$, we have

$$\min_{x \in Y_i} u_i(x) \;\; \geq \;\; \delta_i \sum_{j \in N} \rho_j \min_{y_j \in Y_j} u_i(y_j).$$

In words, $\mathbf{Y}$ is plausible if for each agent $i$, every alternative $x \in Y_i$ is weakly better for $i$ than the worst (discounted) lottery over plausible alternatives $y_j$ for other agents $j$ such that the probability on $y_j$ is equal to the recognition probability of agent $j$. Note that when the interval $Y_i$ is non-degenerate, the inequality holds strictly for all interior alternatives, so that every alternative $x \in \text{int}Y_i$ is strictly better for $i$ than the worst (discounted) lottery over plausible alternatives for other agents.

The following lemma draws out implications for the plausible profiles of subsets defined above and is a key step in deducing the folk theorem for the one-dimensional bargaining model. The proof is relegated to the appendix.

**Lemma 1** *Let $\boldsymbol{\delta} \in (0,1)^n$. If $\mathbf{Y} \in \Omega^n$ is plausible at $\boldsymbol{\delta}$, then $\times_{i \in N} \text{int}Y_i \subseteq X(\boldsymbol{\delta})$.*

The main result of the paper is that when the agents are sufficiently patient, an arbitrarily large sub-interval of alternatives can be supported as outcomes of subgame perfect equilibria in stage-undominated voting strategies. The proof makes use of the implicit function theorem in the following way, where we simplify this informal explanation by assuming a common discount factor $\delta$. We choose alternatives $\underline{y}$ and $\overline{y}$ close to $\underline{x}$ and $\overline{x}$, respectively, such that no agent is indifferent between the two, and we can assume that $L = \{1, \dots, \ell\}$ consists of agents whose worst alternative in $[\underline{y}, \overline{y}]$ is $\underline{y}$, and $R = \{\ell + 1, \dots, n\}$ consists of the remaining agents whose worst alternative in the interval is $\overline{y}$. We define a mapping $F$ from $(n+1)$-tuples $(x_1, \dots, x_n, \delta)$ to $\mathbb{R}^n$ such that if $F(x_1, \dots, x_n, \delta) = 0$, then

$$[x_1, \overline{y}] \times \cdots \times [x_\ell, \overline{y}] \times [\underline{y}, x_{\ell+1}] \times \cdots \times [\underline{y}, x_n] \in \Omega^n$$

is plausible at $\delta < 1$. We note that

$$F(\underbrace{\underline{y}, \dots, \underline{y}}_{\ell \text{ times}}, \underbrace{\overline{y}, \dots, \overline{y}}_{n-\ell \text{ times}}, 1) \quad = \quad 0,$$

and we verify that the derivative with respect to the first $n$ coordinates is non-singular at this element of the domain. Then the implicit function theorem allows us to choose $(x_1, \dots, x_\ell)$ close to $(\underline{y}, \dots, \underline{y})$ and $(x_{\ell+1}, \dots, x_n)$ close to $(\overline{y}, \dots, \overline{y})$ for $\delta$ less than but close to one such that $F(x_1, \dots, x_n, \delta) = 0$ holds, allowing us, via Lemma 1, to support a large set of alternatives as equilibrium outcomes.

**Theorem 1** *For every $\epsilon > 0$, there exists $\underline{\delta} \in (0,1)$ such that if $\delta_i \geq \underline{\delta}$ for every $i \in N$, then we have $[\underline{x} + \epsilon, \overline{x} - \epsilon]^n \subseteq X(\boldsymbol{\delta})$.*

*Proof:* The bulk of the proof is to show that for every $\mathbf{z} = (z_1, \dots, z_n) \in (\text{int}X)^n$, there exist $\delta(\mathbf{z}) \in (0,1)$ and $\epsilon(\mathbf{z}) > 0$ satisfying the following: if $\delta_i \geq \delta(\mathbf{z})$ for every $i \in N$, then the open ball of radius $\epsilon(\mathbf{z})$ around $\mathbf{z}$ is supportable in equilibrium, i.e., $B_{\epsilon(\mathbf{z})}(\mathbf{z}) \subseteq X(\boldsymbol{\delta})$. To prove the claim, consider any $\mathbf{z} = (z_1, \dots, z_n) \in (\text{int}X)^n$. Recall that there exist $i_\ell, j_r \in N$ with $\rho_{i_\ell} > 0$ and $\rho_{j_r} > 0$ such that $u_{i_\ell}(\underline{x}) < u_{i_\ell}(\overline{x})$ and $u_{j_r}(\underline{x}) > u_{j_r}(\overline{x})$. Since $\mathbf{z}$ is interior to $X^n$ and utility functions are continuous, we can choose $\xi > 0$ sufficiently small so that $\mathbf{z} \in (\underline{x} + \xi, \overline{x} - \xi)^n$ and so that for all $x \in (\underline{x}, \underline{x} + \xi)$ and all $y \in (\overline{x} - \xi, \overline{x})$, we have $u_{i_\ell}(x) < u_{i_\ell}(y)$ and $u_{j_r}(x) > u_{j_r}(y)$. By strict quasi-concavity and differentiability, there exist $\underline{y} \in (\underline{x}, \underline{x} + \xi)$ and $\overline{y} \in (\overline{x} - \xi, \overline{x})$ such that for all $k \in N$, we have $u_k(\underline{y}) \neq u_k(\overline{y})$, $u_k'(\underline{y}) \neq 0$, and $u_k'(\overline{y}) \neq 0$. Note that $\mathbf{z} \in (\underline{y}, \overline{y})^n$.

Let $L = \{k \in N \mid u_k(\underline{y}) < u_k(\overline{y})\}$ and $R = \{k \in N \mid u_k(\underline{y}) > u_k(\overline{y})\}$. By construction, $L$ and $R$ are both non-empty, and they partition $N$. Without loss of generality, $L = \{1, \dots, \ell\}$ and $R = \{\ell{+}1, \dots, n\}$. Using continuous differentiability of $u_i$, let $\overline{X}_i$ be an open superset of $X$ such that $u_i$ extends to a $C^1$ function on $\overline{X}_i$, and let $\overline{X} = \bigcap_{i \in N} \overline{X}_i$. Define the mapping

$F\colon \overline{X}^n \times \mathbb{R}^n \to \mathbb{R}^n$ so that for each $\mathbf{x} = (x_1, \ldots, x_n) \in \overline{X}^n$ and each $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n) \in \mathbb{R}^n$,

$$F(\mathbf{x};\boldsymbol{\delta}) \quad = \quad \begin{pmatrix} \delta_1 \left[\sum_{k\in L} \rho_k u_1(x_k) + \sum_{k\in R} \rho_k u_1(\underline{y})\right] - u_1(x_1) \\ \vdots \\ \delta_\ell \left[\sum_{k\in L} \rho_k u_\ell(x_k) + \sum_{k\in R} \rho_k u_\ell(\underline{y})\right] - u_\ell(x_\ell) \\ \delta_{\ell+1} \left[\sum_{k\in L} \rho_k u_{\ell+1}(\overline{y}) + \sum_{k\in R} \rho_k u_{\ell+1}(x_k)\right] - u_{\ell+1}(x_{\ell+1}) \\ \vdots \\ \delta_n \left[\sum_{k\in L} \rho_k u_n(\overline{y}) + \sum_{k\in R} \rho_k u_n(x_k)\right] - u_n(x_n) \end{pmatrix}.$$

Define $\mathbf{x}^*$ so that $x_k^* = \underline{y}$ for every $k \in L$ and $x_k^* = \overline{y}$ for every $k \in R$, and let $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^n$. Note that

$$F(\mathbf{x}^*;\mathbf{1}) \quad = \quad 0.$$

Since $u_k$ is continuously differentiable on $\overline{X}$ for every $k \in N$, $F$ is continuously differentiable.

Differentiating $F$ with respect to $\mathbf{x}$, we obtain for every $(\mathbf{y}, \boldsymbol{\delta}) \in \overline{X}^n \times \mathbb{R}^n$ that

$$D_{\mathbf{x}} F(\mathbf{y};\boldsymbol{\delta}) \quad = \quad \begin{pmatrix} A(\mathbf{y};\boldsymbol{\delta}) & 0 \\ 0 & B(\mathbf{y};\boldsymbol{\delta}) \end{pmatrix},$$

where $A(\mathbf{y};\boldsymbol{\delta})$ is a $\ell \times \ell$ square matrix and $B(\mathbf{y};\boldsymbol{\delta})$ is a $(n-\ell) \times (n-\ell)$ square matrix. The matrix $A(\mathbf{y};\boldsymbol{\delta})$ is

$$\begin{pmatrix} \delta_1\rho_1 u_1'(y_1) - u_1'(y_1) & \delta_1\rho_2 u_1'(y_2) & \cdots & \delta_1\rho_\ell u_1'(y_\ell) \\ \delta_2\rho_1 u_2'(y_1) & \delta_2\rho_2 u_2'(y_2) - u_2'(y_2) & \cdots & \delta_2\rho_\ell u_2'(y_\ell) \\ \vdots & \vdots & \ddots & \vdots \\ \delta_\ell\rho_1 u_\ell'(y_1) & \delta_\ell\rho_2 u_\ell'(y_2) & \cdots & \delta_\ell\rho_\ell u_\ell'(y_\ell) - u_\ell'(y_\ell) \end{pmatrix}.$$

Then

$$A(\mathbf{x}^*;\mathbf{1}) = \begin{pmatrix} (\rho_1 - 1)u_1'(\underline{y}) & \rho_2 u_1'(\underline{y}) & \cdots & \rho_\ell u_1'(\underline{y}) \\ \rho_1 u_2'(\underline{y}) & (\rho_2 - 1)u_2'(\underline{y}) & \cdots & \rho_\ell u_2'(\underline{y}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_1 u_\ell'(\underline{y}) & \rho_2 u_\ell'(\underline{y}) & \cdots & (\rho_\ell - 1)u_\ell'(\underline{y}) \end{pmatrix}.$$

We claim that this matrix is non-singular. To see this, suppose $A(\mathbf{x}^*;\mathbf{1})$ is singular. Then there exists a linear combination of the columns of $A(\mathbf{x}^*;\mathbf{1})$ equal to zero vector in $\mathbb{R}^\ell$. That is, there exist $\alpha_1, \ldots, \alpha_\ell \in \mathbb{R}$ (not all zero) such that for each row $k$,

$$\begin{aligned} &\alpha_1\rho_1 u_k'(\underline{y}) + \cdots + \alpha_{k-1}\rho_{k-1} u_k'(\underline{y}) + \alpha_k(\rho_k - 1)u_k'(\underline{y}) + \alpha_{k+1}\rho_{k+1} u_k'(\underline{y}) + \cdots + \alpha_\ell\rho_\ell u_k'(\underline{y}) \\ &\quad = \quad 0. \end{aligned}$$

Since $u_k'(\underline{y}) \neq 0$ for all $k \in L$, it follows that for all $k \in L$, we have

$$\alpha_1\rho_1 + \cdots + \alpha_{k-1}\rho_{k-1} + \alpha_k(\rho_k - 1) + \alpha_{k+1}\rho_{k+1} + \cdots + \alpha_\ell\rho_\ell \quad = \quad 0, \tag{2}$$

which implies

$$\alpha_k \quad = \quad \sum_{j \in L} \alpha_j \rho_j. \tag{3}$$

Since the righthand side of (3) is independent of $k$, there exists $\alpha \neq 0$ such that $\alpha_k = \alpha$ for all $k \in L$. Then dividing (2) by $\alpha$, we obtain

$$\rho_1 + \cdots + \rho_\ell \quad = \quad 1,$$

contradicting $\rho_{j_r} > 0$. Therefore, $A(\mathbf{x}^*; \mathbf{1})$ is non-singular. A similar argument establishes that $B(\mathbf{x}^*; \mathbf{1})$ is non-singular as well. Thus, we conclude that $D_{\mathbf{x}} F(\mathbf{x}^*; \mathbf{1})$ is non-singular.

The implicit function theorem implies that there exist $\delta(\mathbf{z}) \in (0,1)$ and a continuous mapping $\phi$: $[\delta(\mathbf{z}), 1]^n \to \overline{X}$ such that for every $\boldsymbol{\delta} \in [\delta(\mathbf{z}), 1]^n$, we have $F(\phi(\boldsymbol{\delta}); \boldsymbol{\delta}) = 0$ and such that $\phi(\mathbf{1}) = \mathbf{x}^*$, i.e., for every $i \in N$, we have

$$\phi_i(\mathbf{1}) \quad = \quad \begin{cases} \underline{y} & \text{if } i \in L, \\ \overline{y} & \text{if } i \in R. \end{cases}$$

Taking $\delta(\mathbf{z})$ to be sufficiently close to one, it follows that for every $i \in N$ and for every $\boldsymbol{\delta} \in [\delta(\mathbf{z}), 1]^n$, we have in addition: (i) if $i \in L$, then $z_i \in (\phi_i(\boldsymbol{\delta}), \overline{y})$, and $u_j(\phi_i(\boldsymbol{\delta})) < u_j(\overline{y})$ for every $j \in L$, and $u_j(\phi_i(\boldsymbol{\delta})) > u_j(\overline{y})$ for every $j \in R$; and (ii) if $i \in R$, then $z_i \in (\underline{y}, \phi_i(\boldsymbol{\delta}))$, and $u_j(\phi_i(\boldsymbol{\delta})) > u_j(\underline{y})$ for every $j \in L$, and $u_j(\phi_i(\boldsymbol{\delta})) < u_j(\underline{y})$ for every $j \in R$. Given the choice $\delta(\mathbf{z})$, we then specify $\epsilon(\mathbf{z}) > 0$ sufficiently small that for every $i \in N$ and for every $\boldsymbol{\delta} \in [\delta(\mathbf{z}), 1]^n$, we have: (iii) if $i \in L$, then $\phi_i(\boldsymbol{\delta}) < z_i - \epsilon(\mathbf{z})$ and $z_i + \epsilon(\mathbf{z}) < \overline{y}$, and (iv) if $i \in R$, then $\underline{y} < z_i - \epsilon(\mathbf{z})$ and $z_i + \epsilon(\mathbf{z}) < \phi_i(\boldsymbol{\delta})$.

Now consider any $\boldsymbol{\delta} \in [\delta(\mathbf{z}), 1)^n$. Let $Y_i = [\phi_i(\boldsymbol{\delta}), \overline{y}]$ for each $i \in L$, and let $Y_i = [\underline{y}, \phi_i(\boldsymbol{\delta})]$ for each $i \in R$, so that $z_i \in \text{int} Y_i$ for each $i \in N$. Define $\mathbf{Y} = (Y_1, \ldots, Y_n)$. Clearly $\mathbf{Y} \in \Omega^n$. Strict quasi-concavity of utility functions, together with (i) and (ii) above, implies that for every $i \in L$, we have

$$\min_{y \in Y_j} u_i(y) = \begin{cases} u_i(\phi_j(\boldsymbol{\delta})) & \text{if } j \in L \\ u_i(\underline{y}) & \text{if } j \in R, \end{cases}$$

and for every $i \in R$, we have

$$\min_{y \in Y_j} u_i(y) = \begin{cases} u_i(\overline{y}) & \text{if } j \in L \\ u_i(\phi_j(\boldsymbol{\delta})) & \text{if } j \in R. \end{cases}$$

Then $F(\phi(\boldsymbol{\delta}); \boldsymbol{\delta}) = 0$ implies that for every $i \in N$, we have

$$\min_{x \in Y_i} u_i(x) \quad = \quad \delta_i \sum_{j \in N} \rho_j \min_{y_j \in Y_j} u_i(y_j),$$

and it follows that $\mathbf{Y}$ is plausible at $\boldsymbol{\delta}$ . By Lemma 1, we conclude that $\times_{i\in N}\text{int}Y_i \subseteq X(\boldsymbol{\delta})$. Since $B_{\epsilon(\mathbf{z})}(\mathbf{z}) \subseteq \times_{i\in N}\text{int}Y_i$, by (iii) and (iv), the claim is established.

Finally, consider any $\epsilon > 0$. By the above claim, for each $\mathbf{z} \in [\underline{x}+\epsilon, \overline{x}-\epsilon]^n$, there exist $\delta(\mathbf{z}) \in (0,1)$ and $\epsilon(\mathbf{z}) > 0$ such that if $\delta_i \geq \delta(\mathbf{z})$ for every $i \in N$, then $B_{\epsilon(\mathbf{z})}(\mathbf{z}) \subseteq X(\boldsymbol{\delta})$. The collection $\{B_{\epsilon(\mathbf{z})}(\mathbf{z}) \mid \mathbf{z} \in [\underline{x}+\epsilon, \overline{x}-\epsilon]^n\}$ is an open cover of $[\underline{x}+\epsilon, \overline{x}-\epsilon]^n$. Since $[\underline{x}+\epsilon, \overline{x}-\epsilon]^n$ is compact, there is a finite subcover, say $\{B_{\epsilon(\mathbf{z}^h)}(\mathbf{z}^h) \mid h = 1, \ldots, m\}$, such that $[\underline{x}+\epsilon, \overline{x}-\epsilon]^n \subseteq \bigcup_{h=1}^m B_{\epsilon(\mathbf{z}^h)}(\mathbf{z}^h)$. Letting $\underline{\delta} = \max\{\delta(\mathbf{z}^h) | h = 1, \ldots, m\}$ and $\delta_i \geq \underline{\delta}$ for every $i \in N$, we have $[\underline{x}+\epsilon, \overline{x}-\epsilon]^n \subseteq X(\boldsymbol{\delta})$, completing the proof. ∎

Distributive bargaining models such as Rubinstein (1982) and Baron and Ferejohn (1989) assume that every alternative is weakly preferred to the status quo by every agent; in terms of the notation of this paper, they assume that for all $i \in N$ and all $x \in X$, $u_i(x) \geq 0$. Banks and Duggan (2000) also impose this assumption in their spatial model of bargaining. In the context of the bad status quo model, Lemma 1 has a direct and significant consequence.

**Proposition 1** *Assume that for every $i \in N$ and every $x \in X$, we have $u_i(x) \geq 0$. Then for every $\boldsymbol{\delta} \in (0,1)^n$, we have $(\text{int}X)^n \subseteq X(\boldsymbol{\delta})$.*

Thus, regardless of individual discount factors, every interior alternative is such that for some equilibrium it is proposed by every agent and accepted. This result is a consequence of the simple observation that under the bad status quo assumption, the profile $(X, \ldots, X) \in \Omega^n$ is plausible, so Lemma 1 applies immediately. Note that the bad status quo assumption is essential to the strong folk theorem result of Proposition 1. In the equilibrium construction in the proof of Lemma 1, an agent $i$ is forced to propose undesirable alternatives, because the expected payoff from proposing other alternatives is not expected to be any better. The payoff from proposing another alternative places weight $1-\delta_i$ on the status quo payoff (zero) and weight $\delta_i$ on the continuation payoff; thus, if agent $i$'s status quo payoff is high relative to alternatives, then the agent cannot be forced to propose an undesirable alternative that would pass, and the set of alternatives that are proposed by $i$ and are accepted in some equilibrium is restricted to a proper subset of $X$. Of course, the weight $1-\delta_i$ on the status quo becomes negligible as $\delta_i$ approaches one, so when agents are patient, the restrictions on equilibrium proposals are determined almost entirely by the anticipated future proposals, opening the scope for the folk theorem in Theorem 1.

# A Proof of Lemma 1

**Lemma 1** *Let $\boldsymbol{\delta} \in (0,1)^n$. If $\mathbf{Y} \in \Omega^n$ is plausible at $\boldsymbol{\delta}$, then $\times_{i\in N}\text{int}Y_i \subseteq X(\boldsymbol{\delta})$.*

*Proof:* Let $\boldsymbol{\delta} \in [0,1)^n$ and let $\mathbf{Y} \in \Omega^n$ be plausible at $\boldsymbol{\delta}$. Let $(y_1,\ldots,y_n) \in \times_{i\in N}(\text{int}Y_i)$. It suffices to show that there exists $\sigma \in \Sigma(\boldsymbol{\delta})$ such that for each $i \in N$, agent $i$ proposes $y_i$ in the first period, and it is accepted. Our equilibrium construction relies on the definition of three mappings, denoted $\alpha_i^j$, $\beta_i^k$, $\gamma_i^k$ from alternatives to alternatives for each $i,j,k \in N$.

To define the mappings $\alpha_i^j$, note that for every $j \in N$, strict quasi-concavity of $u_j$ implies that the minimum of $u_j$ in $Y_j$ is attained at a boundary point of the interval. Thus, for all $x \in \text{int}Y_j$, we have $u_j(x) > \min_{z\in Y_j} u_j(z)$, and since $\mathbf{Y}$ is plausible, the inequality $u_j(x) > \delta_j \sum_{i\in N} \rho_i \min_{z_i \in Y_i} u_j(z_i)$ then holds. Continuity of $u_j$ then implies that for each $i,j \in N$, we can define a mapping $\alpha_i^j\colon \text{int}Y_j \to \text{int}Y_i$ such that for each $x \in \text{int}Y_j$ and each $k \in N$, we have $\alpha_i^j(x) \neq \hat{x}_k$, and so that in addition we have

$$u_j(x) \quad > \quad \delta_j \sum_{i\in N} \rho_i u_j(\alpha_i^j(x)). \tag{4}$$

Note that if agent $j$ compares having $x$ as the immediate outcome to rejection and having each $i$ propose and pass $\alpha_i^j(x)$ in the next period, then the agent strictly prefers the former to the latter. In our equilibrium construction, we understand $\alpha_i^j(x)$ to be an alternative proposed by agent $i$ if agent $j$ was supposed to propose $x$ in the preceding period and that alternative is rejected; in equilibrium, agent $j$'s proposal is only rejected when the agent deviates from $x$, so $\alpha_i^j(x)$ acts to punish such deviations.

For each $i,k \in N$, define the mapping $\beta_i^k\colon \text{int}Y_i \to X$ by

$$\beta_i^k(x) \quad = \quad \begin{cases} \frac{x+\min\{\hat{x}_k,\max Y_i\}}{2} & \text{if } x \leq \hat{x}_k, \\ \frac{x+\max\{\hat{x}_k,\min Y_i\}}{2} & \text{if } x > \hat{x}_k. \end{cases} \tag{5}$$

Thus, if $x < \hat{x}_k$, for example, then $\beta_i^k(x)$ is located between $x$ and $\hat{x}_k$, so that $u_k(\beta_i^k(x)) > u_k(x)$. We understand $\beta_i^k(x)$ to be the alternative proposed by agent $i$ to reward agent $k$ if the proposer in the previous period proposed the correct alternative, if this was rejected by all agents but agent $k$, and if $x$ would be the current proposal dictated by $\alpha_i^j$; this essentially moves the current proposal toward the ideal point of agent $k$, while remaining within the set $Y_i$ for the current proposer.

Define the mapping $\gamma_i^k\colon \text{int}Y_i \to X$ by

$$\gamma_i^k(x) \quad = \quad \begin{cases} \frac{x+\min Y_i}{2} & \text{if } x \leq \hat{x}_k, \\ \frac{x+\max Y_i}{2} & \text{if } x > \hat{x}_k. \end{cases} \tag{6}$$

Thus, if $x < \hat{x}_k$, for example, then $\gamma_i^k(x)$ is located between $\min Y_i$ and $x$, so that $u_k(\gamma_i^k(x)) < u_k(x)$. We understand $\gamma_i^k(x)$ to be the alternative proposed by agent $i$ to punish agent $k$ if the proposer in the previous period proposed the wrong alternative, if this was rejected by all agents but agent $k$, and if $x$ would be the current proposal dictated by $\alpha_i^j$; this essentially moves the current proposal away from the ideal point of agent $k$, while remaining within the set $Y_i$ for the current proposer.

To construct an equilibrium $\sigma$ such that each agent $i$ proposes $y_i$ in the first period and it is accepted, we define proposal and voting strategies as follows. For each $i \in N$, we let $\mathbf{b}^i = (b_1^i, \ldots, b_n^i) \in \{a, r\}^n$ denote the voting profile such that $b_i^i = a$ and $b_j^i = r$ for all $j \in N \setminus \{i\}$. Now, for each $i \in N$, define the proposal strategy $p_i \colon H^{pi} \to X$ so that for each $t = 1, 2, \ldots$ and each $h_t \in H_t^{pi}$,

1. $p_i(h_1) = y_i$,
2. if $t = 2, 3 \ldots$ and for some $h_{t-1} \in H_{t-1}^{pj}$, $x \in X$, and $\mathbf{b} \in B$, we have $h_t = (h_{t-1}, x, \mathbf{b}, i)$, then

$$p_i(h_t) = \begin{cases} \alpha_i^j(p_j(h_{t-1})) & \text{if } \mathbf{b} \neq \mathbf{b}^k \text{ for all } k \in N, \\ \beta_i^k(\alpha_i^j(p_j(h_{t-1}))) & \text{if } \mathbf{b} = \mathbf{b}^k \text{ for some } k \in N \text{ and } x = p_j(h_{t-1}), \\ \gamma_i^k(\alpha_i^j(p_j(h_{t-1}))) & \text{if } \mathbf{b} = \mathbf{b}^k \text{ for some } k \in N \text{ and } x \neq p_j(h_{t-1}). \end{cases}$$

In particular, if agent $j$ is supposed to propose $p_j(h_{t-1})$ in period $t-1$, if the agent's proposal is rejected, and if there is not exactly one vote to accept, then agent $i$ proposes $\alpha_i^j(p_j(h_{t-1}))$ to punish $j$. If agent $j$ proposes the correct alternative and exactly one agent, say $k$, accepts $j$'s proposal, then $i$ rewards $k$ with $\beta_i^k(\alpha_i^j(p_j(h_{t-1})))$. And if agent $j$ proposes the incorrect alternative and exactly one agent $k$ accepts it, then $i$ punishes $k$ with $\gamma_i^k(\alpha_i^j(p_j(h_{t-1})))$. The effectiveness of these proposal strategies stems from the fact that by our specification of voting strategies, below, agent $j$'s proposal is rejected if and only if the agent proposes $x \neq p_j(h_{t-1})$, in which case it is rejected by a unanimous vote.

For each $k \in N$, define the voting strategy $v_k \colon H^v \to \{a, r\}$ so that for every $j \in N$, every $h \in H^{pj}$, and every $x \in X$,

$$v_k(h, x) = \begin{cases} a & \text{if } x = p_j(h), \\ r & \text{if } x \neq p_j(h). \end{cases}$$

That is, if the correct alternative $p_j(h)$ is proposed then all agents accept, and otherwise all agents reject.

Let $\sigma = (p_i, v_i)_{i \in N}$. Since each agent $i$ proposes $y_i$ initially and all agents accept the proposal, it remains to be shown that $\sigma \in \Sigma(\boldsymbol{\delta})$. Consider any $j \in N$ and any $h \in H^{pj}$. Note that by construction, $p_j(h) \in \mathrm{int} Y_j$. By the construction of voting strategies, $p_j(h)$ passes if proposed, implying $U_j^\sigma(h, p_j(h)|\boldsymbol{\delta}) = \delta_j^{\tau(h)-1} u_j(p_j(h))$. Now consider any $x \neq p_j(h)$. If agent $j$ deviates and proposes $x$ at $h$, then the outcome of voting is $v(h,x) = (r, \ldots, r)$ and so, in the next period, each $i \in N$ proposes $\alpha_i^j(p_j(h))$, which passes by a unanimous vote. Thus, we have

$$U_j^\sigma(h, x|\boldsymbol{\delta}) \;\;=\;\; \delta_j^{\tau(h)} \sum_{i \in N} \rho_i u_j(\alpha_i^j(p_j(h))).$$

Then, by (4), we have

$$U_j^\sigma(h, p_j(h)|\boldsymbol{\delta}) \;\;>\;\; U_j^\sigma(h, x|\boldsymbol{\delta}),$$

so the deviation is not profitable.

Next, consider any $h \in H^v$, and write $h = (h', x)$ for some $h' \in H^{pj}$ and $x \in X$, and consider any agent $k \in N$. First, suppose $x = p_j(h')$. Then the outcome of voting is $v(h) = (a, \ldots, a)$. Since $N \in \mathcal{D}$ and $N \setminus \{k\} \in \mathcal{D}$, we have

$$U_k^\sigma(h, (v_k(h), v_{-k}(h))|\boldsymbol{\delta}) \;\;=\;\; U_k^\sigma(h, (b_k, v_{-k}(h))|\boldsymbol{\delta}) \;\;=\;\; u_k(x)$$

for every $b_k \in \{a, r\}$. Second, suppose $x \neq p_j(h')$. Then $v(h) = (r, \ldots, r)$. Note that $\emptyset \notin \mathcal{D}$ and $\{k\} \notin \mathcal{D}$. By construction of proposal strategies, we have

$$U_k^\sigma(h, (v_k(h), v_{-k}(h))|\boldsymbol{\delta}) \;\;=\;\; \delta_k^{\tau(h)} \sum_i \rho_i u_k(\alpha_i^j(p_j(h'))),$$

and

$$U_k^\sigma(h, (a, v_{-k}(h))|\boldsymbol{\delta}) \;\;=\;\; \delta_k^{\tau(h)} \sum_i \rho_i u_k(\gamma_i^k(\alpha_i^j(p_j(h')))).$$

Recall that for every $k \in N$, we have $\alpha_i^j(p_j(h')) \neq \hat{x}_k$, so (6) implies $u_k(\alpha_i^j(p_j(h'))) > u_k(\gamma_i^k(\alpha_i^j(p_j(h'))))$. Thus,

$$U_k^\sigma(h, (v_k(h), v_{-k}(h))|\boldsymbol{\delta}) \;\;>\;\; U_k^\sigma(h, (a, v_{-k}(h))|\boldsymbol{\delta}),$$

so again no agent has a profitable deviation. Therefore, $\sigma$ is a subgame perfect equilibrium.

Continuing, we must show that $\sigma$ is in stage-undominated voting strategies. Let $\mathbf{b}_{-k}^k$ be the profile of ballots for all $i \in N \setminus \{k\}$ such that they all choose $r$. In case $x \neq p_j(h')$, we have $v_k(h) = r$. In the previous paragraph, we have already shown that

$$U_k^\sigma(h, (v_k(h), \mathbf{b}_{-k}^k)|\boldsymbol{\delta}) \;\;>\;\; U_k^\sigma(h, (a, \mathbf{b}_{-k}^k)|\boldsymbol{\delta}),$$

so agent $k$'s acceptance leads to a strictly lower payoff. In case $x = p_j(h')$, we have $v_k(h) = a$. Then, by construction, we have

$$U_k^\sigma(h, (v_k(h), \mathbf{b}_{-k}^k)|\boldsymbol{\delta}) = \delta_k^{\tau(h)} \sum_i \rho_i u_k(\beta_i^k(\alpha_i^j(x))),$$

and

$$U_k^\sigma(h, (r, \mathbf{b}_{-k}^k)|\boldsymbol{\delta}) = \delta_k^{\tau(h)} \sum_i \rho_i u_k(\alpha_i^j(x)).$$

Recall that for every $i \in N$, we have $\alpha_i^j(x) \neq \hat{x}_k$, so (5) implies $u_k(\beta_i^k(\alpha_i^j(x))) > u_k(\alpha_i^j(x))$. Thus,

$$U_k^\sigma(h, (v_k(h), \mathbf{b}_{-k})|\boldsymbol{\delta}) > U_k^\sigma(h, (r, \mathbf{b}_{-k})|\boldsymbol{\delta}),$$

establishing that $\sigma$ is in stage-undominated voting strategies. We conclude that $\sigma \in \Sigma(\boldsymbol{\delta})$, as required. ∎

## References


[1] D. Baron and J. Ferejohn (1989) "Bargaining in Legislatures," *American Political Science Review,* 83: 1181–1206.

[2] J. Banks and J. Duggan (2000) "A Bargaining Model of Collective Choice," *American Political Science Review,* 94: 73–88.

[3] J. Banks and J. Duggan (2006) "A General Bargaining Model of Legislative Policy-making," *Quarterly Journal of Political Science,* 1: 49–85.

[4] D. Cardona and C. Ponsati (2011) "Uniqueness of Stationary Equilibria in Bargaining One-dimensional Policies under (Super) Majority Rules," *Games and Economic Behavior,* 73: 65–75.

[5] S. Cho and J. Duggan (2003) "Uniqueness of Stationary Equilibria in a One-dimensional Model of Bargaining," *Journal of Economic Theory,* 113: 118–130.

[6] S. Cho and J. Duggan (2009) "Bargaining Foundations of the Median Voter Theorem," *Journal of Economic Theory,* 144: 851–868.

[7] J. Duggan and M. Fey (2006) "Repeated Downsian Electoral Competition," *International Journal of Game Theory,* 35: 39–69.

[8] J. Duggan (2014) "A Folk Theorem for Repeated Elections with Adverse Selection," *Political Science Research and Methods,* forthcoming.

[9] P. Dutta (1995) "A Folk Theorem for Stochastic Games," *Journal of Economic Theory,* 66: 1–32.

[10] H. Eraslan (2002) "Uniqueness of Stationary Equilibrium Payoffs in the Baron-Ferejohn Model," *Journal of Economic Theory,* 103: 11–30.

[11] H. Eraslan and A. McLennan (2012) "Uniqueness of Stationary Equilibrium Payoffs in Coalitional Bargaining," *Journal of Economic Theory,* forthcoming.

[12] D. Fudenberg, D. Levine, and E. Maskin (1994) "The Folk Theorem with Imperfect Public Information," *Econometrica,* 62: 997–1040.

[13] D. Fudenberg and E. Maskin (1986) "The Folk Theorem for Repeated Games with Discounting or with Incomplete Information," *Econometrica,* 54: 533–554.

[14] D. Fudenberg and Y. Yamamoto (2011) "The Folk Theorem for Irreducible Stochastic Games with Imperfect Public Monitoring," *Journal of Economic Theory,* 146: 1664–1683.

[15] J.-J. Herings, A. Meshalkin, and A. Predtetchinski (2012) "A Folk Theorem for Bargaining Games," Maastrict University working paper RM/12/055.

[16] J. Horner, T. Sugaya, S. Takahasi, and N. Vielle (2011) "Recursive Methods in Discounted Stochastic Games: An Algorithm for $\delta \to 1$ and a Folk Theorem," *Econometrica,* 79: 1277–1318.

[17] P. Norman (2002) "Legislative Bargaining and Coalition Formation," *Journal of Economic Theory,* 102: 322–353.

[18] A. Rubinstein (1982) "Perfect Equilibrium in a Bargaining Model," *Econometrica,* 50: 97–109.

[19] Q. Wen (2002) "A Folk Theorem for Repeated Sequential Games," *Review of Economic Studies,* 69: 493–512.