# Bargaining Foundations of the Median Voter Theorem

Seok-ju Cho
Dept. of Political Science
University of Rochester

John Duggan
Dept. of Political Science
and Dept. of Economics
University of Rochester

September 26, 2003

**Abstract**

We provide an "anti-folk theorem" result for a one-dimensional bargaining model based on Baron and Ferejohn's (1989) model of distributive politics. We prove that, as the agents become arbitrarily patient, the set of proposals that can be passed in any pure strategy subgame perfect equilibrium collapses to the median voter's ideal point. While we leave the possibility of some delay, we prove that the agents' equilibrium continuation payoffs converge to the utility from the median, so that delay, if it occurs, is inconsequential. We do not impose stationarity or any other refinements. This contrasts with the known result for the distributive model that, as agents become patient, any division of the dollar can be supported as a subgame perfect equilibrium outcome, and it provides a strong game-theoretic foundation for Black's (1958) median voter theorem.

## 1 Introduction

Numerous applications in political science and political economy rely on the implications of single-peaked preferences in one-dimensional environments. First noted by Hotelling (1929) and Downs (1957), the median voter theorem dictates that the median of the distribution of voter ideal points bears the preference of a majority of voters to every other alternative. The median, in other words, is the unique element of the core of the majority voting game. This was formalized by Black (1958) and Arrow (1963), who proved that, when the number of voters is odd, the majority preference relation is transitive with the core point being the unique maximal element of the majority ranking. The median voter theorem has facilitated applications by offering concrete predictions in models of committees and elections, indeed, doing so without the prerequisite of defining a non-cooperative game form to describe the strategic calculations of individual decision-makers.

While an advantage in maximizing the flexibility of applied models, the lack of a non-cooperative underpinning of the median voter theorem is a disadvantage in another respect: without a firm game-theoretic foundation, we cannot be sure that the predictions of the median voter theorem are consistent with the incentives of strategically sophisticated agents. More precisely, we do not know what kinds of restrictions on individual preferences and institutional procedures will lead to equilibrium outcomes at or near the median voter's ideal point.[1]

We analyze committee decision-making using a non-cooperative infinite-horizon bargaining model based on a random recognition rule and majority voting: in any period, an agent is randomly selected and proposes an alternative in a one-dimensional policy space, which is then subject to a majority vote; if the proposal passes, then the proposed policy is implemented, and the game ends; and if the proposal fails, then play moves to the next period where this procedure is repeated. Agents' preferences over alternatives are represented by arbitrary strictly concave (and therefore single-peaked) utility functions. We prove two results. First, as the agents become arbitrarily patient, the set of proposals that can pass after any history in any pure strategy subgame perfect equilibrium converges to the median voter's ideal point. Beyond the focus on pure strategies, we do not impose any equilibrium refinements, e.g., we do not impose stationarity. Second, while we do not preclude the possibility of delay in equilibrium as the agents become patient, delay becomes negligible: the set of payoffs for an agent after any history in any pure strategy subgame perfect equilibrium converges to the utility from the median. Thus, we provide strong support for the predictions of the median voter theorem.

The connections between *stationary* subgame perfect equilibrium outcomes of this bargaining model and the majority core have been explored previously. Banks and Duggan (2000) prove existence of a pure strategy stationary equilibrium, and they show that, as agents become patient, the set of proposals that can be passed in any stationary equilibrium collapses to the

[1] See Bergin and Duggan (1999) for a discussion of non-cooperative foundations that separate these two kinds of primitives.

core point. Moreover, unless some agents are perfectly patient, stationary equilibria exhibiting delay do not exist. Whereas the authors assume a "bad status quo" in that paper, Banks and Duggan (2003) assume a common discount factor and allow for the status quo to be an arbitrary element of the policy space. They prove existence and that, again, stationary equilibrium outcomes collapse to the core as agents become arbitrarily patient. Delay can occur, but only if the status quo is the unique core point, and then that alternative is the only proposal that can possibly pass. Thus, delay, if it occurs, cannot affect the alternative realized in any period. Our first result generalizes the core convergence found in these papers by dropping the refinement of stationarity. Non-stationary equilibria can exhibit delay, even when the status quo is not at the core, but our second result shows that the payoffs of the agents, when patient enough, will not be significantly affected by delay.

Our model is related to the literature on infinite-horizon bargaining models initiated by Rubinstein (1982), who considers an alternating-offer protocol for two agents, and Binmore (1987), who assumes the proposer is randomly drawn in each period. In this work, an alternative is an allocation of a private good ("pie") to the agents, and a proposer must obtain the assent of the other agent, so proposals are essentially subject to a unanimity voting rule. Models featuring majority voting among multiple agents trace back to subsequent work by Baron and Ferejohn (1989), who also consider bargaining in the distributive setting. They solve for the unique symmetric stationary subgame perfect equilibrium of their model, in which a proposer offers some of the good to the "cheapest" majority possible and offers zero to the remaining agents. Harrington (1989,1990a,b) examines the effects of risk aversion in this setting. More recently, Eraslan (2002) drops the restriction of symmetry and establishes that the Baron-Ferejohn equilibrium is unique among all stationary equilibria. Eraslan and Merlo (2002) show that this conclusion does not hold generally if the amount of money to be divided varies stochastically over time.

Most work on majority-rule bargaining has focused on stationary equilibria. By virtue of their simplicity, such equilibria may possess a focal effect, lending some justification to stationarity as a refinement of subgame perfect equilibrium.[2] Of course, their relative tractability also makes them a natural first object of study. But the logic of Nash equilibrium alone does not preclude the possibility of other, non-stationary subgame perfect equi-

[2] See Baron and Kalai (1993) for a formalization of these ideas.

libria, in which agents adopt history-dependent strategies. This would seem a problem especially in long-standing institutions, where norms dictating non-stationary behavior may arise over time. The folk theorem suggests that this problem will be exacerbated when the agents are very patient.[3] Indeed, assuming at least three agents, Baron and Ferejohn (1989) prove that *every* allocation of private good can be supported as a subgame perfect equilibrium outcome as bargainers in the distributive model become very patient. Our results show that this folk theorem result does not carry over to the one-dimensional bargaining model, and in fact the opposite occurs: as agents become very patient, the set of subgame perfect equilibrium outcomes converges to the unique core point.

In deriving our characterization result, we are able to extend the framework of Baron and Ferejohn (1989) and Banks and Duggan (2000, 2003) in several ways. First, we allow for both models of the status quo that have been considered, i.e., the status quo may be generally bad for the agents (as when no policy is currently in place) or may itself be an alternative. Second, whereas the probability that a particular agent is selected as proposer is fixed in the standard framework, we allow these recognition probabilities to vary with histories quite arbitrarily. Thus, for example, the probability that one agent is selected in period $t+1$ can depend on the proposal in period $t$ and on the identities of the agents who voted to reject that proposal. We require only that each agent's recognition probability has a positive lower bound. This excludes models in which proposers are chosen in a pre-determined order, but it allows us to approximate such deterministic models to an arbitrary degree.

Third, we modify the basic model by stipulating that voting is sequential, with one agent's vote observed by all later agents. Because the stationarity refinement essentially reduces the voting stage to a binary vote (the proposal vs. continuation play following rejection), the voting stage is usually treated as a simultaneous vote in the majority-rule bargaining literature. This essentially gives each agent a unique "stage-undominated" strategy in every voting subgame,[4] and it is assumed that the agents vote accordingly. Stationary equilibrium outcomes are unchanged if voting is sequential, regardless of the order of voting, and the additional dominance refinement is then unnecessary. When stationarity is dropped, however, continuation

[3]See Fudenberg and Maskin (1986) for folk theorem results for repeated games.

[4]Both votes are undominated if an agent is indifferent, but, under standard convexity conditions, this is not a factor in determining equilibrium voting outcomes.

equilibria may punish particular agents for voting the “wrong” way, and stage-dominance loses its bite. It therefore becomes necessary to model voting as sequential.[5] We do not assume that the order of voting is fixed. Instead, it is determined randomly; we allow the distribution over voting orders to vary with the history, including the current proposal; and we allow for uncertainty regarding the voting order even after a proposal is made. We require only that the probability of a special class of voting orders has a positive lower bound, an exigency that arises because, without any refinement of subgame perfection, we do not restrict the dependence of continuation equilibria on the votes of individual agents.

Last, unlike some work on bargaining with majority-rule voting, we allow for heterogeneous time preferences among the agents. Specifically, our main result holds for sequences of discount factors for the agents that converge to one and satisfy a “convergence condition,” which formalizes the idea that one agent’s discount factor not converge much more quickly than the others’. Thus, the asymptotic core equivalence result of Banks and Duggan (2003), which assumes a common discount factor, extends to the heterogeneous case.

The organization of the paper is as follows. In Section 2, we set up the model. In Section 3, we state our main result and give an overview of the proof. Section 5 contains the formal proof of our theorem. In Section 4, we discuss some of the related literature. Proofs of lemmas are contained in an appendix.

## 2 The Model

We develop the model in a series of steps.

*1. Bargaining protocol*

Let $N = \{1, \ldots, n\}$ denote a set of $n \geq 2$ agents who play an infinite-horizon bargaining game over a set $X$ of alternatives. Assume $X \subseteq \mathbb{R}$ is nonempty, compact, and convex, i.e., $X$ is a nonempty, closed, bounded interval. In any period $t = 1, 2, \ldots$ prior to the choice of an alternative, bargaining is as follows: (1) an agent $i$ is selected by nature to propose an

[5] It is well-known that simultaneous voting in a majority-rule voting game can lead to either decision as a Nash equilibrium, for when all agents vote the same way, no one agent is “pivotal.”

alternative; (2) a "step 1 voting state," denoted $s_1$, is selected by nature from a finite set $S_1$; (3) after observing $s_1$, $i$ makes a proposal, say $x$; (4) a "step 2 voting state," denoted $s_2$, is selected by nature from a finite set $S_2$; and (5) a sequential vote is held in some order $\phi(\cdot|s_1, s_2)\colon N \to N$, where the first voter $\phi(1|s_1, s_2)$ casts a vote $v_{\phi(1|s_1,s_2)} \in \{a, r\}$ to accept or reject the proposal, this is observed by all agents, then $\phi(2|s_1, s_2)$ casts a vote, and so on. The outcome of voting is given by a fixed collection $\mathcal{D} \subseteq 2^N \setminus \{\emptyset\}$ of *decisive coalitions.* If the set of voters voting for the proposal is decisive, i.e., $\{j \in N | v_j = a\} \in \mathcal{D}$, then the proposal is chosen and bargaining ends with outcome $(x, t)$. Otherwise, the above procedure (1)–(5) is repeated in period $t + 1$.

We endow each agent with a continuous and strictly concave utility function $u_i\colon X \to \mathbb{R}$, which we use later to define the payoffs of the agents. We let $u\colon X \to \mathbb{R}^n$ denote the vector-valued utility function defined by $u(x) = (u_1(x), \ldots, u_n(x))$ for all $x \in X$. Each $u_i$ is then maximized by a unique point, denoted $\tilde{x}_i$, the *ideal point* of the agent. Let agent $k$ satisfy

$$\tilde{x}_k \quad = \quad \min\left\{\tilde{x}_i \;\middle|\; |\{j \in N \mid \tilde{x}_i < \tilde{x}_j\}| \leq \frac{n}{2}\right\}.$$

That is, the agents with ideal points to the right of $k$'s do not make up a majority, and $k$ is the "leftmost" agent possessing this property.

We assume that voting is by majority rule, with a minor modification in case $n$ is even. Thus, when $n$ is odd, we define

$$\mathcal{D} \quad = \quad \left\{C \subseteq N \;\middle|\; |C| > \frac{n}{2}\right\}.$$

When $n$ is even, there arises the possibility that the voters are evenly divided between accepting and rejecting a proposal. In this case, we give agent $k$ the power to break ties by extending the above definition as follows:

$$\mathcal{D} \quad = \quad \left\{C \subseteq N \;\middle|\; |C| > \frac{n}{2}\right\} \cup \left\{C \subseteq N \;\middle|\; |C| \geq \frac{n}{2} \text{ and } k \in C\right\}.$$

An implication of our assumptions is that $\mathcal{D}$ is nonempty, *proper* ($C \in \mathcal{D}$ implies $N \backslash C \notin \mathcal{D}$), and *strong* ($C \notin \mathcal{D}$ implies $N \backslash C \in \mathcal{D}$).

## *2. Histories*

A history is a finite or infinite sequence of actions of agents and nature. A *complete history* is a history that is followed by the selection of a proposer and step 1 voting state. These include the initial history, $\emptyset$, and any history ending with $n$ votes of the agents. A *proposer history for $i$* is any history in which agent $i$ and any step 1 voting state have just been selected by nature, and so $i$ must next propose an alternative; a *step 2 history* is any history in which a selected agent has just made a proposal, and so a step 2 voting state must be selected by nature; and a *voting history for $i$* is any history in which it is $i$'s turn to vote. Technically, we let $H_t^c$ denote the set of $t$-period complete histories, $H_t^{pi}$ the set of $t$-period proposer histories for $i$, $H_t^{s_2}$ the set of $t$-period step 2 histories, and $H_t^{vi}$ the set of $t$-period voting histories for $i$. We specify $H_0^c = \{\emptyset\}$, and for $t = 1, 2, \ldots$, we define

$$\begin{aligned}
H_t^{pi} &= H_{t-1}^c \times \{i\} \times S_1 \\
H_t^{s_2} &= \left(\bigcup_{i=1}^{n} H_t^{pi}\right) \times X \\
H_t^{vi} &= H_t^{s_2} \times S_2 \times \left(\bigcup_{C \subseteq N \setminus \{i\}} \{a, r\}^C\right) \\
H_t^c &= H_t^{s_2} \times S_2 \times \{a, r\}^N.
\end{aligned}$$

Then the sets of proposer histories for $i$, of step 2 histories, of voter histories for $i$, and of complete histories are defined as

$$H^{pi} = \bigcup_{t=1}^{\infty} H_t^{pi}, \;\; H^{s_2} = \bigcup_{t=1}^{\infty} H_t^{s_2}, \;\; H^{vj} = \bigcup_{t=1}^{\infty} H_t^{vj}, \;\; H^c = \bigcup_{t=0}^{\infty} H_t^c,$$

respectively. Let

$$H = \bigcup_{i \in N} [H^{pi} \cup H^{s_2} \cup H^{vi} \cup H^c]$$

denote the set of all *finite histories.*

Thus, at any history $h \in H^{pi}$, agent $i$ is the active player at $h$, and $i$'s action set is $A_i(h) = X$; at $h \in H^{s_2}$, nature, denoted $n+1$, is the active player at $h$, and nature's action set is $A_{n+1}(h) = S_2$; at $h \in H^{vi}$, agent $i$ is the active player at $h$, and $i$'s action set is $A_i(h) = \{a, r\}$; and at $h \in H^c$, nature is the active player at $h$, and nature's action set is $A_{n+1}(h) = N \times S_1$. Let $A = X \cup S_2 \cup \{a, r\} \cup N \times S_1$ denote the action space of the bargaining game. Define the mapping $\iota\colon H \to N \cup \{n+1\}$ such that, for each finite history, $\iota(h)$ is the active player at $h$.

Define binary relations $<$ and $\ll$ on $H$ as follows. For any $h, h' \in H$, we say $h'$ *immediately follows* $h$, written as $h < h'$, if $h' \in \{h\} \times A_{\iota(h)}(h)$. That is, $h < h'$ if $h'$ is equal to $h$ with the addition of an action by the active player $h$. We say $h'$ *follows* $h$ if there exist histories $h^1, \ldots, h^T \in H^c$ such that $h < h^1 < \cdots < h^T = h'$. Thus, $\ll$ is the transitive closure of $<$, and $h \ll h'$ holds if and only if $h$ is an initial segment of $h'$.

Let $H^\bullet(x)$ denote the set of complete histories in which $x$ has been proposed and accepted by a decisive coalition, i.e.,

$$H^\bullet(x) \quad = \quad \left\{ h' \in H^c \;\middle|\; \begin{array}{c} \text{there exists } h \in H^c \text{ such that} \\ h' = (h, i, s_1, x, s_2, v_{\varphi(1)}, \ldots, v_{\varphi(n)}), \\ \varphi = \phi(s_1, s_2), \text{ and} \\ \{j \in N \mid v_{\varphi^{-1}(j)} = a\} \in \mathcal{D} \end{array} \right\},$$

and let $H^\bullet = \bigcup_{x \in X} H^\bullet(x)$ denote the set of all *terminal histories.* Define the mapping $\chi \colon H^\bullet \to X$ by $\chi(h) = x$ for any terminal history $h \in H^\bullet(x)$, and define the mapping $\tau \colon H^\bullet \to \mathbb{N}$ by $\tau(h) = t$ for any terminal history $h \in H^c_t$.

Let $H^\infty$ be the set consisting of every *infinite history,* which is an infinite sequence in $\{\emptyset\} \cup N \cup S_1 \cup X \cup S_2 \cup \{a, r\}$ such that every initial segment is a non-terminal history. Of course, every complete truncation of an infinite history must end in the rejection of the proposed alternative, so that any history in $H^\infty$ is characterized by infinite delay. Given $h \in H$ and $h' \in H^\infty$, write $h \ll h'$ if $h$ is an initial segment of $h'$. Finally, let $\overline{H} = H^\bullet \cup H^\infty$ be the set of histories fully describing a play of the bargaining game.

## *3. Payoffs and the core*

Each agent $i$'s preferences over $\overline{H}$ are given by a payoff function $W_i \colon \overline{H} \to \mathbb{R}$ with the following representation: we posit a discount factor $\delta_i \in (0, 1)$, and a "status quo payoff" $\overline{u}_i$ such that

$$W_i(h|\delta) \quad = \quad \begin{cases} (1 - \delta_i^{\tau(h)-1})\overline{u}_i + \delta_i^{\tau(h)-1}, u_i(\chi(h)) & \text{if } h \in H^\bullet, \\ \overline{u}_i & \text{if } h \in H^\infty, \end{cases}$$

for all $h \in \overline{H}$, where $\delta = (\delta_1, \ldots, \delta_n)$ is the vector of discount factors. We interpret these payoffs as generated by a flow, where the agent receives the status quo payoff in every period until a proposal is passed and then receives the utility from that chosen alternative, all payoffs discounted over time.

It will be notationally convenient to append a *status quo* alternative $q \notin X$ to the set of alternatives and to extend the utility functions of the agents to $X \cup \{q\}$ so that $u_i(q) = \overline{u}_i$ for all $i \in N$. This is consistent with the view that some alternative $x^q \in X$ is in place until a proposal is passed, because we do not preclude the possibility that $q$ generates the same utilities as some alternative in $X$. A *lottery* is a Borel probability measure on $X \cup \{q\}$,[6] and we endow the space of lotteries, $\Lambda$, with the weak* topology. We maintain either of two assumptions on the agents' status quo payoffs:

(A1) there exists $x^q \in X$ such that, for all $i \in N$, $\overline{u}_i = u_i(x^q)$,

(A2) for all $i \in N$ and all $x \in X$, $u_i(\tilde{x}_k) > \overline{u}_i$.

The former assumption formalizes the idea that the status quo payoff is generated by an alternative, in place until some other alternative is chosen, and the latter formalizes the idea that delay is unanimously bad for the agents relative to the core point.

The *core*, denoted $K$, consists of the alternatives that are weakly preferred to all others according to the voting rule:

$$K = \left\{ x \in X \;\middle|\; \begin{array}{c} \text{for all } y \in X \text{ and all } C \in \mathcal{D}, \text{ there} \\ \text{exists } i \in C \text{ such that } u_i(x) \geq u_i(y) \end{array} \right\}.$$

That $K$ is nonempty follows because $\mathcal{D}$ is proper, $X$ is one-dimensional, and agents' preferences are "single-peaked." Since $\mathcal{D}$ is also strong, $K$ is actually a singleton and consists of the ideal point $\tilde{x}_k$ of agent $k$, defined above. Thus, our assumption on $\mathcal{D}$ for the $n$ even case amounts to allowing the "core voter" to break ties in case the voters are evenly split. Defining

$$\begin{aligned} C_K &= \{i \in N \mid \tilde{x}_i = \tilde{x}_k\} \\ C_L &= \{i \in N \mid \tilde{x}_i < \tilde{x}_k\} \\ C_R &= \{i \in N \mid \tilde{x}_k < \tilde{x}_i\} \end{aligned}$$

we have $C_K \cup C_L \in \mathcal{D}$ and $C_K \cup C_R \in \mathcal{D}$. Note that $C_K$ includes agent $k$ and may include other agents as well, since we do not assume the agents' ideal points are distinct.

[6] Here, we give $X \cup \{q\}$ the topology in which open sets are of the form $X \cap G$ or $(X \cap G) \cup \{q\}$, where $G$ is open in $\mathbb{R}$. Thus, $X \cup \{q\}$ remains compact.

At any proposer history for $i$, the agent observes the history and has action set $X$, the set of possible proposals. At any voter history for $i$, the agent observes the history and has action set $\{a, r\}$. Thus, a *pure strategy* for $i$ is a pair of mappings, $p_i\colon H^{pi} \to X$ and $v_i\colon H^{vi} \to \{a, r\}$, where $p_i(h)$ describes what $i$ would propose if selected as proposer after history $h$, and $v_i(h)$ describes $i$'s vote after $h$. An alternative representation that will be useful is a probability measure $\sigma_i(\cdot|h)$ that is degenerate on $p_i(h)$ for all $h \in H^{pi}$ and degenerate on $v_i(h)$ for all $h \in H^{vi}$. A pure strategy profile is then denoted $\sigma = (\sigma_1, \ldots, \sigma_n)$.

The selection of a proposer and step 1 voting state after any complete history $h$ is random, given by a probability distribution $\rho(\cdot|h)$ on $N \times S_1$, which can depend on the history quite arbitrarily. The determination of the voting order is also random, and, to maximize generality, we allow the voting order to be determined in two steps: a pair of voting states, $(s_1, s_2)$, uniquely determines a voting order $\phi(\cdot|s_1, s_2)$, where the mapping $\phi\colon S_1 \times S_2 \to N^N$ is onto the set of permutations of $N$. The distribution over $S_2$, and therefore the distribution of voting orders, may depend quite arbitrarily on the preceding complete history, the selected proposer, the step 1 voting state, and the current proposal. We let $\pi(\cdot|h, i, s_1, x)$ denote this distribution on $S_2$, and we let $\Phi \subseteq N^N$ be the set of permutations $\varphi$ on $N$ such that: $\varphi(1) = k$ and either

- for even $j$, $\varphi(j) \in C_K \cup C_L$; and for odd $j > 1$, $\varphi(j) \in C_K \cup C_R$, or
- for even $j$, $\varphi(j) \in C_K \cup C_R$; and for odd $j > 1$, $\varphi(j) \in C_K \cup C_L$.

That is, $\Phi$ is the set of voting orders in which the core voter votes first, and subsequently voters alternate from either side of the core.

Our only restriction on the determination of the proposer and voting order is the following:

$$\mu \;=\; \inf_{h \in H^c, i \in N} \sum_{(s_1, s_2) \in \phi^{-1}(\Phi)} \rho(i, s_1|h)\pi(s_2|h, i, s_1, \tilde{x}_k) \;>\; 0.$$

An implication is that each agent's probability of proposing is bounded strictly above zero as we vary over proposer histories. Thus, we do not capture sequential proposal models, where the agents "take turns" making proposals, but we can approximate them arbitrarily closely. Furthermore, while the proposer observes $(h, i, s_1)$ before choosing $x$, he/she does not observe the realization of $s_2$ before this choice. Thus, we allow for the possibility that the proposer has full, partial, or perhaps no information about the order of voting in which his/her proposal is to be considered. Our restriction on the distribution of voting orders is satisfied if, for example, a minimal amount of uncertainty regarding the order of voting is present before every proposal.

It will be useful to introduce notation for nature's strategy that is consistent with our notation for agents' strategies: let $\sigma_{n+1}(\cdot|h)$ be a probability measure such that $\sigma_{n+1}(\cdot|h) = \rho(\cdot|h)$ for all $h \in H^c$, and let $\sigma_{n+1}(\cdot|h) = \pi(\cdot|h)$ for all $h \in H^{s_2}$.

*5. Distributions over histories and expected payoffs*

Beginning at any finite history $h \in H$, a strategy profile $\sigma$ determines a transition probability for histories following $h$. Specifically, consider $h, h' \in H$ with $h \ll h'$, say $h = h^0 < h^1 < \cdots < h^T = h'$. For each $t = 1, 2, \ldots, T$, let $\alpha_t \in A$ satisfy $h^t = (h_{t-1}, \alpha_t)$. Then define

$$\zeta^\sigma(h'|h) \quad = \quad \prod_{t=1}^{T} \sigma_{\iota(h^{t-1})}(\alpha_t|h^{t-1}),$$

and set $\zeta^\sigma(h'|h) = 0$ if $h'$ does not follow $h$. This can be extended to a probability distribution on histories $\overline{H}$ in the obvious way: for every $h'$ following $h$, the probability of the cylinder set with initial segment $h'$, i.e., $\{h'' \in \overline{H} \mid h' \ll h''\}$, is just $\zeta^\sigma(h'|h)$. As is standard, this probability measure has a unique extension from the ring of such cylinder sets to the $\sigma$-algebra generated by them and is again denoted $\zeta^\sigma(\cdot|h)$.[7]

This allows us to calculate agent $i$'s expected payoff following any history $h \in H$ as

$$U_i^\sigma(h|\delta) \quad = \quad \int_{\overline{H}} W_i(h'|\delta)\, \zeta^\sigma(dh'|h).$$

More transparently, we can write

$$U_i^\sigma(h|\delta) \quad = \quad \sum_{h' \in H^\bullet} \zeta^\sigma(h'|h)[(1 - \delta_i^{\tau(h')-1})\overline{u}_i + \delta_i^{\tau(h')-1} u_i(\chi(h'))]$$

[7] When considering the probability of a singleton, say $\{h'\}$, we just write the argument of $\zeta^\sigma(\cdot|h)$ as $h'$.

$$+(1-\zeta^\sigma(H^\bullet|h))\overline{u}_i,$$

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. Here, of course, $1-\zeta^\sigma(H^\bullet|h)$ is the probability of infinite delay.

## *6. Continuation lotteries*

It is instructive to rewrite agent $i$'s expected payoff starting from $h$ as

$$\begin{aligned} U_i^\sigma(h|\delta) &= \sum_{h'\in H^\bullet} \zeta^\sigma(h'|h)(1-\delta_i)\sum_{t=1}^{\tau(h')-1}\delta_i^{t-1}u_i(q) && (1)\\ &\quad+\sum_{h'\in H^\bullet}\zeta^\sigma(h'|h)\delta_i^{\tau(h')-1}u_i(\chi(h')) && (2)\\ &\quad+(1-\zeta^\sigma(H^\bullet|h))(1-\delta_i)\sum_{t=1}^{\infty}\delta_i^{t-1}u_i(q). && (3)\end{aligned}$$

Since the term $\delta_i^{t-1}u_i(q)$ in (1) appears once for every terminal history following $h$ of length $t+1$ periods or more, we can rewrite (1) as

$$(1-\delta_i)\sum_{t=1}^{\infty}\sum_{h'\in H_t^c\setminus H^\bullet}\zeta^\sigma(\{\hat{h}\in H^\bullet \mid h'\ll\hat{h}\}|h)\delta_i^{t-1}u_i(q),$$

and we can rewrite (3) as

$$(1-\delta_i)\sum_{t=1}^{\infty}\sum_{h'\in H_t^c\setminus H^\bullet}\zeta^\sigma(\{\hat{h}\in H^\infty|h'\ll\hat{h}\}|h)\delta_i^{t-1}u_i(q).$$

Thus, after combining (1) and (3), we arrive at

$$\begin{aligned} U_i^\sigma(h|\delta) &= (1-\delta_i^{\tau(h)})u_i(q)+(1-\delta_i)\sum_{h'\in H^c\setminus H^\bullet}\zeta^\sigma(h'|h)\delta_i^{\tau(h')-1}u_i(q)\\ &\quad+\sum_{x\in X}\sum_{h'\in H^\bullet(x)}\zeta^\sigma(h'|h)\delta_i^{\tau(h')-1}u_i(x),\end{aligned}$$

where we use $\zeta^\sigma(\{\hat{h}\in\overline{H}\mid h'\ll\hat{h}\}|h)=\zeta^\sigma(h'|h)$ for $h'$ following $h$.

Finally, it will be useful to write this expression in terms of an expectation with respect to the *continuation lottery for $i$ at $h$ given $\sigma$*, denoted

$\lambda_i^\sigma(h|\delta)$. This is the discrete probability distribution on $X \cup \{q\}$ defined as follows: for all $x \in X$,

$$\lambda_i^\sigma(h|\delta)(x) \;\; = \;\; \frac{1}{\delta_i^{\tau(h)}} \sum_{h' \in H^\bullet(x)} \zeta^\sigma(h'|h)\delta_i^{\tau(h')-1},$$

and

$$\lambda_i^\sigma(h|\delta)(q) \;\; = \;\; \frac{1-\delta_i}{\delta_i^{\tau(h)}} \sum_{h' \in H^c \setminus H^\bullet} \zeta^\sigma(h'|h)\delta_i^{\tau(h')-1}.$$

For all $i \in N$, define the mapping $V_i \colon \Lambda \to \mathbb{R}$ by

$$V_i(\lambda) \;\; = \;\; \int_{X \cup \{q\}} u_i(z)\,\lambda(dz),$$

which gives $i$'s expected utility from the lottery $\lambda$ on $X \cup \{q\}$. We therefore have

$$U_i^\sigma(h|\delta) \;\; = \;\; (1-\delta_i^{\tau(h)})u_i(q) + \delta_i^{\tau(h)} V_i(\lambda_i^\sigma(h|\delta)),$$

a positive affine transformation of $i$'s expected utility from the continuation lottery at $h$ given $\sigma$. Note that, insofar as the discount factors of the agents may differ, the continuation lottery $\lambda_i^\sigma(h|\delta)$ may vary with $i$.

## *7. Subgame perfect equilibrium*

As is standard, we define a strategy profile $\sigma$ as a *subgame perfect equilibrium* if, for every agent $i \in N$, every strategy $\sigma_i'$ for $i$, and every history $h \in H^{pi} \cup H^{vi}$ at which $i$ is the active player, deviating to $\sigma_i'$ does not increase $i$'s expected payoff:

$$U_i^\sigma(h|\delta) \;\; \geq \;\; U_i^{(\sigma_i',\sigma_{-i})}(h|\delta),$$

where $\sigma_{-i} = (\sigma_1, \ldots, \sigma_{i-1}, \sigma_{i+1}, \ldots, \sigma_n)$ is the strategy profile $\sigma$ less $i$'s strategy. In other words, after every history for $i$, the player weakly prefers the continuation lottery given $\sigma$ to the alternative lottery given $(\sigma_i', \sigma_{-i})$:

$$V_i(\lambda_i^\sigma(h|\delta)) \;\; \geq \;\; V_i(\lambda_i^{(\sigma_i',\sigma_{-i})}(h|\delta)).$$

Let $\Sigma(\delta)$ denote the set of subgame perfect equilibrium strategy profiles. Define

$$X^\sigma(h|\delta) \;\; = \;\; \{x \in X \mid \text{ there exists } h' \in H^\bullet(x) \text{ such that } \zeta^\sigma(h'|h) > 0\}$$

to be the set consisting of all proposals that pass following $h$ in equilibrium $\sigma$, and define

$$X^\sigma(\delta) = \bigcup_{h \in H^c} X^\sigma(h|\delta)$$

to be the set of proposals that pass following any history in equilibrium $\sigma$. Then define

$$X(\delta) = \bigcup_{\sigma \in \Sigma(\delta)} X^\sigma(\delta),$$

the set of all proposals that pass following any history in any subgame perfect equilibrium. Finally, define the set

$$V(\delta) = \{(V_1(\lambda_1^\sigma(h|\delta)), \ldots, V_n(\lambda_1^\sigma(h|\delta))) \mid \sigma \in \Sigma(\delta), h \in H^c \setminus H^\bullet\}$$

of payoff vectors that may arise after any history from any subgame perfect equilibrium.

## *8. Stationary subgame perfect equilibrium*

A subgame perfect equilibrium $\sigma$ is *stationary* if each agent's proposal strategy is history-independent and each agent's voting strategy depends only on the current proposal: for all $i, j, j' \in N$, all $h, h' \in H^c$, all $x \in X$, all $s_1, s_1' \in S_1$, all $s_2, s_2' \in S_2$, all $C, C' \subseteq N \setminus \{i\}$, and $v^C \in \{a, r\}^C$, and all $v^{C'} \in \{a, r\}^{C'}$, it must be that

$$p_i(h) = p_i(h') \quad \text{and} \quad v_i(h, j, s_1, x, s_2, v^C) = v_i(h', j', s_1', x, s_2', v^{C'}).$$

Such equilibria are relatively easy to play and this may confer a focal effect, lending support for the stationarity refinement.

Allowing for a multidimensional space of alternatives and a general voting rule, Banks and Duggan (2000) prove existence of stationary equilibria when recognition probabilities are history-independent, i.e., for all $h, h' \in H^c$ and all $i \in N$, $\rho(i|h) = \rho(i|h')$, under a strengthening of (A2). Banks and Duggan (2003) replace the assumption of a bad status quo with (A1) and assume a common discount factor, and they again prove existence of stationary equilibrium with history-independent recognition probabilities. In the one-dimensional model, under either set of assumptions, they show that there are no stationary equilibria in (non-degenerate) mixed strategies. Cho and Duggan (2002) demonstrate that there may be multiple (non-payoff equivalent) stationary equilibria in one dimension. Moreover, such equilibria must be nested in the sense that the set of alternatives that would pass if proposed in one equilibrium must be contained in the set that would pass in the other. Banks and Duggan (2000, 2003) show that all stationary equilibrium proposals converge to the core point in one dimension, providing a game-theoretic foundation for the median voter theorem in terms of stationary equilibria.

Banks and Duggan (2000) show that delay cannot occur in stationary equilibria under (A2), unless some agents are perfectly patient, i.e., $\delta_i = 1$ for some $i \in N$. Banks and Duggan (2003) prove that, under (A1), delay can occur only if the status quo alternative $x^q$ is in the core, i.e., $\tilde{x}_k = x^q$ here. In this case, all agents must propose the core alternative, which may or not pass (the agent's payoffs are unaffected) and the equilibrium is payoff-equivalent to the unique no-delay equilibrium in which $\tilde{x}_k$ passes. Thus, delay, if it occurs in a stationary equilibrium, is inconsequential.

# 3 The Main Result

Our main result provides a strong game-theoretic foundation for the median voter theorem. First, it shows that the proposals that may pass in any subgame in any subgame perfect equilibrium converge to the unique core point as the agents become patient. The result holds under (A1) or (A2) and therefore generalizes the convergence results of Banks and Duggan (2000, 2003) by dropping the stationarity refinement. Moreover, we consider a sequence $\{\delta^m\}$ of vectors of discount factors satisfying the following *convergence condition:* there exists a sequence $\{c^m\}$ in $\mathbb{R}$ such that $c^m \downarrow 1$ and

$$\left(\max_{i\in N} \delta_i^m\right)^{c^m} \leq \min_{i\in N} \delta_i^m$$

for all $m$. Equivalently, we require that the ratio of logged discount factors converge to one for all agents, i.e., $\ln(\delta_i^m)/\ln(\delta_j^m) \to 1$ for all $i, j \in N$. Thus, we show that the convergence result of Banks and Duggan (2003) is robust with respect to the assumption of a common discount factor.

Second, it shows that the payoffs in any subgame in any subgame perfect equilibrium converge to the utility of the core point. An implication is that, unless (A1) holds and $x^q = \tilde{x}_k$, the probability of delay, measured by the agents' continuation lotteries, must go to zero, i.e.,

$$\sup_{\sigma \in \Sigma(\delta^m), h \in H^c, i \in N} \lambda_i^\sigma(h|\delta^m)(q) \to 0.$$

In any case, the effects of delay are insignificant when the agents are very patient, extending the results of Banks and Duggan (2002, 2003) on the possibility of delay in stationary equilibrium.

In the formal statement of the theorem, when given a sequence $\{Y^m\}$ of subsets of some Euclidean space and a point $x$, we write $Y^m \to x$ if $\sup_{y \in Y^m} ||y - x|| \to 0$.

**Theorem 1** *Assume either (A1) or (A2). Let $\{\delta^m\}$ be a sequence of vectors of discount factors satisfying the convergence condition and such that $X(\delta^m) \neq \emptyset$ for sufficiently large $m$ and $\delta_i^m \to 1$ for all $i \in N$. Then*

*(i)* $X(\delta^m) \to \tilde{x}_k$,

*(ii)* $V(\delta^m) \to u(\tilde{x}_k)$.

In the remainder of this section, we give an overview of the logic of the formal proof of the theorem, given in the next section. The proof proceeds by supposing that $X(\delta^m)$ does not converge to the core point and then deriving necessary conditions that must be satisfied by the infima and suprema of these sets. Ultimately, we show that a contradiction arises, establishing the first part of the theorem. Let $\underline{x}^m$ and $\overline{x}^m$ denote the infimum and supremum of $X(\delta^m)$, and for convenience here assume that these bounds are achieved within the set. Thus, for each $m$, there is some subgame perfect equilibrium and some history after which $\underline{x}^m$ is proposed and passed, and likewise for $\overline{x}^m$. Passing to a subsequence if necessary, we may assume that these bounds converge to $\underline{x}$ and $\overline{x}$, respectively, and in this discussion we focus on the typical case of interest, $\underline{x} < \tilde{x}_k < \overline{x}$.

Because the collection $\mathcal{D}$ of decisive coalitions is proper, it is easy to show that, given any $m$, either the set of agents who weakly prefer $\underline{x}^m$ to $\overline{x}^m$ is decisive or the set of agents with the opposite weak preference is decisive. Passing to a subsequence again, we suppose the former holds for all $m$. In the case we consider, these bounds lie on either side of the core point for large enough $m$, and then strict concavity implies that all agents with the weak preference $u_i(\underline{x}^m) \geq u_i(\overline{x}^m)$ must strictly prefer $\tilde{x}_k$ to the bound $\overline{x}^m$. In fact, because $u_i(\tilde{x}_k) > u_i(\overline{x})$, this strict preference is preserved in the limit as well.

The rest of the analysis must confront the complexity of voting subgames, which, in contrast to work that focuses on stationary equilibria, can no longer be treated as simple binary voting games. In binary sequential voting games of perfect information, it is well-known that the subgame perfect equilibrium outcome must be the majority-preferred of two alternatives.[8] An implication is that, if the core point $\tilde{x}_k$ is one of the alternatives being voted on, then it is the unique subgame perfect equilibrium outcome. It is straightforward to construct binary voting examples in which, along the equilibrium path of play, some agents vote for the winning alternative, despite the fact that it is worse than the other (because changing their vote does not change the outcome). Another implication of the above, however, is that, in equilibrium, every decisive coalition must contain at least one agent who weakly prefers the winning alternative to the remaining one.

In voting subgames of our model, the outcome following acceptance by members of a decisive coalition is fixed: it is just the currently proposed alternative. In contrast, the outcome following rejection is not fixed, as continuation equilibria can conceivably depend on the votes of particular agents. Thus, there are potentially many continuation lotteries following the rejection of any given proposal. While considerably more complex than a binary voting game, we establish, in Lemma 3, that the second implication above extends to our model: in equilibrium, if a proposal passes after some history, then every decisive coalition must contain at least one agent who weakly prefers the proposed alternative (the winner) to *some* continuation lottery following rejection. Returning to the argument above, there must be some agent $i^m$ with a weak preference for $\underline{x}^m$ over $\overline{x}^m$ who weakly prefers $\overline{x}^m$ to some continuation lottery. Passing to a subsequence, we may select an agent $i$ such that $i^m = i$ for all $m$.

Since utility functions are strictly concave, agent $i$'s worst alternative in the interval $[\underline{x}^m, \overline{x}^m]$ is at an endpoint. Since $u_i(\underline{x}^m) \geq u_i(\overline{x}^m)$ by construction, it follows that $\overline{x}^m$ is the worst alternative for $i$ that can possibly pass in any subgame. Thus, by the above, agent $i$ must weakly prefer the worst possible alternative that might pass to some continuation lottery. We arrive at a contradiction if we can show that, when the core point is proposed and

[8] Moreover, iterative elimination of weakly dominated strategies in the strategic form of the voting game produces the same outcomes.

the order of voting alternates on either side of the core (i.e., it lies in $\Phi$), the proposal necessarily passes in equilibrium. By assumption, the probability that agent $i$ is selected and a voting order in $\Phi$ is realized after $\tilde{x}_k$ is proposed is bounded strictly above zero across all non-terminal histories. This means that, when $\overline{x}^m$ is rejected, agent $i$ is guaranteed a payoff of at least $u_i(\tilde{x}_k)$ with a positive probability (that does not go to zero) in every subsequent non-terminal history. We prove in the next section that, since $u_i(\tilde{x}_k) - u_i(\overline{x}^m)$ is positive for all $m$ (and does not go to zero), this produces our contradiction: agent $i$ can guarantee a payoff greater than $u_i(\overline{x}^m)$ for large enough $m$.

To explain why, in equilibrium, the core point will pass if proposed and the voting order lies in $\Phi$, we first illustrate how the usual logic for binary voting games fails unless the order of voting is restricted. For simplicity, we assume a common discount factor for this discussion. Suppose there are five agents, $j = 1, 2, 3, 4, 5$, and that their ideal points are increasing in $j$. Thus, $k = 3$ and the core point is $\tilde{x}_3$. Suppose that the core alternative has been proposed, and that the order of voting is 3, 1, 2, 4, and 5. Assuming agent 3 has voted to accept the core point, the voting subgame takes the form depicted in Figure 1, where we truncate the game form once acceptance or rejection is determined.

Here, $\lambda, \lambda', \lambda''$ are continuation lotteries following rejection of the core point. As we have discussed, these lotteries may be distinct, and we cannot rule out *a priori* the following preferences for agents other than 3, where we only depict preferences needed for this example.

| 1 | 2 | 4 | 5 |
|---|---|---|---|
| $\lambda'$ | $\lambda'$ | $\lambda$ | $\lambda''$ |
| $\tilde{x}_3$ | $\tilde{x}_3$ | $\tilde{x}_3$ | $\lambda$ |
| | $\lambda$ | $\lambda'$ | $\tilde{x}_k$ |
| | | $\lambda''$ | |

Then, as indicated in Figure 1, the unique equilibrium path of play is for agents 1, 2, and 4 to vote to reject the core point in favor of the continuation lottery $\lambda'$. Of course, agents 1 and 2 both prefer $\lambda'$ to the core point, but agent 4 strictly prefers $\tilde{x}_3$ to the outcome of voting; nevertheless, that agent votes to accept the core point because a vote to reject allows agent 5 to obtain $\lambda''$, which is even worse for agent 4. Obviously, this "preference reversal" cannot occur in a binary vote, where $\lambda'$ and $\lambda''$ are necessarily equal. This suggests the possibility of subgame perfect equilibria in which the core point is rejected, even after the first agent votes to accept it. In the

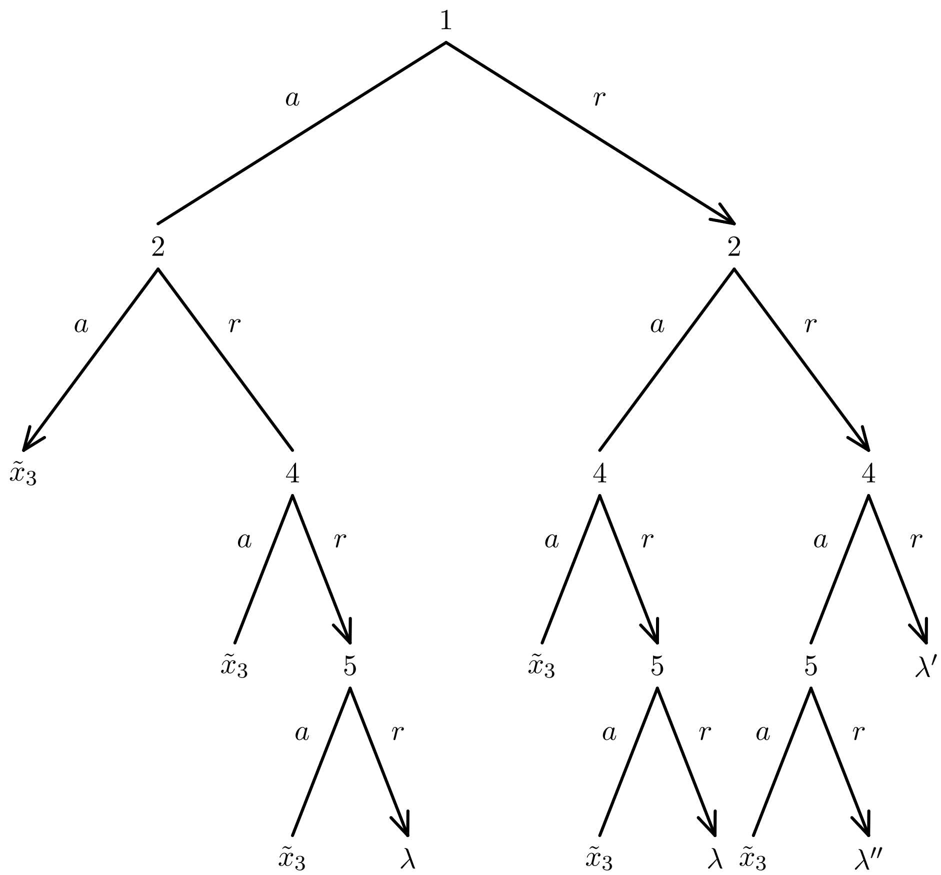


Figure 6: Rejecting the Core Point

corresponding game where agent 3 votes to reject, which we do not take up here, it is even more straightforward to assign continuation lotteries in such a way that the core point is rejected.

The core point necessarily passes, in contrast, when the order of voting alternates, suppose the order of voting is 3, 1, 4, 2, and 5. Consider the voting decision for agent 2 after agent 3 and either 1 or 4 have voted to accept. By voting to accept $\tilde{x}_3$, agent 2 obtains the core point as the outcome, and therefore the agent votes to reject in equilibrium only if doing so results in an outcome at least as good as the core point. In that case, agent 5 may vote to accept, in which case the core point remains the outcome (so agent 2's vote was irrelevant); agent 5 will vote to reject in equilibrium only if doing so results in a continuation lottery at least as good as the core point. But then agents 2 and 5 must each weakly prefer that continuation lottery to the core point. Since $u_2$ and $u_5$ are strictly concave and the agents' ideal points are on opposite sides of $\tilde{x}_3$, we conclude that the continuation lottery must in fact be the point mass on the core point. Thus, the equilibrium outcome starting from agent 2's vote must indeed be the core point. Moving to agent 1's vote, a similar argument applies. We conclude that if agent 3 votes to accept initially, then the core point will obtain, and since this is the agent's ideal point, the proposal of $\tilde{x}_3$ must pass.

The argument is somewhat more involved when we allow for heterogeneous discount factors, for then two agents need not "see" the same continuation lottery. Under the convergence condition, however, Lemma 1 establishes that the continuation lotteries of two agents starting from any history must become close to each other as the agents become arbitrarily patient. This allows us to establish, in Lemma 2, that the continuation payoffs of the agents when the core point is proposed converge to the utility of the core point.

The proof of the second part of the theorem hinges on showing that agent $k$'s continuation payoff converges to $u_k(\tilde{x}_k)$ in every subgame in every subgame perfect equilibrium, which means that the corresponding continuation lotteries must converge to the point mass on $\tilde{x}_k$. And then Lemma 1 implies that the continuation payoffs of all agents must converge to the utility of the core point.

# 4 The Proof

Let $\{\delta^m\}$ be as in the statement of Theorem 1. Before proceeding, we present some preliminary technical results. The first lemma gives a connection between the continuation lotteries of the agents when discount factors converge to one at close to the same rate, as stipulated in the convergence condition: we show that, if one agent's continuation lottery approaches some lottery, then so must the continuation lotteries of all agents. The result is uniform across all subgame perfect equilibria and all complete non-terminal histories.

**Lemma 1** *Let $\{(\sigma^m, h^m)\}$ be an arbitrary sequence such that $\sigma^m \in \Sigma(\delta^m)$ and $h^m \in H^c \setminus H^\bullet$ for all $m$. For all $i, j \in N$, if $\lambda_i^{\sigma^m}(h^m|\delta^m) \to \lambda$ in the weak* topology, then $\lambda_j^{\sigma^m}(h^m|\delta^m) \to \lambda$ in the weak* topology.*

We use this result to establish that, when discount factors are close enough to one, the agents' continuation payoffs approach their utility from the core point whenever that alternative is proposed: essentially, by proposing the core point, any agent can ensure that the core will pass or, at least, that a lottery close to the pointmass on the core will result.

**Lemma 2** *For all $i \in N$, all $s_1 \in S_1$, and all $s_2 \in S_2$ with $\phi(s_1, s_2) \in \Phi$, we have*

$$\sup_{\sigma \in \Sigma(\delta^m), h \in H^c \setminus H^\bullet} |V_i(\lambda_i^\sigma(h, i, s_1, \tilde{x}_k, s_2|\delta^m)) - u_i(\tilde{x}_k)| \to 0.$$

Let $\underline{x}^m = \inf X(\delta^m)$ and $\overline{x}^m = \sup X(\delta^m)$, which are finite since $X$ is compact. To prove (i), suppose $X(\delta^m) \not\to \tilde{x}_k$, so that either $\underline{x}^m \not\to \tilde{x}_k$ or $\overline{x}^m \not\to \tilde{x}_k$. Since $\mathcal{D}$ is proper and strong, it follows that, for each $m$, either $\{i \in N \mid u_i(\underline{x}^m) \geq u_i(\overline{x}^m)\} \in \mathcal{D}$ or $\{i \in N \mid u_i(\underline{x}^m) \leq u_i(\overline{x}^m)\} \in \mathcal{D}$: otherwise, $\{i \in N \mid u_i(\underline{x}^m) < u_i(\overline{x}^m)\}$ and $\{i \in N \mid u_i(\underline{x}^m) > u_i(\overline{x}^m)\}$ would be disjoint decisive coalitions. Without loss of generality, we take a subsequence of $\{\delta^m\}$, still indexed by $m$, such that

$$\{i \in N \mid u_i(\underline{x}^m) \geq u_i(\overline{x}^m)\} \in \mathcal{D} \tag{4}$$

for all $m$. Since $X$ is compact, we can go to a subsequence, still indexed by $m$, such that $\overline{x}^m \to \overline{x}$ for some $\overline{x} \in X$. We claim $\overline{x} \neq \tilde{x}_k$. To see this, suppose $\overline{x} = \tilde{x}_k$. Then, since $X(\delta^m)$ does not converge to $\tilde{x}_k$, we must have $\underline{x}^m \nrightarrow \tilde{x}$. Since $\underline{x}^m \leq \overline{x}^m$, there must then exist $\epsilon > 0$ such that $\underline{x}^m < \tilde{x}_k - \epsilon$ for infinitely many $m$. For such $m$, since each $u_i$ is strictly concave, we have $u_i(\underline{x}^m) < u_i(\tilde{x}_k - \epsilon)$ for all $i \in C_K \cup C_R$. By continuity, we have $u_i(\underline{x}^m) < u_i(\overline{x}^m)$ for infinitely many $m$, but then $\{i \in N \mid u_i(\underline{x}^m) \geq u_i(\overline{x}^m)\} \subseteq C_L$ for such $m$, contradicting (4).

We claim that, for large enough $m$, there exists $C^m \in \mathcal{D}$ such that $u_i(\underline{x}^m) \geq u_i(\overline{x}^m)$ and $u_i(\tilde{x}_k) > u_i(\overline{x}^m)$ for all $i \in C^m$. To see this, first suppose $\overline{x} < \tilde{x}_k$, so $\overline{x}^m < \tilde{x}_k$ for large enough $m$. Then from (4) we have $\overline{x}^m = \underline{x}^m$ for large enough $m$: otherwise, by strict concavity, $\{i \in N \mid u_i(\underline{x}^m) \geq u_i(\overline{x}^m)\} \subseteq C_L$, contradicting (4). Thus, setting $C^m = C_K \cup C_R \in \mathcal{D}$, fulfills the claim. Second, suppose $\overline{x} > \tilde{x}_k$. Then, for large enough $m$, $\overline{x}^m > \tilde{x}_k$. If $\tilde{x}_k \leq \underline{x}^m$, then setting $C^m = C_K \cup C_L \in D$ fulfills the claim. If $\underline{x}^m < \tilde{x}_k$, then, by strict concavity, $u_i(\underline{x}^m) \geq u_i(\overline{x}^m)$ implies $u_i(\tilde{x}_k) > u_i(\overline{x}^m)$. Setting $C^m = \{i \in N \mid u_i(\underline{x}^m) \geq u_i(\overline{x}^m)\}$ fulfills the claim.

The next lemma gives a necessary condition for a proposed alternative to pass: every decisive coalition must contain at least one agent who weakly prefers that alternative to some continuation lottery.

**Lemma 3** *For any $m$, let $\sigma^m \in \Sigma(\delta^m)$. If $x \in X^{\sigma^m}(\delta^m)$, then for all $C \in \mathcal{D}$, there exist $i^m \in C$ and $h^m \in H^c \backslash H^\bullet$ such that $u_{i^m}(x) \geq V_{i^m}(\lambda_{i^m}^{\sigma^m}(h^m|\delta^m))$.*

Since $N$ is finite, we may take a subsequence, still indexed by $m$, for which there exists $C \subseteq N$ such that, for all $m$, $C = C^m$. Since $\overline{x}^m = \sup X(\delta^m)$, we can construct a sequence $x^m$ in $X$ so that $x^m \in [\overline{x}^m - \frac{1}{m}, \overline{x}^m] \cap X(\delta^m)$. For each $m$, there exists $\sigma^m \in \Sigma(\delta^m)$ such that $x^m \in X^{\sigma^m}(\delta^m)$. Since $C \in \mathcal{D}$, Lemma 3 yields $i^m \in C$ and $h^m \in H^c \setminus H^\bullet$ such that $u_{i^m}(x^m) \geq V_{i^m}(\lambda_{i^m}^{\sigma^m}(h^m|\delta^m))$ for all $m$. Again since $N$ is finite, we may take a subsequence, still indexed by $m$, for which there exists $i \in N$ such that, for all $m$, $i = i^m$. Let $\lambda_i^m = \lambda_i^{\sigma^m}(h^m|\delta^m)$ for each $m$, so we have

$$u_i(x^m) \quad \geq \quad V_i(\lambda_i^{\sigma^m}(h^m|\delta^m))$$

for all $m$. Since $X \cup \{q\}$ is compact, $\{\lambda_i^m\}$ has a weak* convergent subsequence, still indexed by $m$, with limit, say, $\lambda$. Taking limits, we have

$$u_i(\overline{x}) \quad \geq \quad V_i(\lambda), \tag{5}$$

by continuity of $u_i$.

For all $j \in N$, let

$$I_j^m = \inf_{\sigma \in \Sigma(\delta^m), h \in H^c} V_j(\lambda_j^\sigma(h|\delta^m))$$

denote the smallest possible equilibrium continuation payoff for agent $j$ after any complete history. Given any strategy profile $\sigma$ and any complete history $h$, let $\hat{\sigma}_j^{\sigma,h}$ denote the strategy for agent $j$ that is identical to $\sigma_j$ with the proviso that $j$ proposes $\tilde{x}_k$ if selected to make a proposal immediately following $h$. Let

$$\begin{aligned}
\hat{I}_j^m &= \inf_{\sigma \in \Sigma(\delta^m), h \in H^c} V_j(\lambda_j^{(\hat{\sigma}_j^{\sigma,h}, \sigma_{-j})}(h|\delta^m)) \\
&= \inf_{\sigma \in \Sigma(\delta^m), h \in H^c} \sum_{s_1 \in S_1, s_2 \in S_2} [\rho(j, s_1|h)\pi(s_2|h, j, s_1, \tilde{x}_k) \\
&\qquad \cdot V_j(\lambda_j^\sigma(h, j, s_1, \tilde{x}_k, s_2|\delta^m))] \\
&\qquad + \sum_{\ell \in N \setminus \{j\}, s_1 \in S_1, s_2 \in S_2} [\rho(\ell, s_1|h)\pi(s_2|h, \ell, s_1, p_\ell(h, \ell, s_1)) \\
&\qquad \cdot V_j(\lambda_j^\sigma(h, \ell, s_1, p_\ell(h, \ell, s_1), s_2|\delta^m))]
\end{aligned}$$

be the smallest possible continuation payoff for agent $j$ when other agents use equilibrium strategies and $j$ proposes the core point. We have substituted $(\hat{\sigma}_j^{\sigma,h}, \sigma_{-j})$ for $\sigma$ in the last expression because the two strategies are identical after a proposal is made. Also let

$$J_j^m = \inf_{\sigma \in \Sigma(\delta^m), h \in H^c, \ell \in N, s_1 \in S_1, s_2 \in S_2} V_j(\lambda_j^\sigma(h, \ell, s_1, p_\ell(h, \ell, s_1), s_2|\delta^m))$$

denote the smallest possible equilibrium continuation payoff for agent $j$ after any history in which some agent is selected to propose.

Since $i \in C$, we have $u_i(\underline{x}^m) \geq u_i(\overline{x}^m)$ and $u_i(\tilde{x}_k) > u_i(\overline{x}^m)$ for all $m$. Since $u_i$ is concave, it follows that

$$\min\{u_i(\underline{x}^m), u_i(\overline{x}^m)\} = \min\{u_i(x) \mid x \in [\underline{x}^m, \overline{x}^m]\},$$

and therefore that, for all $x \in X(\delta^m)$, $u_i(x) \geq u_i(\overline{x}^m)$. Then we have

$$J_i^m \geq \min\{u_i(\overline{x}^m), (1 - \delta_i^m)u_i(q) + \delta_i^m I_i^m\}.$$

Suppose that $(1 - \delta_i^m)u_i(q) + \delta_i^m I_i^m \geq u_i(\overline{x}^m)$ for infinitely many $m$. Going to this subsequence, still indexed by $m$, we have

$$V_i(\lambda_i^m) \geq I_i^m \geq \hat{I}_i^m$$

$$\begin{aligned}
&\geq \inf_{\sigma\in\Sigma(\delta^m),h\in H^c} \sum_{(s_1,s_2)\in\phi^{-1}(\Phi)} [\rho(i,s_1|h)\pi(s_2|h,i,s_1,\tilde{x}_k) \\
&\qquad\qquad \cdot V_i(\lambda_i^\sigma(h,i,s_1,\tilde{x}_k,s_2|\delta^m))] \\
&\qquad + \sum_{(s_1,s_2)\notin\phi^{-1}(\Phi)} \rho(i,s_1|h)\pi(s_2|h,i,s_1,\tilde{x}_k)J_i^m \\
&\qquad + \rho(N\setminus\{i\}|h)J_i^m \\
&\geq \inf_{\sigma\in\Sigma(\delta^m),h\in H^c} \sum_{(s_1,s_2)\in\phi^{-1}(\Phi)} [\rho(i,s_1|h)\pi(s_2|h,i,s_1,\tilde{x}_k) \\
&\qquad\qquad \cdot V_i(\lambda_i^\sigma(h,i,s_1,\tilde{x}_k,s_2|\delta^m))] \\
&\qquad + \sum_{(s_1,s_2)\notin\phi^{-1}(\Phi)} \rho(i,s_1|h)\pi(s_2|h,i,s_1,\tilde{x}_k)u_i(\overline{x}^m) \\
&\qquad + \rho(N\setminus\{i\}|h)u_i(\overline{x}^m)
\end{aligned}$$

for all $m$, where the second inequality is implied by subgame perfection. Taking limits and applying Lemma 2, this implies

$$V_i(\lambda) \quad \geq \quad \mu u_i(\tilde{x}_k) + (1-\mu)u_i(\overline{x}),$$

where, by assumption, $\mu > 0$. By construction, however, $u_i(\tilde{x}_k) > u_i(\overline{x})$, which then yields $V_i(\lambda) > u_i(\overline{x})$, contradicting (5).

Now suppose that $u_i(\overline{x}^m) \geq (1-\delta_i^m)u_i(q) + \delta_i^m I_i^m$ for infinitely many $m$. Going to such a subsequence, still indexed by $m$, we have

$$\begin{aligned}
I_i^m &\geq \hat{I}_i^m \\
&\geq \inf_{\sigma\in\Sigma(\delta^m),h\in H^c} \sum_{(s_1,s_2)\in\phi^{-1}(\Phi)} [\rho(i,s_1|h)\pi(s_2|h,i,s_1,\tilde{x}_k) \\
&\qquad\qquad \cdot V_i(\lambda_i^\sigma(h,i,s_1,\tilde{x}_k,s_2|\delta^m))] \\
&\qquad + \sum_{(s_1,s_2)\notin\phi^{-1}(\Phi)} \rho(i,s_1|h)\pi(s_2|h,i,s_1,\tilde{x}_k)J_i^m \\
&\qquad + \rho(N\setminus\{i\}|h)J_i^m \\
&\geq \inf_{\sigma\in\Sigma(\delta^m),h\in H^c} \sum_{(s_1,s_2)\in\phi^{-1}(\Phi)} [\rho(i,s_1|h)\pi(s_2|h,i,s_1,\tilde{x}_k) \\
&\qquad\qquad \cdot V_i(\lambda_i^\sigma(h,i,s_1,\tilde{x}_k,s_2|\delta^m))] \\
&\qquad + \sum_{(s_1,s_2)\notin\phi^{-1}(\Phi)} \rho(i,s_1|h)\pi(s_2|h,i,s_1,\tilde{x}_k) \\
&\qquad\qquad \cdot [(1-\delta_i^m)u_i(q) + \delta_i^m I_i^m] \\
&\qquad + \rho(N\setminus\{i\}|h)[(1-\delta_i^m)u_i(q) + \delta_i^m I_i^m].
\end{aligned}$$

Taking limits and applying Lemma 2, this implies

$$\liminf I_i^m \quad \geq \quad \mu u_i(\tilde{x}_k) + (1-\mu) \liminf I_i^m.$$

Using $\mu > 0$, we conclude that

$$V_i(\lambda) \ \geq \ \liminf I_i^m \ \geq \ u_i(\tilde{x}_k) \ > \ u_i(\overline{x}),$$

which again contradicts (5). This contradiction completes the proof of (i).

We have shown that $\underline{x}^m \to \tilde{x}_k$ and $\overline{x}^m \to \tilde{x}_k$. To prove (ii), let $\{(\sigma^m, h^m)\}$ be an arbitrary sequence such that $\sigma^m \in \Sigma(\delta^m)$ and $h^m \in H^c \setminus H^\bullet$ for all $m$. Let $\lambda^m = \lambda_k^{\sigma^m}(h^m|\delta^m)$ for all $m$. Since $X \cup \{q\}$ is compact, $\{\lambda^m\}$ has a weak* convergent subsequence, still indexed by $m$, with limit, say, $\lambda$. Let $\hat{x}^m \in \arg\min\{u_k(x) \mid x \in [\underline{x}^m, \overline{x}^m]\}$, and note that $\hat{x}^m \to \tilde{x}_k$. Using the notation from the proof of (i), we have

$$J_k^m \quad \geq \quad \min\{u_k(\hat{x}^m), (1-\delta_k^m)u_k(q) + \delta_k^m I_k^m\}.$$

As in the proof of (i), if $(1-\delta_k^m)u_k(q) + \delta_k^m I_k^m \geq u_k(\hat{x}^m)$ for infinitely many $m$, then we have

$$\lim V_k(\lambda^m) \quad \geq \quad \mu u_k(\tilde{x}_k) + (1-\mu) \lim u_k(\hat{x}^m),$$

implying $V_k(\lambda) \geq u_k(\tilde{x}_k)$. As in the proof of (i), if $u_k(\hat{x}^m) \geq (1-\delta_k^m)u_k(q) + \delta_k^m I_k^m$ for infinitely many $m$, then we have

$$V_k(\lambda) \ \geq \ \liminf I_k^m \ \geq \ u_k(\tilde{x}_k).$$

We conclude that $V_k(\lambda) \geq u_k(\tilde{x}_k)$.

Under (A1), define the probability measure $\lambda'$ on $X$ by transferring probability mass from $q$ to the alternative $x^q$, as follows: for every Borel measurable set $Y \subseteq \mathbb{R}$,

$$\lambda'(Y) \quad = \quad \begin{cases} \lambda(Y) + \lambda(q) & \text{if } x^q \in Y, \\ \lambda(Y) & \text{else.} \end{cases} \tag{6}$$

Note that $V_j(\lambda') = V_j(\lambda)$ for all $j \in N$. Thus, $V_k(\lambda') = u_k(\tilde{x}_k)$, and it follows that $\lambda'$ is the pointmass on $\tilde{x}_k$. Using Lemma 1, this implies

$$V_j(\lambda_j^{\sigma^m}(h^m|\delta^m)) \to V_j(\lambda) \ = \ V_j(\lambda') \ = \ u_j(\tilde{x}_k)$$

for all $j \in N$, as required. Under (A2), it follows that $\lambda(q) = 0$, and that $\lambda$ is the pointmass on $\tilde{x}_k$, with a similar conclusion. Since our argument applies to all subsequences of $\{\lambda^m\}$, we have shown that

$$\sup_{\sigma \in \Sigma(\delta^m), h \in H^c \setminus H^\bullet, j \in N} |V_j(\lambda_j^\sigma(h^m|\delta^m) - u_j(\tilde{x}_k)| \to 0,$$

which completes the proof.

# 5 Related Literature

Other examples of non-cooperative games with connections to the median voter theorem can be found in the existing literature. It is well-known that, in an election with two office-motivated candidates who can commit to their campaign platforms, the unique Nash equilibrium is for both candidates to locate at the median voter's ideal point.[9] In general environments, allowing for an arbitrary set of alternatives and arbitrary preferences, Bergin and Duggan (1999) propose a simple game form to implement the core in subgame perfect equilibrium. That game form involves simultaneous proposals by all agents, including a time at which a proposal should be voted on; the earliest proposal is voted on first, and if it is rejected, a default alternative obtains. Thus, while perhaps lending some support to the median voter theorem, their model does not fully capture the dynamics of many examples of committee decision-making. Moldovanu and Winter (1995) provide a "non-cooperative foundation" of the core of a general NTU game that is consistent with the one-dimensional environment. They consider a model in which voting is sequential and the first agent to reject a proposal becomes the next proposer: they show that the payoffs from "order independent" subgame perfect equilibria must lie in the core.[10]

The asymptotic uniqueness results we find are reminiscent of results for bargaining under unanimity rule. Rubinstein (1982) analyzes a two-agent model, where, in contrast to the models based on majority voting, the role of proposer alternates between the two agents. He proves a strong uniqueness result: regardless of the discount factors of the agents, there is a unique subgame perfect equilibrium, and delay does not occur in this equilibrium. Binmore (1987) analyzes the two-agent model in which the proposer is determined randomly in each period, as in Baron and Ferejohn (1989), and, in contrast to the results of that paper, he finds a general uniqueness result. Furthermore, as the agents become very patient in these two-agent models, the subgame perfect equilibrium payoffs converge to the Nash solutions. Merlo and Wilson (1995) assume random proposer selection and show that there is a unique stationary subgame perfect equilibrium regardless of the discount factor, even if there are multiple agents and the amount of private good varies stochastically over time. As shown by Shaked (See Sutton, 1986)

[9]This characterization can be extended to mixed strategies (Banks, Duggan, and Le Breton, 2002) and policy-motivated candidates (Duggan and Fey, 2003).

[10]A number of other papers also consider non-cooperative foundations of the core in TU games or in economic environments.

in a three-player example, however, general uniqueness does not extend to more than two agents with deterministic proposer selection.

Our results contrast with the standard intuition drawn from the folk theorem for repeated games, and they suggest an "anti-folk theorem" for an important class of bargaining games. As such, they are similar in spirit to the results of Lagunoff and Matsui (1997), who analyze a two-player pure coordination game of asynchronous timing. Those authors show that, for a sufficiently high common discount factor, the unique subgame perfect equilibrium outcome following any history is the payoff-maximizing strategy pair. While there are many technical differences between their model and ours, a feature common to both is that the set of utility imputations achievable after any history is of lower dimension. In their model, this is due to the assumption of pure coordination, so that the two players' payoff functions are identical; in our model, it follows from our assumption of strictly concave utilities defined over a one-dimensional space. Thus, the ability to construct punishments of deviating players is restricted, and the folk theorem of Dutta (1995) for stochastic games does not apply.[11]

# A Proofs of Lemmas

**Lemma 1** *Let $\{(\sigma^m, h^m)\}$ be an arbitrary sequence such that $\sigma^m \in \Sigma(\delta^m)$ and $h^m \in H^c \setminus H^\bullet$ for all $m$. For all $i, j \in N$, if $\lambda_i^{\sigma^m}(h^m|\delta^m) \to \lambda$ in the weak* topology, then $\lambda_j^{\sigma^m}(h^m|\delta^m) \to \lambda$ in the weak* topology.*

*Proof:* For each $m$, let $\overline{\delta}^m = \max\{\delta_j^m \mid j \in N\}$, and let $\underline{\delta}^m = \min\{\delta_j^m \mid j \in N\}$. Let $\alpha^m$ solve

$$\max_{\alpha \geq 0} (\overline{\delta}^m)^\alpha - (\underline{\delta}^m)^\alpha. \tag{7}$$

Note that $\alpha^m$ is well-defined, positive, and satisfies the first order condition

$$[\ln(\overline{\delta}^m)](\overline{\delta}^m)^{\alpha^m} \quad = \quad [\ln(\underline{\delta}^m)](\underline{\delta}^m)^{\alpha^m}.$$

[11]Several other of Dutta's (1995) assumptions are violated in our bargaining model: he analyzes a finite state and action stochastic game, whereas our model requires a continuum of states and actions; he assumes a common discount factor; because our game "ends" in some states, his asymptotic state independence conditions are violated; and he uses for joint randomization, something we do not allow.

Define $r^m$ by the equality $(\overline{\delta}^m)^{r^m} = \underline{\delta}^m$, and note that $1 \leq r^m \leq c^m$. Then the first order condition reduces to

$$\frac{1}{r^m} = ((\overline{\delta}^m)^{r^m-1})^{\alpha^m}.$$

Then

$$\begin{aligned}(\overline{\delta}^m)^{\alpha^m} - (\underline{\delta}^m)^{\alpha^m} &= (\overline{\delta}^m)^{\alpha^m} - (\overline{\delta}^m)^{r^m\alpha^m} \\ &= (\overline{\delta}^m)^{\alpha^m}(1 - (\overline{\delta}^m)^{r^m\alpha^m - \alpha^m}) \\ &= (\overline{\delta}^m)^{\alpha^m}(1 - \frac{1}{r^m}).\end{aligned}$$

By the convergence condition, $r^m \to 1$, which implies $1 - \frac{1}{r^m} \to 0$. Thus, the maximized value in (7) goes to zero as $m$ goes to infinity.

Take any continuous (and therefore bounded) function $f\colon X \cup \{q\} \to \mathbb{R}$, so that

$$\begin{aligned}&\int f(z)\, \lambda_i^{\sigma^m}(h^m|\delta^m)(dz) \\ &= \sum_{x\in X} f(x)\lambda_i^{\sigma^m}(h^m|\delta^m)(x) + f(q)\lambda_i^{\sigma^m}(h^m|\delta^m)(q) \\ &\to \int f(z)\, \lambda(dz).\end{aligned}$$

Letting $b \geq |f|$ denote a bound for $f$, note that

$$\begin{aligned}&\left|\int f(z)\, \lambda_i^{\sigma^m}(h^m|\delta^m)(dz) - \int f(z)\, \lambda_j^{\sigma^m}(h^m|\delta^m)(dz)\right| \\ &\leq \sum_{x\in X} |f(x)\lambda_i^{\sigma^m}(h^m|\delta^m)(x) - \lambda_j^{\sigma^m}(h^m|\delta^m)(x)| \\ &\quad + |f(q)\lambda_i^{\sigma^m}(h^m|\delta^m)(q) - \lambda_j^{\sigma^m}(h^m|\delta^m)(q)| \\ &\leq \sum_{x\in X} |f(x)| \sum_{h'\in H^\bullet(x)} \zeta^{\sigma^m}(h'|h) \left|(\delta_i^m)^{\alpha(h')} - (\delta_j^m)^{\alpha(h')}\right| \\ &\quad + |f(q)| \sum_{h'\in H^c\setminus H^\bullet} \zeta^{\sigma^m}(h'|h) \left|(1-\delta_i^m)(\delta_i^m)^{\alpha(h')} - (1-\delta_j^m)(\delta_j^m)^{\alpha(h')}\right| \\ &\leq \sum_{x\in X} |f(x)| \sum_{h'\in H^\bullet(x)} \zeta^{\sigma^m}(h'|h) \left|(\delta_i^m)^{\alpha(h')} - (\delta_j^m)^{\alpha(h')}\right| \\ &\quad + |f(q)| \sum_{h'\in H^c\setminus H^\bullet} \zeta^{\sigma^m}(h'|h) \Big[\left|(\delta_i^m)^{\alpha(h')} - (\delta_j^m)^{\alpha(h')}\right|\end{aligned}$$

$$
\begin{aligned}
& +\left|(\delta_i^m)^{\alpha(h')+1} - (\delta_j^m)^{\alpha(h')+1}\right|\Big] \\
\leq\ & b[(\overline{\delta}^m)^{\alpha^m} - (\underline{\delta}^m)^{\alpha^m}] \sum_{x\in X} \sum_{h'\in H^\bullet(x)} \zeta^{\sigma^m}(h'|h) \\
& +2b[(\overline{\delta}^m)^{\alpha^m} - (\underline{\delta}^m)^{\alpha^m}] \sum_{h'\in H^c\setminus H^\bullet} \zeta^{\sigma^m}(h'|h) \\
\leq\ & b[(\overline{\delta}^m)^{\alpha^m} - (\underline{\delta}^m)^{\alpha^m}][\zeta^{\sigma^m}(H^c) + \zeta^{\sigma^m}(H^c\setminus H^\bullet)],
\end{aligned}
$$

where $\alpha(h') = \tau(h') - \tau(h) - 1$. We have shown that this goes to zero, and we conclude that $\lambda_j^{\sigma^m}(h^m|\delta^m) \to \lambda$ in the weak* topology. ∎

**Lemma 2** *For all $i \in N$, all $s_1 \in S_1$, and all $s_2 \in S_2$ with $\phi(s_1, s_2) \in \Phi$, we have*

$$
\sup_{\sigma\in\Sigma(\delta^m),h\in H^c\setminus H^\bullet} |V_i(\lambda_i^\sigma(h, i, s_1, \tilde{x}_k, s_2|\delta^m)) - u_i(\tilde{x}_k)| \to 0.
$$

*Proof:* Take any sequence $\{(\sigma^m, h^m)\}$ such that $\sigma^m \in \Sigma(\delta^m)$ and $h^m \in H^c \setminus H^\bullet$ for all $m$. We first prove the lemma for the $n$ odd case, and we then modify the argument for the $n$ even case. Let $\varphi = \phi(\cdot|s_1, s_2)$ for all $m$, and assume without loss of generality that $\varphi$ is the identity mapping on $N$, and that $k = 1$, $\{2, 4, \ldots, n-3, n-1\} \subseteq C_L \cup C_K$, and $\{3, 5, \ldots, n-2, n\} \subseteq C_R \cup C_K$. Let $v^\ell \in \{a, r\}^\ell$ denote a sequence of votes of length $\ell$. We prove by induction that the following statement is true for all $\ell = 2, 4, \ldots, n-1$:

(H1) for all $v^\ell$ such that $|\{j \leq \ell \mid v_j^\ell = a\}| \geq \frac{\ell}{2}$, it must be that

$$
\liminf V_{\ell+1}(\lambda_{\ell+1}^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^\ell|\delta^m)) \quad \geq \quad u_{\ell+1}(\tilde{x}_k),
$$

(H2) for all $v^{\ell-1}$ such that $|\{j < \ell \mid v_j^{\ell-1} = a\}| \geq \frac{\ell}{2}$, it must be that

$$
V_j(\lambda_j^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{\ell-1}|\delta^m)) \to u_j(\tilde{x}_k),
$$

for all $j \in N$.

To prove the statement for $\ell = n-1$, take any $v^{n-1}$ such that $|\{j \leq n-1 \mid v_j^{n-1} = a\}| \geq \frac{n-1}{2}$. Since voting is by majority rule, we have

$(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-1}, a) \in H^\bullet(\tilde{x}_k)$. Letting $\sigma_n^a$ be identical to $\sigma_n^m$ with the proviso that $\sigma_n^a(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-1}) = a$, we have

$$\lambda_n^{(\sigma_n^a, \sigma_{-n}^m)}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-1}|\delta^m)(\tilde{x}_k) \;\; = \;\; 1,$$

and therefore

$$V_n(\lambda_n^{(\sigma_n^a, \sigma_{-n}^m)}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-1}|\delta^m)) \;\; = \;\; u_n(\tilde{x}_k).$$

By subgame perfection, it follows that

$$V_n(\lambda_n^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-1}|\delta^m)) \;\; \geq \;\; u_n(\tilde{x}_k)$$

for all $m$, establishing (H1).

Take any $v^{n-2}$ such that $|\{j < n-1 \mid v_j^{n-2} = a\}| \geq \frac{n-1}{2}$. Since voting is by majority rule, we have $(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-2}, a, v_n) \in H^\bullet(\tilde{x}_k)$ for each $v_n \in \{a, r\}$. Letting $\sigma_{n-1}^a$ be identical to $\sigma_{n-1}^m$ with the proviso that $\sigma_{n-1}^a(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-2}) = a$, we have

$$\lambda_j^{(\sigma_{n-1}^a, \sigma_{-(n-1)}^m)}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-2}|\delta^m)(\tilde{x}_k) \;\; = \;\; 1$$

for all $j \in N$, and therefore

$$V_j(\lambda_j^{(\sigma_{n-1}^a, \sigma_{-(n-1)}^m)}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-2}|\delta^m)) \;\; = \;\; u_j(\tilde{x}_k)$$

for all $j \in N$. If $\sigma_{n-1}^m = \sigma_{n-1}^a$, then (H2) is fulfilled.

Suppose instead that $\sigma_{n-1}^m(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-2}) = r$. Then subgame perfection implies

$$V_{n-1}(\lambda_{n-1}^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-2}|\delta^m)) \;\; \geq \;\; u_{n-1}(\tilde{x}_k).$$

By (H1), we have

$$\liminf V_n(\lambda_n^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-2}|\delta^m)) \;\; \geq \;\; u_n(\tilde{x}_k).$$

Since $X \cup \{q\}$ is compact, $\{\lambda_{n-1}^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-2}|\delta^m)\}$ has a weak* convergent subsequence, still indexed by $m$, with limit, say, $\lambda$. By Lemma 1, we have $\lambda_j^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-2}|\delta^m) \to \lambda$ for all $j \in N$. By continuity, it follows that

$$V_{n-1}(\lambda) \geq u_{n-1}(\tilde{x}_k) \quad \text{and} \quad V_n(\lambda) \geq u_n(\tilde{x}_k). \tag{8}$$

We next consider two possible cases.

First, under (A1), define the probability measure $\lambda'$ on $X$ by transferring probability mass from $q$ to the alternative $x^q$, as in (6). Note that $V_j(\lambda') = V_j(\lambda)$ for all $j \in N$. Thus, (8) and concavity imply

$$u_{n-1}(E\lambda') \geq u_{n-1}(\tilde{x}_k) \quad \text{and} \quad u_n(E\lambda') \geq u_n(\tilde{x}_k),$$

where $E\lambda' = \int x\lambda'(dx)$ is the mean of $\lambda'$. Furthermore, since $\tilde{x}_{n-1} \leq \tilde{x}_k \leq \tilde{x}_n$, we conclude that $E\lambda' = \tilde{x}_k$. Then (8) and strict concavity imply that $\lambda'$ is the pointmass on $\tilde{x}_k$. Therefore, we have

$$V_j(\lambda_j^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-2}|\delta^m)) \to V_j(\lambda) = V_j(\lambda') = u_j(\tilde{x}_k)$$

for all $j \in N$, as required. Second, under (A2), we claim that $\lambda(q) = 0$. Otherwise, if $\lambda(q) > 0$, then define $\lambda'$ as $\lambda'(Y) = \frac{\lambda(Y)}{1-\lambda(q)}$ for every Borel measurable set $Y \subseteq \mathbb{R}$. Since $u_j(q) < u_j(\tilde{x}_k)$ for all $j \in N$, we have $V_{n-1}(\lambda') > u_n(\tilde{x}_k)$ and $V_n(\lambda') > u_n(\tilde{x}_k)$ by (8). By concavity and $\tilde{x}_{n-1} \leq \tilde{x}_k \leq \tilde{x}_n$, however, this implies $E\lambda' < \tilde{x}_k$ and $\tilde{x}_k < E\lambda'$, a contradiction. Thus, $\lambda(q) = 0$. Setting $\lambda' = \lambda$, the argument of the previous paragraph carries over. Since our argument applies to all subsequences of $\{\lambda_{n-1}^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-2}|\delta^m)\}$, we have established (H2).

Suppose the statement is true for $\ell+2, \ldots, n-1$. We claim that it is true for $\ell$. Take any $v^\ell$ such that $|\{j \leq \ell \mid v_j^\ell = a\}| \geq \frac{\ell}{2}$. If agent $\ell+1$ votes to accept, then, letting $v^{\ell+1} = (v^\ell, a)$, we have $|\{j < \ell+2 \mid v_j^{\ell+1} = a\}| \geq \frac{\ell}{2}+1$, and so the antecedent of (H2) holds for $\ell+2$. Therefore, letting $\sigma_{\ell+1}^a$ be identical to $\sigma_{\ell+1}^m$ with the proviso that $\sigma_{\ell+1}^a(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^\ell) = a$, we must have

$$V_{\ell+1}(\lambda_{\ell+1}^{(\sigma_{\ell+1}^a, \sigma_{-(\ell+1)}^m)}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^\ell|\delta^m)) \to u_{\ell+1}(\tilde{x}_k). \tag{9}$$

By subgame perfection, it follows that

$$\begin{aligned} & V_{\ell+1}(\lambda_{\ell+1}^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^\ell|\delta^m)) \\ \geq \ & V_{\ell+1}(\lambda_{\ell+1}^{(\sigma_{\ell+1}^a, \sigma_{-(\ell+1)}^m)}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^\ell|\delta^m)) \end{aligned}$$

for all $m$. Taking limits and using (9), this implies (H1).

Take any $v^{\ell-1}$ such that $|\{j < \ell \mid v_j^{\ell-1} = a\}| \geq \frac{\ell}{2}$. If agent $\ell$ votes to accept, then, letting $v^{\ell+1} = (v^{\ell-1}, a, v_{\ell+1})$, we have $|\{j < \ell+2 \mid v_j^{\ell+1} = a\}| \geq \frac{\ell}{2}+1$ for each $v_{\ell+1} \in \{a, r\}$. Thus, the antecedent of (H2) holds for $\ell+2$. Therefore, letting $\sigma_\ell^a$ be identical to $\sigma_\ell^m$ with the proviso that $\sigma_\ell^a(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{\ell-1}) = a$, we have

$$V_j(\lambda_j^{(\sigma_\ell^a, \sigma_{-\ell}^m)}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{\ell-1}|\delta^m)) \to u_j(\tilde{x}_k) \tag{10}$$

for all $j \in N$. If $\sigma_\ell^a = \sigma_\ell^m$, then (H2) is fulfilled.

Suppose instead that $\sigma_\ell^m(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{\ell-1}) = r$. Subgame perfection implies

$$\begin{aligned} & V_\ell(\lambda_\ell^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{\ell-1}|\delta^m)) \\ & \quad \geq \; V_n(\lambda_\ell^{(\sigma_\ell^a, \sigma_{-\ell}^m)}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{\ell-1}|\delta^m)) \end{aligned}$$

for all $m$. With (10), this implies

$$\liminf V_\ell(\lambda_\ell^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{\ell-1}|\delta^m)) \;\; \geq \;\; u_\ell(\tilde{x}_k).$$

By (H1), we have

$$\liminf V_{\ell+1}(\lambda_{\ell+1}^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{\ell}|\delta^m)) \;\; \geq \;\; u_{\ell+1}(\tilde{x}_k).$$

Since $X$ is compact, $\{\lambda_\ell^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{\ell-1}|\delta^m)\}$ has a weak* convergent subsequence, still indexed by $m$, with limit, say, $\lambda$. By Lemma 1, we have $\lambda_j^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{n-2}, r|\delta^m) \to \lambda$ for all $j \in N$. By continuity, it follows that

$$V_\ell(\lambda) \; \geq \; u_\ell(\tilde{x}_k) \quad \text{ and } \quad V_{\ell+1}(\lambda) \; \geq \; u_{\ell+1}(\tilde{x}_k). \tag{11}$$

Then the argument for two possible cases, either (A1) or (A2), proceeds as above. Thus,

$$V_j(\lambda_j^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{\ell-1}|\delta^m)) \to u_j(\tilde{x}_k)$$

for all $j \in N$. Since our argument applies to all convergent subsequences of $\{\lambda_\ell^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{\ell-1}|\delta^m)\}$, we have established (H2).

We conclude that the induction statement is true for $\ell = 2$. Letting $\sigma_1^a$ be identical to $\sigma_1^m$ with the proviso that $\sigma_1^a(h^m, i, s_1^m, \tilde{x}_k, s_2^m) = a$, it follows that

$$V_1(\lambda_1^{(\sigma_1^a, \sigma_{-1}^m)}(h^m, i, s_1^m, \tilde{x}_k, s_2^m|\delta^m)) \to u_1(\tilde{x}_k).$$

Subgame perfection then implies

$$\liminf V_1(\lambda_1^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m|\delta^m)) \;\; \geq \;\; u_1(\tilde{x}_k).$$

Since $X \cup \{q\}$ is compact, $\{\lambda_1^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m|\delta^m)\}$ has a weak* convergent subsequence, still indexed by $m$, with limit, say, $\lambda$. By continuity, $V_1(\lambda) \geq u_1(\tilde{x}_k)$, and recall that $k = 1$, so $\tilde{x}_k$ is the ideal point of agent 1. Under (A1), define $\lambda'$ as in (6), so that $V_1(\lambda') = u_1(\tilde{x}_k)$. It follows that $\lambda'$ is the pointmass on $\tilde{x}_k$. Using Lemma 1, this implies

$$V_j(\lambda_j^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m|\delta^m)) \to V_j(\lambda) \;=\; V_j(\lambda') \;=\; u_j(\tilde{x}_k)$$

for all $j \in N$, as required. Under (A2), it follows that $\lambda(q) = 0$, and that $\lambda$ is the pointmass on $\tilde{x}_k$, with a similar conclusion. Since our argument applies to all subsequences of $\{\lambda_1^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m|\delta^m)\}$, the $n$ odd case is proved.

For the $n$ even case, we assume that $k = 1$, $\{3, 5, \ldots, n-3, n-1\} \subseteq C_L \cup C_K$, and $\{2, 4, \ldots, n-2, n\} \subseteq C_R \cup C_K$. We prove that the following version of the induction statement is true for all $\ell = 3, 5, \ldots, n-1$:

(H1) for all $v^\ell$ such that $v_1^\ell = a$ and $|\{j \leq \ell \mid v_j^\ell = a\}| \geq \frac{\ell-1}{2}$, it must be that

$$\liminf V_{\ell+1}(\lambda_{\ell+1}^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^\ell|\delta^m)) \;\;\geq\;\; u_{\ell+1}(\tilde{x}_k),$$

(H2) for all $v^{\ell-1}$ such that $v_1^{\ell-1} = a$ and $|\{j < \ell \mid v_j^{\ell-1} = a\}| \geq \frac{\ell-1}{2}$, it must be that

$$V_j(\lambda_j^{\sigma^m}(h^m, i, s_1^m, \tilde{x}_k, s_2^m, v^{\ell-1}|\delta^m)) \to u_j(\tilde{x}_k),$$

for all $j \in N$.

The argument is then as above, where now, when agents $n$ and $n-1$ vote to accept in the first step of the induction proof, at least half of the agents are in agreement; and because agent $k = 1$ is among them, this coalition is decisive. Once the induction statement is proved for $\ell = 3, 5, \ldots, n-1$, we skip agent 2. Agent 1 can obtain a payoff arbitrarily close to $u_1(\tilde{x}_k)$, and the conclusion follows as above. This completes the proof of the lemma. ∎

**Lemma 3** *For any $m$, let $\sigma^m \in \Sigma(\delta^m)$. If $x \in X^{\sigma^m}(\delta^m)$, then for all $C \in \mathcal{D}$, there exist $i^m \in C$ and $h^m \in H^c \backslash H^\bullet$ such that $u_{i^m}(x) \geq V_{i^m}(\lambda_{i^m}^{\sigma^m}(h^m|\delta^m))$.*

*Proof:* Suppose that $\sigma = ((p_i)_{i \in N}, (v_i)_{i \in N}) \in \Sigma(\delta^m)$, that $x \in X^\sigma(\delta^m)$, and take any $C \in \mathcal{D}$. Let $\tilde{h}^m \in H^c \setminus H^\bullet$ and $\hat{h}^m \in H^\bullet(x)$ satisfy $\tilde{h}^m < \hat{h}^m$

and $\zeta^\sigma(\hat{h}^m|\tilde{h}^m) > 0$. That is, after $\tilde{h}^m$, there is a positive probability that the agent selected proposes $x$, which then passes with a positive probability. Thus, there exist $i$, $s_1$, and $s_2$ such that $\rho(i, s_1|\tilde{h}^m)\pi(s_2|\tilde{h}^m, i, s_1, x) > 0$, $p_i(\tilde{h}^m, i, s_1) = x$, and $\zeta^\sigma(\hat{h}^m|\tilde{h}^m, i, s_1, x, s_2) > 0$. Without loss of generality, suppose that $\phi(\cdot|s_1, s_2)$ is the identity mapping on $N$, so that agent 1 votes first, then agent 2, and so on. Let $v_1^0 = v_1(\tilde{h}^m, i, s_1, x, s_2)$, and let

$$v_j^0 \quad = \quad v_j(\tilde{h}^m, i, s_1, x, s_2, v_1^0, \ldots, v_{j-1}^0)$$

denote agent $j$'s vote along the equilibrium path starting from $(\tilde{h}^m, i, s_1, x, s_2)$, for $j = 2, \ldots, n$.

Let $\ell = |C|$. Let $h^0 = (\tilde{h}^m, i, s_1, x, s_2, v_1^0, \ldots, v_n^0)$. For any $t = 1, \ldots, \ell$, we recursively define $i^t$ and $h^t = (\tilde{h}^m, i, s_1, x, s_2, v_1^t, \ldots, v_n^t)$ by changing the vote of each member of $C$, in order, to reject and letting all other agents vote according to their equilibrium strategies, until we generate a non-terminal history. More precisely, if $h^{t-1} \in H^\bullet$, then let $i^t = \min\{j \in C \mid v_j^{t-1} = a\}$. Note that this minimum is well-defined since $x$ passes at $h^{t-1}$ and $N \setminus C \notin \mathcal{D}$, implying that at least one member of $C$ must accept $x$. And define $h^t$ by the following: for all $j$ with $j < i^t$, let $v_j^t = v_j^{t-1}$; let $v_{i^t}^t = r$; and, for all $j$ with $j > i^t$, let

$$v_j^t \quad = \quad v_j(\tilde{h}^m, i, s_1, x, s_2, v_1^t, \ldots, v_{j-1}^t).$$

If $h^{t-1} \in H^c \setminus H^\bullet$, then let $i^t = i^{t-1}$ and $h^t = h^{t-1}$.

Let $h^m = h^\ell$ and $i^m = i^\ell$. It is clear that $h^m \in H^c \setminus H^\bullet$: otherwise, for all $t = 1, \ldots, \ell$, we have $h^t \in H^\bullet$, so by construction $v_j^\ell = r$ for all $j \in C$; this implies $\{j \in N \mid v_j^\ell = a\} \subseteq N \setminus C$ is not decisive, contradicting $h^\ell \in H^\bullet$. By construction, agent $i^m$ votes to accept after previous members of $C$ have been changed to reject, i.e.,

$$v_{i^m}(\tilde{h}^m, i, s_1, x, s_2, v_1^\ell, \ldots, v_{i^m-1}^\ell) \quad = \quad a.$$

Define the strategy $\sigma_{i^m}^r$ for $i^m$ that is identical to $\sigma_i$ except that $i^m$ votes to reject after this history:

$$v_{i^m}^r(\tilde{h}^m, i, s_1, x, s_2, v_1^\ell, \ldots, v_{i^m-1}^\ell) \quad = \quad r.$$

By construction, $x$ passes when $i^m$ votes to accept, which yields

$$\lambda_{i^m}^\sigma(\tilde{h}^m, i, s_1, x, s_2, v_1^\ell, \ldots, v_{i^m-1}^\ell, a|\delta^m)(x) \quad = \quad 1.$$

Also by construction, $x$ fails when $i^m$ votes to reject, leading to history $h^m$, so that

$$\lambda_{i^m}^{(\sigma_i^r,\sigma_{-i})}(\tilde{h}^m, i, s_1, x, s_2, v_1^\ell, \ldots, v_{i^m-1}^\ell, r|\delta^m) \quad = \quad \lambda_{i^m}^\sigma(h^m|\delta^m).$$

Then subgame perfection requires that

$$\begin{aligned} u_{i^m}(x) &= V_{i^m}(\lambda_{i^m}^\sigma(\tilde{h}^m, i, s_1, x, s_2, v_1^\ell, \ldots, v_{i^m}^\ell, a)|\delta^m) \\ &\geq V_{i^m}(\lambda_{i^m}^\sigma(h^m|\delta^m)), \end{aligned}$$

as desired. ∎

# References


[1] K. Arrow. 1963. *Social Choice and Individual Values.* New Haven: Yale University Press.

[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. 2003. "A Bargaining Model of Legislative Policy-Making." mimeo.

[4] J. Banks, J. Duggan, and M. Le Breton. 2002. "Bounds for Mixed Strategy Equilibria and the Spatial Model of Elections." *Journal of Economic Theory* 103:88-105.

[5] D. Baron and J. Ferejohn. 1989. "Bargaining in Legislatures." *American Political Science Review* 83:1181-1206.

[6] D. Baron and K. Ehud. 1993. "The Simplest Equilibrium of a Majority Rule Division Game." *Journal of Economic Theory* 61:290-301.

[7] J. Bergin and J. Duggan. 1999. "An Implementation-Theoretic Approach to Non-Cooperative Foundations." *Journal of Economic Theory* 86:50-76.

[8] K. Binmore. 1987. "Perfect Equilibria in Bargaining Models." In *The Economics of Bargaining,* ed. K. Binmore and P. Dasgupta. Oxford: Basil Blackwell. Pp.77-105.

[9] D. Black. 1958. *The Theory of Committees and Elections.* Cambridge: Cambridge University Press.

[10] S. Cho and J. Duggan. 2002. "Uniqueness of Stationary Equilibria in a One-dimensional Model of Bargaining." *Journal of Economic Theory,* forthcoming.

[11] A. Downs. 1957. *An Economic Theory of Democracy.* New York: Harper and Row.

[12] J. Duggan and M. Fey. 2002. "Electoral Competition with Policy-Motivated Candidates." mimeo.

[13] P. Dutta. 1995. "A Folk Theorem for Stochastic Games." *Journal of Economic Theory* 66:1-32.

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

[15] H. Eraslan and A. Merlo. "Majority Rule in a Stochastic Model of Bargaining." *Journal of Economic Theory* 103:31-48.

[16] D. Fudenberg and E. Maskin. 1986. "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica* 54:533-556.

[17] J. Harrington. 1989. "The Advantageous Nature of Risk Aversion in a Three-player Bargaining Game where Acceptance of a Proposal Requires a Simple Majority." *Economics Letters* 30:195-200.

[18] J. Harrington. 1990a. "The Power of the Proposal Maker in a Model of Endogenous Agenda Formation." *Public Choice* 64:1-20.

[19] J. Harrington. 1990b. "The Role of Risk Preferences in Bargaining when Acceptance of a Proposal Requires less than Unanimous Approval." *Journal of Risk and Uncertainty* 3:135-154.

[20] H. Hotelling. 1929. "Stability in Competition." *Economic Journal* 39:41-57.

[21] R. Lagunoff and A. Matsui. 1997. "Asynchronous Choice in Repeated Coordination Games." *Econometrica* 65:1467-1477.

[22] A. Merlo and C. Wilson. 1995. "A Stochastic Model of Sequential Bargaining with Complete Information." *Econometrica* 63:371-399.

[23] B. Moldovanu and E. Winter. 1995. “Order Independent Equilibria.” *Games and Economic Behavior* 9:21-34.

[24] A. Rubinstein. 1982. “Perfect Equilibrium in a Bargaining Model.” *Econometrica* 50:97-109.

[25] J. Sutton (1986) “Non-Cooperative Bargaining Theory: An Introduction.” *Review of Economic Studies* 53:709-724.