PROPORTIONALITY IN MARKOV TEMPORAL ELECTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Temporal voting assumes that the alternative selected in one round does not affect the alternatives available later. We present *Markov temporal elections*, where the selected alternative instead determines the next state and changes future voting opportunities. We ask how proportionality should be defined when earlier decisions affect the opportunities for representation that remain. We show that the population shares used at the initial state cannot simply be reused at every later state: this can make proportionality impossible or force a Pareto-dominated policy that makes every voter worse off by a factor approaching the number of voters. We instead derive state-dependent proportional claims that account for the remaining representation opportunities. Among separable welfare objectives, these claims uniquely select Nash welfare, whose optimal policies satisfy a proportionality guarantee against coalitional deviations at every state. We show that deciding whether a coalition of a given size has an improving deviation is NP-complete, so we develop an efficiently computable certificate that bounds such deviations without enumerating coalitions. We extend the certificate to unknown transitions, where the proportional claims must also be inferred from data. Even unlimited on-policy data can leave the consequences of unchosen alternatives unidentified, making valid certificates highly conservative under limited coverage. Experiments show that action-dependent transitions increase the deviations available to coalitions against a fixed policy and that the certificate is much tighter than its worst-case guarantee on the tested instances. With unknown transitions, how informative the certificate is depends on coverage and uncertainty in the proportional claims.

1 INTRODUCTION

Temporal voting studies sequences of collective decisions by the same electorate, with one alternative selected per round. Proportionality notions from approval-based committee voting have been adapted to this setting, including justified representation (JR), proportional justified representation (PJR), extended justified representation (EJR), and the Core (Chandak et al., 2026; Phillips et al., 2026). The Core prevents a group from using a share of the decisions proportional to its size to obtain an outcome every member strictly prefers. We use its fractional counterpart, where a coalition's population share scales its utility from an alternative randomized outcome. These notions protect groups across time but assume future alternatives are independent of earlier choices.

Many repeated decisions violate this assumption. Funding a transport corridor can change future projects, electing a council member can make that alternative temporarily unavailable, and approving a research direction can reshape future proposals. We thus introduce *Markov temporal elections*, where the current state determines the available alternatives and the selected alternative determines the next-state distribution, allowing decisions to reshape future voting opportunities.

This raises a question absent when future alternatives are fixed: *after a state is reached, what proportional claim should a coalition have on the continuation?* Carrying initial population shares forward ignores the representation opportunities already used and those that remain, while allowing shares to change arbitrarily makes proportionality meaningless. Our goal is therefore to identify a well-founded state-dependent notion of proportionality, study its consequences, and develop certificates for the resulting coalitional guarantees under known and unknown transitions.

Contributions. Our main contributions are as follows.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Fixed population shares are insufficient (Section 3). We show that population shares used by the fractional Core at the initial state cannot generally be reused at continuation states. In some elections, no policy can satisfy the fractional Core initially and at every reached state using the original shares; in others, imposing them at continuation states forces a Pareto-dominated profile that can make every voter worse off by a factor approaching the number of voters. Thus, proportional claims must adapt as future representation opportunities change.

Continuation claims and Nash welfare (Sections 4 and 5). We derive state-dependent proportional claims that track remaining representation opportunities. Dynamic consistency determines the continuation weights among separable welfare objectives, and our proportional update uniquely selects Nash welfare; an independent axiomatic characterization yields the same rule. Policies optimal for the resulting weighted objective at every state are protected against coalitional deviations and recover the fractional Core at the initial state.

Efficient assessment of coalitional deviations (Section 6). Deciding whether a coalition of a given size has an improving deviation is NP-complete. We develop an efficiently computable certificate that bounds the strongest proportional deviation without enumerating coalitions. Its looseness depends on the state-dependent claims, and its guarantees degrade under approximate optimality.

Unknown dynamics and coverage (Section 7). With unknown transitions, the policy's utilities and hence its proportional claims are uncertain. We develop robust and sample-based certificates for both sources of uncertainty and study coverage of state-alternative pairs. Even unlimited on-policy data can leave unchosen alternatives indistinguishable despite very different coalitional deviations. Our certificate adapts to coverage and remains valid under arbitrary coverage but becomes conservative, or reports insufficient coverage, when the proportional claims cannot be identified.

Empirical evaluation (Section 8). On synthetic elections, the certificate is substantially tighter than its worst-case guarantee and tightens as policies approach a Nash-welfare optimum while coalitional deviations shrink. Action-dependent transitions increase deviations against a fixed policy, while limited coverage under unknown transitions makes the certificate more conservative, partly through uncertainty in the proportional claims.

Related work. We discuss the closest related work; a broader comparison appears in Appendix A.

Temporal voting and constrained collective decisions. Perpetual and temporal voting study fairness in repeated collective decisions (Lackner, 2020; Bulteau et al., 2021; Lackner & Maly, 2023). Temporal voting has adapted JR, PJR, EJR, and stronger Core notions (Chandak et al., 2026; Phillips et al., 2026), with subsequent work studying verification and utilitarian cost (Elkind et al., 2025; Teh, 2026). Existing models may also constrain which decision sequences are feasible (Elkind et al., 2024). Our setting instead allows the selected alternative to change the distribution over future states and opportunities, making proportional claims state-dependent. Related work on constrained public decisions studies static dependencies among feasible outcomes (Brill et al., 2023; Chingoma et al., 2024; Masařík et al., 2024); ours arise dynamically from current choices and state transitions.

Nash welfare, the Core, and proportional fairness. Nash welfare and Core stability are closely connected in public-decision problems (Conitzer et al., 2017; Fain et al., 2016; 2018). Ebadian et al. (2024) show that approximate proportional fairness implies an approximate Core guarantee in randomized social choice. Our initial-state certificate recovers their proportional-fairness quantity, while our analysis extends this connection to continuation states using state-dependent claims and bounds the gap between the scalar certificate and the strongest coalitional deviation.

Fairness and social choice in MDPs. Dynamic social choice has been formulated as an MDP (Parkes & Procaccia, 2013), and occupancy measures have been used to aggregate policies through social-choice rules (Alamdari et al., 2024). Most closely, Blair & Larson (2026) combine an MDP with Nash-welfare-based ex-ante Core stability; our guarantees are statewise and use state-dependent proportional claims. Fair reinforcement learning studies welfare objectives in known and unknown MDPs (Mandal & Gan, 2022; Cousins et al., 2024; Kim et al., 2025). Our statistical problem instead assesses coalitional deviations when both the transition kernel and policy-induced claims are uncertain, using robust and offline-MDP tools (Iyengar, 2005; Wiesemann et al., 2013; Petrik et al., 2016; Kidambi et al., 2020; Jin et al., 2021). Related offline fairness verification considers constraints specified independently of unknown transitions (Metevier et al., 2019).

2 MARKOV TEMPORAL ELECTIONS

For $z \in \mathbb{N}$, write $[z] := \{1, \dots, z\}$. We consider collective decisions over a finite horizon $H \in \mathbb{N}$, with voters $N = \{1, \dots, n\}$ and a finite state set $\mathcal{S}_h$ in each round $h \in [H]$. Each state $s \in \mathcal{S}_h$ has a nonempty finite set of alternatives $C_h(s)$, which need not come from a common set. A *Markov temporal election* is specified by $\mathcal{E} = (N, H, \{\mathcal{S}_h\}_{h=1}^{H+1}, \{C_h\}_{h=1}^{H}, \mu, \{P_h\}_{h=1}^{H}, \{a_{i,h}\}_{i \in N, h \in [H]})$. Here, $\mu \in \Delta(\mathcal{S}_1)$ is the initial-state distribution and $P_h(\cdot \mid s, c) \in \Delta(\mathcal{S}_{h+1})$ the next-state distribution after selecting $c \in C_h(s)$, where $\Delta(X)$ denotes distributions over finite X. Classical temporal voting is the special case in which $P_h(\cdot \mid s, c)$ is independent of c. Voter i receives approval utility $a_{i,h}(s, c) \in \{0, 1\}$ from that selection. Thus, the state determines the available alternatives, while the selected alternative determines current utilities and future opportunities. Results not relying on approval utilities extend to nonnegative cardinal utilities $a_{i,h}(s, c) \in [0, \bar{a}]$ for finite $\bar{a}$; nonnegativity is assumed throughout. For convenience, we add an artificial round 0 containing a single state z_0 and zero-utility alternative whose transition is μ. Thus, the process starts deterministically at z_0, and whole-election quantities can be viewed as continuation quantities at z_0. Round $H + 1$ is terminal, with no decision or utility. We write $\mathcal{Z}$ for the resulting round-state pairs and abbreviate (h, s) by z.

A *randomized Markov policy* assigns, at each nonterminal state $s \in \mathcal{S}_h$, a distribution $\pi_h(\cdot \mid s) \in \Delta(C_h(s))$ over its available alternatives. At the artificial initial state z_0, the unique alternative is selected with probability one. Let Π denote the set of such policies. For a policy $\pi \in \Pi$, voter i's *ex-ante utility* is the total expected utility over the election $U_i(\pi) = \mathbb{E}_\pi[\sum_{h=1}^{H} a_{i,h}(s_h, c_h)]$. The corresponding *continuation utility* at state $s \in \mathcal{S}_h$ is $V_i^\pi(h, s) = \mathbb{E}_\pi[\sum_{t=h}^{H} a_{i,t}(s_t, c_t) \mid s_h = s]$. We set $V_i^\pi(H + 1, s) = 0$. By construction, $V_i^\pi(z_0) = U_i(\pi)$, so ex-ante utility evaluates the policy before the initial state is realized, and continuation utility evaluates it from a particular state onward.

We represent policies by their probabilities of visiting state-alternative pairs. For $h \le H$, $s \in \mathcal{S}_h$, and $c \in C_h(s)$, define the *state-action occupancy* $x_h^\pi(s, c) = \Pr_\pi(s_h = s, c_h = c)$ and let $\mathbf{x}^\pi = (x_h^\pi(s, c))_{h,s,c}$ be π's occupancy vector. The feasible occupancy set $\mathcal{X}$ is a compact convex polytope whose flow constraints appear in Appendix B. As utilities are linear in occupancy, the attainable utility profiles also form a compact convex polytope. Randomized Markov policies attain the same utility profiles as history-dependent policies (Derman & Strauch, 1966; Puterman, 1994; Altman, 1999), in the layered form of Rosenberg & Mansour (2019). Summing these occupancies over the alternatives available at a state gives its reach probability, $\sum_{c \in C_h(s)} x_h^\pi(s, c) = \Pr_\pi(s_h = s)$. We call (h, s) *reached* by π if this probability is positive and *off the support* otherwise. Policies remain defined at off-support states, which matters for guarantees applying at every state.

We now define proportionality at the initial state. A *coalition* is a nonempty subset $S \subseteq N$, with population share $\alpha(S) = |S|/n$. The fractional Core requires that no coalition can improve every member's utility by switching to another randomized policy after scaling its utilities by this share.

Definition 2.1 (Fractional Core)**.** A policy $\pi \in \Pi$ is in the *fractional Core* if there are no nonempty coalition $S \subseteq N$ and policy $\sigma \in \Pi$ such that $\alpha(S)U_i(\sigma) > U_i(\pi)$ for every $i \in S$.

The scaling in Definition 2.1 applies to utilities, not the deviating policy: σ may be any policy in Π, with no additional feasibility or budget constraint. See Appendix B for the occupancy interpretation.

3 WHY FIXED POPULATION SHARES ARE INSUFFICIENT

The fractional Core in Definition 2.1 assigns each coalition S its population share $|S|/n$ when evaluating the election as a whole. A natural statewise extension is to reuse this share after every reached state. We show that this fails: the initial and statewise requirements can be incompatible and, even when they are compatible, enforcing the original shares can force a Pareto-dominated outcome that makes every voter worse off by a factor approaching n. Thus, the issue extends beyond feasibility.

Theorem 3.1. *For every $n \ge 2$ and integer $K \ge 2$, there is a Markov temporal election with deterministic transitions, approval preferences, $H = K + 2$ rounds, and $O(n^2 K)$ states in which no policy is simultaneously in the fractional Core at the initial state and in the fractional Core with the original shares $|S|/n$ at every state reached with positive probability.*

The incompatibility arises because the initial decision determines which continuation state is reached, with each such state favoring a different voter. In the state favoring voter j, that voter

can secure continuation value K, while any other voter can secure only 1. Reusing the ex-ante share $1/n$ after opportunities become unequal ignores the changed representation opportunities, making initial and statewise proportionality incompatible. See Appendix C.1 for the construction and proof.

Infeasibility, however, might reflect an overly demanding statewise requirement. We next show that even when the initial and statewise requirements are jointly feasible, reusing population shares can force a Pareto-dominated outcome that makes every voter worse off by a factor approaching n.

Theorem 3.2. *For every $n \geq 2$ and $K \geq 2$, there is a stochastic approval election in which imposing the original shares at every continuation state forces the unique utility profile $U_i = (K+n-1)/n^2$ for every voter i, while another policy gives every voter K/n and lies in the fractional Core at the initial state. The ratio between these utilities is $nK/(K+n-1)$, which approaches n as $K \to \infty$.*

Proof sketch. Modify the construction of Theorem 3.1 so that the continuation state favoring each voter is reached uniformly at random. Reusing the original shares forces a uniform continuation lottery, giving every voter $(K+n-1)/n^2$. Instead, always selecting the alternative favoring the advantaged voter gives every voter K/n and lies in the fractional Core: any deviation gives a coalition total utility at most K, while blocking would require more than K. The resulting welfare ratio approaches n. See Appendix C.2 for the full proof.

4 State-dependent proportionality

Section 3 shows that population shares cannot simply be carried forward after a state is reached. We therefore ask how proportional claims should adapt when earlier decisions change the opportunities for representation that remain. We approach this through welfare optimization: if a policy is optimal ex ante, how should the same welfare criterion evaluate its continuation after conditioning on a state?

Welfare objectives. For positive utilities, consider the *separable welfare objective* $F_\phi(\pi) = \sum_i w_i \phi(U_i(\pi))$, where $\mathbf{w} \in \mathbb{R}^n_{>0}$ and $\phi : (0,\infty) \to \mathbb{R}$ is strictly increasing, concave, and continuously differentiable. Here, w_i is voter i's initial weight and $\phi'(x)$ its marginal welfare value at utility x. To accommodate zero continuation utility, extend ϕ by $\bar{\phi}(0) = \lim_{t \downarrow 0} \phi(t)$, possibly $-\infty$; its boundary properties are given in Appendix D.3. Compactness of the attainable utility set and upper semicontinuity of F_ϕ ensure a maximizer. We assume every voter can receive positive utility: $\max_{\pi \in \Pi} U_i(\pi) > 0$ for every i, which by Lemma D.1 implies that some policy gives all voters positive utility. Hence, if $\bar{\phi}(0) = -\infty$, every maximizer is strictly positive; more generally, we consider a maximizer $\pi^\star$ with $U(\pi^\star) > 0$. Voters who can never receive positive utility can be included but not simply deleted, since coalition shares are measured relative to the full electorate (Lemma D.1 and Proposition D.1).

Scalarization and tightness. Let $u_i = U_i(\pi^\star)$ denote voter i's ex-ante utility under $\pi^\star$. At this optimum, voter i has marginal welfare weight $\lambda_i = w_i \phi'(u_i) > 0$. Lemma D.5 shows that $\pi^\star$ also maximizes the scalarized objective $\sum_i \lambda_i U_i$. Letting $\boldsymbol{\lambda} = (\lambda_i)_{i \in N}$, the corresponding single-objective MDP has scalar reward $r^{\boldsymbol{\lambda}}_h(s,c) = \sum_i \lambda_i a_{i,h}(s,c)$. Let $y^{\boldsymbol{\lambda}}(z)$ denote its maximum expected reward from state z, with $y^{\boldsymbol{\lambda}}(H+1,s) = 0$. We call a policy $\boldsymbol{\lambda}$-*tight* if $\sum_i \lambda_i V_i^\pi(z) = y^{\boldsymbol{\lambda}}(z)$ at every state z, meaning that it remains optimal for the same scalar objective from every state, not only ex ante. By Lemma D.3, a scalar-optimal policy can be completed at off-support states to become $\boldsymbol{\lambda}$-tight without changing its ex-ante utilities, so we may take $\pi^\star$ to be $\boldsymbol{\lambda}$-tight. Thus, $\boldsymbol{\lambda}$ is the marginal-weight vector determined by F_ϕ at $\pi^\star$ rather than an arbitrary scalarization.

Dynamic consistency. We now ask how the nonlinear welfare objective should be evaluated after conditioning on a state. Fix z and let $v_i = V_i^{\pi^\star}(z)$ denote voter i's continuation utility under $\pi^\star$. A *continuation-weight rule* assigns a positive weight $b_i(z)$ to each voter with $v_i > 0$. We consider rules that depend on $\pi^\star$ and z only through the initial weights $\mathbf{w}$, ex-ante utilities $\mathbf{u} = (u_i)_{i \in N}$, and continuation utilities $\mathbf{v} = (v_i)_{i \in N}$. This defines the scope of our characterization; it does not require any state-dependent proportionality notion to have this form.

Such a rule is *dynamically consistent* with F_ϕ if the continuation of $\pi^\star$ from every state z maximizes $\sum_{i: v_i > 0} b_i(z) \bar{\phi}(V_i^\sigma(z))$ over continuation policies σ. Thus, the welfare transformation remains fixed while the voter weights may change after conditioning on a state. We next identify these weights and, within the stated $(\mathbf{w}, \mathbf{u}, \mathbf{v})$ scope, characterize them up to a common positive scaling.

Theorem 4.1. *Let $\pi^\star$ maximize F_ϕ with $U(\pi^\star) > 0$, $\lambda_i = w_i \phi'(u_i) > 0$, and suppose $\pi^\star$ is $\boldsymbol{\lambda}$-tight.*

(i) *For any voter i with $v_i > 0$, let $b_i^\phi(z) = \lambda_i/\phi'(v_i) = w_i\phi'(u_i)/\phi'(v_i)$. At any state z, the continuation of $\pi^\star$ maximizes $\sum_{i:v_i>0} b_i^\phi(z)\bar{\phi}(V_i^\sigma(z))$ over all continuation policies σ, in the extended-real sense, including when some voters have zero continuation utility under either $\pi^\star$ or σ.*

(ii) *For $n \geq 2$, any positive continuation-weight rule depending only on $(\mathbf{w}, \mathbf{u}, \mathbf{v})$ and guaranteeing the local optimality in part* (i) *on every convex continuation set must coincide with $\mathbf{b}^\phi$ up to a common positive factor, for every $(\mathbf{w}, \mathbf{u}, \mathbf{v}) \in (\mathbb{R}^n_{>0})^3$, with no restriction relating $\mathbf{v}$ to $\mathbf{u}$.*

Proof sketch. For part *(i)*, $\boldsymbol{\lambda}$-tightness makes the policy optimal for the objective with weights $\lambda_i = w_i\phi'(u_i)$ at every state. Since $b_i^\phi(z)\phi'(v_i) = \lambda_i$, the tangent inequality for $\bar{\phi}$ gives local optimality. For part *(ii)*, local optimality along feasible segments forces the marginal coefficients to be proportional to $\boldsymbol{\lambda}$, yielding $b_i(z) \propto \lambda_i/\phi'(v_i)$. See Appendix D.4 for the full proof.

The continuation weights have a direct interpretation. The marginal welfare value of voter i's utility changes from $w_i\phi'(u_i)$ at the ex ante optimum to $\phi'(v_i)$ after conditioning on z, and $b_i^\phi(z)$ compensates for this change. The continuation objective therefore preserves the trade-offs that make the policy optimal ex ante while adapting to the utilities that remain.

Proportional interpretations. Requiring continuation weights to remain proportional to the initial weights $\mathbf{w}$ characterizes weighted utilitarian welfare within our separable class (Appendix D.5). Instead, scaling voter i's initial claim w_i by the continuation-to-ex-ante utility ratio $V_i^{\pi^\star}(z)/U_i(\pi^\star)$ leads to Nash welfare, the product of voters' utilities. The next result shows that Nash welfare is the unique separable objective inducing this update. The full proof appears in Appendix D.5.

Corollary 4.1. *If $b_i^\phi \propto w_i V_i^{\pi^\star}(z)/U_i(\pi^\star)$ for any voter i at every state, then $\phi(x) = a\log x + e$ for some $a > 0$. Conversely, the Nash-welfare objective induces exactly these continuation weights.*

The characterization concerns positive continuation utilities. For Nash welfare, the proportional rule extends naturally to $V_i^{\pi^\star}(z) = 0$ by assigning that voter zero continuation claim; Appendix D.7 gives the corresponding boundary treatment. Thus, voter i's continuation claim is their initial claim w_i scaled by $V_i^{\pi^\star}(z)/U_i(\pi^\star)$, tracking the representation opportunities remaining after conditioning on the state. An independent axiomatic characterization in Appendix D.7 derives the same update rule directly from conditions on proportional claims, while Appendix D.6 gives a complementary price characterization of Nash-optimal policies.

5 Continuation stability and representation guarantees

Section 4 derived state-dependent proportional claims that track the representation opportunities remaining at a state. We now show that a $\boldsymbol{\lambda}$-tight policy satisfies *continuation stability*, protecting against coalitional deviations at every state and recovering the fractional Core at the initial state. This requires less structure: fix any $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$ and a $\boldsymbol{\lambda}$-tight policy π, so that $\sum_i \lambda_i V_i^\pi(z) = y^{\boldsymbol{\lambda}}(z)$ at every state. By Lemma D.3*(ii)*, any policy maximizing $\sum_i \lambda_i U_i$ can be completed at off-support states to satisfy this condition without changing its ex-ante utilities.

Now, fix a state z and define voter i's *continuation entitlement* as $b_i(z) = \lambda_i V_i^\pi(z)$, their contribution to the scalarized continuation value. For a coalition S, let $b(S) = \sum_{i\in S} b_i(z)$ and $B(z) = \sum_{i\in N} b_i(z)$ denote the coalition's and the electorate's total continuation entitlement, respectively. Since π is $\boldsymbol{\lambda}$-tight, $B(z) = y^{\boldsymbol{\lambda}}(z)$, the maximum scalarized continuation value from z. When $B(z) > 0$, define S's *continuation share* as $\alpha_z(S) = b(S)/B(z)$; when $B(z) = 0$, set $\alpha_z(S) = 0$. We call voter i *active* at z if $V_i^\pi(z) > 0$ and write $A(z) = \{i \in N : V_i^\pi(z) > 0\}$.

These continuation shares are a state-dependent analogue of the population shares used by the fractional Core. The corresponding stability question is whether a coalition can use its share $\alpha_z(S)$ to find a continuation that weakly improves every member and strictly improves at least one. We next show that $\boldsymbol{\lambda}$-tightness rules out such deviations at every state. See Appendix E.1 for the full proof.

Theorem 5.1. *Let π be $\boldsymbol{\lambda}$-tight for $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$. At any state z, no $\varnothing \neq S \subseteq N$ and continuation policy σ satisfy $\alpha_z(S)V_i^\sigma(z) \geq V_i^\pi(z)$ for every $i \in S$, with strict inequality for some $i \in S$.*

The weak-blocking guarantee of Theorem 5.1 immediately implies the all-strict condition used by the fractional Core. Under the reciprocal normalization $\lambda_i = 1/U_i(\pi)$, $V_i^\pi(z_0) = U_i(\pi)$ gives

$b_i(z_0) = 1$ for any voter, so $B(z_0) = n$ and $\alpha_{z_0}(S) = |S|/n$. Thus, a $\boldsymbol{\lambda}$-tight policy with reciprocal weights lies in the fractional Core at the initial state; the formal corollary appears in Appendix E.2.

More generally, the fractional Core uses the population share $|S|/n$ at the initial state, whereas continuation stability uses the policy-dependent share $\alpha_z(S)$ at each state. We use *continuation Core* for policies satisfying continuation stability at every state. Under reciprocal normalization, the shares coincide at z_0 but can differ thereafter; Section 3 shows why continuing to impose population shares can be infeasible or highly inefficient. The identity $b_i(z_0) = 1$ under reciprocal weights holds algebraically for any positive-utility policy and is distinct from $\boldsymbol{\lambda}$-tightness, the additional optimality condition yielding the statewise guarantee of Theorem 5.1. Additional weighted-mean and quantile representation guarantees appear in Appendix E.3.

6 Assessment of continuation stability

The guarantees of Section 5 apply to $\boldsymbol{\lambda}$-tight policies but do not assess a given policy or quantify what remains when tightness fails. We therefore compare the strongest coalitional deviation at a state with an efficiently computable scalar certificate. We show that a related exact blocking problem is NP-complete, while the certificate is computable by dynamic programming, and quantify its degradation under approximate optimality. Throughout, fix a policy π with $U(\pi) > 0$ and $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$, defining the scalarized objective $\sum_i \lambda_i U_i$. The canonical choice is $\lambda_i = 1/U_i(\pi)$, but none of the results assumes that π is Nash-optimal or $\boldsymbol{\lambda}$-tight.

Exact deviation ratio. At a state z with $B(z) > 0$, define the *worst deviation ratio* $\mathrm{Dev}_z(\pi; \boldsymbol{\lambda}) = \max_{\varnothing \neq S \subseteq A(z)} \max_\sigma \alpha_z(S) \min_{i \in S} V_i^\sigma(z)/V_i^\pi(z)$. For a coalition S and continuation policy σ, the inner minimum is its least improvement ratio, while $\alpha_z(S)$ accounts for its continuation share. Thus, Dev_z is the strongest proportional deviation from z. The maximization is over active voters, so the ratios are well defined; by Lemma F.1, removing inactive members does not affect all-strict blocks, and $\mathrm{Dev}_z = 1$ exactly when no such block exists ($\mathrm{Dev}_z \geq 1$ when $B(z) > 0$). This is weaker than continuation stability, which also rules out weak improvements with at least one strict improvement.

Scalar certificate. Computing Dev_z requires optimizing over coalitions and continuation policies. To avoid this search, define the *scalar continuation gap* $\Delta_\pi(z) = y^{\boldsymbol{\lambda}}(z) - \sum_i \lambda_i V_i^\pi(z) \geq 0$, measuring how far π falls short of the maximum scalarized continuation value from z. When $B(z) > 0$, define the *proportional-fairness ratio* $\mathrm{PF}_z(\pi; \boldsymbol{\lambda}) = y^{\boldsymbol{\lambda}}(z)/B(z) = 1 + \Delta_\pi(z)/B(z)$. Unlike Dev_z, PF_z requires no coalitional search and is computed by one policy evaluation and one Bellman pass. When $B(z) = 0$, Dev_z is undefined, while $\mathrm{PF}_z = 1$ if $y^{\boldsymbol{\lambda}}(z) = 0$ and $+\infty$ otherwise. Thus, PF_z is defined at every state and $\mathrm{PF}_z \equiv 1$ exactly when π is $\boldsymbol{\lambda}$-tight. For $B(z) > 0$, $\mathrm{PF}_z = 1$ assesses continuation stability, whereas $\mathrm{Dev}_z = 1$ only excludes all-strict blocks.

We suppress the dependence on $\boldsymbol{\lambda}$ when the weights are fixed. In Section 7, where the kernel and weights may vary, we write $\mathrm{PF}_z^Q(\pi; \boldsymbol{\lambda})$ and $\mathrm{Dev}_z^Q(\pi; \boldsymbol{\lambda})$, together with $B^\pi_{\boldsymbol{\lambda},Q}(z)$ and $y^{\boldsymbol{\lambda},Q}(z)$. Under reciprocal weights, $V_i^\pi(z_0) = U_i(\pi)$ gives $b_i(z_0) = 1$, $B(z_0) = n$, and $\mathrm{PF}_{z_0}(\pi) = \max_\sigma \frac{1}{n} \sum_{i \in N} \frac{U_i(\sigma)}{U_i(\pi)}$. This is the proportional-fairness ratio of Ebadian et al. (2024); PF_z extends it to continuation states. None of these definitions requires π to be Nash-optimal or $\boldsymbol{\lambda}$-tight: departures from scalar optimality are measured by $\Delta_\pi(z)$.

Computational hardness. We do not resolve the complexity of computing Dev_z in general, but even deciding whether a coalition of a given size has an improving deviation is NP-complete (Appendix F.1). Hardness holds already for $H = 1$ with rational cardinal utilities and a deterministic reference policy, and for stochastic approval elections with $H = 2n+1$ and $O(nd)$ states. Membership in NP follows because a coalition is a witness and an improving continuation can be checked by linear programming; hardness is by reduction from Open Hemisphere (Johnson & Preparata, 1978; Garey & Johnson, 1979). The formal statement and proof appear in Appendix F.1.

This motivates using the efficiently computable PF_z in place of coalitional search. Suppose every voter is active at z and let $\sigma^\star$ attain $y^{\boldsymbol{\lambda}}(z)$. Writing $r_i = V_i^{\sigma^\star}(z)/V_i^\pi(z)$ gives $\mathrm{PF}_z = \sum_i b_i(z) r_i / B(z)$, the entitlement-weighted average improvement under a scalar-optimal continuation, whereas Dev_z evaluates a coalition by its least-improved member. The next theorem bounds the gap between them.

Theorem 6.1. *Let $B(z) > 0$. Then, $\mathrm{Dev}_z(\pi) \le \mathrm{PF}_z(\pi)$. If every voter is active at z, then $\mathrm{PF}_z(\pi) \le (1 + \ln \frac{B(z)}{\min_{i \in N} b_i(z)})\mathrm{Dev}_z(\pi)$. At the initial state under reciprocal normalization, the upper factor can be improved to the nth harmonic number H_n.*

Proof sketch. A coalitional deviation contributes to the scalar objective defining PF_z, giving $\mathrm{Dev}_z \le \mathrm{PF}_z$. With all voters active, comparing this entitlement-weighted average with the best coalition's worst-member improvement gives the logarithmic factor; see Appendix F.3.

At the initial state under reciprocal weights, the lower bound recovers the proportional-fairness bound of Ebadian et al. (2024), while the converse shows that the looseness of the certificate depends on how unequal continuation entitlements are. Appendix F.3 gives the tighter profile-dependent factor, its tightness, and the characterization of $\mathrm{PF}_z = \mathrm{Dev}_z$. The activity assumption is necessary: with inactive voters, $\mathrm{PF}_z/\mathrm{Dev}_z$ can be arbitrarily large (Example F.1).

Approximate coalitional guarantees. The certificate also gives quantitative guarantees without $\boldsymbol{\lambda}$-tightness. If $\Delta_\pi(z) \le \varepsilon_z$ and $B(z) > 0$, then Theorem 6.1 gives $\mathrm{Dev}_z(\pi; \boldsymbol{\lambda}) \le 1 + \varepsilon_z/B(z)$, so a small scalar gap limits the proportional improvement available to coalitions of active voters. When some continuation utilities are zero, an additive guarantee remains: for any coalition S with $b(S) > 0$ and continuation policy σ, some $i \in S$ satisfies $\alpha_z(S)V_i^\sigma(z) - V_i^\pi(z) \le \varepsilon_z \alpha_z(S)/\sum_{j \in S} \lambda_j$. Appendix F.4 gives the full result and a performance-difference identity expressing $\Delta_\pi(z)$ as a sum of local Bellman residuals, computed from the same Bellman pass used for $y^{\boldsymbol{\lambda}}$ and PF_z.

7 Assessing continuation stability with unknown transitions

The certificates of Section 6 assume known transitions. With unknown transitions, both continuation values and the canonical weights $\lambda_i = 1/U_i(\pi)$ are uncertain, so assessment must account for both transitions and proportional claims. In this section, we thus develop a robust certificate for kernels in ambiguity sets and weights in known intervals (Section 7.1). We then construct these sets from sampled transitions and study coverage of the state-alternative pairs that determine future opportunities (Section 7.2). Full coverage is unnecessary for validity, but poor coverage can make certificates uninformative and leave canonical continuation entitlements unidentified.

7.1 Robust assessment under model uncertainty

Suppose the transition kernel lies in a set $\mathcal{P}$ of plausible kernels. For fixed weights, a conservative certificate maximizes its numerator and minimizes its denominator over $\mathcal{P}$; uncertain weights are handled by componentwise bounds $\boldsymbol{\lambda}^- \le \boldsymbol{\lambda}^+$. Let $\mathcal{P} = \prod_{h,s,c} \mathcal{P}_h(s,c)$ be an (s,a)-rectangular ambiguity set, with each $\mathcal{P}_h(s,c) \subseteq \Delta(\mathcal{S}_{h+1})$ nonempty and compact. Define the optimistic numerator $\bar{y}^+(z) = \sup_{Q \in \mathcal{P}} \sup_\sigma \sum_i \lambda_i^+ V_{i,Q}^\sigma(z)$ and pessimistic denominator $\underline{B}^-(z) = \inf_{Q \in \mathcal{P}} \sum_i \lambda_i^- V_{i,Q}^\pi(z)$. Each is computed by one robust backward pass (Lemma G.2).

Definition 7.1 (Robust certificate)**.** Define $\mathrm{Cert}_z(\pi; \boldsymbol{\lambda}^-, \boldsymbol{\lambda}^+, \mathcal{P}) = \bar{y}^+(z)/\underline{B}^-(z)$ when $\underline{B}^-(z) > 0$. If $\underline{B}^-(z) = 0 < \bar{y}^+(z)$, set the certificate to $+\infty$, and if $\underline{B}^-(z) = \bar{y}^+(z) = 0$, set it to 1.

Bounding the numerator and denominator in opposite directions makes the certificate conservative and tractable under rectangular ambiguity. When $\mathcal{P} = \{P\}$ and $\boldsymbol{\lambda}^- = \boldsymbol{\lambda}^+ = \boldsymbol{\lambda}$, it reduces to $\mathrm{PF}_z(\pi; \boldsymbol{\lambda})$. The next result formalizes the guarantee: if the true kernel and weights lie in the specified sets, the certificate upper-bounds PF_z at every state and hence Dev_z wherever it is defined.

Theorem 7.1. *If $P \in \mathcal{P}$ and $\boldsymbol{\lambda}^- \le \boldsymbol{\lambda} \le \boldsymbol{\lambda}^+$, then, at every state z, $\mathrm{PF}_z^P(\pi; \boldsymbol{\lambda}) \le \mathrm{Cert}_z(\pi; \boldsymbol{\lambda}^-, \boldsymbol{\lambda}^+, \mathcal{P})$, and $\mathrm{Dev}_z^P(\pi; \boldsymbol{\lambda}) \le \mathrm{Cert}_z(\pi; \boldsymbol{\lambda}^-, \boldsymbol{\lambda}^+, \mathcal{P})$ whenever $B_{\boldsymbol{\lambda},P}^\pi(z) > 0$.*

See Appendix G.3 for the full proof. The first guarantee holds at every state because PF_z is defined in the extended reals, while the second requires $B_{\boldsymbol{\lambda},P}^\pi(z) > 0$. With fixed weights, Theorem 7.1 gives a robust certificate uniformly over $\mathcal{P}$. Canonical claims require more care: their weights are $\lambda_i^P = 1/U_i^P(\pi)$, so reciprocals computed under an estimated kernel generally correspond to different continuation shares. Moreover, the assumption that each voter can receive positive utility does not ensure it under a given policy. Canonical assessment must therefore establish positive lower bounds on the policy's utilities and bound their reciprocals.

7.2 Assessing continuation stability from sampled transitions

Consider a generative model with m successor-state samples from every transition row $P_h(\cdot \mid s, c)$, utility values in $[0, \bar{a}]$, and known initial distribution. Fix $\delta \in (0,1)$ and let $\hat{P}$ be the empirical kernel. Writing $\mathsf{S} = \max_h |\mathcal{S}_h|$ and $\mathsf{A} = \max_{h,s} |C_h(s)|$, define $\varepsilon_P = \sqrt{2(\mathsf{S}\ln 2 + \ln(H\mathsf{SA}/\delta))/m}$ and let $\hat{\mathcal{P}}$ be the rowwise L_1 ambiguity set of radius ε_P around $\hat{P}$. By Appendix G.7, $P \in \hat{\mathcal{P}}$ with probability at least $1-\delta$. On this event, Lemma G.3 bounds each voter's initial-state error by $\delta_U = \frac{1}{4}\varepsilon_P \bar{a} H(H-1)$. Thus, if $\hat{U}_i(\pi) > \delta_U$ for every voter i, the true canonical weight $\lambda_i^P = 1/U_i^P(\pi)$ satisfies $\lambda_i^- = \frac{1}{\hat{U}_i(\pi)+\delta_U} \le \lambda_i^P \le \frac{1}{\hat{U}_i(\pi)-\delta_U} = \lambda_i^+$. Hence, the same high-probability event places both the true kernel and the true canonical weights in the sets required by Theorem 7.1, yielding the following certificate for the policy's true canonical entitlements.

Theorem 7.2. *With probability at least $1-\delta$, simultaneously for every policy π satisfying $\hat{U}_i(\pi) > \delta_U$ for all i, $\mathrm{PF}_z^P(\pi; \boldsymbol{\lambda}^P) \le \mathrm{Cert}_z(\pi; \boldsymbol{\lambda}^-, \boldsymbol{\lambda}^+, \hat{\mathcal{P}})$ at every state z, and $\mathrm{Dev}_z^P(\pi; \boldsymbol{\lambda}^P) \le \mathrm{Cert}_z(\pi; \boldsymbol{\lambda}^-, \boldsymbol{\lambda}^+, \hat{\mathcal{P}})$ wherever the deviation ratio is defined. At the initial state, $B^{\pi}_{\boldsymbol{\lambda}^P, P}(z_0) = n$ exactly, so the deviation bound uses the population shares $|S|/n$.*

See Appendix G.7 for the full proof. The initial-state identity is exact because $\lambda_i^P = 1/U_i^P(\pi)$ and $V^{\pi}_{i,P}(z_0) = U_i^P(\pi)$ under the same true kernel, so $\lambda_i^P V^{\pi}_{i,P}(z_0) = 1$ for every voter. Hence, there is no approximation error in the population shares. Theorem 7.2 evaluates fractional-Core membership for any positive-utility policy whenever the reported bound is at most 1; it evaluates a given policy rather than learning a Nash-optimal one.

Coverage and informativeness. The preceding result assumes equal sampling of every state-alternative pair. With incomplete coverage, poorly observed rows can receive larger ambiguity sets, preserving validity but reducing informativeness: uncertainty can enlarge possible deviations and, under canonical weights, leave proportional claims unidentified. In some elections, more on-policy data cannot resolve this uncertainty. For every $R \ge 2$, Appendix G.8 constructs two elections that generate identical data under the same behaviour policy at every sample size, yet have canonical deviation ratios 1 and R at a reached state. The elections differ only after an alternative the behaviour policy never selects.

Hence, even unlimited on-policy data may not identify deviations through unchosen alternatives. We thus assign observed transition rows confidence radii based on visit counts and unobserved rows the full successor-state simplex, yielding an ambiguity set valid under arbitrary coverage. For fixed weights, Theorem 7.1 then applies directly. For canonical weights, define $\underline{U}_i = \inf_{Q \in \hat{\mathcal{P}}} U^{\pi}_{i,Q}$ and $\overline{U}_i = \sup_{Q \in \hat{\mathcal{P}}} U^{\pi}_{i,Q}$. If $\underline{U}_i > 0$ for every voter i, then $1/\overline{U}_i \le \lambda_i^P \le 1/\underline{U}_i$; otherwise, the procedure reports INSUFFICIENT COVERAGE. See Appendix G.9 for the full guarantees and procedure.

8 Experiments

Experimental setup. We generate finite layered Markov temporal elections from seeded distributions. The default model uses binary approvals, independently equal to 1 with probability 0.35; one family instead uses independent $U[0,1]$ cardinal values. Transition rows use diffuse or concentrated Dirichlet distributions. Analyses requiring the exact deviation ratio use five small-election families with $n \in \{2,3,4\}$ and $H \in \{2,3\}$, varying the reward model, layer width, number of alternatives, or transition concentration. Each contains 115–120 eligible elections, for 590 total. See Appendices H.1 and H.5 for the generator, eligibility criteria, and parameter grids.

Policies. Most analyses hold the evaluated policy fixed while varying the feature of interest. We use a reproducible positive-utility reference policy π, uniform over alternatives whenever this gives every voter positive ex-ante utility and otherwise drawn from a seeded Dirichlet distribution. The experiment on action dependence uses the uniform policy to keep its policy-averaged transition kernel fixed as action dependence varies. There, we use the canonical reciprocal weights $\lambda_i = 1/U_i(\pi)$, which give $b_i(z_0) = 1$ and recover the population shares $|S|/n$ at the initial state. Another analysis computes a Nash-welfare optimum in each known election and evaluates policies along a path from the reference policy to this endpoint, which serves as a theoretical benchmark rather than a policy-learning algorithm. Further details appear in Appendices H.2 and H.11.

Quantities evaluated. We compare the exact deviation ratio Dev_z, scalar certificate PF_z, and, under unknown transitions, robust certificate Cert_z. Computing Dev_z enumerates active coalitions and solves a continuation linear program for each (Appendix H.3), whereas PF_z requires one policy evaluation and scalar Bellman pass. The robust certificate uses robust backward dynamic programming over an estimated kernel and ambiguity set (Appendix H.4). Exact-deviation experiments thus use small elections, while robust verification is also evaluated on larger instances.

The efficiently computable certificate is substantially tighter than its worst-case guarantee. Appendix F.3 gives $\mathrm{Dev}_z \le \mathrm{PF}_z \le L_z \mathrm{Dev}_z$, where L_z depends on the continuation-entitlement profile. Across the five families, PF_z exceeds Dev_z by less than 13% at the median, while L_z is roughly two (Table 1, Appendix H.9); this factor is attained exactly by Proposition F.1. Equality $\mathrm{PF}_z = \mathrm{Dev}_z$ occurs at 23–29$\%$ of states in the binary-approval families and none in the cardinal family, although their different active-set structures mean that this contrast cannot be attributed to the reward model alone.

Action-dependent transitions increase the deviations available to coalitions against a fixed policy. For each seed, we hold fixed the graph, approvals, and candidate successor distributions and interpolate $P_\theta(\cdot \mid s, c) = (1-\theta)\bar{q}_s + \theta q_{s,c}$ from action-independent ($\theta = 0$) to fully action-dependent ($\theta = 1$) transitions (Appendix H.10). Under the uniform reference policy, the policy-averaged kernel remains $\bar{q}_s$, so continuation utilities, canonical weights, and entitlements are unchanged. Changes in Dev_z therefore reflect only the additional opportunities to influence transitions. At $\theta = 1$, the per-election average deviation ratio increases by a median of 0.145 across 40 paired elections, with the middle half between 0.096 and 0.178 and all 40 increasing (Figure 2). Controls show that the effect vanishes when actions cannot change future opportunities, grows with the remaining horizon, and is much smaller when states differ only in realized approvals (Appendix H.10).

Coalitional deviations become small well before all-strict blocks disappear. For each of the 590 elections, we follow an occupancy path from the reference policy to a numerically Nash-optimal endpoint. The interpolation parameter $t \in [0, 1]$ indexes this path, with $t = 0$ the reference policy and $t = 1$ the Nash endpoint; canonical weights are recomputed at each evaluated t (Appendix H.11, Figure 3). At the endpoint, Lemma D.3 and theorem 5.1 imply $\mathrm{Dev}_z = 1$ wherever $B(z) > 0$. At the initial state, the deviation ratio falls to a median of 1.025 by $t = 0.9$ and never increases between consecutive evaluated points; the certificate also follows the exact ratio more closely. Yet, at $t = 0.9$, only 4 of 590 policies have $\mathrm{Dev}_z = 1$ at every state with $B(z) > 0$, and 3 already did at $t = 0$.

Coverage, rather than sample count alone, determines how informative the certificates are. On 27 fixed elections, sparse behaviour-policy sampling leaves four of eight transition rows unobserved, while uniform sampling increasingly resolves all rows (Appendix H.12). The certificate under sparse sampling exceeds the one under uniform sampling by a paired median of 0.419, 0.646, and 0.754 at 800, 4000, and 16000 samples (Figure 4). Coverage also affects canonical claims: using the true weights removes 58–66% of the certificate's excess over the true $\mathrm{PF}^P_{z_0}$ at the median and yields a finite certificate in 13 of 675 cases otherwise reporting INSUFFICIENT COVERAGE (Appendix H.12.1). Finally, we implement the indistinguishability construction of Appendix G.8: unchosen alternatives remain unresolved with more data from the same behaviour policy, and indistinguishable elections can have deviation ratios 1 and $H - 1$ (Figure 5, Appendix H.12.2).

9 Conclusion and Future Work

We introduced Markov temporal elections for settings where current decisions reshape future representation opportunities. We first proved that reusing initial population shares at continuation states can be infeasible or severely inefficient. Instead, we derived state-dependent claims that track remaining opportunities and uniquely select Nash welfare among separable welfare objectives. Nash-optimal policies can be completed to guarantee statewise protection against coalitional deviations. Since exact coalitional verification is computationally hard, we developed an efficient certificate for these guarantees that extends to unknown transitions and uncertain claims. Experiments show that the certificate closely tracks exact deviations, while action dependence enlarges deviation opportunities and limited coverage makes the certificate less informative. Open questions include the complexity of unrestricted fractional-Core membership, tighter welfare-cost bounds, and data collection methods for evaluating rarely chosen alternatives. Future work should also extend our framework to richer temporal constraints, evolving preferences, and larger state spaces.

AI USE STATEMENT

Generative AI tools were used to assist with writing and presentation of the manuscript, identifying potentially relevant related work, generating synthetic experimental instances, and generating and editing code for the experiments. The synthetic election generator and experimental parameter regimes are described in Appendices H.1 and H.5, respectively. The authors consulted the original sources and independently verified all citations and characterizations of prior work. All AI-assisted text was reviewed and revised by the authors, and all AI-assisted instances, code, and reported results were independently checked.

Generative AI was not used to develop the theoretical results or proofs. The indistinguishability experiment in Appendix H.12.2 instantiates the construction underlying Proposition G.1; the impossibility result itself is established theoretically and is independent of the experimental instantiation. All theoretical results and proofs were developed and independently verified by the authors. All empirical analyses, interpretations, and scientific conclusions are the authors' own, and the authors take full responsibility for the content of the paper.

ETHICS STATEMENT

This work is theoretical and uses only synthetic elections; it involves no human participants, personal data, or real-world deployment. The proposed proportionality notions formalize one approach to representation in sequential collective decisions and should not be interpreted as prescribing how representation or entitlements should be defined in a particular application. Real-world uses may involve additional legal, institutional, and distributive considerations not captured by the model, including how preferences and voter groups are defined and whose interests are represented. Applying the framework in such settings would therefore require application-specific assessment and engagement with the affected stakeholders.

REPRODUCIBILITY STATEMENT

The assumptions, definitions, theorem statements, and complete proofs are provided in the main text and Appendices B–G. Appendix H specifies the experimental methodology, including the synthetic election generators, parameter grids, policy construction, computation of exact deviations and certificates, evaluation metrics, controls, and additional results. Appendix H.15 specifies the computational environment and cost and maps the reported tables and figures to their experimental outputs.

We retain the complete experimental configuration, including parameter settings, random seeds, and experimental regimes, together with the generated outputs, intermediate results, implementation and validation code, and figure-generation scripts. All stochastic components are reproducibly seeded, and matched comparisons use the same underlying election instances. These materials, together with the documented computational environment, permit reproduction of the reported computational analyses without external datasets or data preprocessing.

REFERENCES

Parand A. Alamdari, Soroush Ebadian, and Ariel D. Procaccia. Policy aggregation. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Charalambos D Aliprantis and Kim C Border. *Infinite dimensional analysis: a hitchhiker's guide*. Springer, 2006.

Eitan Altman. *Constrained Markov Decision Processes*. Chapman and Hall/CRC, 1999.

Edoardo Amaldi and Viggo Kann. The complexity and approximability of finding maximum feasible subsystems of linear relations. *Theoretical Computer Science*, 147(1–2):181–210, 1995.

Konstantin Avrachenkov, Laura Cottatellucci, and Lorenzo Maggi. Cooperative Markov decision processes: Time consistency, greedy players satisfaction, and cooperation maintenance. *International Journal of Game Theory*, 42(1):239–262, 2013.

Siddhartha Banerjee, Vasilis Gkatzelis, Safwan Hossain, Billy Jin, Evi Micha, and Nisarg Shah. Proportionally fair online allocation of public goods with predictions. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence*, pp. 20–28, 2023.

Carter Blair and Kate Larson. Generating fair consensus statements with social choice on token-level MDPs. In *Proceedings of the 25th International Conference on Autonomous Agents and Multiagent Systems*, pp. 1692–1700, 2026.

Markus Brill, Evangelos Markakis, Georgios Papasotiropoulos, and Jannik Peters. Proportionality guarantees in elections with interdependent issues. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence*, pp. 2537–2545, 2023.

Laurent Bulteau, Noam Hazon, Rutvik Page, Ariel Rosenfeld, and Nimrod Talmon. Justified representation for perpetual voting. *IEEE Access*, 9:96598–96612, 2021.

Nikhil Chandak, Shashwat Goel, and Dominik Peters. Proportional aggregation of preferences for sequential decision making. *Journal of Artificial Intelligence Research*, 85, 2026.

Julian Chingoma, Umberto Grandi, and Arianna Novaro. Proportionality for constrained public decisions. *arXiv preprint arXiv:2409.02609*, 2024.

Vincent Conitzer, Rupert Freeman, and Nisarg Shah. Fair public decision making. In *Proceedings of the 18th ACM Conference on Economics and Computation*, pp. 629–646, 2017.

Cyrus Cousins, Kavosh Asadi, Elita Lobo, and Michael L. Littman. On welfare-centric fair reinforcement learning. *Reinforcement Learning Journal*, 3:1124–1137, 2024.

Gabriel Cramer. *Introduction à l'analyse des lignes courbes algébriques*. Chez les frères Cramer et C. Philibert, 1750.

Cyrus Derman and Ralph E. Strauch. A note on memoryless rules for controlling sequential control processes. *The Annals of Mathematical Statistics*, 37(1):276–278, 1966.

Soroush Ebadian, Anson Kahng, Dominik Peters, and Nisarg Shah. Optimized distortion and proportional fairness in voting. *ACM Transactions on Economics and Computation*, 12(1):1–39, 2024.

Edith Elkind, Svetlana Obraztsova, and Nicholas Teh. Temporal fairness in multiwinner voting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 22633–22640, 2024.

Edith Elkind, Svetlana Obraztsova, Jannik Peters, and Nicholas Teh. Verifying proportionality in temporal voting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 13805–13813, 2025.

Brandon Fain, Ashish Goel, and Kamesh Munagala. The core of the participatory budgeting problem. In *Web and Internet Economics (WINE)*, pp. 384–399, 2016.

Brandon Fain, Kamesh Munagala, and Nisarg Shah. Fair allocation of indivisible public goods. In *Proceedings of the 19th ACM Conference on Economics and Computation*, pp. 575–592, 2018.

Julius Farkas. Theorie der einfachen ungleichungen. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1902(124):1–27, 1902.

Rupert Freeman, Seyed Majid Zahedi, and Vincent Conitzer. Fair and efficient social choice in dynamic settings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 4580–4587, 2017.

Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

Garud N. Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2): 257–280, 2005.

Thomas Jaksch, Ronald Ortner, and Peter Auer. Near-optimal regret bounds for reinforcement learning. *Journal of Machine Learning Research*, 11:1563–1600, 2010.

Ying Jin, Zhuoran Yang, and Zhaoran Wang. Is pessimism provably efficient for offline RL? In *Proceedings of the 38th International Conference on Machine Learning*, pp. 5084–5096, 2021.

David S. Johnson and Franco P. Preparata. The densest hemisphere problem. *Theoretical Computer Science*, 6(1):93–107, 1978.

Peizhong Ju, Arnob Ghosh, and Ness B. Shroff. Achieving fairness in multi-agent MDP using reinforcement learning. In *International Conference on Learning Representations*, 2024.

Sham Kakade and John Langford. Approximately optimal approximate reinforcement learning. In *Proceedings of the 19th International Conference on Machine Learning*, pp. 267–274, 2002.

Rahul Kidambi, Aravind Rajeswaran, Praneeth Netrapalli, and Thorsten Joachims. MOReL: Model-based offline reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Woosung Kim, Jinho Lee, Jongmin Lee, and Byung-Jun Lee. FairDICE: Fairness-driven offline multi-objective reinforcement learning. In *Proceedings of the 39th Conference on Neural Information Processing Systems (NeurIPS)*, 2025.

Antoni A Kosinski. Cramer's rule is due to cramer. *Mathematics Magazine*, 74(4):310–312, 2001.

Martin Lackner. Perpetual voting: Fairness in long-term decision making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 2103–2110, 2020.

Martin Lackner and Jan Maly. Proportional decisions in perpetual voting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 5722–5729, 2023.

Simon Lacoste-Julien and Martin Jaggi. On the global linear convergence of frank-wolfe optimization variants. *Advances in neural information processing systems*, 28:496–504, 2015.

Debmalya Mandal and Jiarui Gan. Socially fair reinforcement learning. *arXiv preprint arXiv:2208.12584*, 2022.

Tomáš Masařík, Grzegorz Pierczyński, and Piotr Skowron. A generalised theory of proportionality in collective decision making. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pp. 734–754, 2024.

Blossom Metevier, Stephen Giguere, Sarah Brockman, Ari Kobren, Yuriy Brun, Emma Brunskill, and Philip S. Thomas. Offline contextual bandits with high probability fairness guarantees. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Arnab Nilim and Laurent El Ghaoui. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.

Elena Parilina and Georges Zaccour. Node-consistent core for games played over event trees. *Automatica*, 53:304–311, 2015.

David C. Parkes and Ariel D. Procaccia. Dynamic social choice with evolving preferences. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 27, pp. 767–773, 2013.

Dominik Peters, Grzegorz Pierczyński, Nisarg Shah, and Piotr Skowron. Market-based explanations of collective decisions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 5656–5663, 2021a.

Dominik Peters, Grzegorz Pierczyński, and Piotr Skowron. Proportional participatory budgeting with additive utilities. In *Advances in Neural Information Processing Systems*, volume 34, pp. 12726–12737, 2021b.

Marek Petrik, Mohammad Ghavamzadeh, and Yinlam Chow. Safe policy improvement by minimizing robust baseline regret. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Bradley Phillips, Edith Elkind, Nicholas Teh, and Tomasz Wąs. Strengthening proportionality in temporal voting. In *Proceedings of the 25th International Conference on Autonomous Agents and Multiagent Systems*, pp. 3823–3832, 2026.

Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, 1994.

Aviv Rosenberg and Yishay Mansour. Online convex optimization in adversarial Markov decision processes. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 5478–5486, 2019.

Alexander Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1986.

Nicholas Teh. The price of proportional representation in temporal voting. In *Proceedings of the 35th International Joint Conference on Artificial Intelligence*, 2026.

Tsachy Weissman, Erik Ordentlich, Gadiel Seroussi, Sergio Verdú, and Marcelo J. Weinberger. Inequalities for the L1 deviation of the empirical distribution. Technical Report HPL-2003-97R1, Hewlett-Packard Laboratories, 2003.

Wolfram Wiesemann, Daniel Kuhn, and Berç Rustem. Robust Markov decision processes. *Mathematics of Operations Research*, 38(1):153–183, 2013.

Yuancheng Xu, Chenghao Deng, Yanchao Sun, Ruijie Zheng, Xiyao Wang, Jieyu Zhao, and Furong Huang. Adapting static fairness to sequential decision-making: Bias mitigation strategies towards equal long-term benefit rate. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 54962–54982, 2024.

Chongjie Zhang and Julie A. Shah. Fairness in multi-agent sequential decision-making. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, volume 27, pp. 2636–2644, 2014.

A Extended related work

This appendix provides a more detailed comparison with work on temporal voting, constrained collective choice, social choice in sequential decision problems, and fairness in reinforcement learning.

A.1 Temporal and perpetual voting

Perpetual voting was introduced as a framework for long-term collective decision making in which the same electorate participates in a sequence of approval-based elections. The central idea is to use information from earlier rounds to compensate voters who have been repeatedly dissatisfied, giving temporal fairness guarantees that may be impossible in a single election (Lackner, 2020). Subsequent work studies proportionality more directly. Lackner & Maly (2023) characterize perpetual voting rules with proportionality guarantees and compare different approaches to distributing influence over time. These models are temporal in the sense that past decisions affect how the rule treats voters in later rounds, but the set of alternatives available in a round is not generated by an action-dependent stochastic state process.

A particularly close precursor is Freeman et al. (2017), who study repeated social choice with cardinal utilities and use long-run Nash welfare to balance fairness and efficiency across decisions. Their model, like perpetual voting, uses repeated decisions to compensate agents over time, but the current choice does not change a stochastic state that determines which alternatives become available later. Our use of Nash welfare therefore addresses a different issue: not only how utility should be balanced across a sequence of decisions, but how proportional claims should change when the decisions themselves reshape the future opportunities for representation.

Temporal voting develops the same long-run perspective while adapting familiar approval-based representation axioms. Chandak et al. (2026) introduce temporal versions of justified representation (JR), proportional justified representation (PJR), and extended justified representation (EJR). Their framework treats the election as a sequence of approval-based decisions and evaluates whether voters

that are sufficiently cohesive over time receive adequate representation. Phillips et al. (2026) go beyond these basic adaptations by introducing stronger notions, including EJR+, full JR (FJR), full PJR (FPJR), and several temporal variants of Core stability. Phillips et al. (2026) then study their relationships and existence. Thus, the literature already contains a rich family of proportionality notions for repeated collective decisions.

The setting in those papers nevertheless differs from ours in how decisions depend on one another over time. In standard temporal voting, the available alternatives in future rounds are specified independently of the alternative selected in the current round, even when the fairness rule itself depends on the history. A broader survey by Elkind et al. (2024) reviews temporal models in which decisions may constrain later choices, for example by limiting how often a project can be selected, restricting how frequently committee membership can change, or imposing other temporal feasibility conditions. These are constraints on the sequence of selected alternatives. In our model, by contrast, the selected alternative changes the state of the process and hence changes the distribution over future alternatives. What matters is therefore not only which alternatives have already been selected, but which representation opportunities remain conditional on the current state.

Relatedly, Banerjee et al. (2023) study proportional fairness in an online public-goods problem in which goods arrive sequentially and current allocations must be made before all future values are known. Their results show how information about future opportunities can improve proportional fairness. The temporal uncertainty is different from ours, however: future goods arrive independently of current choices, whereas in a Markov temporal election the current collective choice changes the distribution of future opportunities themselves.

This distinction matters for proportionality. A history-dependent feasibility constraint can prevent some sequences of choices, but the proportional claim assigned to a coalition can remain a fixed population fraction. In a Markov temporal election, an earlier decision can instead change the future opportunity set itself. Two states reached after different decisions can therefore give different voters very different remaining opportunities, creating the need for state-dependent proportional claims. This is the feature that our continuation entitlements are designed to capture.

The complexity and welfare consequences of temporal proportionality have also been studied. Elkind et al. (2025) investigate the complexity of verifying temporal proportionality properties, while Teh (2026) studies the utilitarian welfare cost of temporal proportionality. These works therefore address questions that arise once a temporal proportionality notion has been fixed. Our focus is complementary: we first ask what the proportional claim should be when the future depends on the current decision, and then study the guarantees and efficiently computable certificates that follow from that claim.

A.2 Static feasibility constraints and interdependent collective choice

There is also a growing literature on proportionality when the feasible alternatives are subject to dependencies. Brill et al. (2023) study approval-based voting with interdependent alternatives, where whether an alternative can be selected may depend on other choices. Chingoma et al. (2024) develop proportionality notions for constrained approval-based voting and study the effect of feasibility dependencies on proportional representation. Masařík et al. (2024) provide a general feasibility-based framework in which generalized proportionality properties of Proportional Approval Voting and Phragmén's sequential rule are characterized by matroid structure.

These papers are close in spirit because they relax the assumption that every subset of alternatives is feasible. However, their feasibility relation is static: one instance specifies a common ground set together with a collection of feasible sets. By contrast, our model is dynamic. There need not be a common alternative set at all, and which alternatives can be considered later is determined by the current state and the transition generated by the selected alternative. The distinction is therefore between static dependence among feasible outcomes and stochastic dependence of future opportunities on current actions.

The difference also affects what proportionality must track. In a static constrained election, a coalition's claim can be evaluated relative to a fixed feasible-choice system. In a Markov temporal election, the feasible future itself changes as the election evolves. Our state-dependent claims therefore

measure the representation opportunities remaining at a state rather than only the coalition's position in a static feasible set.

A.3 Nash welfare, the Core, and proportional fairness

Our use of Nash welfare builds on a well-established connection between Nash welfare and Core stability in divisible collective decision problems. Conitzer et al. (2017) develop the public-decision framework in which voters have cardinal utilities over alternatives, while Fain et al. (2016) study Core-based notions in participatory budgeting. Fain et al. (2018) establish a direct connection between Nash-welfare optimization and Core-type guarantees for divisible public decisions. We use this classical first-order argument in the root case rather than claim it as a new result. Our contribution is to extend the same logic to continuation states whose available alternatives depend on the preceding collective decisions, using state-dependent claims derived from the policy rather than a fixed population fraction.

A related line of work connects proportional fairness to Core guarantees. Ebadian et al. (2024) study proportional fairness in randomized social choice and show that a proportional-fairness bound implies a corresponding universal approximate-Core guarantee. At the initial state, under our canonical normalization, the scalar quantity used in our certificate from Section 6 reduces to the same proportional-fairness ratio: it compares the best average proportional improvement with the utilities of the current outcome. Our result differs in two ways. First, we apply the comparison at every continuation state, where the relevant shares depend on the policy rather than being fixed population fractions. Second, we characterize the gap between this efficiently computable scalar quantity and the strongest coalition deviation, including a tight bound that depends on the distribution of continuation entitlements.

Price-based interpretations of proportionality are another relevant connection. Peters et al. (2021a) develop market-based characterizations of proportionality in multiwinner elections, and Peters et al. (2021b) study analogous price systems in participatory budgeting. Our state-dependent prices have the same basic role of justifying a proportional outcome through local prices, but they must additionally account for the future consequences of an alternative. In a Markov temporal election, the price of an alternative therefore contains both its immediate value and the continuation value it makes possible. This yields the state-dependent price characterization of Nash welfare proved in our Section 5.

A.4 Social choice and sequential decision making

Several papers explicitly combine social choice with Markov decision processes. Parkes & Procaccia (2013) formulate dynamic social choice as an MDP in which the state is a preference profile and a deterministic stationary policy corresponds to a social choice function. Their work uses social-choice axioms to constrain the policy and develops algorithms that use symmetries among preference profiles. Our model starts from a different source of state dependence: the state is not itself a preference profile, but a representation of the current voting opportunity set, and the selected alternative changes the stochastic distribution of future opportunities. Our central question is therefore not how a social-choice rule should react to evolving preferences, but how proportional claims should evolve when collective choices reshape the future decision problem.

Alamdari et al. (2024) study policy aggregation in an MDP by representing policies through the state-action occupancy polytope and applying social-choice rules, including approval-based and proportional-veto-Core approaches, in the policy-aggregation setting. This work is closely related to our use of occupancy measures and scalarized MDP objectives. The distinction is the source and role of the sequential structure: policy aggregation treats the occupancy polytope as a representation of complete policies and applies a social-choice rule to them, whereas our social-choice problem is temporal and asks for proportional protection after conditioning on states whose future opportunities are themselves action-dependent.

Blair & Larson (2026) provide another close MDP-based comparison. They formulate consensus generation over token-level MDPs and use Nash welfare to construct a stochastic generation policy with an ex-ante Core guarantee. The common feature is the combination of Nash welfare, MDP structure, and coalitional stability. The difference is the level at which stability is required: their

Core guarantee is an ex-ante statement about complete trajectories from the initial state, whereas our main object is statewise stability after arbitrary continuation states have been reached. This distinction is why we need a state-dependent entitlement rule.

A.5 Fairness in reinforcement learning

Fairness objectives in MDPs predate the recent fair-reinforcement-learning literature. Zhang & Shah (2014) study multi-agent sequential decision making in which agents have distinct local interests and optimize a fairness criterion that emphasizes the worst-performing agents while retaining concern for overall performance. This work already treats fairness as an objective over long-run outcomes of an MDP. Our setting differs in the fairness notion: rather than optimizing a fixed welfare or fairness criterion across agents, we derive state-dependent proportional claims and ask whether a given policy is protected against coalitional deviations under those claims.

A related long-term-fairness literature emphasizes that applying a static fairness criterion separately at each decision point can be inappropriate when current actions affect future populations. For example, Xu et al. (2024) formulate long-term group fairness in an MDP and account explicitly for the future effects of current decisions. The concern is conceptually related to ours, but the fairness object is different: they study disparities between demographic groups, whereas we derive coalitional proportional claims from the representation opportunities remaining at each state.

More recent work studies how such welfare and fairness objectives can be optimized through reinforcement learning. Mandal & Gan (2022) consider socially fair policy optimization in unknown MDPs and study welfare objectives including Nash welfare. Ju et al. (2024) develop fairness-oriented policy optimization in reinforcement learning, Cousins et al. (2024) develop a welfare-based framework for fair reinforcement learning with explicit welfare objectives, and Kim et al. (2025) introduce FairDICE for offline multi-objective reinforcement learning with nonlinear welfare objectives such as Nash welfare.

These papers primarily ask how to *learn or optimize* a policy for a specified fairness or welfare objective. We instead study *assessment* of a given policy. In particular, our certificate concerns a coalitional property that must hold at every state, and under the canonical normalization the weights defining that property are themselves functions of the policy's utility under the unknown transition kernel. Thus, uncertainty affects not only the value of the certificate but also the proportional claims that the certificate is meant to protect. This distinction is especially important in offline settings. A policy can be evaluated or optimized under a learned model while still leaving its state-dependent proportional claims poorly identified. Our statistical results thus treat uncertainty in the transition kernel and uncertainty in the induced entitlement weights jointly. We do not present a new method for learning Nash-welfare-optimal policies.

A.6 Dynamic cooperative games

State-indexed stability concepts also appear in dynamic cooperative games. Avrachenkov et al. (2013) study cooperative Markov decision processes and develop time-consistency notions for distributing cooperative value across states. Parilina & Zaccour (2015) study node-consistent Core allocations for cooperative games played over event trees. Both works are relevant because they ask whether a cooperative solution remains appropriate after the stochastic process reaches a later node.

The underlying cooperative object is nevertheless different. These papers formulate transferable-utility games in which a joint payoff can be distributed among players, whereas our voter utilities are non-transferable and arise directly from the sequence of selected alternatives. Our state-dependent shares also depend on the policy: they are derived from the continuation utilities of the policy being evaluated. The question is therefore not how to redistribute a fixed cooperative surplus at different nodes, but how the proportional entitlement itself should change when the chosen actions alter future representation opportunities.

A.7 Robust MDPs and offline verification

The robust and statistical parts of our analysis use standard tools from robust and offline reinforcement learning. Robust MDPs with uncertainty sets over transition kernels were developed by Iyen-

gar (2005) and Nilim & El Ghaoui (2005), and Wiesemann et al. (2013) provide a general treatment of rectangular uncertainty sets and efficient robust dynamic programming. We rely on these ideas only as computational tools. Our contribution is the particular coalitional quantity that is robustly bounded.

Several offline-RL methods also use uncertainty to obtain conservative guarantees. Petrik et al. (2016) study safe policy improvement relative to a baseline under uncertain transition dynamics, emphasizing the need to account jointly for model uncertainty and the baseline policy. Kidambi et al. (2020) develop MOReL, which constructs a pessimistic model from offline data so that optimization remains conservative at state-action pairs with limited data. Jin et al. (2021) develop a general pessimistic framework for offline reinforcement learning. These methods motivate the use of ambiguity sets and pessimistic evaluation in our unknown-dynamics analysis, but they do not address a coalitional proportionality constraint whose defining weights depend on the unknown value of the policy itself.

Finally, Metevier et al. (2019) study high-probability verification of fairness constraints from offline contextual-bandit data. Their problem is closer to our statistical goal because the objective is to verify a fairness property from data rather than only to optimize it. The distinction is that their fairness constraint is specified independently of the unknown environment dynamics, whereas our canonical proportional claims depend on the policy's own utilities and therefore on the transition kernel. As a result, our confidence procedure must account simultaneously for uncertainty in the environment and uncertainty in the claims themselves. This dependence on the unknown kernel also explains the effect of coverage studied in Section 7: validity can be guaranteed without coverage assumptions, while obtaining a finite canonical certificate may require enough information to identify every voter's positive utility.

B Occupancy measures and randomized Markov policies

The main text represents policies through their state-action occupancies. This appendix describes the standard flow constraints that define this representation and explains why randomized Markov policies can attain every feasible utility profile. These facts also give the convexity and compactness properties used throughout the paper.

Lemma B.1 (Occupancy representation and Markov sufficiency)**.** *Let $\mathcal{X}$ be the set of nonnegative vectors $x = (x_h(s,c))_{0\le h\le H}$ satisfying the initial condition $x_0(s_0,c_0) = 1$, the round-1 flow constraints $\sum_c x_1(s,c) = P_0(s \mid s_0,c_0) = \mu(s)$ for every $s \in \mathcal{S}_1$, and, for $1 \le h \le H-1$, the flow constraints $\sum_c x_{h+1}(s',c) = \sum_{s,c} x_h(s,c)P_h(s' \mid s,c)$.*

Then, $\mathcal{X}$ is exactly the set of state-action occupancy vectors induced by randomized Markov policies. In particular, every $\mathbf{x} \in \mathcal{X}$ is induced by the policy $\pi_h(c \mid s) = x_h(s,c)/\sum_{c'} x_h(s,c')$ whenever the denominator is positive, with the policy chosen arbitrarily at states where it is zero. Moreover, $\mathcal{X}$ is a polytope, and the attainable utility set $\mathcal{U} = \{(\langle x, a_i\rangle)_i : x \in \mathcal{X}\}$ is a compact convex polytope. The attainable utility set is unchanged if randomized Markov policies are replaced by arbitrary history-dependent randomized policies.

Proof. We first characterize the occupancy vectors. For any randomized Markov policy π, its occupancy x^π satisfies the stated constraints by the law of total probability: the initial flow is μ, and the probability of reaching a state in round $h+1$ is obtained by multiplying the round-h state-action occupancies by the corresponding transition probabilities and summing.

Conversely, fix $\mathbf{x} \in \mathcal{X}$ and define π from x as in the statement. We show by forward induction that $x^\pi = x$. At the artificial root, $x_0(s_0,c_0) = 1$. Suppose at round h that the probability of reaching state s under π is $d_h^\pi(s) = \sum_c x_h(s,c)$. If this quantity is positive, then

$$x_h^\pi(s,c) = d_h^\pi(s)\pi_h(c \mid s) = \left(\sum_{c'} x_h(s,c')\right) \frac{x_h(s,c)}{\sum_{c'} x_h(s,c')} = x_h(s,c).$$

If $d_h^\pi(s) = 0$, nonnegativity implies $x_h(s,c) = 0$ for every c, so the same equality holds regardless of how π is defined at that state. The flow constraints then imply that the reach probabilities at round $h+1$ also agree. Induction therefore gives $x^\pi = x$ at every round.

The set $\mathcal{X}$ is defined by finitely many linear equalities and inequalities, and is bounded because its coordinates are probabilities. Hence, it is a polytope and, in particular, compact and convex. Each voter's utility is linear in the occupancy vector, $U_i(x) = \langle x, a_i \rangle$, so the attainable utility set $\mathcal{U}$ is the linear image of $\mathcal{X}$ and is therefore also a compact convex polytope. None of these claims requires rational transition probabilities.

No rationality assumption is needed for the polytope or utility-set conclusions above, and none is imposed in the general model. Rationality is required only under the input encoding convention of Section 6. Under that convention, the transition kernels and initial distribution μ are rational, so the flow constraints have rational coefficients and can be written with encoding length polynomial in that of the election.

Finally, allowing history dependence does not enlarge the attainable utility set. By the Derman–Strauch result, every history-dependent randomized policy has the same state-action marginals at each round as some randomized Markov policy (Derman & Strauch, 1966; Puterman, 1994; Altman, 1999). Since every U_i depends on a policy only through these marginals, the two policy classes attain the same utility profiles. The layered occupancy-polytope formulation follows Rosenberg & Mansour (2019). □

Fractional scaling. The scaling in Definition 2.1 applies to utilities, not to the deviating policy. In particular, σ may be any policy in Π; the coalition faces no additional feasibility or budget constraint. The occupancy representation makes this explicit: since utility is linear in occupancy, $\alpha(S)U_i(\sigma) = \langle \alpha(S)x^\sigma, a_i \rangle$, but for $\alpha(S) < 1$, the scaled occupancy $\alpha(S)x^\sigma$ is not feasible because its initial flow is $\alpha(S)$ rather than 1. Thus, the fractional scaling is applied only to utilities, not to the occupancy of a feasible policy.

C Omitted Proofs for Section 3

C.1 Proof of Theorem 3.1

Theorem 3.1. *For every $n \geq 2$ and integer $K \geq 2$, there is a Markov temporal election with deterministic transitions, approval preferences, $H = K + 2$ rounds, and $O(n^2K)$ states in which no policy is simultaneously in the fractional Core at the initial state and in the fractional Core with the original shares $|S|/n$ at every state reached with positive probability.*

Proof. We construct the election first. Round 1 has a single state with alternatives $g_1, \ldots, g_n$, all of which have zero approval utility. Selecting g_j leads deterministically to a state s_j in round 2. At s_j, there are again n zero-approval alternatives $c_1, \ldots, c_n$. Selecting c_i starts a deterministic chain of forced rounds in which only voter i receives positive utility. The chain gives voter i continuation utility K if $i = j$ and 1 otherwise; every other voter receives zero continuation utility. Equivalently, if $K_i = K$ for $i = j$ and $K_i = 1$ otherwise, selecting c_i at s_j gives voter i continuation utility K_i and every other voter utility 0. The construction has $1 + n + n^2K + 1 = O(n^2K)$ states and horizon $H = K + 2$.

Suppose that a policy π satisfies both the fractional Core at the initial state and the fractional Core with the original population shares at every state it reaches. Let q_j be the probability that π selects g_j in round 1. For every state s_j reached with positive probability, let $p_{j,i}$ be the probability that π selects c_i there.

Consider such a reached state s_j. By selecting c_i, the singleton coalition $\{i\}$ can obtain continuation utility K_i. Under π, its continuation utility is $p_{j,i}K_i$. Since the singleton's population share is $1/n$, statewise proportionality requires $p_{j,i}K_i \geq K_i/n$. Since $K_i > 0$, this implies $p_{j,i} \geq 1/n$ for every i. The probabilities $p_{j,1}, \ldots, p_{j,n}$ sum to one, so in fact $p_{j,i} = 1/n$ for every i. Thus, at every reached continuation state, the statewise requirement forces the uniform lottery over $c_1, \ldots, c_n$.

We now examine the initial-state requirement. If s_i is reached, the uniform continuation lottery gives voter i expected utility K/n; otherwise, if any other state s_j is reached, it gives voter i expected utility $1/n$. Therefore,

$$U_i(\pi) = q_i \frac{K}{n} + (1 - q_i)\frac{1}{n}.$$

This expression also covers alternatives g_j selected with probability zero, since their corresponding states make no contribution to the ex-ante expectation.

At the initial state, singleton $\{i\}$ can deviate by selecting g_i and then c_i, obtaining utility K. Its population share is again $1/n$, so the fractional-Core condition requires $U_i(\pi) \geq K/n$. Substituting the expression above yields

$$1 + (K-1)q_i \geq K.$$

Since $K \geq 2$, this implies $q_i \geq 1$. The same argument applies to every $i \in N$, but the probabilities $q_1, \ldots, q_n$ must sum to one. For $n \geq 2$, they therefore cannot all satisfy $q_i \geq 1$, which gives the desired contradiction. □

C.2 Proof of Theorem 3.2

Theorem 3.2. *For every $n \geq 2$ and $K \geq 2$, there is a stochastic approval election in which imposing the original shares at every continuation state forces the unique utility profile $U_i = (K+n-1)/n^2$ for every voter i, while another policy gives every voter K/n and lies in the fractional Core at the initial state. The ratio between these utilities is $nK/(K+n-1)$, which approaches n as $K \to \infty$.*

Proof. We modify the construction of Theorem 3.1 only at the initial decision. Round 1 now has a single zero-approval alternative that leads to each state s_j with probability $1/n$. Rounds $2, \ldots, K+2$ are unchanged. Thus, every s_j is reached with positive probability.

As in the proof of Theorem 3.1, imposing the original singleton shares at s_j forces $p_{j,i} \geq 1/n$ for every i, and hence $p_{j,i} = 1/n$. Therefore, the continuation lottery is uniform at every s_j. For voter i, state s_i contributes expected utility K/n, while each of the other $n-1$ states contributes $1/n$. Since every s_j is reached with probability $1/n$,

$$U_i = \frac{1}{n} \cdot \frac{K}{n} + \frac{n-1}{n} \cdot \frac{1}{n} = \frac{K+n-1}{n^2}.$$

The fixed-share requirement therefore forces this unique utility profile.

Now, consider the policy that selects c_j at state s_j. Each voter then receives utility K/n. This strictly exceeds the forced utility because

$$\frac{K}{n} > \frac{K+n-1}{n^2} \quad \Longleftrightarrow \quad (K-1)(n-1) > 0,$$

which holds for $K, n \geq 2$.

It remains to show that this comparison policy lies in the fractional Core. Fix a coalition S of size k and any deviating policy σ. At state s_j, only one alternative is selected and each alternative benefits exactly one voter. Hence, the largest total utility that σ can give the members of S is K when $j \in S$ and 1 when $j \notin S$. Averaging over the uniformly reached states gives

$$\sum_{i \in S} U_i(\sigma) \leq \frac{kK + (n-k)}{n} \leq K,$$

where the last inequality follows from $K \geq 1$. If S blocked the comparison policy, then, since its population share is k/n and every member receives K/n under the comparison policy, blocking would require $U_i(\sigma) > K/k$ for every $i \in S$. Summing over S would give $\sum_{i \in S} U_i(\sigma) > K$, contradicting the bound above. Thus, the comparison policy lies in the fractional Core. □

D Omitted Proofs for Section 4

D.1 Positivity and the treatment of null voters

The main text assumes that every voter can obtain positive utility under some policy. This subsection justifies the positivity conditions used for Nash welfare and explains how null voters should be treated. For positive utilities, maximizing Nash welfare $\prod_i U_i(\pi)$ is equivalent to maximizing its logarithm $\sum_i \log U_i(\pi)$; the latter form is more convenient for the analysis below. We first show that

if every voter can obtain positive utility under some policy, then there exists a policy that simultaneously gives all voters positive utility, and hence Nash welfare attains a strictly positive optimum. We then consider null voters, who receive zero utility under every policy. Such voters cannot participate in a strict block, but they still count toward the electorate size used to determine coalition shares and therefore cannot in general be deleted.

Lemma D.1 (Positivity and null voters)**.** *Suppose that* $\max_\pi U_i(\pi) > 0$ *for every* $i \in N$*. Then there exists a policy with* $U_i > 0$ *for every voter simultaneously. Moreover, Nash welfare is maximized by a unique, strictly positive utility profile; equivalently,* $\max_\pi \sum_{i\in N} \log U_i(\pi)$ *is attained and has a unique maximizing utility profile.*

More generally, let $N^0 = \{i \in N : \max_\pi U_i(\pi) = 0\}$ *and* $N^+ = N \setminus N^0$*. No coalition containing a voter in* N^0 *can block, and any policy maximizing Nash welfare over the non-null voters, equivalently* $\sum_{i\in N^+} \log U_i(\pi)$*, lies in the fractional Core of the original* n*-voter election.*

Proof. We first show that all voters can simultaneously obtain positive utility under one policy. For each voter i, choose an occupancy $x^{(i)} \in \mathcal{X}$ with $U_i(x^{(i)}) > 0$. By convexity of $\mathcal{X}$, the average $\bar{x} = \frac{1}{n}\sum_i x^{(i)}$ also belongs to $\mathcal{X}$. Since utilities and occupancies are nonnegative, $U_i(\bar{x}) \geq \frac{1}{n}U_i(x^{(i)}) > 0$ for every i. Thus, a strictly positive utility profile is attainable.

We next show that Nash welfare attains a strictly positive maximum. We work with its equivalent log form, $\sum_i \log U_i$. The function $u \mapsto \sum_i \log u_i$, extended to take value $-\infty$ whenever some coordinate is zero, is upper semicontinuous on the compact attainable set $\mathcal{U}$. Indeed, if $u^k \to u$ and $u_i = 0$ for some i, boundedness of $\mathcal{U}$ gives a finite $\bar{u}$ such that $\sum_j \log u_j^k \leq \log u_i^k + (n-1)\log \bar{u} \to -\infty$. A maximizing sequence therefore has a convergent subsequence whose limit attains the supremum. Since the strictly positive profile constructed above has finite Nash welfare, the maximizing profile cannot have a zero coordinate. Finally, strict concavity of $\sum_i \log u_i$ on the positive orthant, together with convexity of $\mathcal{U}$, makes the maximizing utility profile unique.

Now, suppose that N^0 is nonempty. If $i \in N^0$, then $U_i(\sigma) = U_i(\pi) = 0$ for every pair of policies σ, π. Hence, the strict inequality required for blocking cannot hold for voter i, so every blocking coalition is contained in N^+.

Let $\pi^\star$ maximize Nash welfare over the non-null voters, equivalently $\sum_{i\in N^+} \log U_i(\pi)$. For any policy σ, consider the attainable segment $u(t) = (1-t)U(\pi^\star) + tU(\sigma)$. Its coordinates in N^+ remain positive for $t \in [0, 1)$, and the function $t \mapsto \sum_{i\in N^+} \log u_i(t)$ is maximized at $t = 0$. Its right derivative therefore gives

$$\sum_{i\in N^+} \frac{U_i(\sigma) - U_i(\pi^\star)}{U_i(\pi^\star)} \leq 0,$$

or equivalently

$$\sum_{i\in N^+} \frac{U_i(\sigma)}{U_i(\pi^\star)} \leq |N^+|.$$

This is a directional first-order condition, obtained by differentiating along the attainable segment.

Suppose for contradiction that a coalition $S \subseteq N^+$ of size k blocks $\pi^\star$ in the original election. Since its population share is k/n, blocking requires $U_i(\sigma)/U_i(\pi^\star) > n/k$ for every $i \in S$. By nonnegativity,

$$\sum_{i\in N^+} \frac{U_i(\sigma)}{U_i(\pi^\star)} > k\frac{n}{k} = n \geq |N^+|,$$

contradicting the first-order bound above. Thus, $\pi^\star$ lies in the fractional Core of the original n-voter election. □

The last part of Lemma D.1 keeps the null voters in the electorate. This distinction matters. Although a null voter can never participate in a strict block, removing such voters changes the denominator used to determine every remaining coalition's population share.

Proposition D.1 (Deleting null voters can change the fractional Core)**.** *Let* N^0 *be the set of null voters and* $N^+ = N \setminus N^0$*. Write* $E|_{N^+}$ *for the same election restricted to voter set* N^+*. Identifying policies across the two elections,* $\mathrm{Core}\big(E|_{N^+}\big) \subseteq \mathrm{Core}(E)$*, and the inclusion can be strict.*

Proof. For every nonempty coalition $S \subseteq N^+$,

$$\alpha_{N^+}(S) = \frac{|S|}{|N^+|} \geq \frac{|S|}{n} = \alpha_N(S),$$

with strict inequality whenever $N^0 \neq \varnothing$. Thus deleting null voters increases the population share of every coalition consisting of non-null voters.

By Lemma D.1, any coalition that blocks in the original election is contained in N^+. If such a coalition S blocks a policy π via σ, then nonnegativity of utilities gives $\alpha_{N^+}(S)U_i(\sigma) \geq \alpha_N(S)U_i(\sigma) > U_i(\pi)$ for every $i \in S$. Hence, the same coalition also blocks π in $E|_{N^+}$. Therefore, every policy in the fractional Core of the restricted election also belongs to the fractional Core of the original election, which proves the inclusion.

To see that the inclusion can be strict, consider a one-round election with three voters and two alternatives having cardinal utility vectors $(2, 0, 0)$ and $(0, 2, 0)$. Voter 3 is null, and the attainable profiles are $\{(2p, 2(1-p), 0) : p \in [0, 1]\}$. Consider the policy with utility profile $(0.8, 1.2, 0)$. It lies in the fractional Core of the three-voter election. No coalition containing voter 3 can block. A singleton coalition can obtain utility at most 2, whose one-third share is $2/3$, below both 0.8 and 1.2. Coalition $\{1, 2\}$ cannot block either: every policy satisfies $U_1 + U_2 \leq 2$, so after multiplication by its population share $2/3$, it cannot make both voters strictly better than a profile whose first two utilities sum to 2.

After voter 3 is deleted, however, singleton $\{1\}$ has population share $1/2$ and can choose the first alternative, obtaining scaled utility $\frac{1}{2} \cdot 2 = 1 > 0.8$. It therefore blocks the same policy in the restricted two-voter election. The inclusion can consequently be strict. □

Thus, null voters need not be removed for the Nash-welfare argument, and removing them can change the proportionality requirement. Throughout the paper, coalition shares are therefore defined relative to the electorate as given.

D.2 Subgame completion

A policy that maximizes a weighted objective from the initial state need not be specified optimally at states it never reaches. Our later results, however, require scalar optimality at every state. This subsection shows that we can always meet this requirement. We first express a policy's ex-ante optimality gap as the expected sum of its Bellman gaps. This implies that a scalar-optimal policy must also be optimal from every state it reaches. We can then replace its arbitrary choices at unreached states by optimal ones without changing its occupancy or any voter's ex-ante utility.

For a positive weight vector $\boldsymbol{\lambda}$, recall that $r_h^{\boldsymbol{\lambda}}(s, c) = \sum_i \lambda_i a_{i,h}(s, c)$ and that $y_h^{\boldsymbol{\lambda}}(s)$ is the optimal scalar continuation value. Define the *Bellman gap* of alternative c at (h, s) by

$$G_h^{\boldsymbol{\lambda}}(s, c) = y_h^{\boldsymbol{\lambda}}(s) - r_h^{\boldsymbol{\lambda}}(s, c) - \sum_{s'} P_h(s' \mid s, c) y_{h+1}^{\boldsymbol{\lambda}}(s').$$

Bellman optimality gives $G_h^{\boldsymbol{\lambda}}(s, c) \geq 0$: the gap measures how much scalar value is lost by choosing c at (h, s) rather than an optimal alternative.

Lemma D.2 (Occupancy decomposition of the optimality gap)**.** *For every policy π and $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$,*

$$y^{\boldsymbol{\lambda}}(z_0) - \sum_i \lambda_i U_i(\pi) = \sum_{h,s,c} x_h^\pi(s, c) G_h^{\boldsymbol{\lambda}}(s, c).$$

Proof. The right-hand side is the expected Bellman gap accumulated under π. Expanding $G_h^{\boldsymbol{\lambda}}(s, c)$ and weighting each term by its occupancy gives three terms. By Lemma B.1, $\sum_c x_h^\pi(s, c) = d_h^\pi(s)$, while $\sum_{s,c} x_h^\pi(s, c) P_h(s' \mid s, c) = d_{h+1}^\pi(s')$. Hence,

$$\sum_{h,s,c} x_h^\pi(s, c) G_h^{\boldsymbol{\lambda}}(s, c) = \sum_h \left(\mathbb{E}_\pi[y_h^{\boldsymbol{\lambda}}(s_h)] - \mathbb{E}_\pi[r_h^{\boldsymbol{\lambda}}(s_h, c_h)] - \mathbb{E}_\pi[y_{h+1}^{\boldsymbol{\lambda}}(s_{h+1})] \right).$$

The intermediate value terms cancel. Since $y_{H+1}^{\boldsymbol{\lambda}} \equiv 0$ and the artificial round 0 is deterministic and carries no approval utility, the remaining terms are $y^{\boldsymbol{\lambda}}(z_0)$ minus the policy's expected scalar reward. The latter equals $\sum_i \lambda_i U_i(\pi)$, which proves the identity. □

Since every term in the decomposition is nonnegative, a scalar-optimal policy has zero Bellman gap on every state-alternative pair that it uses with positive probability. This yields the two properties needed later: optimality on every reached subgame and an optimal completion at states outside the policy's support.

Lemma D.3 (Subgame optimality and completion)**.** *Let π maximize $\sum_i \lambda_i U_i$. Then:*

(i) at every state z reached with positive probability, the continuation of π attains the optimal scalar value $y^{\boldsymbol{\lambda}}(z)$;

(ii) there exists a policy $\tilde{\pi}$ that agrees with π at every state reached by π, satisfies $U(\tilde{\pi}) = U(\pi)$, and is $\boldsymbol{\lambda}$-tight.

Proof. Since π is scalar-optimal, the left-hand side of Lemma D.2 is zero. Every term on the right-hand side is nonnegative, so $G_h^{\boldsymbol{\lambda}}(s,c) = 0$ whenever $x_h^{\pi}(s,c) > 0$.

For part *(i)*, we use backward induction on h. The claim is immediate at the terminal layer $h = H+1$, where both the continuation utility and $y^{\boldsymbol{\lambda}}$ are zero. Suppose it holds at round $h+1$, and consider a state s reached with positive probability at round h. Every alternative c selected with positive probability at s has $G_h^{\boldsymbol{\lambda}}(s,c) = 0$. Moreover, every successor s' reached with positive probability after such an alternative is itself reached by π, so the induction hypothesis gives $\sum_i \lambda_i V_i^{\pi}(h+1,s') = y_{h+1}^{\boldsymbol{\lambda}}(s')$. Therefore, each action used by π at (h,s) achieves the Bellman-optimal value. Averaging over the policy's action distribution yields $\sum_i \lambda_i V_i^{\pi}(h,s) = \sum_c \pi_h(c|s)\big[y_h^{\boldsymbol{\lambda}}(s) - G_h^{\boldsymbol{\lambda}}(s,c)\big] = y_h^{\boldsymbol{\lambda}}(s)$, proving part *(i)*.

For part *(ii)*, define $\tilde{\pi}$ to agree with π at every state reached by π and to choose a Bellman-optimal alternative at every other state. Changing the policy only at states outside its support cannot make any of those states reachable: a forward induction gives $d^{\tilde{\pi}} = d^{\pi}$, and hence $x^{\tilde{\pi}} = x^{\pi}$. Consequently, $U(\tilde{\pi}) = U(\pi)$.

It remains to verify $\boldsymbol{\lambda}$-tightness. At states reached by π, part *(i)* shows that the unchanged continuation is scalar-optimal. At all other states, $\tilde{\pi}$ uses Bellman-optimal alternatives by construction. A backward induction over all states therefore gives $\sum_i \lambda_i V_i^{\tilde{\pi}}(z) = y^{\boldsymbol{\lambda}}(z)$ for every z. Thus $\tilde{\pi}$ is $\boldsymbol{\lambda}$-tight while preserving the original utility profile. □

The two parts of Lemma D.3 play different roles. Part *(i)* shows that ex-ante scalar optimality already guarantees scalar optimality after every state the policy actually reaches. Part *(ii)* extends this property to unreached states without changing the policy's utility profile. This distinction is important because the statewise guarantees and certificates later in the paper apply to every state, including states outside the reference policy's support.

D.3 Analytic preliminaries for continuation weights

This appendix proves two analytic results used in the proof of Theorem 4.1. Recall that we extend ϕ to $[0,\infty)$ by setting $\bar{\phi}(x) = \phi(x)$ for $x > 0$ and $\bar{\phi}(0) = \lim_{t \downarrow 0} \phi(t) = \inf_{t>0} \phi(t)$, where the boundary value may be $-\infty$. The corresponding welfare objective is then $F_\phi(\pi) = \sum_i w_i \bar{\phi}(U_i(\pi))$, which is defined for every policy in the extended reals and agrees with the expression above whenever all utilities are positive.

We first show that extending ϕ to zero preserves the concavity and boundary properties needed to evaluate continuation policies with zero utility.

Lemma D.4. *For every $x > 0$, $\phi'(x) > 0$. Moreover, $\bar{\phi}$ is concave and upper semicontinuous on $[0,\infty)$, and for every $v > 0$ and $V \geq 0$, $\bar{\phi}(V) - \phi(v) \leq \phi'(v)(V - v)$ in $\mathbb{R} \cup \{-\infty\}$.*

Proof. We prove the three properties in turn.

Positivity of ϕ'. Since ϕ is differentiable and concave, ϕ' is nonincreasing. Strict monotonicity of ϕ implies $\phi'(x) \geq 0$ for every $x > 0$. If $\phi'(x_0) = 0$ at some $x_0 > 0$, then monotonicity of the derivative gives $0 \leq \phi'(x) \leq \phi'(x_0) = 0$ for every $x \geq x_0$. Hence, ϕ would be constant on $[x_0,\infty)$, contradicting strict increase. Therefore, $\phi'(x) > 0$ for every $x > 0$. In particular, the denominators $\phi'(v_i)$ appearing in the continuation weights of Theorem 4.1 are positive.

Concavity and upper semicontinuity of $\bar{\phi}$. Away from zero, both properties follow directly from the corresponding properties of ϕ. It remains to check the boundary. Fix $x > 0$ and $t \in (0, 1)$. For every $\eta > 0$, concavity of ϕ gives

$$\phi(tx + (1-t)\eta) \geq t\phi(x) + (1-t)\phi(\eta).$$

Letting $\eta \downarrow 0$, continuity of ϕ at $tx > 0$ and the definition of $\bar{\phi}(0)$ yield

$$\bar{\phi}(tx) \geq t\bar{\phi}(x) + (1-t)\bar{\phi}(0),$$

with the inequality interpreted in the extended reals when $\bar{\phi}(0) = -\infty$. The case in which both endpoints are zero is immediate, so $\bar{\phi}$ is concave on $[0, \infty)$. Upper semicontinuity at zero follows from $\bar{\phi}(0) = \lim_{t\downarrow 0} \phi(t)$ by construction, while at every positive point $\bar{\phi} = \phi$ is continuous.

Tangent inequality at the boundary. For $V > 0$, concavity and differentiability of ϕ give the usual tangent inequality $\phi(V) - \phi(v) \leq \phi'(v)(V - v)$. To include $V = 0$, apply this inequality with $V = x > 0$ and let $x \downarrow 0$. The left-hand side converges to $\bar{\phi}(0) - \phi(v)$ in $\mathbb{R} \cup \{-\infty\}$, while the right-hand side converges to $-\phi'(v)v$. Hence, $\bar{\phi}(0) - \phi(v) \leq -\phi'(v)v$, which is the desired inequality at $V = 0$. Since $v > 0$, $\phi(v)$ is finite, so this limiting argument never involves an indeterminate difference of infinities. □

The next lemma shows that a positive maximizer of the separable welfare objective is also optimal for the linear objective defined by its marginal welfare weights at an optimum of F_ϕ. A competing attainable profile may contain zero coordinates, where F_ϕ can equal $-\infty$. We therefore differentiate along the segment leaving the strictly positive optimum; near that endpoint all coordinates remain positive.

Lemma D.5 (Linear optimality of a welfare optimum)**.** *Let $\pi^\star$ maximize F_ϕ and suppose $u^\star = U(\pi^\star) > 0$. Define $\lambda_i = w_i\phi'(u_i^\star)$ for every i. Then,*

$$\sum_i \lambda_i U_i(\sigma) \leq \sum_i \lambda_i U_i(\pi^\star)$$

for every policy σ. Hence, $\pi^\star$ maximizes the scalar objective $\sum_i \lambda_i U_i$, and it can be completed, without changing its utility profile, to a $\boldsymbol{\lambda}$-tight policy.

Proof. Fix any attainable utility profile $u \in \mathcal{U}$ and consider the segment $u(t) = u^\star + t(u - u^\star)$ for $t \in [0, 1]$. Convexity of $\mathcal{U}$ makes $u(t)$ attainable, and because $u^\star > 0$, every coordinate of $u(t)$ remains positive for all sufficiently small $t > 0$. Thus $g(t) = F_\phi(u(t))$ is differentiable from the right at $t = 0$, even if $\mathbf{u}$ itself has zero coordinates and $F_\phi(u) = -\infty$.

Since $u^\star$ maximizes F_ϕ over the attainable set, g is maximized at $t = 0$. Its right derivative must therefore be nonpositive:

$$0 \geq g'(0^+) = \sum_i w_i\phi'(u_i^\star)(u_i - u_i^\star) = \sum_i \lambda_i(u_i - u_i^\star).$$

Since $\mathbf{u}$ was arbitrary, $u^\star$ maximizes the linear functional $\sum_i \lambda_i u_i$ over $\mathcal{U}$. Equivalently, $\pi^\star$ maximizes $\sum_i \lambda_i U_i$. Lemma D.3 then allows us to modify $\pi^\star$ only at states outside its support so that it becomes $\boldsymbol{\lambda}$-tight without changing its utility profile. □

Together with the subgame-completion result of Lemma D.3, Lemmas D.4 and D.5 allow the ex-ante first-order condition to be applied at every continuation state.

D.4 Proof of Theorem 4.1

We now prove the two parts of Theorem 4.1. Part *(i)* combines the linear optimality result of Lemma D.5 with the tangent inequality of Lemma D.4. The converse in part *(ii)* requires one additional construction. To characterize the continuation weights at an arbitrary positive triple $(\mathbf{w}, \mathbf{u}, \mathbf{v})$, we must be able to realize a continuation set that locally varies around $\mathbf{v}$ in any direction that leaves the corresponding linear objective unchanged. The following lemma constructs such an election.

Lemma D.6 (Realization of a continuation segment)**.** *Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n_{>0}$ be arbitrary, with no restriction relating $\mathbf{v}$ to $\mathbf{u}$. Let $\lambda_i = w_i\phi'(u_i)$, and let $\mathbf{d} \neq 0$ satisfy $\boldsymbol{\lambda} \cdot \mathbf{d} = 0$. For every sufficiently small $\varepsilon > 0$, there exists a Markov temporal election with nonnegative cardinal utilities and $H = 2$, together with a state z reached with positive probability, such that:*

1. *the attainable continuation-utility set at z is exactly $\{\mathbf{v} + t\mathbf{d} : |t| \leq \varepsilon\}$; and*
2. *a $\boldsymbol{\lambda}$-tight maximizer of F_ϕ has ex-ante utility profile $\mathbf{u}$ and continuation profile $\mathbf{v}$ at z.*

In particular, the construction permits $v_i > u_i$.

Proof. The construction allows us to choose the continuation profile $\mathbf{v}$ independently of the ex-ante profile $\mathbf{u}$ by making the relevant state sufficiently unlikely. Choose $q \in (0, 1)$ such that

$$q \leq \frac{\min_i u_i}{2 \max_i v_i}.$$

Then, $\mathbf{u} - q\mathbf{v} > 0$. In round 1, there is a single alternative with zero utility that leads to a state z with probability q and to another state z' with probability $1 - q$. At z, there are two alternatives with utility vectors $\mathbf{v} + \varepsilon\mathbf{d}$ and $\mathbf{v} - \varepsilon\mathbf{d}$. Since $\mathbf{v} > 0$, both vectors are nonnegative for sufficiently small ε. At z', there is a single alternative with utility vector $(\mathbf{u} - q\mathbf{v})/(1 - q) \geq 0$.

Suppose the policy selects the first alternative at z with probability p. Its continuation profile at z is then $\mathbf{v} + (2p - 1)\varepsilon\mathbf{d}$. As p ranges over $[0, 1]$, these are exactly the profiles $\{\mathbf{v} + t\mathbf{d} : |t| \leq \varepsilon\}$. The corresponding ex-ante utility profile is

$$q\big(\mathbf{v} + (2p - 1)\varepsilon\mathbf{d}\big) + (\mathbf{u} - q\mathbf{v}) = \mathbf{u} + q(2p - 1)\varepsilon\mathbf{d}.$$

It remains to show that the equal mixture $p = 1/2$ is an F_ϕ-maximizer. Along this family of policies, welfare is

$$F_\phi(p) = \sum_i w_i\phi\big(u_i + q(2p - 1)\varepsilon d_i\big).$$

For sufficiently small ε, every argument of ϕ remains positive. Hence, $F_\phi(p)$ is differentiable and concave in p. Its derivative at $p = 1/2$ is

$$F'_\phi(1/2) = 2q\varepsilon \sum_i w_i\phi'(u_i)d_i = 2q\varepsilon\, \boldsymbol{\lambda} \cdot \mathbf{d} = 0.$$

A concave differentiable function with a stationary interior point is maximized there, so $p = 1/2$ is a global maximizer. At this policy the ex-ante utility profile is $\mathbf{u}$ and the continuation profile at z is $\mathbf{v}$.

Finally, the policy is $\boldsymbol{\lambda}$-tight. At state z, it holds that $\boldsymbol{\lambda} \cdot (\mathbf{v} + t\mathbf{d}) = \boldsymbol{\lambda} \cdot \mathbf{v}$ for every t because $\boldsymbol{\lambda} \cdot \mathbf{d} = 0$. Thus, every feasible continuation at z has the same scalar value, so the equal mixture is scalar-optimal there. Every other state offers only one alternative, and scalar optimality therefore propagates backward through the remaining states. Hence, the policy is $\boldsymbol{\lambda}$-tight everywhere.

The construction imposes no ordering between $\mathbf{u}$ and $\mathbf{v}$. The state z is reached only with probability q, so its conditional continuation utility can exceed a voter's ex-ante utility. This is why the converse of Theorem 4.1 can range over arbitrary positive $(\mathbf{w}, \mathbf{u}, \mathbf{v})$. □

We can now prove both directions of the continuation-weight characterization of Theorem 4.1. Part *(i)* shows that the proposed weights preserve the welfare optimum after every state. Part *(ii)* uses Lemma D.6 to show that no other rule of the stated $(\mathbf{w}, \mathbf{u}, \mathbf{v})$ form can do so on every convex continuation set, except for a common rescaling.

Theorem 4.1. *Let $\pi^\star$ maximize F_ϕ with $U(\pi^\star) > 0$, $\lambda_i = w_i\phi'(u_i) > 0$, and suppose $\pi^\star$ is $\boldsymbol{\lambda}$-tight.*

(i) *For any voter i with $v_i > 0$, let $b_i^\phi(z) = \lambda_i/\phi'(v_i) = w_i\phi'(u_i)/\phi'(v_i)$. At any state z, the continuation of $\pi^\star$ maximizes $\sum_{i:v_i>0} b_i^\phi(z)\bar{\phi}(V_i^\sigma(z))$ over all continuation policies σ, in the extended-real sense, including when some voters have zero continuation utility under either $\pi^\star$ or σ.*

(ii) *For $n \geq 2$, any positive continuation-weight rule depending only on $(\mathbf{w}, \mathbf{u}, \mathbf{v})$ and guaranteeing the local optimality in part* (i) *on every convex continuation set must coincide with $\mathbf{b}^\phi$ up to a common positive factor, for every $(\mathbf{w}, \mathbf{u}, \mathbf{v}) \in (\mathbb{R}^n_{>0})^3$, with no restriction relating $\mathbf{v}$ to $\mathbf{u}$.*

Proof. By Lemma D.5, the ex-ante maximizer $\pi^\star$ is optimal for the scalar weights $\lambda_i = w_i\phi'(u_i)$. By the hypothesis of the theorem, we take $\pi^\star$ to be $\boldsymbol{\lambda}$-tight.

Part (i): continuation optimality. Fix a state z and any continuation policy σ. Write $V_i = V_i^\sigma(z) \geq 0$. For every voter with $v_i > 0$, Lemma D.4 gives

$$\bar{\phi}(V_i) \leq \phi(v_i) + \phi'(v_i)(V_i - v_i),$$

with the inequality interpreted in $\mathbb{R} \cup \{-\infty\}$. This includes both $V_i > 0$ and $V_i = 0$, regardless of whether $\bar{\phi}(0)$ is finite or $-\infty$.

Multiplying by the positive continuation weight $b_i^\phi(z) = \lambda_i/\phi'(v_i)$ and summing over the voters with $v_i > 0$ gives

$$\sum_{i:v_i>0} b_i^\phi(z)\bar{\phi}(V_i) \leq \sum_{i:v_i>0} b_i^\phi(z)\phi(v_i) + \sum_{i:v_i>0} \lambda_i(V_i - v_i). \tag{1}$$

This additive form avoids taking differences of extended-real quantities: the two terms on the right are finite, while the left-hand side may equal $-\infty$.

We now bound the right-hand sum in Equation (1). If $v_i = 0$, then $\lambda_i(V_i - v_i) = \lambda_i V_i \geq 0$ by nonnegativity of utilities. Therefore,

$$\sum_{i:v_i>0} \lambda_i(V_i - v_i) \leq \sum_{i\in N} \lambda_i(V_i - v_i).$$

Since $\pi^\star$ is $\boldsymbol{\lambda}$-tight, $\sum_i \lambda_i v_i = y^{\boldsymbol{\lambda}}(z)$, while scalar optimality gives $\sum_i \lambda_i V_i \leq y^{\boldsymbol{\lambda}}(z)$. Hence, $\sum_i \lambda_i(V_i - v_i) \leq 0$. Combining the inequalities yields

$$\sum_{i:v_i>0} b_i^\phi(z)\bar{\phi}\big(V_i^\sigma(z)\big) \leq \sum_{i:v_i>0} b_i^\phi(z)\phi\big(V_i^{\pi^\star}(z)\big).$$

The right-hand side is exactly the continuation objective attained by $\pi^\star$. Thus, its continuation is optimal at z. Since z and σ were arbitrary, the conclusion holds at every state, including against continuations that give a voter with $v_i > 0$ zero utility.

Part (ii): uniqueness. Assume $n \geq 2$ and fix an arbitrary positive triple $(\mathbf{w}, \mathbf{u}, \mathbf{v})$. Let $\lambda_i = w_i\phi'(u_i)$. To identify the continuation weights, take any nonzero direction $\mathbf{d}$ satisfying $\boldsymbol{\lambda} \cdot \mathbf{d} = 0$. Lemma D.6 constructs an election in which an F_ϕ-maximizing and $\boldsymbol{\lambda}$-tight policy has ex-ante profile $\mathbf{u}$, continuation profile $\mathbf{v}$, and attainable continuation set $\{\mathbf{v} + t\mathbf{d} : |t| \leq \varepsilon\}$.

Let $\beta_i > 0$ denote the weights assigned by any continuation-weight rule satisfying the hypothesis of part *(ii)*. Dynamic consistency requires $t = 0$ to maximize $t \mapsto \sum_i \beta_i\phi(v_i + td_i)$ on this segment. Since $\mathbf{v} > 0$, the objective is differentiable at $t = 0$, which is an interior point. Its derivative must therefore be zero:

$$\sum_i \beta_i\phi'(v_i)d_i = 0.$$

This holds for every $\mathbf{d}$ orthogonal to $\boldsymbol{\lambda}$. Consequently, the vector $(\beta_i\phi'(v_i))_{i\in N}$ is orthogonal to $\boldsymbol{\lambda}^\perp$ and hence belongs to $\mathrm{span}(\boldsymbol{\lambda})$. Thus, for some scalar κ,

$$\beta_i\phi'(v_i) = \kappa\lambda_i \qquad \text{for every } i.$$

Since β_i, λ_i, and $\phi'(v_i)$ are all positive, it holds that $\kappa > 0$. Therefore,

$$\beta_i = \kappa\frac{\lambda_i}{\phi'(v_i)} = \kappa b_i^\phi(z),$$

so the continuation-weight rule agrees with $\mathbf{b}^\phi$ up to a common positive factor, as claimed. □

D.5 Characterizing utilitarian welfare and proof of Corollary 4.1

We now specialize the continuation-weight characterization of Theorem 4.1 to the two requirements specified in the main text. We first assume that a voter's claim should remain tied to their initial weight, so that conditioning on a state does not change voters' relative weights. This corresponds to *weighted utilitarian welfare*, which evaluates a policy by the weighted sum $\sum_i w_i U_i(\pi)$. Within our separable class, this is obtained when ϕ is affine, and the next result shows that it is exactly the objective with this invariance property.

Corollary D.1. *The continuation weights remain proportional to the initial weights* $\mathbf{w}$ *at every state if and only if* ϕ *is affine.*

Proof. By Theorem 4.1*(ii)*, the dynamically consistent continuation weights are $b_i^\phi \propto w_i \frac{\phi'(u_i)}{\phi'(v_i)}$ up to a common positive factor. Suppose these weights are proportional to the initial weights $\mathbf{w}$ for every positive $(\mathbf{w}, \mathbf{u}, \mathbf{v})$. Then $\phi'(u_i)/\phi'(v_i)$ must be the same for every voter.

Fix arbitrary $x, y > 0$ and take $n = 2$ with $\mathbf{u} = (x, x)$ and $\mathbf{v} = (x, y)$. The first voter has derivative ratio $\phi'(x)/\phi'(x) = 1$, so the second must satisfy $\phi'(x)/\phi'(y) = 1$. Hence, $\phi'(x) = \phi'(y)$. Since x and y were arbitrary, ϕ' is constant on $(0, \infty)$, and therefore ϕ is affine.

Conversely, if ϕ is affine, then ϕ' is a positive constant. The formula above therefore gives $b_i^\phi \propto w_i$ at every state, so the relative voter weights remain unchanged. □

We next assume that the continuation weights equal the initial weights scaled by the continuation-to-ex-ante utility ratio. This leads to Nash welfare. The *Nash welfare* of a policy π is the product of voters' utilities $\prod_i U_i(\pi)$, whose maximization over positive profiles is equivalent to maximizing $\sum_i \log U_i(\pi)$. More generally, with positive weights w_i, the weighted Nash-welfare objective is $\prod_i U_i(\pi)^{w_i}$, equivalently $\sum_i w_i \log U_i(\pi)$.

Corollary 4.1. *If* $b_i^\phi \propto w_i V_i^{\pi^\star}(z)/U_i(\pi^\star)$ *for any voter* i *at every state, then* $\phi(x) = a \log x + e$ *for some* $a > 0$*. Conversely, the Nash-welfare objective induces exactly these continuation weights.*

Proof. Again, Theorem 4.1*(ii)* gives $b_i^\phi \propto w_i \phi'(u_i)/\phi'(v_i)$. Suppose instead that the continuation weights are required to satisfy $b_i^\phi \propto w_i v_i/u_i$. Comparing these two expressions shows that $\frac{u_i \phi'(u_i)}{v_i \phi'(v_i)}$ must be the same for every voter i.

Fix arbitrary $x, y > 0$ and again take $\mathbf{u} = (x, x)$ and $\mathbf{v} = (x, y)$. For the first voter the ratio above equals 1, so for the second we obtain $x\phi'(x) = y\phi'(y)$. Since x and y were arbitrary, $x\phi'(x)$ is a positive constant, say $a > 0$. Thus, $\phi'(x) = a/x$ with $a > 0$, and integration gives $\phi(x) = a \log x + c$ for some constant c.

Conversely, if $\phi(x) = a \log x + c$, then $\phi'(x) = a/x$, and the continuation weights become

$$b_i^\phi \propto w_i \frac{a/u_i}{a/v_i} = w_i \frac{v_i}{u_i}.$$

Thus, the Nash-welfare objective induces exactly the required continuation weights. □

D.6 A STATE-DEPENDENT PRICE CHARACTERIZATION OF NASH WELFARE

The reciprocal weights associated with Nash welfare also have a price interpretation. We assign each alternative a state-dependent price that accounts for both the utility obtained when the alternative is selected and the continuation opportunities induced by its transition. For values $y_h(s)$ with $y_{H+1} \equiv 0$, define the price of alternative c at (h, s) as $p_h(s, c) = y_h(s) - \mathbb{E}_{s' \sim P_h(\cdot|s,c)}[y_{h+1}(s')]$. Thus, the price is the current value $y_h(s)$ minus the expected continuation value after selecting c. Alternatives available at the same state may therefore have different prices when they induce different distributions over future states.

We say that a positive-utility policy π *admits a state-dependent price system* if there are $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$ and values y such that $\lambda_i U_i(\pi) = 1$ for every voter i, $p_h(s, c) \geq \sum_i \lambda_i a_{i,h}(s, c)$ for every state s and alternative c, and equality holds whenever (h, s, c) occurs with positive probability under π. The first condition gives the reciprocal weights $\lambda_i = 1/U_i(\pi)$. The price inequality requires the immediate weighted utility from an alternative to be no greater than its price, while alternatives used with positive probability under π attain equality.

Theorem D.1. *A policy with positive utilities maximizes Nash welfare if and only if it admits a state-dependent price system. Equivalently, these price conditions are the KKT conditions of the Nash occupancy program in reduced-cost form, with* $\lambda_i = 1/U_i(\pi)$ *as the utility-constraint multipliers,* y *as the flow-constraint multipliers, and the price slacks as the nonnegative multipliers for the occupancy constraints.*

Proof. Round 0 contributes no approvals and has fixed occupancy $x_0(s_0, c_0) = 1$, so we omit it from the occupancy program and restore it below. Consider the convex minimization problem

$$\min\Big\{F(x,t) = -\sum_{i\in N} \log t_i : (x,t) \in \mathcal{F},\ t > 0\Big\},$$
$$\mathcal{F} = \{(x,t) : x \in \mathcal{X},\ t_i - U_i(x) \le 0 \text{ for every } i \in N\}.$$

We use the normal-cone convention $N_C(z) = \{v : \langle v, z' - z\rangle \le 0 \text{ for every } z' \in C\}$. Since $U_i(x) = \langle x, a_i\rangle$ is linear and $\mathcal{X}$ is the occupancy polytope of Lemma B.1, $\mathcal{F}$ is a polyhedron described by the flow equalities, $x \ge 0$, and $t_i - \langle x, a_i\rangle \le 0$. The set $\mathcal{F} \cap \{t > 0\}$ itself is not a polyhedron. The normal cone used below is $N_{\mathcal{F}}(z^\star)$ for some minimizer $z^\star = (x^\star, t^\star)$ of the above program; this is valid because $t^\star > 0$ makes the open constraint $t > 0$ locally inactive at the optimum.

Attainment and positivity. By Lemma D.1, the Nash problem $\max_{x\in\mathcal{X}} \sum_i \log U_i(x)$ is attained at some $x^\star$ with $U(x^\star) > 0$. Setting $t^\star = U(x^\star)$ gives $z^\star = (x^\star, t^\star) \in \mathcal{F}$, which minimizes F over $\mathcal{F} \cap \{t > 0\}$ because feasibility requires $t_i \le U_i(x)$ and $-\log$ is decreasing. Since $t^\star > 0$, F is differentiable at $z^\star$, with $\nabla F(z^\star) = \big(0, (-1/t_i^\star)_i\big)$.

Optimality conditions. We do not assume that $x^\star$, or any occupancy giving every voter positive utility, lies in the relative interior of $\mathcal{X}$. Such an occupancy may lie on a proper face. Instead, fix any $z \in \mathcal{F}$ and set $z_s = z^\star + s(z - z^\star)$. By convexity, $z_s \in \mathcal{F}$ for $s \in [0,1]$, and for sufficiently small $s > 0$ its t-coordinates remain positive because $t^\star > 0$. Hence, $g(s) = F(z_s)$ is defined on a right neighborhood of 0, and the optimality of $z^\star$ gives

$$g'(0^+) = \langle \nabla F(z^\star), z - z^\star\rangle \ge 0.$$

Since $z \in \mathcal{F}$ was arbitrary, $-\nabla F(z^\star) \in N_{\mathcal{F}}(z^\star)$, equivalently $0 \in \nabla F(z^\star) + N_{\mathcal{F}}(z^\star)$.

For a polyhedron $\mathcal{P} = \{z : g_j(z) \le 0,\ Az = b\}$ with affine functions g_j, Farkas' lemma (Farkas, 1902) gives

$$N_{\mathcal{P}}(z^\star) = \left\{\sum_j \mu_j \nabla g_j + A^\top y : \mu \ge 0,\ \mu_j = 0 \text{ for every inactive constraint } j,\ y \text{ free}\right\}.$$

We apply this representation to the polyhedron $\mathcal{F}$ defined above. Its inequality constraints are the utility constraints $g_i(x,t) = t_i - \langle x, a_i\rangle \le 0$ and the occupancy constraints $g^0_{h,s,c}(x,t) = -x_h(s,c) \le 0$. The gradient of g_i with respect to x is $-a_i$ and with respect to t_j is $\mathbf{1}\{j = i\}$, while $g^0_{h,s,c}$ has derivative -1 with respect to $x_h(s,c)$ and 0 with respect to every other coordinate.

Hence, there are multipliers $\boldsymbol{\lambda} \in \mathbb{R}^n_{\ge 0}$ and $\nu \ge 0$ for the two families of inequality constraints, respectively, and unrestricted multipliers $y_h(s)$ for the flow equalities, such that

$$\nabla F(z^\star) + \sum_i \lambda_i \nabla g_i + \sum_{h,s,c} \nu_h(s,c) \nabla g^0_{h,s,c} + A^\top y = 0.$$

The t_i- and $x_h(s,c)$-coordinates of this equation give

$$\frac{\partial}{\partial t_i}: \quad -\frac{1}{t_i^\star} + \lambda_i = 0, \qquad \frac{\partial}{\partial x_h(s,c)}: \quad -\sum_i \lambda_i a_{i,h}(s,c) - \nu_h(s,c) + p_h(s,c) = 0,$$

and complementary slackness gives $\lambda_i(t_i^\star - U_i(x^\star)) = 0$ and $\nu_h(s,c) x_h^\star(s,c) = 0$.

It remains to identify the contribution $p_h(s,c)$ of the flow multipliers in the second stationarity equation. The variable $x_h(s,c)$ enters the flow constraint at (h,s) with coefficient $+1$ and the flow constraint at each successor $(h+1, s')$ with coefficient $-P_h(s' \mid s, c)$. Its coefficient from the flow constraints is therefore

$$y_h(s) - \sum_{s'} P_h(s' \mid s, c) y_{h+1}(s') = p_h(s,c),$$

where $y_{H+1} \equiv 0$ because the terminal layer has no flow constraint.

These conditions are also sufficient. If $z^\star \in \mathcal{F}$ with $t^\star > 0$ admits such multipliers, then $-\nabla F(z^\star) \in N_{\mathcal{F}}(z^\star)$, so $\langle \nabla F(z^\star), z - z^\star \rangle \geq 0$ for every $z \in \mathcal{F}$. By convexity of F,

$$F(z) \geq F(z^\star) + \langle \nabla F(z^\star), z - z^\star \rangle \geq F(z^\star)$$

for every $z \in \mathcal{F}$ with $t > 0$. Hence, $z^\star$ is a global minimizer.

Deriving the price system. Stationarity in t_i gives $\lambda_i = 1/t_i^\star > 0$. Complementary slackness then gives $t_i^\star = U_i(x^\star)$, and hence $\lambda_i U_i(\pi) = 1$. Stationarity in $x_h(s,c)$ gives $p_h(s,c) = \sum_i \lambda_i a_{i,h}(s,c) + \nu_h(s,c)$ and therefore $p_h(s,c) \geq \sum_i \lambda_i a_{i,h}(s,c)$ at every state-alternative pair. This includes pairs that are unreachable, or whose occupancy is zero in every feasible occupancy, because no such coordinate was removed from the program. Complementary slackness gives equality whenever $x_h^\star(s,c) > 0$.

These are exactly the state-dependent price conditions presented before Theorem D.1. In reduced-cost form, the entries of $\boldsymbol{\lambda}$ are the multipliers of the utility constraints, y the multipliers of the flow constraints, and the nonnegative multipliers $\nu_h(s,c)$ of $x \geq 0$ are precisely the slacks in the price inequalities. Thus, the price inequalities express reduced-cost nonnegativity, while their tightness on the support of π is complementary slackness.

Restoring round 0. The artificial round 0 has a single alternative c_0, no approvals, and transition distribution μ. Setting $y(z_0) = \sum_s \mu(s) y_1(s) = \mathbb{E}_{s_1 \sim \mu}[y_1(s_1)]$ gives

$$p_0(s_0, c_0) = y(z_0) - \mathbb{E}_{s_1 \sim \mu}[y_1(s_1)] = 0 = \sum_i \lambda_i a_{i,0}(s_0, c_0).$$

Thus, the round-0 price condition holds with equality automatically, which justifies omitting round 0 from the occupancy program. The identity $y(z_0) = \mathbb{E}_{s_1 \sim \mu}[y_1(s_1)]$ will also be used below.

Price system $\Rightarrow$ Nash. Suppose that π admits a state-dependent price system $(\boldsymbol{\lambda}, y)$. For any policy σ and any $1 \leq h \leq H$, the definition of p_h gives

$$\mathbb{E}_\sigma[p_h(s_h, c_h) \mid \mathcal{F}_h] = y_h(s_h) - \mathbb{E}_\sigma[y_{h+1}(s_{h+1}) \mid \mathcal{F}_h].$$

Summing over h, applying the tower property, and using $y_{H+1} \equiv 0$ yields

$$\mathbb{E}_\sigma\left[\sum_{h=1}^{H} p_h(s_h, c_h)\right] = \mathbb{E}_\sigma[y_1(s_1)] - \mathbb{E}_\sigma[y_{H+1}(s_{H+1})] = \mathbb{E}_{s_1 \sim \mu}[y_1(s_1)] = y(z_0),$$

where the second equality uses the fact that $s_1 \sim \mu$ under every policy.

Since $p_h(s,c) \geq \sum_i \lambda_i a_{i,h}(s,c)$ everywhere,

$$\sum_i \lambda_i U_i(\sigma) = \mathbb{E}_\sigma\left[\sum_{h=1}^{H} \sum_i \lambda_i a_{i,h}(s_h, c_h)\right] \leq y(z_0)$$

for every σ. Equality holds for $\sigma = \pi$, because every state-alternative pair visited with positive probability under π satisfies the price condition with equality. Hence, π maximizes the scalar objective $\sum_i \lambda_i U_i$, with $\lambda_i = 1/U_i(\pi)$.

This scalar optimality implies Nash optimality. For every σ with $U(\sigma) > 0$, concavity of $\log$ gives

$$\sum_i \log U_i(\sigma) - \sum_i \log U_i(\pi) \leq \sum_i \lambda_i \big(U_i(\sigma) - U_i(\pi)\big) \leq 0.$$

If $U_i(\sigma) = 0$ for some voter i, then σ has log-Nash welfare $-\infty$ and cannot improve upon π, whose log-Nash welfare is finite. Thus, π maximizes Nash welfare.

Nash $\Rightarrow$ price system. Conversely, suppose that π maximizes Nash welfare and $U(\pi) > 0$. Set $\lambda_i = 1/U_i(\pi) > 0$, and let $y = y^{\boldsymbol{\lambda}}$ be the optimal value function for the scalar objective, whose per-round reward is $r_h^{\boldsymbol{\lambda}}(s,c) = \sum_i \lambda_i a_{i,h}(s,c)$.

We first show that π maximizes the scalar objective $\sum_i \lambda_i U_i$. The tangent inequality for $\log$ gives an upper bound on the difference in log welfare, but it does not determine the sign of the corresponding linear difference. We instead use a directional derivative. Fix any policy σ. By convexity of the

attainable utility set (Lemma B.1), $u(t) = (1-t)U(\pi) + tU(\sigma)$ is attainable for every $t \in [0,1]$. Moreover, since $U(\pi) > 0$ and $U(\sigma) \geq 0$, $u_i(t) \geq (1-t)U_i(\pi) > 0$ for every $t \in [0,1)$. Hence, $g(t) = \sum_i \log u_i(t)$ is defined and differentiable on $[0,1)$. Nash optimality gives $g(t) \leq g(0)$, and therefore

$$0 \geq g'(0^+) = \sum_i \frac{U_i(\sigma) - U_i(\pi)}{U_i(\pi)} = \sum_i \lambda_i \big(U_i(\sigma) - U_i(\pi)\big).$$

Since σ was arbitrary, $\pi \in \arg\max_\sigma \sum_i \lambda_i U_i(\sigma)$. This is the same directional argument as the first-order condition in Lemma D.5, specialized to $\mathbf{w} = \mathbf{1}$ and $\phi = \log$.

It remains to verify the price conditions. By Lemma D.2,

$$y^{\boldsymbol{\lambda}}(z_0) - \sum_i \lambda_i U_i(\pi) = \sum_{h,s,c} x_h^\pi(s,c) G_h^{\boldsymbol{\lambda}}(s,c),$$

where $G_h^{\boldsymbol{\lambda}}(s,c) \geq 0$. The left-hand side is zero because π maximizes the scalar objective. Since every term on the right is nonnegative, it follows that $G_h^{\boldsymbol{\lambda}}(s,c) = 0$ whenever $x_h^\pi(s,c) > 0$. This gives equality in the price condition on the support of π. Bellman optimality of $y^{\boldsymbol{\lambda}}$ gives $G_h^{\boldsymbol{\lambda}}(s,c) \geq 0$ at every state-alternative pair, reachable or not, which is exactly the price inequality everywhere.

Finally, by construction, $\lambda_i U_i(\pi) = 1$ for every voter. Thus, π admits the required state-dependent price system. The policy π need not be changed outside its support: only scalar optimality of π is used here, so the off-support completion in Lemma D.3*(ii)* is not needed. □

When transitions are independent of the selected alternative, the continuation term in $p_h(s,c)$ is common to all alternatives at a state, recovering the usual temporal price interpretation. With action-dependent transitions, prices can differ because current alternatives induce different future opportunities.

D.7 Axiomatic characterization of continuation claims

Section 4 derives the update $w_i v_i / u_i$ from dynamic consistency of a separable welfare objective. This leaves a natural question: does the same rule have a justification that does not begin with welfare optimization? We answer this by imposing conditions directly on how proportional claims should change. The resulting characterization uses no welfare objective, and therefore provides an independent foundation for the rule derived in the main text.

A *continuation-claim rule* maps the initial claims $\mathbf{w} \in \mathbb{R}^n_{>0}$, ex-ante utilities $\mathbf{u} \in \mathbb{R}^n_{>0}$, and continuation utilities $\mathbf{v} \in \mathbb{R}^n_{\geq 0}$ to nonnegative continuation claims: $\beta : \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0} \times \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{\geq 0}$. Thus, $\beta_i(\mathbf{w}, \mathbf{u}, \mathbf{v})$ is voter i's claim when their initial claim is w_i, the policy gives them ex-ante utility u_i, and the continuation profile is $\mathbf{v}$. As in Section 4, restricting the rule to these three inputs specifies the scope of the characterization. It does not require every possible notion of state-dependent proportionality to have this form.

The domain above allows arbitrary continuation profiles, with no requirement that $v_i \leq u_i$. This is important because continuation utility is conditional on reaching a state and can exceed ex-ante utility. The following lemma shows that the full domain, as well as the chance mixtures used below, can be realized by actual elections.

Lemma D.7 (Realisability of the axiom domain)**.** *Fix* $\mathbf{u} \in \mathbb{R}^n_{>0}$.

(i) *For every* $\mathbf{v} \in \mathbb{R}^n_{\geq 0}$*, there is a Markov temporal election with nonnegative cardinal utilities, a policy* π*, and a state* z *reached with positive probability such that* $U(\pi) = \mathbf{u}$ *and* $V^\pi(z) = \mathbf{v}$.

(ii) *For every* $\mathbf{v}^1, \mathbf{v}^2 \in \mathbb{R}^n_{\geq 0}$ *and* $p \in (0,1)$*, there is such an election containing a state* z *with no utility at that state, whose successor is* z^1 *with probability* p *and* z^2 *otherwise, such that* $V^\pi(z^j) = \mathbf{v}^j$ *and* $V^\pi(z) = p\mathbf{v}^1 + (1-p)\mathbf{v}^2$*. Hence, every continuation profile and chance decomposition appearing in the axioms below can arise in a Markov temporal election.*

Proof. For part *(i)*, choose $q \in (0,1)$ sufficiently small that $\mathbf{u} - q\mathbf{v} \geq 0$ componentwise, which is possible because $\mathbf{u} > 0$. In round 1, a single alternative gives no utility and leads to z with probability q and to another state z' otherwise. At z, the unique alternative gives utility vector $\mathbf{v}$; at

z', the unique alternative gives $(\mathbf{u} - q\mathbf{v})/(1-q) \geq 0$. There is no decision to make, and the unique policy therefore satisfies $V^\pi(z) = \mathbf{v}$ and $U(\pi) = q\mathbf{v} + (1-q)(\mathbf{u} - q\mathbf{v})/(1-q) = \mathbf{u}$.

For part *(ii)*, set $\bar{\mathbf{v}} = p\mathbf{v}^1 + (1-p)\mathbf{v}^2$ and apply the construction above with continuation profile $\bar{\mathbf{v}}$. Replace the payoff at z by a zero-utility chance transition to z^1 with probability p and to z^2 with probability $1-p$, where the unique alternatives give utility vectors $\mathbf{v}^1$ and $\mathbf{v}^2$, respectively. Then, $V^\pi(z^j) = \mathbf{v}^j$ and $V^\pi(z) = p\mathbf{v}^1 + (1-p)\mathbf{v}^2$, while the ex-ante utility remains $\mathbf{u}$. □

We impose three requirements on β. The first specifies what happens when the policy leaves a voter no continuation utility. This boundary condition concerns the claim induced by the policy being evaluated. It does not say that voter i could obtain no utility under another continuation, only that their claim under the current continuation profile is zero.

Axiom 1 (Boundary)**.** If $v_i = 0$, then $\beta_i(\mathbf{w}, \mathbf{u}, \mathbf{v}) = 0$.

The second requirement on β fixes the claims at the beginning of the election.

Axiom 2 (Root consistency)**.** $\beta(\mathbf{w}, \mathbf{u}, \mathbf{u}) = \mathbf{w}$.

Thus, when the entire ex-ante utility profile remains to be realized, the continuation claims equal the initial claims. Finally, the third requirement on β specifies how claims change through pure chance transitions.

Axiom 3 (Claim conservation)**.** For every $p \in [0,1]$ and $\mathbf{v}^1, \mathbf{v}^2 \in \mathbb{R}^n_{\geq 0}$,

$$\beta(\mathbf{w}, \mathbf{u}, p\mathbf{v}^1 + (1-p)\mathbf{v}^2) = p\beta(\mathbf{w}, \mathbf{u}, \mathbf{v}^1) + (1-p)\beta(\mathbf{w}, \mathbf{u}, \mathbf{v}^2).$$

Claim conservation concerns a state at which no collective decision is made and no utility is obtained before a chance transition. The claim before that transition is the expected claim afterwards: chance alone neither creates nor destroys claims.

These three requirements already determine the continuation-claim rule.

Theorem D.2 (Characterization of continuation claims)**.** *A continuation-claim rule satisfies Axioms 1–3 if and only if $\beta_i(\mathbf{w}, \mathbf{u}, \mathbf{v}) = w_i v_i / u_i$ for every voter i. Moreover, the three axioms are independent.*

Proof. First consider sufficiency. The rule $\beta_i(\mathbf{w}, \mathbf{u}, \mathbf{v}) = w_i v_i / u_i$ is zero whenever $v_i = 0$, equals w_i when $\mathbf{v} = \mathbf{u}$, and is linear in $\mathbf{v}$. It therefore satisfies the boundary, root-consistency, and claim-conservation axioms.

For necessity, fix $(\mathbf{w}, \mathbf{u})$ and write $f(\mathbf{v}) = \beta(\mathbf{w}, \mathbf{u}, \mathbf{v})$. By Lemma D.7, the axioms apply throughout $\mathbb{R}^n_{\geq 0}$ and to every chance mixture used below. We first show that f is linear in $\mathbf{v}$, without assuming continuity.

The boundary axiom gives $f(\mathbf{0}) = \mathbf{0}$. Claim conservation applied to $\mathbf{v}$ and $\mathbf{0}$ therefore yields $f(p\mathbf{v}) = pf(\mathbf{v})$ for every $p \in [0,1]$. Next, evaluating the midpoint of $\mathbf{v}^1$ and $\mathbf{v}^2$ in two ways gives $f(\mathbf{v}^1 + \mathbf{v}^2) = f(\mathbf{v}^1) + f(\mathbf{v}^2)$. Additivity gives $f(k\mathbf{v}) = kf(\mathbf{v})$ for positive integers k, and combining this with the $[0,1]$ homogeneity extends the identity to $f(t\mathbf{v}) = tf(\mathbf{v})$ for every $t \geq 0$. Hence, writing $\mathbf{v} = \sum_j v_j \mathbf{e}_j$, $f(\mathbf{v}) = \sum_j v_j f(\mathbf{e}_j)$.

It follows that $\beta_i(\mathbf{w}, \mathbf{u}, \mathbf{v}) = \sum_j c_{ij} v_j$, where the coefficients depend only on $(\mathbf{w}, \mathbf{u})$. If $j \neq i$, then the ith coordinate of $\mathbf{e}_j$ is zero, so the boundary axiom gives $c_{ij} = 0$. Thus $\beta_i(\mathbf{w}, \mathbf{u}, \mathbf{v}) = c_{ii} v_i$. Root consistency now gives $c_{ii} u_i = w_i$, and therefore $c_{ii} = w_i / u_i$, proving $\beta_i(\mathbf{w}, \mathbf{u}, \mathbf{v}) = w_i v_i / u_i$.

Finally, the axioms are independent. The rule $\beta_i = w_i$ satisfies root consistency and claim conservation but violates the boundary condition. The rule $\beta \equiv 0$ satisfies the boundary condition and claim conservation but violates root consistency. The rule $\beta_i = w_i (v_i / u_i)^2$ satisfies the boundary condition and root consistency but violates claim conservation. □

The characterization uses only these three axioms. It assumes neither a welfare objective nor separability, scale invariance, or continuity of the claim rule. In particular, the ratio $w_i v_i / u_i$ is not obtained here by reproducing the welfare argument of Section 4. It follows directly from how claims behave at the boundary, at the root, and across chance transitions.

Claim conservation is the strongest structural requirement among these three axioms. It is therefore useful to ask what remains if conservation is replaced by more local conditions on how an individual voter's claim updates. We use three such conditions. First, *separability* requires $\beta_i(\mathbf{w}, \mathbf{u}, \mathbf{v}) = w_i g(u_i, v_i)$ for a common continuous function g. Second, *unit invariance* requires $g(cx, cy) = g(x, y)$ for every $c > 0$, so changing the units in which utility is measured does not change the proportional update. Third, *path independence* requires $\beta(\beta(\mathbf{w}, \mathbf{u}, \mathbf{v}), \mathbf{v}, \mathbf{v}') = \beta(\mathbf{w}, \mathbf{u}, \mathbf{v}')$ whenever $\mathbf{u}, \mathbf{v}, \mathbf{v}' > 0$ and the intermediate claim vector is strictly positive. Thus, updating claims through an intermediate continuation profile must agree with updating them directly.

Unlike claim conservation, these requirements do not uniquely select the ratio rule. They identify a one-parameter family.

Theorem D.3 (Isoelastic continuation claims)**.** *Suppose a continuation-claim rule satisfies Axioms 1 and 2, together with separability, unit invariance, and path independence as defined above. Then, for some $p > 0$,*

$$\beta_i(\mathbf{w}, \mathbf{u}, \mathbf{v}) = w_i \left(\frac{v_i}{u_i} \right)^p .$$

These are exactly the continuation weights associated in Theorem 4.1 with the isoelastic welfare transformations satisfying $\phi_p'(x) = x^{-p}$. If claim conservation is imposed as well, then $p = 1$, giving the Nash rule. If the boundary axiom is instead dropped, the limiting case $p = 0$ gives the fixed-weight rule associated with utilitarian welfare.

Proof. By separability, $\beta_i(\mathbf{w}, \mathbf{u}, \mathbf{v}) = w_i g(u_i, v_i)$. Unit invariance implies that g depends only on the ratio of its arguments: writing $\gamma(t) = g(1, t)$ gives $g(x, y) = \gamma(y/x)$. Root consistency gives $\gamma(1) = 1$, while the boundary axiom gives $\gamma(0) = 0$.

We next use path independence to identify γ on $(0, \infty)$. Let $A = \{t > 0 : \gamma(t) > 0\}$. Continuity and $\gamma(1) = 1$ imply that A contains an interval around 1. Fix $a \in A$ and $b > 0$. Choose positive $\mathbf{w}, \mathbf{u}$ and set $v_j = a u_j$ for every voter j. Then every intermediate claim equals $w_j \gamma(a) > 0$, so path independence applies. For a fixed coordinate i, choose $v_i' = ab\, u_i$ and $v_j' = a u_j$ for $j \neq i$. Comparing the ith claim obtained directly with the claim obtained through $\mathbf{v}$ gives $\gamma(ab) = \gamma(a)\gamma(b)$.

This identity implies that A is closed under products and inverses. Since it contains an interval around 1, it must equal $(0, \infty)$. Thus γ is positive and multiplicative everywhere on the positive real line. Define $h(\xi) = \log \gamma(e^\xi)$. Then h is continuous and satisfies $h(\xi + \zeta) = h(\xi) + h(\zeta)$. The continuous solutions of this additive Cauchy equation are linear, so $h(\xi) = p\xi$ and hence $\gamma(t) = t^p$ for $t > 0$. Continuity at zero together with $\gamma(0) = 0$ forces $p > 0$. Therefore $\beta_i(\mathbf{w}, \mathbf{u}, \mathbf{v}) = w_i (v_i/u_i)^p$.

Finally, imposing claim conservation makes the rule linear in the continuation profile by the argument of Theorem D.2. Within the power family, this is possible only for $p = 1$. Conversely, $p = 1$ is linear in $\mathbf{v}$ and therefore satisfies claim conservation. □

The exponent p controls how strongly claims respond to changes in the continuation-to-ex-ante utility ratio. The Nash rule is $p = 1$. Values $0 < p < 1$ give a weaker response, while $p > 1$ gives a stronger response to differences in the ratio. Formally, each member of the family corresponds to an isoelastic welfare transformation with $\phi_p'(x) = x^{-p}$, namely $\phi_p(x) = x^{1-p}/(1-p)$ for $p \neq 1$ and $\phi_1(x) = \log x$. For every $p > 0$, these transformations are strictly increasing and strictly concave. Their behavior at zero differs: the extended value is finite for $p < 1$ and equals $-\infty$ for $p \geq 1$. Accordingly, the positivity of a welfare maximizer is automatic in the latter case under the assumption that each voter can obtain positive utility, whereas for $p < 1$ it remains a separate hypothesis in Theorem 4.1.

The limiting case $p = 0$ leaves claims unchanged, $\beta_i(\mathbf{w}, \mathbf{u}, \mathbf{v}) = w_i$, and corresponds to weighted utilitarian welfare. It violates the boundary axiom and gives exactly the fixed-weight behavior that motivates the comparison with the Nash rule. In the equal-weight case, carrying these fixed claims forward is also the behavior whose population-share interpretation does not hold in Section 3.

The welfare and axiomatic characterizations therefore answer different questions. Section 4 begins with an optimized separable welfare objective and asks which continuation weights preserve its

optimality after conditioning. The characterization here assumes no welfare objective and instead asks how proportional claims themselves should evolve. Both arguments restrict attention to rules defined by $(\mathbf{w}, \mathbf{u}, \mathbf{v})$; neither set of assumptions implies the other. Their agreement on the Nash update $w_i v_i / u_i$ therefore shows that the same rule arises both from dynamic consistency of welfare and from direct requirements on the evolution of proportional claims.

E Omitted Proofs and Additional Results for Section 5

E.1 Proof of Theorem 5.1

Theorem 5.1. *Let π be $\boldsymbol{\lambda}$-tight for $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$. At any state z, no $\varnothing \neq S \subseteq N$ and continuation policy σ satisfy $\alpha_z(S) V_i^\sigma(z) \geq V_i^\pi(z)$ for every $i \in S$, with strict inequality for some $i \in S$.*

Proof. Fix a state z and a nonempty coalition $S \subseteq N$. Recall that $B(z) = \sum_{i \in N} b_i(z)$. We consider three cases:

1. *$B(z) = 0$:* Then $b_i(z) = 0$ for every voter i. Since $b_i(z) = \lambda_i V_i^\pi(z)$, with $\lambda_i > 0$ and $V_i^\pi(z) \geq 0$, we have $V_i^\pi(z) = 0$ for every voter i. By definition, $\alpha_z(S) = 0$ for every coalition S. The inequalities in Theorem 5.1 would therefore require $0 \cdot V_i^\sigma(z) = 0 \geq 0$ for every $i \in S$, with $0 > 0$ for at least one member, which is impossible.

2. *$B(z) > 0$ and $b(S) = 0$:* Again, nonnegativity and $\lambda_i > 0$ imply $\alpha_z(S) = 0$ and $V_i^\pi(z) = 0$ for every $i \in S$. Since $\alpha_z(S) = 0$, a weak block would again require $0 > 0$ for at least one member of S, which is impossible.

3. *$B(z) > 0$ and $b(S) > 0$:* Now, $\alpha_z(S) = b(S)/B(z) > 0$. Suppose that a continuation policy σ satisfies $\alpha_z(S) V_i^\sigma(z) \geq V_i^\pi(z)$ for every $i \in S$, with strict inequality for at least one member. Multiplying by $\lambda_i / \alpha_z(S) > 0$ and summing over S gives

$$\sum_{i \in S} \lambda_i V_i^\sigma(z) > \frac{b(S)}{\alpha_z(S)} = B(z).$$

   On the other hand, nonnegativity and the definition of $y^{\boldsymbol{\lambda}}(z)$ imply

$$\sum_{i \in S} \lambda_i V_i^\sigma(z) \leq \sum_{i \in N} \lambda_i V_i^\sigma(z) \leq y^{\boldsymbol{\lambda}}(z) = B(z),$$

   where the last equality follows from $\boldsymbol{\lambda}$-tightness. This is a contradiction. Hence, no such coalition and continuation policy exist. □

E.2 Proof of Corollary E.1

Corollary E.1. *Let π be $\boldsymbol{\lambda}$-tight for $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$.* (i) *At every state z, there are no $\varnothing \neq S \subseteq N$ and continuation policy σ with $\alpha_z(S) V_i^\sigma(z) > V_i^\pi(z)$ for every $i \in S$.* (ii) *If $\lambda_i = 1/U_i(\pi)$ for every i, then $\alpha_{z_0}(S) = |S|/n$ for every coalition S, and π lies in the fractional Core of Definition 2.1.*

Proof. For part *(i)*, an all-strict block satisfies the weak inequalities of Theorem 5.1 and, since S is nonempty, has at least one strict inequality. Thus, Theorem 5.1 rules it out.

For part *(ii)*, suppose $\lambda_i = 1/U_i(\pi)$ for every voter i. These weights are well defined because $U_i(\pi) > 0$. The identity $V_i^\pi(z_0) = U_i(\pi)$ holds because round 0 has a single alternative with no approvals. Thus, $b_i(z_0) = \lambda_i V_i^\pi(z_0) = 1$ for every voter i. Hence, $B(z_0) = n > 0$ and $\alpha_{z_0}(S) = |S|/n$ for every coalition S. Applying part *(i)* at z_0 and using $V_i^\sigma(z_0) = U_i(\sigma)$ shows that there are no nonempty $S \subseteq N$ and policy σ such that

$$\frac{|S|}{n} U_i(\sigma) > U_i(\pi) \qquad \text{for every voter } i \in S.$$

This is exactly the fractional-Core condition of Definition 2.1. □

The reciprocal normalization gives voters equal entitlements at the initial state. This property is distinct from $\boldsymbol{\lambda}$-tightness. For any policy that gives every voter positive utility, $\lambda_i = 1/U_i(\pi)$ implies $b_i(z_0) = 1$, regardless of whether the policy is $\boldsymbol{\lambda}$-tight. More generally, since $V_i^\pi(z_0) = U_i(\pi)$, whenever $B(z_0) > 0$ the initial-state share of coalition S is $\alpha_{z_0}(S) = \big(\sum_{i\in S} \lambda_i U_i(\pi)\big)/\big(\sum_{j\in N} \lambda_j U_j(\pi)\big)$. Thus, the initial-state shares are determined by the normalized quantities $\lambda_i U_i(\pi)$, and the reciprocal normalization makes these quantities equal across voters and gives $|S|/n$. For Nash-optimal policies, the reciprocal weights have an additional optimality property described by the state-dependent price system of Theorem D.1.

E.3 Representation guarantees

Continuation stability rules out coalitional deviations, but it does not say how much representation a coalition receives under the reference policy. We next derive two stronger statements: one bounds the coalition's average continuation utility, while the other bounds how much of its entitlement can belong to poorly represented members.

For a coalition S, let $K_S(z) = \max_\sigma \min_{i\in S} V_i^\sigma(z)$ be the largest continuation utility that one policy can guarantee simultaneously to every member of S. Both representation guarantees below rely on the inequality $K_S(z)\sum_{i\in S}\lambda_i \le B(z)$, which we prove below. A policy attaining $K_S(z)$ gives every member of S at least $K_S(z)$, while no continuation can exceed $y^{\boldsymbol{\lambda}}(z) = B(z)$ for the scalarized objective.

Lemma E.1 (Coalitional budget lemma)**.** *For every state z and coalition S, $K_S(z)\sum_{i\in S}\lambda_i \le B(z)$.*

Proof. Let σ attain $K_S(z)$. Such a continuation exists because the attainable continuation set is compact and the function $\min_{i\in S} V_i^\sigma(z)$ is continuous. By definition, $V_i^\sigma(z) \ge K_S(z)$ for every $i \in S$. Therefore,

$$K_S(z)\sum_{i\in S}\lambda_i \le \sum_{i\in S}\lambda_i V_i^\sigma(z) \le \sum_{j\in N}\lambda_j V_j^\sigma(z) \le y^{\boldsymbol{\lambda}}(z) = B(z).$$

The second inequality uses nonnegativity of continuation utilities, the third follows from the definition of $y^{\boldsymbol{\lambda}}(z)$, and the final equality follows from $\boldsymbol{\lambda}$-tightness. □

The first consequence bounds the coalition's representation on average.

Theorem E.1 (Weighted-mean representation)**.** *For every state z and coalition S, define the $\boldsymbol{\lambda}$-weighted arithmetic mean of the coalition's continuation utilities as $\big(\sum_{i\in S}\lambda_i V_i^\pi(z)\big)/\big(\sum_{i\in S}\lambda_i\big)$. Then, this mean is at least $\alpha_z(S)K_S(z)$.*

Proof. If $K_S(z) = 0$ or $B(z) = 0$, the claim is immediate. Otherwise, Lemma E.1 gives $B(z)/K_S(z) \ge \sum_{i\in S}\lambda_i > 0$. Hence,

$$\alpha_z(S)K_S(z) = \frac{b(S)}{B(z)/K_S(z)} \le \frac{b(S)}{\sum_{i\in S}\lambda_i},$$

and the right-hand side is the $\boldsymbol{\lambda}$-weighted arithmetic mean in Theorem E.1, since $b(S) = \sum_{i\in S}\lambda_i V_i^\pi(z)$.

When every member of S is active, $b_i(z)/V_i^\pi(z) = \lambda_i$, so the same quantity can also be written as the b-weighted harmonic mean $b(S)/\sum_{i\in S}\big(b_i(z)/V_i^\pi(z)\big)$. □

When all members of S are active, the mean in Theorem E.1 can also be written as the b-weighted harmonic mean $b(S)/\sum_{i\in S}(b_i(z)/V_i^\pi(z))$. It should not be confused with the b-weighted arithmetic mean $\sum_{i\in S} b_i(z)V_i^\pi(z)/b(S)$; for example, with $\boldsymbol{\lambda} = (1,1)$ and continuation utilities $(1,4)$, these are 2.5 and 3.4, respectively. Since an arithmetic mean cannot exceed the largest value, Theorem E.1 also implies that some coalition member receives at least $\alpha_z(S)K_S(z)$. The mean guarantee is stronger than this conclusion about one member: for equal weights and utilities $(0, 1.5\theta)$, for instance, some member exceeds θ while the mean is only 0.75θ.

A high average, however, can result from high utility for only a few members. The next result therefore bounds the total entitlement of underrepresented members. For $\tau \in (0, 1]$, call voter $i \in S$ *underrepresented* at threshold τ if $V_i^\pi(z) < \tau\alpha_z(S)K_S(z)$.

Theorem E.2 (Quantile representation)**.** *Fix $\tau \in (0, 1]$ and a coalition S with $b(S) > 0$, and let $T = \{i \in S : V_i^\pi(z) < \tau\alpha_z(S)K_S(z)\}$. Then, $b(T) < \tau b(S)$.*

Proof. Let $\theta = \tau\alpha_z(S)K_S(z)$. If $T = \varnothing$, then $b(T) = 0 < \tau b(S)$ because $b(S) > 0$. Suppose therefore that $T \neq \varnothing$. Then some $i \in T$ satisfies $V_i^\pi(z) < \theta$. Since continuation utilities are nonnegative, $\theta > 0$, and hence $K_S(z) > 0$.

Since $T \subseteq S$, guaranteeing a given utility to every member of T is no harder than guaranteeing it to every member of S, so $K_T(z) \geq K_S(z)$. Applying Lemma E.1 to T therefore gives

$$\sum_{i\in T} \lambda_i \leq \frac{B(z)}{K_T(z)} \leq \frac{B(z)}{K_S(z)}.$$

By the definition of T, every $i \in T$ satisfies $V_i^\pi(z) < \theta$. Hence

$$b(T) = \sum_{i\in T} \lambda_i V_i^\pi(z) < \theta \sum_{i\in T} \lambda_i \leq \tau\alpha_z(S)B(z) = \tau b(S),$$

where the first inequality is strict because $T \neq \varnothing$ and $\lambda_i > 0$ for every i. □

At the initial state under reciprocal weights, every voter has unit entitlement. Thus, Theorem E.2 says that fewer than $\tau|S|$ members of S can receive less than $\tau(|S|/n)K_S(z_0)$. At $\tau = 1$, this implies that some coalition member receives at least $(|S|/n)K_S(z_0)$. For $\tau < 1$, the result also bounds the fraction of the coalition that can fall below a scaled representation benchmark, beyond identifying one well-represented member. The statements hold simultaneously for all $\tau \in (0, 1]$. The following result shows that the constant is asymptotically tight.

Proposition E.1 (Tightness of the quantile constant)**.** *For every $n \geq 2$ and $\delta > 0$, there is a deterministic approval election, a Nash-optimal policy, a coalition S, and $\tau \in (0, 1)$ such that $b(T) > (1 - \delta)\tau b(S)$.*

Proof. Fix an integer $M \geq n$, to be chosen below, and let $H = M + 1$. In round 1, alternative A is approved by voters $1, \ldots, n - 1$ and leads to M forced rounds with no approvals. Alternative B is approved by nobody and leads to M forced rounds, each approved only by voter n.

If A is selected with probability t, the resulting utility profile is

$$\big(t, \ldots, t, (1 - t)M\big).$$

Maximizing Nash welfare is therefore equivalent to maximizing $(n - 1)\log t + \log((1 - t)M)$, whose unique maximizer is $t^\star = (n - 1)/n$. Thus,

$$U_i(\pi) = \frac{n-1}{n} \quad (i < n), \qquad U_n(\pi) = \frac{M}{n}.$$

There are no choices after the first round, so the Nash-optimal policy is $\boldsymbol{\lambda}$-tight. Under the reciprocal normalization, $b_i(z_0) = 1$ for every voter and hence $B(z_0) = n$.

Take $S = N$. Then, $\alpha_{z_0}(S) = 1$, and the largest utility that a single policy can guarantee to every voter is

$$K_N(z_0) = \max_{t\in[0,1]} \min\{t, (1 - t)M\} = \frac{M}{M + 1}.$$

Let $\tau_M = \frac{(n-1)(M+1)}{nM}$. Since $M \geq n$, we have $\tau_M < 1$. Moreover, $\tau_M > (n - 1)/n$, and $M/n \geq M/(M + 1) = K_N(z_0)$.

Choose any $\tau \in (\tau_M, 1)$. For each voter $i < n$,

$$U_i(\pi) = \frac{n-1}{n} < \tau\frac{M}{M+1} = \tau\alpha_{z_0}(S)K_N(z_0),$$

whereas

$$U_n(\pi) = \frac{M}{n} \geq K_N(z_0) > \tau K_N(z_0).$$

Hence, the underrepresented set is exactly $T = \{1, \ldots, n-1\}$. Since every voter has unit entitlement at z_0, $b(T) = n - 1$ and $b(S) = n$, so

$$\frac{b(T)}{\tau b(S)} = \frac{n-1}{\tau n}.$$

As $\tau \downarrow \tau_M$, this ratio converges to $M/(M+1)$. Choose $M \geq n$ large enough that $M/(M+1) > 1 - \delta$, and then choose $\tau > \tau_M$ sufficiently close to τ_M that $b(T)/(\tau b(S)) > 1 - \delta$. This proves the claim. □

F Omitted Proofs and Additional Results for Section 6

F.1 Proof of Theorem F.1

For the complexity result below, we use the standard finite binary encoding of the election and reference policy. The utility values $a_{i,h}(s, c)$, transition probabilities $P_h(s' \mid s, c)$, initial probabilities $\mu(s)$, and action probabilities $\pi_h(c \mid s)$ of the reference policy are binary-encoded rationals. The integers n, H, and k are encoded in binary, while the layered state sets $\mathcal{S}_h$ and alternative sets $C_h(s)$ are listed explicitly. Let L denote the total encoding length. Since every layer is nonempty and explicitly represented, $H \leq L$.

The reference policy is part of the input and is represented explicitly by its rational action probabilities. Consequently, its utilities $U_i(\pi)$ are rational numbers of polynomial encoding length and can be computed exactly in polynomial time by backward induction. The decision problem is therefore a finite rational decision problem, and comparisons appearing in the fractional blocking condition can be performed exactly. As a deterministic policy is rationally described, restricting the input policy to rational descriptions does not weaken the hardness result. The hardness constructions below use rational data throughout; in particular, the one-round cardinal reduction uses a deterministic reference policy, so the rational-policy restriction does not weaken the hardness result.

Theorem F.1. *Deciding, given a Markov temporal election with rational data, a rationally described positive-utility policy π, and an integer k, whether some coalition of size exactly k has a fractionally improving deviation is NP-complete. Hardness holds already for $H = 1$ with rational cardinal values and a deterministic reference policy, and for stochastic approval elections with $H = 2n + 1$ and $O(nd)$ states.*

Proof. **Membership.** By the encoding convention above, every numerical entry in the instance, including the approval or cardinal values, the transition probabilities, and the action probabilities of π, is a binary-encoded rational. Thus, all arithmetic below can be carried out exactly.

Step 1: $U_i(\pi)$ is rational with polynomial encoding length. List the positive denominators of all rational numerical entries of the election and of π as $d_1, \ldots, d_M$, and let $D = \prod_{j \leq M} d_j$. Every input rational then has a representation whose denominator divides D, and

$$\operatorname{size}(D) \leq \sum_{j \leq M} \operatorname{size}(d_j) \leq L.$$

The first inequality follows because the bit length of a product is at most the sum of the bit lengths of its factors, and the second because every d_j is explicitly encoded in the instance. The total input length bounds $\operatorname{size}(D)$, not the encoding length of any single entry. With pairwise coprime denominators, for example, $\operatorname{lcm}(d_1, \ldots, d_M)$ can already have about $\sum_j \operatorname{size}(d_j)$ bits, so no corresponding bound in terms of a single denominator is available.

We next prove by downward induction on h that $V_{i,h}(s)$ has a representation whose denominator divides $D^{2(H-h+1)}$. At $h = H + 1$, the exponent is 0 and $V_{i,H+1} \equiv 0$. Suppose the claim holds at $h + 1$. In $V_{i,h}(s) = \sum_c \pi_h(c \mid s) \left[a_{i,h}(s, c) + \sum_{s'} P_h(s' \mid s, c) V_{i,h+1}(s')\right]$, each product $P_h(s' \mid s, c) V_{i,h+1}(s')$ has denominator dividing $D^{2(H-h)+1}$, and summing over s' preserves this property. The value $a_{i,h}(s, c)$ has denominator dividing D, which in turn divides $D^{2(H-h)+1}$, so the

entire bracket has denominator dividing $D^{2(H-h)+1}$. Multiplication by $\pi_h(c \mid s)$ adds at most one further factor of D, giving denominator dividing $D^{2(H-h)+2} = D^{2(H-h+1)}$; summing over c again preserves the bound.

Round 0 has a single alternative with no approvals and transition distribution μ, so $U_i(\pi) = \sum_s \mu(s) V_{i,1}(s)$ has denominator dividing $D \cdot D^{2H} = D^{2H+1}$. Hence,

$$\operatorname{size}\big(\text{denominator of } U_i(\pi)\big) \le (2H+1)\operatorname{size}(D) \le (2H+1)L.$$

The numerators require no separate recursion. Since values are nonnegative and bounded, $0 \le V_{i,h}(s) \le H\bar{a}$. Therefore, $N_{i,h}(s) = V_{i,h}(s) D^{2(H-h+1)}$ is an integer satisfying $0 \le N_{i,h}(s) \le H\bar{a}D^{2(H-h+1)}$, and therefore has encoding length at most $\operatorname{size}(H\bar{a}) + 2H\operatorname{size}(D) + O(1) = O(HL)$. The same argument applies to $U_i(\pi)D^{2H+1}$. Thus, every intermediate quantity has $O(HL) = O(L^2)$ bits, and the recursion uses $O(H|\mathcal{S}|^2|C|)$ arithmetic operations per voter. It follows that every $U_i(\pi)$ can be computed exactly in time polynomial in L and has encoding length polynomial in L.

Step 2: checking a proposed blocking pair. A witness is a proposed blocking pair (S, x) with $S \subseteq N$, $|S| = k$, and rational $\mathbf{x} \in \mathcal{X}$ satisfying $\frac{k}{n} U_i(x) > U_i(\pi)$ for every voter $i \in S$. Here, $\mathcal{X}$ is the occupancy polytope of Lemma B.1. Since $U_i(x) = \langle x, a_i \rangle$ is linear with rational coefficients and, under the encoding convention above, $\mathcal{X}$ is described by finitely many rational linear equalities together with $x \ge 0$, with total encoding length polynomial in L (Lemma B.1), membership in $\mathcal{X}$ and all the strict inequalities can be checked by exact rational arithmetic in time polynomial in the sizes of the instance and of x. This check does not require the additional variable θ introduced next.

Step 3: a polynomial-size witness exists whenever a witness exists. It remains to show that a block satisfying the strict inequalities can be described by a rational pair with polynomial encoding length. Fix a coalition S and consider

$$\max\Big\{\theta : x \in \mathcal{X},\ \tfrac{k}{n} U_i(x) - U_i(\pi) \ge \theta\ (i \in S),\ -\Gamma \le \theta \le 1\Big\}, \qquad \Gamma := \max_{i \in N} U_i(\pi).$$

The lower bound on θ makes the feasible region a *polytope* without changing whether its optimum is positive. Its data are rational and have encoding length polynomial in L, by Step 1 and Lemma B.1.

The polytope is nonempty. Since $\mathcal{X} \neq \varnothing$, take any $\mathbf{x} \in \mathcal{X}$ and set $\theta = -\Gamma$. Nonnegativity gives $\frac{k}{n} U_i(x) \ge 0$, while $U_i(\pi) \le \Gamma$, and hence $\frac{k}{n} U_i(x) - U_i(\pi) \ge -\Gamma$ for every $i \in S$. It is also bounded: $\mathcal{X}$ is bounded and $\theta \in [-\Gamma, 1]$. Thus, the linear objective attains its maximum at a vertex.

Every such vertex has polynomial encoding length. Indeed, a vertex is the unique solution of $A'z = b'$ for some nonsingular square subsystem of the defining constraints. By Cramer's rule (Kosinski, 2001; Cramer, 1750), each coordinate is a quotient of determinants of submatrices of the rational input data, and Hadamard's inequality bounds the encoding lengths of these determinants polynomially in the encoding length of the data. This is the standard size bound for rational polyhedra (Schrijver, 1986).

The optimum is strictly positive exactly when S blocks. If S blocks via some occupancy vector x, then the same x together with $\theta = \min\{1, \min_{i \in S}(\frac{k}{n} U_i(x) - U_i(\pi))\} > 0$ is feasible. Conversely, if an optimal solution $(x^\star, \theta^\star)$ has $\theta^\star > 0$, then $\frac{k}{n} U_i(x^\star) - U_i(\pi) \ge \theta^\star > 0$ for every $i \in S$. Choosing $x^\star$ from an optimal vertex therefore gives a blocking pair $(S, x^\star)$ with polynomial encoding length, making $(S, x^\star)$ a polynomial-size witness. Hence, the problem belongs to NP.

Hardness. We reduce from OPEN HEMISPHERE: given $\mathbf{g}_1, \ldots, \mathbf{g}_n \in \mathbb{Z}^d$ and k, decide whether there exists $\mathbf{v} \in \mathbb{R}^d$ such that $\mathbf{g}_i \cdot \mathbf{v} > 0$ for at least k indices. This problem is NP-complete when d is part of the input (Johnson & Preparata, 1978; Garey & Johnson, 1979); its optimization version is MAX-FS for homogeneous strict inequalities (Amaldi & Kann, 1995).

For each row, let $M_i = \max\{1, \max_j |g_{ij}|\}$ and set $\tilde{\mathbf{g}}_i = \mathbf{g}_i / M_i$. This rescales each row by a positive rational number, so $|\tilde{g}_{ij}| \le 1$ and $\operatorname{sign}(\tilde{\mathbf{g}}_i \cdot \mathbf{v}) = \operatorname{sign}(\mathbf{g}_i \cdot \mathbf{v})$ for every x. Hence, the normalized instance has the same answer. The term $\max\{1, \cdot\}$ is necessary because OPEN HEMISPHERE does not require every row to be nonzero, so dividing a zero row by $\max_j |g_{ij}|$ would be undefined. A zero row satisfies no strict homogeneous inequality before or after normalization, so it is not counted in either case. Write $\mathbf{g}$ for $\tilde{\mathbf{g}}$ from now on.

Set $a = \frac{1}{2}$ and $\eta = \frac{1}{4}$. Construct a one-round election with alternatives c_0 and $c_j^{\pm}$, and cardinal values $u_i(c_0) = ak/n$ and $u_i(c_j^{\pm}) = a \pm \eta g_{ij}$. The latter values lie in $[\frac{1}{4}, \frac{3}{4}]$, while $u_i(c_0) = k/(2n) \in (0, \frac{1}{2}]$. Thus, all values are rational and lie in $[0, 1]$, and the reference policy has positive utility for every voter. Let π be the deterministic policy selecting c_0, so $U_i(\pi) = ak/n$ for every i.

A coalition S with $|S| = k$ blocks via a lottery q exactly when $\frac{k}{n}U_i(q) > U_i(\pi) = a\frac{k}{n}$ for every $i \in S$, equivalently, when $U_i(q) > a$ for every $i \in S$. For a lottery q, define $v_j = q(c_j^+) - q(c_j^-)$. Then,

$$U_i(q) = q_0\, a\tfrac{k}{n} + \sum_j \big[q(c_j^+)(a + \eta g_{ij}) + q(c_j^-)(a - \eta g_{ij})\big] = a - aq_0\big(1 - \tfrac{k}{n}\big) + \eta\, \mathbf{g}_i \cdot \mathbf{v}.$$

We next prove that a coalition S of size k blocks via q if and only if $\mathbf{g}_i \cdot \mathbf{v} > 0$ for at least k indices.

(⇒) Suppose a coalition S of size k blocks via q. For every $i \in S$,

$$\eta \mathbf{g}_i \cdot \mathbf{v} = U_i(q) - a + aq_0\big(1 - \tfrac{k}{n}\big) > 0,$$

because $U_i(q) > a$, $q_0 \geq 0$, and $k \leq n$. Since $\eta > 0$, we have $\mathbf{g}_i \cdot \mathbf{v} > 0$ for every $i \in S$. Thus, at least k rows satisfy the strict halfspace inequality.

(⇐) Conversely, suppose $\mathbf{g}_i \cdot \mathbf{v} > 0$ for at least k indices. Then, $\mathbf{v} \neq 0$, so we may rescale it by the positive factor $1/\sum_j |v_j|$. Homogeneity preserves all the strict inequalities, and after rescaling $\sum_j |v_j| = 1$. Set $q_0 = 0$ and $q(c_j^+) = (v_j)^+$, $q(c_j^-) = (-v_j)^+$. These probabilities are nonnegative and sum to 1, so they define a lottery, and $q(c_j^+) - q(c_j^-) = v_j$. Consequently, $U_i(q) = a + \eta \mathbf{g}_i \cdot \mathbf{x} > a$ for every satisfied row. There are at least k such voters, so any k of them form a blocking coalition of size exactly k.

The two formulations therefore agree: a solution satisfying at least k rows produces a blocking coalition of size exactly k, and a blocking coalition of size exactly k produces k satisfied rows. The argument also covers $k = n$. In that case, $1 - \frac{k}{n} = 0$, so the forward implication becomes $\eta \mathbf{g}_i \cdot \mathbf{v} = U_i(q) - a > 0$, which remains strict because blocking is strict; the converse is unchanged. Allowing $q_0 > 0$ also preserves the argument, since it only strengthens the forward implication. All encoding lengths remain polynomial, as each M_i has encoding length polynomial in that of $\mathbf{g}_i$. This proves NP-hardness already for horizon $H = 1$ with rational cardinal values and a deterministic reference policy.

Approval-only hardness. It remains to prove hardness for approval elections. Replace each rational value $u_i(c) \in [0, 1]$ by the following stochastic chain. After c is selected in round 1, which has no approvals, process voters $i = 1, \ldots, n$ in sequence. At round $2i$, a state with a single alternative approved by nobody moves to a "yes" state with probability $u_i(c)$ and to a "no" state otherwise. At round $2i + 1$, the unique alternative in the "yes" state is approved only by voter i, while the unique alternative in the "no" state is approved by nobody; both then continue to the next voter's draw.

The resulting election has horizon $H = 2n + 1$ and $1 + 3n|C| + 1 = O(nd)$ states. Its transition probabilities are the values $u_i(c)$ themselves and therefore have polynomial encoding length. Conditional on selecting c, voter i receives expected utility exactly $u_i(c)$, while no other voter receives utility in rounds $2i$ and $2i + 1$. Since every state after the first offers only one alternative, the only policy choice is the initial lottery over C. Thus, the approval election has exactly the same attainable utility profiles and preserves the reduction. □

The reduction in Theorem F.1 fixes the coalition size $|S| = k$. In the constructed instance, the grand coalition blocks, so the reduction does not establish the complexity of deciding fractional-Core membership when coalition size is unrestricted. We leave that problem open.

F.2 Reducing a block to its active part

Lemma F.1 (Blocks reduce to their active part)**.** *Fix a state z with $B(z) > 0$, so $A(z) \neq \varnothing$ and $\mathrm{Dev}_z(\pi)$ is defined. For $\eta \geq 0$, say that a coalition S blocks at z* with factor *η via a continuation policy σ if $\alpha_z(S)V_i^{\sigma}(z) > \eta V_i^{\pi}(z)$ for every voter $i \in S$. Here, S is required to be nonempty, since the empty set would satisfy these inequalities without improving any voter. Then:*

(i) no coalition $S \subseteq N \setminus A(z)$ blocks at any factor;

(ii) if S blocks at factor η via σ, then $T = S \cap A(z)$ is nonempty and blocks at the same factor via the same σ;

(iii) the maximum defining $\mathrm{Dev}_z(\pi)$ is attained, and for every $\eta \geq 0$, some coalition blocks at factor η if and only if $\eta < \mathrm{Dev}_z(\pi)$. Consequently, the set of blocking factors is $[0, \mathrm{Dev}_z(\pi))$; in particular, because blocking is strict, $\eta = \mathrm{Dev}_z(\pi)$ is not itself a blocking factor;

(iv) $\mathrm{Dev}_z(\pi) \geq 1$. Hence, $\mathrm{Dev}_z(\pi) = 1$ if and only if no coalition blocks at factor 1, equivalently, if and only if π has no all-strict continuation block at z in the sense of Corollary E.1(i).

Proof. Since $B(z) = \sum_i \lambda_i V_i^\pi(z) > 0$, while $\lambda_i > 0$ and $V_i^\pi(z) \geq 0$ for every i, at least one voter i satisfies $V_i^\pi(z) > 0$. Hence, $A(z) \neq \varnothing$.

(i) If $S \subseteq N \setminus A(z)$, then $V_i^\pi(z) = 0$ for every $i \in S$. Thus, $b(S) = 0$ and, since $B(z) > 0$, $\alpha_z(S) = 0$. The blocking inequalities would therefore require $0 = \alpha_z(S) V_i^\sigma(z) > \eta V_i^\pi(z) = 0$ for every $i \in S$, which is impossible for any $\eta \geq 0$.

(ii) By part *(i)*, a blocking coalition S cannot consist entirely of inactive voters, so $T = S \cap A(z) \neq \varnothing$. Inactive voters have $b_i(z) = 0$, and therefore $b(T) = b(S)$ and $\alpha_z(T) = \alpha_z(S)$. The blocking inequalities for the members of $T \subseteq S$ are thus unchanged, so T blocks at the same factor via the same continuation policy σ.

(iii) For every nonempty $T \subseteq A(z)$, define

$$d_z(T, \sigma) = \alpha_z(T) \min_{i \in T} \frac{V_i^\sigma(z)}{V_i^\pi(z)}.$$

The denominators are positive because $T \subseteq A(z)$, and by definition

$$\mathrm{Dev}_z(\pi) = \max_{\varnothing \neq T \subseteq A(z)} \max_\sigma d_z(T, \sigma).$$

There are finitely many nonempty subsets $T \subseteq A(z)$. For each fixed T, $d_z(T, \sigma)$ depends continuously on the continuation-utility profile attained by σ at z. Since the set of attainable continuation-utility profiles is compact, each inner maximum is attained, and therefore so is the outer maximum.

Suppose first that some coalition S blocks at factor η via σ. By part *(ii)*, $T = S \cap A(z)$ also blocks at factor η via σ. Dividing its blocking inequalities by $V_i^\pi(z) > 0$ gives $\alpha_z(T) V_i^\sigma(z) / V_i^\pi(z) > \eta$ for every $i \in T$. Hence, $d_z(T, \sigma) > \eta$, and therefore $\mathrm{Dev}_z(\pi) > \eta$.

Conversely, suppose $\eta < \mathrm{Dev}_z(\pi)$, and let $(T^\star, \sigma^\star)$ attain the maximum. For every $i \in T^\star$,

$$\alpha_z(T^\star) \frac{V_i^{\sigma^\star}(z)}{V_i^\pi(z)} \geq d_z(T^\star, \sigma^\star) = \mathrm{Dev}_z(\pi) > \eta.$$

Multiplying by $V_i^\pi(z) > 0$ shows that $T^\star$ blocks at factor η via $\sigma^\star$. This proves the equivalence. In particular, a block at factor $\mathrm{Dev}_z(\pi)$ would imply $\mathrm{Dev}_z(\pi) > \mathrm{Dev}_z(\pi)$, so the threshold itself is not a blocking factor.

(iv) Inactive voters contribute zero entitlement, so $b(A(z)) = B(z)$ and hence $\alpha_z(A(z)) = 1$. Taking $T = A(z)$ and $\sigma = \pi$ gives

$$d_z(A(z), \pi) = \min_{i \in A(z)} \frac{V_i^\pi(z)}{V_i^\pi(z)} = 1.$$

Therefore, $\mathrm{Dev}_z(\pi) \geq 1$. Applying part *(iii)* with $\eta = 1$ shows that an all-strict continuation block exists exactly when $\mathrm{Dev}_z(\pi) > 1$. Since $\mathrm{Dev}_z(\pi) \geq 1$, it follows that $\mathrm{Dev}_z(\pi) = 1$ if and only if π has no all-strict continuation block at z. □

Part *(ii)* does not imply that inactive voters cannot belong to a blocking coalition. An inactive voter may gain positive utility from a continuation. The result shows that removing inactive members preserves the block, which justifies restricting the definition of $\mathrm{Dev}_z(\pi)$ to coalitions of active voters. Part *(iv)* explains why $\mathrm{Dev}_z(\pi) = 1$ is the natural benchmark for the absence of an all-strict continuation block: the coalition of all active voters, by following π itself, attains deviation ratio 1, and therefore blocks at every factor $\eta < 1$.

F.3 THE TIGHT TWO-SIDED CERTIFICATE

The main-text version of Theorem 6.1 gives an upper estimate of the scalar certificate using only the total continuation entitlement and the smallest individual entitlement. We now derive a tighter factor that depends on the entire entitlement profile. The following ordering lemma gives the largest value of the expression that appears in the proof.

Lemma F.2 (Ascending order gives the maximum)**.** *Let $b_1, \dots, b_m > 0$. For a permutation ρ, define*

$$L_\rho(b) = \sum_{k=1}^{m} \frac{b_{\rho(k)}}{\sum_{j \le k} b_{\rho(j)}}.$$

This quantity is maximized by ordering the weights in ascending order. Writing $b_{[1]} \le \cdots \le b_{[m]}$ and $W_k = \sum_{j=1}^{k} b_{[j]}$, the maximum is therefore $L(b) = \sum_{k=1}^{m} b_{[k]}/W_k$, and

$$1 \le L(b) \le 1 + \ln \frac{\sum_j b_j}{\min_j b_j}.$$

If all weights are equal, then $L(b) = \mathrm{H}_m$.

Proof. It suffices to consider swapping two adjacent weights, since this changes only the corresponding two terms. Suppose the preceding weights sum to P and the adjacent weights satisfy $a > c > 0$. The value obtained by placing a before c, minus the value obtained by placing c before a, is

$$(a-c)\left[\frac{P}{(P+a)(P+c)} - \frac{1}{P+a+c}\right].$$

Since $(P+a)(P+c) = P(P+a+c) + ac > P(P+a+c)$, the expression in brackets is negative. Thus, placing the smaller weight first weakly increases the expression. Repeated adjacent swaps therefore show that ascending order gives the maximum.

For this order, the first term is 1. For every $k \ge 2$,

$$\frac{b_{[k]}}{W_k} = 1 - \frac{W_{k-1}}{W_k} \le \ln \frac{W_k}{W_{k-1}}.$$

Summing gives

$$L(b) \le 1 + \ln \frac{W_m}{W_1} = 1 + \ln \frac{\sum_j b_j}{\min_j b_j}.$$

The lower bound $L(b) \ge 1$ follows from the first term. If all weights are equal, then $b_{[k]}/W_k = 1/k$, so $L(b) = \mathrm{H}_m$. □

We can now state the tighter version of the two-sided certificate.

Theorem F.2 (Full statement of Theorem 6.1: Two-sided certificate)**.** *Let $B(z) > 0$. Then $\mathrm{Dev}_z(\pi) \le \mathrm{PF}_z(\pi)$. If every voter is active at z, then $\mathrm{PF}_z(\pi) \le L_z \mathrm{Dev}_z(\pi)$, where, writing the entitlements at z in ascending order $b_{[1]} \le \cdots \le b_{[n]}$ with prefix sums $W_k = \sum_{\ell=1}^{k} b_{[\ell]}$,*

$$L_z = \sum_{k=1}^{n} \frac{b_{[k]}}{W_k} \le 1 + \ln \frac{B(z)}{\min_{i \in N} b_i(z)}.$$

At the initial state under the reciprocal normalization, $b_i(z_0) = 1$ for every voter and $L_{z_0} = \mathrm{H}_n$.

Proof. For the first inequality, fix a nonempty coalition $S \subseteq A(z)$ and a continuation policy σ, and let $\mu_S = \min_{i \in S} V_i^\sigma(z)/V_i^\pi(z)$. Since continuation utilities are nonnegative,

$$y^{\boldsymbol{\lambda}}(z) \ge \sum_{i \in S} b_i(z) \frac{V_i^\sigma(z)}{V_i^\pi(z)} \ge b(S)\mu_S.$$

Dividing by $B(z) > 0$ gives $\mathrm{PF}_z(\pi) \ge \alpha_z(S)\mu_S$. Maximizing over S and σ yields $\mathrm{Dev}_z(\pi) \le \mathrm{PF}_z(\pi)$.

For the converse, suppose every voter is active at z, and let $\sigma^\star$ attain $y^{\boldsymbol{\lambda}}(z)$. Define the improvement ratios $r_i = V_i^{\sigma^\star}(z)/V_i^{\pi}(z)$, which are finite because $V_i^{\pi}(z) > 0$ for every voter. Order the voters by decreasing r_i, and write $r_{(k)}$ and $b_{(k)}$ for the ratio and entitlement of the voter in position k. Let $W_k = \sum_{j \le k} b_{(j)}$ and set

$$d = \max_k \frac{W_k r_{(k)}}{B(z)}.$$

For every k, the first k voters form an admissible coalition, and the minimum improvement ratio within that coalition is $r_{(k)}$. Its deviation value under $\sigma^\star$ is therefore $W_k r_{(k)}/B(z)$, so $d \le \mathrm{Dev}_z(\pi)$.

By the definition of d, $r_{(k)} \le B(z)d/W_k$ for every k. Since $\sigma^\star$ attains the maximum value of the scalarized objective,

$$\mathrm{PF}_z(\pi) = \frac{1}{B(z)} \sum_k b_{(k)} r_{(k)} \le d \sum_k \frac{b_{(k)}}{W_k}.$$

This order follows the improvement ratios and need not put the entitlements in ascending order. By Lemma F.2, however, the final sum is at most its value when the entitlements are arranged in ascending order, namely L_z. Hence, $\mathrm{PF}_z(\pi) \le dL_z \le L_z \mathrm{Dev}_z(\pi)$.

The logarithmic bound on L_z follows directly from Lemma F.2. At the initial state under reciprocal weights, $b_i(z_0) = 1$ for every voter, so the same lemma gives $L_{z_0} = \mathrm{H}_n$. □

The factor L_z makes the dependence on the continuation-entitlement profile more precise than the logarithmic bound stated in the main text. With equal entitlements, $L_z = \mathrm{H}_n$. More generally, it depends on the full distribution of the entitlements that remain, rather than only their total and smallest value. Thus, how much the scalar certificate can overestimate the deviation ratio away from the initial state depends on the continuation entitlements as well as the population size.

The assumption that every voter is active in the upper bound is necessary. The following example shows that, once an inactive voter is present, the scalar certificate can be arbitrarily larger than the exact deviation ratio.

Example F.1 (The upper bound requires every voter to be active)**.** Let $n = 2$, $H = 2$, and $M > 1$. Round 1 has a single alternative with values $(0, 1)$ that leads to a state z, where the two alternatives have values $(1, 0)$ and $(0, M)$. Let π choose the first alternative at z. Then, $U(\pi) = (1, 1)$, so the canonical weights are $\boldsymbol{\lambda} = (1, 1)$, while $V^\pi(z) = (1, 0)$. Hence, $B(z) = 1$ and $A(z) = \{1\}$. The scalarized objective can obtain value M at z by choosing the second alternative, so $\mathrm{PF}_z = M$. By contrast, the only nonempty coalition of active voters is $\{1\}$, whose best deviation ratio is 1, and therefore $\mathrm{Dev}_z = 1$. Thus, $\mathrm{PF}_z/\mathrm{Dev}_z = M$ can be arbitrarily large. The scalar certificate remains valid. However, when a voter has zero continuation utility under π, the certificate can count utility for a voter who belongs to no coalition allowed in the definition of Dev_z.

When every voter is active, the profile-dependent factor L_z cannot in general be improved. In fact, the next result shows that it is attained for every positive entitlement profile, even when the policy has no all-strict continuation block. It also expresses L_z as an optimization problem independent of the election construction.

Proposition F.1 (L_z is attained and measures the loss from scalarization)**.** *For every* $\mathbf{b} \in \mathbb{R}^n_{>0}$*, without any normalization, including when* $b_i > 1$*, there is an election with* $H = 2$ *and a state* z *reached with positive probability whose entitlements under the canonical weights are exactly* $\mathbf{b}$*, and for which* $\mathrm{Dev}_z(\pi) = 1$ *and* $\mathrm{PF}_z(\pi) = L_z$*. Moreover,*

$$L(\mathbf{b}) = \max_{\substack{\mathbf{r} \ge 0 \\ \mathbf{r} \ne 0}} \frac{\sum_i b_i r_i}{\max_{\varnothing \ne S \subseteq N} b(S) \min_{i \in S} r_i},$$

and the maximum is attained by $r_{[k]} = B/W_k$ *when the weights are ordered increasingly.*

Proof. Order the entries of $\mathbf{b}$ increasingly, with prefix sums $W_k = \sum_{j \le k} b_{[j]}$ and $B = W_n$, and set $r_{[k]} = B/W_k$. These ratios are decreasing in k.

Choose any $\mathbf{v} \in \mathbb{R}^n_{>0}$ and set $\lambda_i = b_i/v_i$. The target expected utilities from the initial state are $u_i = v_i/b_i$. Let $q = \min\{\frac{1}{2}, 1/\max_i b_i\}$. Then, $q \in (0, 1)$ and $q \le 1/b_i$ for every i, so $u_i - qv_i = v_i(1/b_i - q) \ge 0$.

In round 1, a single alternative with no utility leads to z with probability q and to a state z' otherwise. At z, there are two alternatives, with value profiles $\mathbf{v}$ and $(r_i v_i)_i$, respectively. At z', the unique alternative gives value profile $(\mathbf{u} - q\mathbf{v})/(1-q) \geq 0$. Let π choose the first alternative at z. Then,

$$V_i^{\pi}(z) = v_i, \qquad U_i(\pi) = qv_i + (1-q)\frac{u_i - qv_i}{1-q} = u_i = \frac{v_i}{b_i}.$$

Thus, the canonical weights are $\lambda_i = 1/U_i(\pi) = b_i/v_i$, and the continuation entitlements at z are $\lambda_i V_i^{\pi}(z) = b_i$, as required. The random transition is needed to obtain arbitrary positive entitlement profiles, including $b_i > 1$. If instead $\mathbf{b} \in (0,1]^n$, the same profile can be obtained with a deterministic transition by giving voter i immediate value $v_i(1/b_i - 1) \geq 0$ before reaching z. Simply setting $q = 1$ in the construction above would not work: it would give $U_i(\pi) = v_i$ and hence continuation entitlement 1, rather than b_i.

At z, the attainable continuations are mixtures of the two alternatives. For the second alternative,

$$\sum_i \lambda_i r_i v_i = \sum_i b_i r_i = B\, L(\mathbf{b}),$$

whereas the first has scalar value B. Hence, $y^{\boldsymbol{\lambda}}(z) = B\, L(\mathbf{b})$ and $\mathrm{PF}_z = L(\mathbf{b})$.

Now consider a mixture that assigns probability p to the second alternative. Voter i's improvement ratio is $1 - p + pr_i$. For any coalition S, let k be the largest index of one of its members in the increasing order of the entitlements. Since the ratios r_i decrease in this order, $\min_{i \in S} r_i = r_k$, while $b(S) \leq W_k$. Therefore,

$$\alpha_z(S) \min_{i \in S}(1 - p + pr_i) \leq (1-p)\frac{W_k}{B} + p\frac{W_k r_k}{B} = (1-p)\frac{W_k}{B} + p \leq 1.$$

Thus, $\mathrm{Dev}_z(\pi) \leq 1$. Since $\mathrm{Dev}_z(\pi) \geq 1$ by Lemma F.1*(iv)*, we obtain $\mathrm{Dev}_z(\pi) = 1$.

For the optimization identity, the argument used in the proof of Theorem F.2 gives, for every $\mathbf{r} \geq 0$,

$$\sum_i b_i r_i \leq L(\mathbf{b}) \max_{\varnothing \neq S \subseteq N} b(S) \min_{i \in S} r_i.$$

The profile $r_{[k]} = B/W_k$ attains equality: the denominator is B, while the numerator is $B\, L(\mathbf{b})$. The ratio is well defined whenever $\mathbf{r} \geq 0$ and $\mathbf{r} \neq 0$, since some $r_j > 0$ and the singleton $\{j\}$ makes the denominator at least $b_j r_j > 0$. □

Thus, L_z measures how much precision can be lost by replacing the exact search over coalitions with the scalar certificate. The full factor can arise even when $\mathrm{Dev}_z = 1$, so that no all-strict continuation block exists. The optimization identity shows that this loss comes from comparing the weighted-mean certificate with the best coalitional value, independently of the particular election constructed above.

The scalar certificate can also be exact. The next result states exactly when this occurs.

Proposition F.2 (When the certificate is exact)**.** *At a state where every voter is active,* $\mathrm{PF}_z(\pi) = \mathrm{Dev}_z(\pi)$ *if and only if some continuation* $\sigma^\star$ *maximizing the scalarized objective has a two-valued improvement profile: for some nonempty coalition* S *and some* $\rho > 0$*,* $V_i^{\sigma^\star}(z)/V_i^{\pi}(z) = \rho$ *for every* $i \in S$*, while* $V_i^{\sigma^\star}(z) = 0$ *for every* $i \notin S$*.*

Proof. Suppose first that such a continuation $\sigma^\star$ exists. Its scalar value gives $\mathrm{PF}_z = \rho b(S)/B(z)$. Taking coalition S and continuation $\sigma^\star$ in the definition of Dev_z gives $\mathrm{Dev}_z \geq \alpha_z(S)\rho = \mathrm{PF}_z$. Since $\mathrm{Dev}_z \leq \mathrm{PF}_z$ by Theorem 6.1, equality follows.

Conversely, suppose $\mathrm{PF}_z = \mathrm{Dev}_z$. By Lemma F.1*(iii)*, let (S°, σ°) attain Dev_z, and define $r_i = V_i^{\sigma^\circ}(z)/V_i^{\pi}(z)$. Then,

$$\mathrm{Dev}_z = \frac{b(S^\circ)}{B(z)} \min_{i \in S^\circ} r_i \leq \frac{1}{B(z)} \sum_{i \in S^\circ} b_i r_i \leq \frac{1}{B(z)} \sum_{i \in N} b_i r_i \leq \mathrm{PF}_z.$$

Since every voter is active, $b_i > 0$ for every i. Since the two endpoints are equal, every inequality in the chain must be an equality.

Equality in the final inequality means $\sum_i \lambda_i V_i^{\sigma^\circ}(z) = y^{\boldsymbol{\lambda}}(z)$, so σ° maximizes the scalarized objective. Equality in the second inequality gives $\sum_{i \notin S^\circ} b_i r_i = 0$; since $b_i > 0$ and $r_i \geq 0$, this implies $r_i = 0$, or equivalently $V_i^{\sigma^\circ}(z) = 0$, for every $i \notin S^\circ$. Equality in the first inequality requires $r_i = \min_{j \in S^\circ} r_j =: \rho$ for every $i \in S^\circ$.

Finally, $\rho > 0$. Since π itself is a feasible continuation, $y^{\boldsymbol{\lambda}}(z) \geq B(z)$ and hence $\mathrm{PF}_z \geq 1$. Therefore, $\mathrm{Dev}_z = \alpha_z(S^\circ)\rho = \mathrm{PF}_z \geq 1$, which implies $\rho > 0$. Thus, σ° has the required two-valued improvement profile. □

F.4 Approximate coalitional guarantees

We now give the full approximation result summarized in Section 6. A small scalar continuation gap yields two complementary bounds on coalitional deviations. The first is multiplicative and applies to coalitions of active voters when $B(z) > 0$. The second is additive and requires only that the coalition have positive entitlement, so it also covers coalitions containing inactive voters. We then express the scalar continuation gap as the expected sum of Bellman residuals of the reference policy.

Theorem F.3 (Approximate coalitional guarantees)**.** *Fix a policy π, weights $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$, and a state z, and suppose $\Delta_\pi(z) \leq \varepsilon_z$.*

- *(i)* Multiplicative guarantee. *If $B(z) > 0$, then* $\mathrm{Dev}_z(\pi; \boldsymbol{\lambda}) \leq \mathrm{PF}_z(\pi; \boldsymbol{\lambda}) = 1 + \Delta_\pi(z)/B(z) \leq 1 + \varepsilon_z/B(z)$. *Equivalently, for every $\varnothing \neq S \subseteq A(z)$ and every continuation policy σ, some $i \in S$ satisfies $\alpha_z(S)V_i^\sigma(z) \leq (1 + \varepsilon_z/B(z))V_i^\pi(z)$.*
- *(ii)* Additive guarantee. *For every coalition S with $b(S) > 0$ and every continuation policy σ, some $i \in S$ satisfies $\alpha_z(S)V_i^\sigma(z) - V_i^\pi(z) \leq \varepsilon_z \alpha_z(S)/\sum_{j \in S} \lambda_j$. Inactive voters may belong to S.*
- *(iii)* Residual identity. *Define $\epsilon_h(s) = y_h^{\boldsymbol{\lambda}}(s) - \sum_c \pi_h(c \mid s)\big[r_h^{\boldsymbol{\lambda}}(s,c) + \sum_{s'} P_h(s' \mid s,c) y_{h+1}^{\boldsymbol{\lambda}}(s')\big] \geq 0$. Then $\Delta_\pi(h,s) = \mathbb{E}_\pi[\sum_{t=h}^H \epsilon_t(s_t) \mid s_h = s]$.*

Proof. (i) Since $B(z) > 0$, both PF_z and Dev_z are defined. By the definition of $\Delta_\pi(z)$ and the assumption,

$$\mathrm{PF}_z(\pi; \boldsymbol{\lambda}) = \frac{y^{\boldsymbol{\lambda}}(z)}{B(z)} = 1 + \frac{\Delta_\pi(z)}{B(z)} \leq 1 + \frac{\varepsilon_z}{B(z)}.$$

By Theorem 6.1, $\mathrm{Dev}_z(\pi; \boldsymbol{\lambda}) \leq \mathrm{PF}_z(\pi; \boldsymbol{\lambda})$. Hence, for every nonempty $S \subseteq A(z)$ and every continuation policy σ,

$$\alpha_z(S) \min_{i \in S} \frac{V_i^\sigma(z)}{V_i^\pi(z)} \leq 1 + \frac{\varepsilon_z}{B(z)}.$$

Since every $i \in S$ is active, the denominators are positive. Thus, a voter attaining the minimum satisfies $\alpha_z(S)V_i^\sigma(z) \leq (1 + \varepsilon_z/B(z))V_i^\pi(z)$.

(ii) Fix a coalition S with $b(S) > 0$. Then $B(z) \geq b(S) > 0$, and $\alpha := \alpha_z(S) = b(S)/B(z) \in (0, 1]$. Suppose, for contradiction, that for some continuation policy σ and some $\theta \geq 0$,

$$\alpha V_i^\sigma(z) - V_i^\pi(z) > \theta \qquad \text{for every } i \in S.$$

Equivalently, $V_i^\sigma(z) > (V_i^\pi(z) + \theta)/\alpha$ for every $i \in S$. Multiplying by $\lambda_i > 0$, summing over S, and using $b_i(z) = \lambda_i V_i^\pi(z)$ gives

$$\sum_{i \in S} \lambda_i V_i^\sigma(z) > \frac{1}{\alpha}\left(b(S) + \theta \sum_{i \in S} \lambda_i\right) = B(z) + \frac{\theta}{\alpha} \sum_{i \in S} \lambda_i,$$

where the equality uses $b(S) = \alpha B(z)$.

Since $y^{\boldsymbol{\lambda}}(z)$ is the maximum continuation value of the scalarized objective from z, and continuation utilities are nonnegative,

$$y^{\boldsymbol{\lambda}}(z) \geq \sum_{i \in N} \lambda_i V_i^\sigma(z) \geq \sum_{i \in S} \lambda_i V_i^\sigma(z).$$

Combining the two inequalities and subtracting $B(z)$ yields

$$\varepsilon_z \geq \Delta_\pi(z) = y^{\boldsymbol{\lambda}}(z) - B(z) > \frac{\theta}{\alpha} \sum_{i \in S} \lambda_i.$$

Therefore, $\theta < \varepsilon_z \alpha_z(S) / \sum_{i \in S} \lambda_i$. It follows that, for every continuation policy σ, some $i \in S$ satisfies

$$\alpha_z(S) V_i^\sigma(z) - V_i^\pi(z) \leq \frac{\varepsilon_z \alpha_z(S)}{\sum_{j \in S} \lambda_j}.$$

The argument requires only $b(S) > 0$ and $\lambda_i > 0$; it does not require each member of S to be active. Thus, inactive voters may belong to S.

(iii) Define

$$\epsilon_h(s) = y_h^{\boldsymbol{\lambda}}(s) - \sum_c \pi_h(c \mid s) \left[r_h^{\boldsymbol{\lambda}}(s, c) + \sum_{s'} P_h(s' \mid s, c) y_{h+1}^{\boldsymbol{\lambda}}(s') \right] \geq 0.$$

Subtracting the corresponding Bellman expansion of $\sum_i \lambda_i V_i^\pi(h, s)$ gives

$$\Delta_\pi(h, s) = \epsilon_h(s) + \sum_c \pi_h(c \mid s) \sum_{s'} P_h(s' \mid s, c) \Delta_\pi(h + 1, s').$$

Since $\Delta_\pi(H + 1, \cdot) = 0$, applying this recursion through the final layer yields

$$\Delta_\pi(h, s) = \mathbb{E}_\pi \left[\sum_{t=h}^{H} \epsilon_t(s_t) \,\middle|\, s_h = s \right].$$

This is the performance-difference identity (Kakade & Langford, 2002) for the scalarized reward. □

G Omitted Proofs and Additional Results for Section 7

Throughout this appendix, let $\mathcal{P} = \prod_{h,s,c} \mathcal{P}_h(s, c)$ be an (s, a)-rectangular family of nonempty compact sets $\mathcal{P}_h(s, c) \subseteq \Delta(\mathcal{S}_{h+1})$. For a weight vector $\mathbf{w} \in \mathbb{R}^n_{\geq 0}$, define the weighted reward

$$r_h^{\mathbf{w}}(s, c) = \sum_i w_i a_{i,h}(s, c).$$

We allow general nonnegative utility values $a_{i,h}(s, c) \in [0, \bar{a}]$, where $\bar{a} < \infty$ is the uniform upper bound from Section 2. Unlike the entitlement weights used in Sections 6 and 7, the additional weight vectors in this appendix may have zero coordinates. This is needed for the per-voter robust evaluations used to bound the canonical weights, which take $\mathbf{w} = \mathbf{e}_i$. The results below are therefore stated for $\mathbf{w} \geq 0$. The arguments use only two properties that remain valid when some coordinates are zero: $r^{\mathbf{w}} \geq 0$ pointwise, because utilities are nonnegative, and $r^{\mathbf{w}} \leq \|\mathbf{w}\|_1 \bar{a}$. The first ensures that the relevant value functions are nonnegative and preserves the required monotonicity, while the second controls the value spans used in Lemma G.3. Strict positivity is needed only when the weights serve as entitlement weights, as in Sections 6 and 7.

G.1 The exact L_1 inner problem

The robust Bellman recursions below require optimizing a linear continuation value over an L_1 ball of transition distributions. For a vector $\mathbf{y} \in \mathbb{R}^K$, write $\operatorname{span}(\mathbf{y}) = \max_x y_x - \min_x y_x$. The next lemma solves this inner problem exactly and gives the span bound used later in the sensitivity analysis.

Lemma G.1 (Exact inner problem)**.** *Let $K \geq 1$, $\hat{\mathbf{p}} \in \Delta([K])$, $\mathbf{y} \in \mathbb{R}^K$, and $\varepsilon \geq 0$, and define $\mathcal{B} = \{\mathbf{p} \in \Delta([K]) : \|\mathbf{p} - \hat{\mathbf{p}}\|_1 \leq \varepsilon\}$. Set $\beta = \min\{\varepsilon, 2\}/2$, and let x^* be any element of $\arg\max_{x \in [K]} y_x$. Then, $\max_{\mathbf{p} \in \mathcal{B}} \langle \mathbf{p}, \mathbf{y} \rangle$ is attained by moving $\min\{\beta, 1 - \hat{p}_{x^*}\}$ units of probability mass to x^*, taking that mass greedily from states with the smallest values of $\mathbf{y}$. For the corresponding*

minimization, choose any minimizer of $\mathbf{y}$ *as the destination and take mass greedily from states with the largest values of* $\mathbf{y}$.

If $\mathbf{y}$ *has several maximizers, the optimal way to move mass need not be unique: mass may instead be distributed among them. However, every choice of* $x^* \in \arg\max_x y_x$ *in the construction above yields the same optimal value. Moreover,*

$$\left| \max_{\mathbf{p} \in \mathcal{B}} \langle \mathbf{p}, \mathbf{y} \rangle - \langle \hat{\mathbf{p}}, \mathbf{y} \rangle \right| \leq \frac{\varepsilon}{2} \operatorname{span}(\mathbf{y}). \tag{2}$$

Proof. Write $\mathbf{p} = \hat{\mathbf{p}} + \mathbf{u} - \mathbf{v}$, where $\mathbf{u}, \mathbf{v} \geq 0$ have disjoint supports, i.e., they have no coordinate in which both are positive. Since both $\mathbf{p}$ and $\hat{\mathbf{p}}$ are probability distributions,

$$\sum_x u_x = \sum_x v_x =: m.$$

Moreover, $\mathbf{v} \leq \hat{\mathbf{p}}$ and $\|\mathbf{p} - \hat{\mathbf{p}}\|_1 = 2m \leq \varepsilon$, so $m \leq \beta$.

For a fixed amount m of mass moved, the gain relative to $\hat{\mathbf{p}}$ is $\langle \mathbf{u}, \mathbf{y} \rangle - \langle \mathbf{v}, \mathbf{y} \rangle$. It is optimal to place the added mass on a state $x^* \in \arg\max_x y_x$ and to remove mass greedily from states with the smallest values of $\mathbf{y}$. If

$$g(m) = \min \left\{ \langle \mathbf{v}, \mathbf{y} \rangle : 0 \leq \mathbf{v} \leq \hat{\mathbf{p}},\ \sum_x v_x = m \right\},$$

the largest gain obtained by moving mass m to x^* is therefore

$$h(m) = m y_{x^*} - g(m).$$

The function h is concave and piecewise linear. As the greedy removal proceeds through states in increasing order of their $\mathbf{y}$-values, its slopes are of the form $y_{x^*} - y_{(k)}$ and are therefore nonnegative and nonincreasing.

It remains to account for ties among the maximizers of $\mathbf{y}$. Let $A = \arg\max_x y_x$ and $\mu_A = \hat{\mathbf{p}}(A)$. A slope becomes zero precisely when the greedy removal has moved all probability mass from outside A, which occurs at $m = 1 - \mu_A$. In particular, when A contains more than one state, this need not occur at $1 - \hat{p}_{x^*}$. Hence, h is nondecreasing on $[0, 1 - \mu_A]$ and constant on $[1 - \mu_A, 1]$. Therefore, every $m \geq \min\{\beta, 1 - \mu_A\}$ that is feasible attains $\max_{m \leq \beta} h(m)$.

Now, set $m^* = \min\{\beta, 1 - \hat{p}_{x^*}\}$. Moving this amount to a single destination is feasible because the total mass outside x^* is $1 - \hat{p}_{x^*}$. It is also optimal. If $\beta \leq 1 - \mu_A$, then $\beta \leq 1 - \hat{p}_{x^*}$ and hence $m^* = \beta$. If $\beta > 1 - \mu_A$, then $1 - \mu_A \leq 1 - \hat{p}_{x^*}$, so $m^* \in [1 - \mu_A, \beta]$, where h is already constant. Thus, the stated movement of mass is optimal for any choice of maximizer x^*. The minimization statement follows in the same way.

Finally, because $\sum_x (p_x - \hat{p}_x) = 0$, we may center $\mathbf{y}$ without changing $\langle \mathbf{p} - \hat{\mathbf{p}}, \mathbf{y} \rangle$. Taking $\bar{y} = \frac{1}{2}(\max_x y_x + \min_x y_x)$ gives

$$|\langle \mathbf{p} - \hat{\mathbf{p}}, \mathbf{y} \rangle| = |\langle \mathbf{p} - \hat{\mathbf{p}}, \mathbf{y} - \bar{y}\mathbf{1} \rangle| \leq \|\mathbf{p} - \hat{\mathbf{p}}\|_1 \|\mathbf{y} - \bar{y}\mathbf{1}\|_\infty \leq \frac{\varepsilon}{2} \operatorname{span}(\mathbf{y}),$$

which proves the bound. □

Equation (2) is only an inequality. For example, if $\hat{\mathbf{p}} = \delta_{x^*}$, the left-hand side is zero while the right-hand side may be positive. Thus, $\langle \hat{\mathbf{p}}, \mathbf{y} \rangle + \frac{\varepsilon}{2} \operatorname{span}(\mathbf{y})$ is not, in general, the robust Bellman value; we use this expression only as an upper bound in Lemma G.3.

G.2 Robust dynamic programming

Rectangularity allows the optimizations for each row in Lemma G.1 to be combined across states into a single transition kernel. The next lemma gives the resulting dynamic programming recursions for both an optimal continuation and the fixed reference policy, under optimistic and pessimistic choices of the transition kernel.

Lemma G.2 (Robust dynamic programming)**.** *Fix $\mathbf{w} \in \mathbb{R}^n_{\geq 0}$ and a policy π. Define the following four functions, all equal to zero at $h = H + 1$:*

$$\bar{y}_h^{\mathbf{w}}(s) = \max_c \left\{ r_h^{\mathbf{w}}(s,c) + \max_{p \in \mathcal{P}_h(s,c)} \langle p, \bar{y}_{h+1}^{\mathbf{w}} \rangle \right\},$$

$$\overline{B}_h^{\mathbf{w}}(s) = \sum_c \pi_h(c \mid s) \left\{ r_h^{\mathbf{w}}(s,c) + \max_{p \in \mathcal{P}_h(s,c)} \langle p, \overline{B}_{h+1}^{\mathbf{w}} \rangle \right\},$$

$$\underline{y}_h^{\mathbf{w}}(s) = \max_c \left\{ r_h^{\mathbf{w}}(s,c) + \min_{p \in \mathcal{P}_h(s,c)} \langle p, \underline{y}_{h+1}^{\mathbf{w}} \rangle \right\},$$

$$\underline{B}_h^{\mathbf{w}}(s) = \sum_c \pi_h(c \mid s) \left\{ r_h^{\mathbf{w}}(s,c) + \min_{p \in \mathcal{P}_h(s,c)} \langle p, \underline{B}_{h+1}^{\mathbf{w}} \rangle \right\}.$$

Then, for every state z,

$$\bar{y}^{\mathbf{w}}(z) = \max_{Q \in \mathcal{P}} \max_\sigma \sum_i w_i V_{i,Q}^\sigma(z), \qquad \underline{B}^{\mathbf{w}}(z) = \min_{Q \in \mathcal{P}} \sum_i w_i V_{i,Q}^\pi(z),$$

and similarly

$$\underline{y}^{\mathbf{w}}(z) = \min_{Q \in \mathcal{P}} \max_\sigma \sum_i w_i V_{i,Q}^\sigma(z), \qquad \overline{B}^{\mathbf{w}}(z) = \max_{Q \in \mathcal{P}} \sum_i w_i V_{i,Q}^\pi(z).$$

Each quantity is obtained by one backward pass over the layered state space. When every $\mathcal{P}_h(s,c)$ is an L_1 ball, its inner optimization is solved exactly by Lemma G.1.

All four quantities are also nondecreasing in the weight vector. If $\mathbf{w} \leq \mathbf{w}'$ componentwise, then

$$\bar{y}^{\mathbf{w}} \leq \bar{y}^{\mathbf{w}'}, \qquad \underline{y}^{\mathbf{w}} \leq \underline{y}^{\mathbf{w}'}, \qquad \overline{B}^{\mathbf{w}} \leq \overline{B}^{\mathbf{w}'}, \qquad \underline{B}^{\mathbf{w}} \leq \underline{B}^{\mathbf{w}'}.$$

Proof. We first prove the identity for $\bar{y}^{\mathbf{w}}$. Fix $Q \in \mathcal{P}$ and a continuation policy σ. We show by backward induction that

$$\sum_i w_i V_{i,Q}^\sigma(h,s) \leq \bar{y}_h^{\mathbf{w}}(s)$$

at every state (h,s). The claim is immediate at the terminal layer. Suppose it holds at $h+1$. Then

$$\begin{aligned}
\sum_i w_i V_{i,Q}^\sigma(h,s) &= \sum_c \sigma_h(c \mid s) \left[r_h^{\mathbf{w}}(s,c) + \sum_{s'} Q_h(s' \mid s,c) \sum_i w_i V_{i,Q}^\sigma(h+1,s') \right] \\
&\leq \sum_c \sigma_h(c \mid s) \left[r_h^{\mathbf{w}}(s,c) + \max_{p \in \mathcal{P}_h(s,c)} \langle p, \bar{y}_{h+1}^{\mathbf{w}} \rangle \right] \\
&\leq \bar{y}_h^{\mathbf{w}}(s).
\end{aligned}$$

The first inequality uses the induction hypothesis, $Q_h(\cdot \mid s,c) \in \mathcal{P}_h(s,c)$, and the nonnegativity of the transition probabilities. The second uses the fact that the average over alternatives under $\sigma_h(\cdot \mid s)$ cannot exceed the largest term. Hence,

$$\max_{Q \in \mathcal{P}} \max_\sigma \sum_i w_i V_{i,Q}^\sigma(z) \leq \bar{y}^{\mathbf{w}}(z).$$

For the reverse inequality, compactness of every $\mathcal{P}_h(s,c)$ ensures that each inner maximization in the recursion has a maximizer $p^\star_{h,s,c}$. By (s,a)-rectangularity, these independently selected rows form a single kernel $Q^\star \in \mathcal{P}$. Let $\sigma^\star$ choose, at every state, an alternative attaining the outer maximum in the recursion. Repeating the preceding backward induction with $Q^\star$ and $\sigma^\star$ makes every inequality an equality. Therefore,

$$\bar{y}^{\mathbf{w}}(z) = \max_{Q \in \mathcal{P}} \max_\sigma \sum_i w_i V_{i,Q}^\sigma(z).$$

The identities for $\underline{y}^{\mathbf{w}}$, $\overline{B}^{\mathbf{w}}$, and $\underline{B}^{\mathbf{w}}$ follow by the same argument. For $\underline{y}^{\mathbf{w}}$, replace the maximization for each row by a minimization; for the two policy-value recursions, replace the outer maximization over alternatives by the fixed average under $\pi_h(\cdot \mid s)$.

Finally, suppose $\mathbf{w} \leq \mathbf{w}'$ componentwise. Since the utilities are nonnegative, $r_h^{\mathbf{w}}(s,c) \leq r_h^{\mathbf{w}'}(s,c)$ at every state-alternative pair. Each of the four Bellman operators is nondecreasing in both its immediate reward and its continuation function: maxima and minima over a fixed ambiguity set are nondecreasing, as is averaging over alternatives. Backward induction therefore gives the four stated monotonicity inequalities. □

G.3 Proof of Theorem 7.1

Theorem 7.1. *If $P \in \mathcal{P}$ and $\boldsymbol{\lambda}^- \leq \boldsymbol{\lambda} \leq \boldsymbol{\lambda}^+$, then, at every state z, $\mathrm{PF}_z^P(\pi;\boldsymbol{\lambda}) \leq \mathrm{Cert}_z(\pi;\boldsymbol{\lambda}^-,\boldsymbol{\lambda}^+,\mathcal{P})$, and $\mathrm{Dev}_z^P(\pi;\boldsymbol{\lambda}) \leq \mathrm{Cert}_z(\pi;\boldsymbol{\lambda}^-,\boldsymbol{\lambda}^+,\mathcal{P})$ whenever $B^\pi_{\boldsymbol{\lambda},P}(z) > 0$.*

Proof. Fix $P \in \mathcal{P}$ and a state z, and write the two quantities in the election with kernel P as

$$y = y^{\boldsymbol{\lambda},P}(z) = \sup_\sigma \sum_i \lambda_i V^\sigma_{i,P}(z), \qquad B = B^\pi_{\boldsymbol{\lambda},P}(z) = \sum_i \lambda_i V^\pi_{i,P}(z).$$

Thus, $\mathrm{PF}_z^P(\pi;\boldsymbol{\lambda}) = y/B$ when $B > 0$, equals 1 when $B = y = 0$, and equals $+\infty$ when $B = 0 < y$.

We use three observations. First,

$$0 \leq B \leq y, \tag{3}$$

because continuation utilities are nonnegative and π is one of the continuation policies over which y is optimized. Second, since $\boldsymbol{\lambda} \leq \boldsymbol{\lambda}^+$ componentwise and $P \in \mathcal{P}$, the monotonicity and optimistic-value identity of Lemma G.2 give

$$y \leq \bar{y}^{\boldsymbol{\lambda}}(z) \leq \bar{y}^{\boldsymbol{\lambda}^+}(z) = \bar{y}^+(z). \tag{4}$$

Third, since $\boldsymbol{\lambda} \geq \boldsymbol{\lambda}^-$ componentwise and continuation utilities are nonnegative,

$$B \geq \sum_i \lambda_i^- V^\pi_{i,P}(z) \geq \underline{B}^{\boldsymbol{\lambda}^-}(z) = \underline{B}^-(z) \tag{5}$$

where the second inequality follows from the pessimistic policy-value identity of Lemma G.2 and $P \in \mathcal{P}$.

We now consider the three cases in Definition 7.1.

If $\underline{B}^-(z) > 0$, then $B \geq \underline{B}^-(z) > 0$ by Equation (5), so $\mathrm{PF}_z^P(\pi;\boldsymbol{\lambda}) = y/B$. The preceding bounds yield

$$\mathrm{PF}_z^P(\pi;\boldsymbol{\lambda}) = \frac{y}{B} \leq \frac{\bar{y}^+(z)}{\underline{B}^-(z)} = \mathrm{Cert}_z(\pi;\boldsymbol{\lambda}^-,\boldsymbol{\lambda}^+,\mathcal{P}).$$

If $\underline{B}^-(z) = 0 < \bar{y}^+(z)$, then $\mathrm{Cert}_z = +\infty$, so the inequality holds immediately in the extended reals.

If $\underline{B}^-(z) = \bar{y}^+(z) = 0$, then Equation (4) gives $0 \leq y \leq \bar{y}^+(z) = 0$, and hence $y = 0$. Equation (3) then gives $0 \leq B \leq y = 0$, so $B = 0$. By the convention of Section 6,

$$\mathrm{PF}_z^P(\pi;\boldsymbol{\lambda}) = 1 = \mathrm{Cert}_z(\pi;\boldsymbol{\lambda}^-,\boldsymbol{\lambda}^+,\mathcal{P}),$$

which proves the first inequality at every state.

Now, also suppose that $B^\pi_{\boldsymbol{\lambda},P}(z) > 0$. Then, $\mathrm{Dev}_z^P(\pi;\boldsymbol{\lambda})$ is defined, and Theorem 6.1, applied to the election with kernel P, gives

$$\mathrm{Dev}_z^P(\pi;\boldsymbol{\lambda}) \leq \mathrm{PF}_z^P(\pi;\boldsymbol{\lambda}).$$

Combining this with the first part proves the second inequality. No statement about Dev_z^P is made when $B = 0$, because the deviation ratio is then undefined. Indeed, since the weights are positive and continuation utilities are nonnegative, $B = 0$ implies $V^\pi_{i,P}(z) = 0$ for every voter, so there are no active voters at the state.

Finally, consider the two special cases of the certificate. If $\mathcal{P} = \{P\}$ and $\boldsymbol{\lambda}^- = \boldsymbol{\lambda}^+ = \boldsymbol{\lambda}$, the robust recursions of Lemma G.2 reduce to their nominal counterparts. Hence $\bar{y}^+(z) = y^{\boldsymbol{\lambda}}(z)$ and $\underline{B}^-(z) = B(z)$, and the three cases of Definition 7.1 coincide exactly with the convention defining PF_z. If instead $\boldsymbol{\lambda}^- = \boldsymbol{\lambda}^+ = \boldsymbol{\lambda}$ while $\mathcal{P}$ is arbitrary, the same argument applies uniformly to every $P \in \mathcal{P}$, giving the fixed-weight robust certificate. □

The fixed-weight and canonical versions of the certificate differ in an important way. Under the true kernel P, the canonical weights are $\lambda_i^P = 1/U_i^P(\pi)$, and thus $b_i(z_0) = \lambda_i^P U_i^P(\pi) = 1$ for every voter i. Hence, $B(z_0) = n$ and $\alpha_{z_0}(S) = |S|/n$, recovering the population shares of the fractional Core at the initial state. Weights computed from an estimated kernel do not generally have this property. If $\hat{\lambda}_i = 1/\hat{U}_i(\pi)$, then under the true kernel the initial entitlement is $b_i(z_0) = U_i^P(\pi)/\hat{U}_i(\pi)$, which need not equal 1.[1] Using the estimated reciprocal weights would therefore establish certificates for different continuation shares, rather than the canonical shares under the true kernel. This is why certificates for the canonical shares must account for uncertainty in the weights as well as in the transition kernel.

The guarantee in Theorem 7.1 is also deterministic. The theorem requires only that $P \in \mathcal{P}$ and $\boldsymbol{\lambda} \in [\boldsymbol{\lambda}^-, \boldsymbol{\lambda}^+]$; it makes no probabilistic assumption about how these sets were obtained. Thus, whenever a data-dependent event guarantees both inclusions, Theorem 7.1 applies on that event to every weight vector in the interval, including weights determined from the same data. This observation allows the results for sampled transitions below to apply the theorem to the unknown canonical weights $\lambda_i^P = 1/U_i^P(\pi)$ rather than requiring the weights to be fixed in advance.

G.4 Limits of the robust certificate

The robust certificate relies on joint optimization of the weighted continuation value over a common transition kernel. The two examples below clarify why this construction is necessary and why the resulting certificate need not be finite. The first shows that separately minimizing each voter's continuation utility does not, in general, produce a continuation profile associated with any single kernel in the ambiguity set. The second shows that even when the nominal certificate is finite, transition uncertainty can drive the pessimistic denominator to zero.

Example G.1 (Per-voter robust values are not a profile)**.** Let $n = 2$ and $\boldsymbol{\lambda} = (1, 1)$. At state z, suppose there is one alternative with no immediate utility and two possible successor states, with continuation-utility profiles $(1, 0)$ and $(0, 1)$. Let the nominal transition assign probability $\frac{1}{2}$ to each successor, and let $\mathcal{P}$ be an L_1 ambiguity set of radius ε. Then, for every $Q \in \mathcal{P}$,

$$\min_{Q \in \mathcal{P}} V_{1,Q}^{\pi}(z) = \min_{Q \in \mathcal{P}} V_{2,Q}^{\pi}(z) = \frac{1 - \varepsilon}{2},$$

whereas

$$V_{1,Q}^{\pi}(z) + V_{2,Q}^{\pi}(z) = 1.$$

Thus, the vector formed by taking each voter's worst-case continuation utility separately is not the continuation-utility profile generated by any single kernel in $\mathcal{P}$. This distinction matters for the entitlement construction. Combining per-voter worst cases would assign entitlements using values that cannot occur jointly and would therefore invalidate the calculation using a common kernel that gives the identity $b(S)/\alpha_z(S) = B(z)$. This is why Definition 7.1 keeps the weighted sum inside the optimization over Q, and why uncertainty in the weights is represented by an interval for $\boldsymbol{\lambda}$ rather than by separately constructed worst-case entitlements.

The preceding example concerns how uncertainty must be combined even when the relevant robust values remain positive. A separate issue arises when the ambiguity set contains kernels under which the reference policy has no continuation value. In that case, the certificate remains well defined under the extended-real convention of Definition 7.1, but it can become infinite.

Example G.2 (The certificate is not always finite)**.** Let $n = 2$ and $H = 3$. The initial state has a single alternative with no immediate utility whose transition reaches states g and b with nominal probability $\frac{1}{2}$ each. From g, both voters receive utility 1 in each of the two remaining rounds; from b, neither voter receives any utility. With $\boldsymbol{\lambda} = (1, 1)$, the nominal values satisfy

$$B(z_0) = \bar{y}(z_0) = 2.$$

Now, place an L_1 ambiguity set of radius $\varepsilon = 1$ around the initial transition row. In the pessimistic policy-value calculation, the adversary can move all probability mass to b, giving $\underline{B}^-(z_0) = 0$. In

[1] For example, in a two-voter, two-round election, suppose the initial chance move reaches a voter-1 state with true probability $\frac{1}{2}$ but estimated probability 0.6. Then $U^P = (0.5, 0.5)$, $\hat{U} = (0.6, 0.4)$, and $\hat{\lambda}_i U_i^P = (0.833, 1.25)$.

the optimistic calculation, it can instead move all mass to g, giving $\bar{y}^+(z_0) = 4$. Hence,

$$\mathrm{Cert}_{z_0} = +\infty.$$

Thus, the optimistic numerator remains bounded while the pessimistic denominator can be zero. A deterministic initial transition would not behave this way: with only one possible successor, there is no probability mass to move and the ambiguity set contains only that transition. The random transition is therefore essential to the example.

G.5 Transition sensitivity

The robust certificate will later be constructed from empirical transition kernels. We therefore need to bound how far the robust values can differ from their nominal counterparts when each transition row lies in an L_1 ball around its estimate. The following lemma gives a uniform bound for all four robust recursions in Lemma G.2.

Lemma G.3 (Transition sensitivity)**.** *Suppose each $\mathcal{P}_h(s,c)$ is the L_1 ball of radius ε around $\hat{P}_h(\cdot \mid s,c)$, and let $\hat{y}^{\mathbf{w}}$ and $\hat{B}^{\mathbf{w}}$ denote the corresponding nominal optimal and policy values computed under $\hat{P}$. Then, for every state $z = (h,s)$ and every $\mathbf{w} \in \mathbb{R}^n_{\geq 0}$,*

$$\hat{y}^{\mathbf{w}}(z) - \rho_h(\|\mathbf{w}\|_1) \leq \underline{y}^{\mathbf{w}}(z) \leq \hat{y}^{\mathbf{w}}(z) \leq \bar{y}^{\mathbf{w}}(z) \leq \hat{y}^{\mathbf{w}}(z) + \rho_h(\|\mathbf{w}\|_1),$$
$$\hat{B}^{\mathbf{w}}(z) - \rho_h(\|\mathbf{w}\|_1) \leq \underline{B}^{\mathbf{w}}(z) \leq \hat{B}^{\mathbf{w}}(z) \leq \overline{B}^{\mathbf{w}}(z) \leq \hat{B}^{\mathbf{w}}(z) + \rho_h(\|\mathbf{w}\|_1),$$

where

$$\rho_h(W) = \frac{\varepsilon W \bar{a}}{2} \cdot \frac{(H-h)(H-h+1)}{2} \leq \frac{\varepsilon W \bar{a} H^2}{4}.$$

In particular, $\rho_H(W) = 0$. For the weighted sum of utilities, taking $\mathbf{w} = \boldsymbol{\lambda}$ gives $W = \|\boldsymbol{\lambda}\|_1 = \sum_i \lambda_i$; for a single voter, taking $\mathbf{w} = \mathbf{e}_i$ gives $W = 1$. At the initial state, the transition μ is known rather than estimated, so the corresponding bound is

$$\rho_{z_0}(W) = \rho_1(W) = \frac{1}{4}\varepsilon W \bar{a} H(H-1).$$

Proof. By Lemma G.1, for every transition distribution $\mathbf{p}$ in the L_1 ball around $\hat{\mathbf{p}}$ and every continuation-value vector $\boldsymbol{\zeta}$,

$$|\langle \mathbf{p} - \hat{\mathbf{p}}, \boldsymbol{\zeta} \rangle| \leq \frac{\varepsilon}{2} \operatorname{span}(\boldsymbol{\zeta}).$$

We first consider the optimistic optimal value $\bar{y}^{\mathbf{w}}$. Define

$$\delta_h = \max_s \big(\bar{y}_h^{\mathbf{w}}(s) - \hat{y}_h^{\mathbf{w}}(s) \big) \geq 0.$$

Comparing the robust and nominal Bellman recursions alternative by alternative gives

$$\bar{y}_h^{\mathbf{w}}(s) \leq \hat{y}_h^{\mathbf{w}}(s) + \delta_{h+1} + \frac{\varepsilon}{2} \operatorname{span}(\bar{\mathbf{y}}_{h+1}^{\mathbf{w}}),$$

because each robust continuation term is bounded by its nominal counterpart plus the change in the transition and the error from the induction hypothesis, and taking the maximum over alternatives preserves this inequality.

The weighted reward is nonnegative and bounded above by $\|\mathbf{w}\|_1 \bar{a}$ in each round. From round $h+1$, at most $H-h$ rounds with rewards remain. Hence,

$$0 \leq \bar{y}_{h+1}^{\mathbf{w}}(s') \leq (H-h)\|\mathbf{w}\|_1 \bar{a}$$

for every successor state s', and therefore

$$\operatorname{span}(\bar{\mathbf{y}}_{h+1}^{\mathbf{w}}) \leq (H-h)\|\mathbf{w}\|_1 \bar{a}.$$

It follows that

$$\delta_h \leq \delta_{h+1} + \frac{\varepsilon \|\mathbf{w}\|_1 \bar{a}}{2}(H-h).$$

Since $\delta_{H+1} = 0$, iterating this inequality gives

$$\begin{aligned}\delta_h &\le \frac{\varepsilon\|\mathbf{w}\|_1\bar{a}}{2}\sum_{j=h}^{H}(H-j)\\ &= \frac{\varepsilon\|\mathbf{w}\|_1\bar{a}}{2}\frac{(H-h)(H-h+1)}{2}\\ &= \rho_h(\|\mathbf{w}\|_1).\end{aligned}$$

This proves $\bar{y}^{\mathbf{w}}(z) \le \hat{y}^{\mathbf{w}}(z) + \rho_h(\|\mathbf{w}\|_1)$.

The remaining three bounds follow from the same comparison. For $\underline{y}^{\mathbf{w}}$, the transition optimization for each row is a minimum rather than a maximum, giving the corresponding lower bound relative to the nominal value. For $\overline{B}^{\mathbf{w}}$ and $\underline{B}^{\mathbf{w}}$, the outer maximization over alternatives is replaced by the fixed average under $\pi_h(\cdot \mid s)$, which preserves the same bounds. The middle inequalities

$$\underline{y}^{\mathbf{w}} \le \hat{y}^{\mathbf{w}} \le \bar{y}^{\mathbf{w}}, \qquad \underline{B}^{\mathbf{w}} \le \hat{B}^{\mathbf{w}} \le \overline{B}^{\mathbf{w}}$$

hold because the nominal transition row $\hat{P}_h(\cdot \mid s, c)$ belongs to its own L_1 ambiguity set.

Finally, round 0 has no utility and its transition μ is known exactly. Thus, each robust or nominal value at z_0 is the same μ-weighted average of the corresponding values at the round-1 states. Averaging the bound for $h = 1$ therefore gives

$$\rho_{z_0}(W) = \rho_1(W) = \frac{1}{4}\varepsilon W\bar{a}H(H-1),$$

as claimed. □

G.6 Uniform model error

We first establish a single high-probability event on which every empirical transition row is close to its true counterpart. Since this event depends only on the transition samples, the same event can later be used simultaneously for every policy being evaluated.

Lemma G.4 (Uniform model error)**.** *Suppose m i.i.d. successor-state samples are drawn from $P_h(\cdot \mid s, c)$ for every transition row. Write $\mathsf{S} = \max_h |\mathcal{S}_h|$ and $\mathsf{A} = \max_{h,s} |C_h(s)|$, and let*

$$\varepsilon_P = \sqrt{\frac{2\big(\mathsf{S}\ln 2 + \ln(H\mathsf{SA}/\delta)\big)}{m}}.$$

Then, with probability at least $1 - \delta$,

$$\|\hat{P}_h(\cdot \mid s, c) - P_h(\cdot \mid s, c)\|_1 \le \varepsilon_P$$

simultaneously for every transition row (h, s, c).

Proof. For a fixed transition row with true distribution $\mathbf{p}$ and empirical distribution $\hat{\mathbf{p}}$, the L_1 concentration inequality of Weissman et al. (2003) gives

$$\mathbb{P}(\|\hat{\mathbf{p}} - \mathbf{p}\|_1 \ge \varepsilon) \le 2^{\mathsf{S}} e^{-m\varepsilon^2/2},$$

where using $\mathsf{S} = \max_h |\mathcal{S}_h|$ gives a common bound for all successor-state spaces. Substituting $\varepsilon = \varepsilon_P$ yields

$$2^{\mathsf{S}} e^{-m\varepsilon_P^2/2} = \frac{\delta}{H\mathsf{SA}}.$$

There are at most $H\mathsf{SA}$ transition rows. A union bound therefore gives

$$\mathbb{P}\Big(\exists (h, s, c) : \|\hat{P}_h(\cdot \mid s, c) - P_h(\cdot \mid s, c)\|_1 > \varepsilon_P\Big) \le \delta,$$

which proves the claim. □

G.7 Proof of Theorem 7.2

The preceding lemma gives the ambiguity set for the transition kernel. Lemma G.3 then converts the same transition error into uniform bounds on the policy's utilities, which in turn bound the canonical reciprocal weights. Combining these two results with the robust certificate of Theorem 7.1 gives the full guarantee from sampled transitions.

Theorem G.1 (Full statement of Theorem 7.2: Canonical certificate from sampled transitions)**.** *Assume m samples of $P_h(\cdot \mid s, c)$ are available for every state-alternative pair from a generative model, the utility values are known and lie in $[0, \bar{a}]$, and the initial distribution is known. Fix $\delta \in (0, 1)$, write $\mathsf{S} = \max_h |\mathcal{S}_h|$ and $\mathsf{A} = \max_{h,s} |C_h(s)|$, and let*

$$\varepsilon_P = \sqrt{\frac{2\big(\mathsf{S}\ln 2 + \ln(H\mathsf{S}\mathsf{A}/\delta)\big)}{m}}.$$

Let $\hat{\mathcal{P}}$ be the (s, a)-rectangular ambiguity set consisting, at each state-alternative pair, of the L_1 ball of radius ε_P around the empirical transition kernel, and define

$$\delta_U = \tfrac{1}{4}\varepsilon_P \bar{a} H(H-1).$$

Then, with probability at least $1 - \delta$, the following statements hold simultaneously for every policy π satisfying $\hat{U}_i(\pi) > \delta_U$ for every voter i, including a policy computed from the same transition samples. Define

$$\lambda_i^- = \frac{1}{\hat{U}_i(\pi) + \delta_U}, \qquad \lambda_i^+ = \frac{1}{\hat{U}_i(\pi) - \delta_U},$$

and let $\lambda_i^P = 1/U_i^P(\pi)$ denote the true canonical weights. Then:

- *(i) the true kernel satisfies $P \in \hat{\mathcal{P}}$, and $\lambda_i^- \le \lambda_i^P \le \lambda_i^+$ for every voter i;*
- *(ii) $\mathrm{PF}_z^P(\pi; \boldsymbol{\lambda}^P) \le \mathrm{Cert}_z(\pi; \boldsymbol{\lambda}^-, \boldsymbol{\lambda}^+, \hat{\mathcal{P}})$ at every state z, and $\mathrm{Dev}_z^P(\pi; \boldsymbol{\lambda}^P) \le \mathrm{Cert}_z(\pi; \boldsymbol{\lambda}^-, \boldsymbol{\lambda}^+, \hat{\mathcal{P}})$ at every state with $B^\pi_{\boldsymbol{\lambda}^P, P}(z) > 0$;*
- *(iii) $B^\pi_{\boldsymbol{\lambda}^P, P}(z_0) = \sum_i \lambda_i^P U_i^P(\pi) = n$ at the initial state. Consequently, $\mathrm{Dev}_{z_0}^P(\pi; \boldsymbol{\lambda}^P)$ is always defined and satisfies $\mathrm{Dev}_{z_0}^P(\pi; \boldsymbol{\lambda}^P) \le \frac{\bar{y}^+(z_0)}{n}$. Since $b_i(z_0) = 1$ for every voter, this is a bound on deviations measured against the population shares $|S|/n$. It is a quantitative bound for an arbitrary positive-utility policy; it establishes membership in the fractional Core of Definition 2.1 only when the reported bound is at most 1.*

Proof. Let E denote the event in Lemma G.4. Then, $\mathbb{P}(E) \ge 1 - \delta$, and on E,

$$\|\hat{P}_h(\cdot \mid s, c) - P_h(\cdot \mid s, c)\|_1 \le \varepsilon_P$$

for every transition row. In particular, $P \in \hat{\mathcal{P}}$.

The event E depends only on the true and empirical transition kernels, not on the policy being evaluated. Once E occurs, Lemma G.3 and theorem 7.1 apply deterministically to the same pair $(P, \hat{P})$ for every policy. Hence, all of the conclusions below hold on this single event simultaneously for every policy satisfying the stated positive-utility margin, including a policy selected using the same transition samples. No additional union bound over policies is required. Fix any such policy π.

(i) It remains to establish the bounds on the canonical weights. Apply Lemma G.3 to the ambiguity set $\hat{\mathcal{P}}$ with $\mathbf{w} = \mathbf{e}_i$, so that $\|\mathbf{w}\|_1 = 1$, while holding π fixed. The optimistic and pessimistic policy-value bounds at the initial state give, for every $Q \in \hat{\mathcal{P}}$,

$$\hat{U}_i(\pi) - \delta_U \le \underline{B}^{\mathbf{e}_i}(z_0) \le V_{i,Q}^\pi(z_0) \le \overline{B}^{\mathbf{e}_i}(z_0) \le \hat{U}_i(\pi) + \delta_U,$$

because $\rho_{z_0}(1) = \frac{1}{4}\varepsilon_P \bar{a} H(H-1) = \delta_U$ and $\hat{B}^{\mathbf{e}_i}(z_0) = \hat{U}_i(\pi)$.

By construction of the artificial initial state, round 0 is deterministic with no approvals, so $V^{\pi}_{i,Q}(z_0) = U^Q_i(\pi)$. Since $P \in \hat{\mathcal{P}}$ on E, taking $Q = P$ therefore gives

$$|U^P_i(\pi) - \hat{U}_i(\pi)| \le \delta_U.$$

The hypothesis $\hat{U}_i(\pi) > \delta_U$ then implies

$$U^P_i(\pi) \ge \hat{U}_i(\pi) - \delta_U > 0,$$

so the canonical weight $\lambda^P_i = 1/U^P_i(\pi)$ is well defined. Taking reciprocals of the positive utility bounds reverses their order and yields

$$\frac{1}{\hat{U}_i(\pi) + \delta_U} \le \lambda^P_i \le \frac{1}{\hat{U}_i(\pi) - \delta_U},$$

which is exactly $\lambda^-_i \le \lambda^P_i \le \lambda^+_i$.

(ii) Part *(i)* establishes both hypotheses of Theorem 7.1 with $\mathcal{P} = \hat{\mathcal{P}}$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^P$. Therefore, at every state z,

$$\mathrm{PF}^P_z(\pi; \boldsymbol{\lambda}^P) \le \mathrm{Cert}_z(\pi; \boldsymbol{\lambda}^-, \boldsymbol{\lambda}^+, \hat{\mathcal{P}}).$$

Whenever $B^{\pi}_{\boldsymbol{\lambda}^P, P}(z) > 0$, the second conclusion of Theorem 7.1 also gives

$$\mathrm{Dev}^P_z(\pi; \boldsymbol{\lambda}^P) \le \mathrm{Cert}_z(\pi; \boldsymbol{\lambda}^-, \boldsymbol{\lambda}^+, \hat{\mathcal{P}}).$$

(iii) At the initial state, $V^{\pi}_{i,P}(z_0) = U^P_i(\pi) > 0$. By the definition of the canonical weights,

$$B^{\pi}_{\boldsymbol{\lambda}^P, P}(z_0) = \sum_i \lambda^P_i U^P_i(\pi) = \sum_i \frac{U^P_i(\pi)}{U^P_i(\pi)} = n.$$

This identity is exact and involves no estimation. In particular, $\mathrm{Dev}^P_{z_0}(\pi; \boldsymbol{\lambda}^P)$ is defined, and

$$\mathrm{PF}^P_{z_0}(\pi; \boldsymbol{\lambda}^P) = \frac{y^{\boldsymbol{\lambda}^P, P}(z_0)}{n}.$$

Applying Theorem 6.1 and the optimistic numerator bound used in the proof of Theorem 7.1 gives

$$\mathrm{Dev}^P_{z_0}(\pi; \boldsymbol{\lambda}^P) \le \frac{y^{\boldsymbol{\lambda}^P, P}(z_0)}{n} \le \frac{\bar{y}^+(z_0)}{n}.$$

This gives a tighter bound than using Cert_{z_0} directly because the denominator is known exactly rather than replaced by its pessimistic robust lower bound. Finally, $b_i(z_0) = \lambda^P_i U^P_i(\pi) = 1$ for every voter i, so $\alpha_{z_0}(S) = |S|/n$. The bound therefore evaluates initial-state coalitional deviations using the population shares of the fractional Core. □

The guarantee of Theorem 7.2 also allows the evaluated policy to depend on the same transition samples used to construct the bound. Indeed, the event in Lemma G.4 depends only on $(P, \hat{P})$ and not on π. On this event, the sensitivity results and robust certificates are deterministic and apply simultaneously to every policy satisfying the required utility margin. Thus, π may itself be computed from the data, for example as an approximately Nash-optimal policy in the empirical election, without sample splitting or an additional union bound over policies. The same observation will apply to the coverage-adaptive certificate below.

The guarantee in Theorem 7.2 is computed after observing the samples: for a fixed sample size, it gives the certificate obtained from the resulting ambiguity set. The sensitivity result also gives a sample size requirement when a target certificate is specified in advance.

Corollary G.1 (Sample size for a target certificate slack)**.** *Fix weights $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$ and write $\Lambda = \|\boldsymbol{\lambda}\|_1 = \sum_i \lambda_i$. Suppose that, in the empirical election, π is $\boldsymbol{\lambda}$-tight and $\hat{B}(z) \ge \beta > 0$ at every state. Fix $\eta \in (0, 1]$. If $m \ge \frac{9}{8}\big(\mathsf{S}\ln 2 + \ln(H\mathsf{SA}/\delta)\big)\frac{\Lambda^2 \bar{a}^2 H^4}{\eta^2 \beta^2}$, then, on the event of Lemma G.4, the fixed-weight robust certificate satisfies $\mathrm{Cert}_z(\pi; \boldsymbol{\lambda}, \boldsymbol{\lambda}, \hat{\mathcal{P}}) \le 1 + \eta$ at every state z. More generally, let $\xi = \rho_1(\Lambda)$. For an arbitrary policy π, at every state satisfying $\hat{B}(z) > \xi$,*

$$\mathrm{Cert}_z(\pi; \boldsymbol{\lambda}, \boldsymbol{\lambda}, \hat{\mathcal{P}}) \le 1 + \frac{\hat{\Delta}(z) + 2\xi}{\hat{B}(z) - \xi}.$$

Proof. By Lemma G.3, $\rho_h(\Lambda) \le \rho_1(\Lambda) =: \xi$ for every $h \ge 1$. Hence, at every state,

$$\bar{y}^{\boldsymbol{\lambda}}(z) \le \hat{y}^{\boldsymbol{\lambda}}(z) + \xi, \qquad \underline{B}^{\boldsymbol{\lambda}}(z) \ge \hat{B}(z) - \xi.$$

If π is $\boldsymbol{\lambda}$-tight in the empirical election, then $\hat{y}^{\boldsymbol{\lambda}}(z) = \hat{B}(z)$, so whenever $\xi < \beta$,

$$\mathrm{Cert}_z \le \frac{\hat{B}(z) + \xi}{\hat{B}(z) - \xi} = 1 + \frac{2\xi}{\hat{B}(z) - \xi} \le 1 + \frac{2\xi}{\beta - \xi}.$$

If $\xi \le \beta/3$, then

$$\frac{2\xi}{\beta - \xi} \le \frac{3\xi}{\beta}.$$

It is therefore sufficient that $3\xi/\beta \le \eta$. Since $\eta \le 1$, this condition also implies $\xi \le \beta/3$.

Now,

$$\xi = \rho_1(\Lambda) = \frac{1}{4}\varepsilon_P \Lambda \bar{a} H(H-1) \le \frac{1}{4}\varepsilon_P \Lambda \bar{a} H^2.$$

Thus, $3\xi/\beta \le \eta$ is ensured by

$$\varepsilon_P \le \frac{4\eta\beta}{3\Lambda\bar{a}H^2}.$$

Substituting

$$\varepsilon_P = \sqrt{\frac{2\big(\mathsf{S}\ln 2 + \ln(H\mathsf{SA}/\delta)\big)}{m}}$$

and rearranging gives

$$m \ge \frac{9}{8}\big(\mathsf{S}\ln 2 + \ln(H\mathsf{SA}/\delta)\big)\frac{\Lambda^2\bar{a}^2H^4}{\eta^2\beta^2},$$

as claimed.

For the more general statement, use $\hat{y}^{\boldsymbol{\lambda}}(z) = \hat{B}(z) + \hat{\Delta}(z)$ instead. Whenever $\hat{B}(z) > \xi$,

$$\mathrm{Cert}_z \le \frac{\hat{B}(z) + \hat{\Delta}(z) + \xi}{\hat{B}(z) - \xi} = 1 + \frac{\hat{\Delta}(z) + 2\xi}{\hat{B}(z) - \xi},$$

which proves the second claim. □

The dependence on Λ shows how a lower bound on voter utilities enters when the weights are reciprocal. If the empirical utilities satisfy $\hat{U}_i(\pi) \ge u_{\min} > 0$, then $\hat{\Lambda} \le n/u_{\min}$. If instead $U_i^P(\pi) \ge u_{\min}$ is assumed for the true utilities, then on the event of Lemma G.4, $\hat{U}_i(\pi) \ge u_{\min} - \delta_U$. Once $\delta_U \le u_{\min}/2$, this gives $\hat{\Lambda} \le 2n/u_{\min}$, increasing the corresponding sample requirement by a factor of four.

The two parts of Corollary G.1 also clarify the role of empirical tightness. Exact $\boldsymbol{\lambda}$-tightness is needed only to obtain the simple target bound $1 + \eta$. For an arbitrary policy, its computed empirical gap $\hat{\Delta}(z)$ is retained explicitly. If $\hat{B}(z) \le \xi$, the sensitivity bound does not establish that the robust denominator is positive, so the simplified finite bound is unavailable; the robust certificate itself remains the relevant result and may equal $+\infty$.

Finally, Corollary G.1 is a fixed-sample statement, just like Theorem 7.2 and Lemma G.4: the sample size m is chosen in advance. A guarantee that holds simultaneously across sample sizes is given in Proposition G.2.

G.8 Proof of Proposition G.1 and its high-probability extension

We next prove the coverage limitation stated in Proposition G.1. The construction gives two elections with identical data distributions under the behaviour policy, even though an alternative that the policy never selects gives coalitions very different opportunities to deviate. We first establish the deterministic statement and then show that the same argument gives a high-probability lower bound for any uniformly valid randomized verification procedure.

Proposition G.1 (On-policy data cannot give an informative certificate)**.** *For every integer $R \geq 2$, there are two Markov temporal elections $\mathcal{E}_1$ and $\mathcal{E}_2$ with $n = 2$, $H = R + 1$, and $O(R)$ states, identical in every respect except the transition of one alternative $c^\star$ at one state $z^\star$, together with a behaviour policy π_b, such that:*

(i) π_b reaches $z^\star$ with probability one and never selects $c^\star$;

(ii) the distribution of the data generated by π_b is identical under $\mathcal{E}_1$ and $\mathcal{E}_2$ at every sample size;

(iii) $\mathrm{Dev}_{z^\star}^{\mathcal{E}_1}(\pi_b; \boldsymbol{\lambda}) = 1$ and $\mathrm{Dev}_{z^\star}^{\mathcal{E}_2}(\pi_b; \boldsymbol{\lambda}) = R$ at the canonical weights.

Consequently, let C map data sets to $[0, \infty]$ and suppose that C is pointwise valid in both elections, meaning that $\mathrm{Dev}_{z^\star}^{\mathcal{E}}(\pi_b; \boldsymbol{\lambda}) \leq C(D)$ for every data set D in the common support of the data generated by π_b and every $\mathcal{E} \in \{\mathcal{E}_1, \mathcal{E}_2\}$. Then, $C(D) \geq R$ for every such data set. Thus, in $\mathcal{E}_1$, where the true deviation ratio is 1, no pointwise-valid certificate can report value below $R = H - 1$, regardless of the sample size.

Proof. **Construction.** Fix $R \geq 2$ and set $H = R + 1$. Round 1 contains a single state $z^\star$ with two alternatives, c_1 and $c^\star$, both giving zero immediate utility. Alternative c_1 leads deterministically to a state g at the beginning of a forced chain through rounds $2, \ldots, R + 1$. Along this chain, the alternative in round 2 is approved by both voters and all subsequent alternatives are approved by neither voter. Thus, the continuation utility profile from g is $(1, 1)$.

Both elections also contain two additional forced chains, beginning at states d_0 and d_R. Along the chain from d_0, neither voter approves any alternative, so its continuation profile is $(0, 0)$. Along the chain from d_R, both voters approve the alternative in each of the R remaining rounds, so its continuation profile is (R, R). The elections have exactly the same states, alternatives, and approval values and differ in only one transition row:

$$P(\cdot \mid z^\star, c^\star) = \delta_{d_0} \quad \text{in } \mathcal{E}_1, \qquad P(\cdot \mid z^\star, c^\star) = \delta_{d_R} \quad \text{in } \mathcal{E}_2.$$

The construction uses $O(R)$ states. Let the behaviour policy π_b select c_1 at $z^\star$; at every subsequent state there is only one available alternative.

Part (i). Since $z^\star$ is the unique state in round 1, π_b reaches it with probability one. By construction, π_b selects c_1 there and never selects $c^\star$.

Part (ii). Under π_b, the process follows the chain beginning at g and never enters either the d_0 or the d_R chain. In particular, the transition row $P(\cdot \mid z^\star, c^\star)$ that distinguishes the two elections is never used. Every finite trajectory generated by π_b therefore has exactly the same probability under $\mathcal{E}_1$ and $\mathcal{E}_2$. Consequently, the distribution of any data generated by π_b, including complete trajectories or transition counts, is identical in the two elections at every sample size.

Part (iii). In both elections,

$$U(\pi_b) = (1, 1), \qquad V^{\pi_b}(z^\star) = (1, 1).$$

The canonical weights are therefore $\boldsymbol{\lambda} = (1, 1)$, so $b_1(z^\star) = b_2(z^\star) = 1$ and $B(z^\star) = 2 > 0$. Hence, $\alpha_{z^\star}(N) = 1$ and $\alpha_{z^\star}(\{i\}) = \frac{1}{2}$.

At $z^\star$, every continuation is a mixture of the continuation obtained through c_1, whose profile is $(1, 1)$, and the continuation obtained through $c^\star$. If p denotes the probability assigned to $c^\star$, the attainable profile is $p\boldsymbol{\theta} + (1 - p)(1, 1)$, where $\boldsymbol{\theta} = (0, 0)$ in $\mathcal{E}_1$ and $\boldsymbol{\theta} = (R, R)$ in $\mathcal{E}_2$.

In $\mathcal{E}_1$, the corresponding improvement-ratio profile is $(1 - p, 1 - p)$. Thus, for every nonempty coalition S,

$$\alpha_{z^\star}(S) \min_{i \in S}(1 - p) \leq 1,$$

with equality attained by the grand coalition at $p = 0$. Therefore,

$$\mathrm{Dev}_{z^\star}^{\mathcal{E}_1}(\pi_b; \boldsymbol{\lambda}) = 1.$$

In $\mathcal{E}_2$, the improvement-ratio profile is $(1 - p + pR,\ 1 - p + pR)$, which is maximized at $p = 1$. The grand coalition then attains deviation value R, whereas a singleton attains only $R/2$. Hence,

$$\mathrm{Dev}_{z^\star}^{\mathcal{E}_2}(\pi_b; \boldsymbol{\lambda}) = R.$$

Deterministic conclusion. Let C be any function of the observed data that is pointwise valid in both elections, so that $\mathrm{Dev}_{z^\star}^{\mathcal{E}}(\pi_b;\boldsymbol{\lambda}) \le C(D)$ for every data set D in the common support and for $\mathcal{E} \in \{\mathcal{E}_1, \mathcal{E}_2\}$. By part *(ii)*, the same data sets occur in both elections. Since the true deviation ratio in $\mathcal{E}_2$ is R, validity there requires $C(D) \ge R$ for every data set in the common support. The same reported value must therefore be at least R in $\mathcal{E}_1$, even though its true deviation ratio is only 1.

Valid certificates do exist: the constant R, for example, is valid in both elections. Such certificates, however, cannot be informative in both elections. The data cannot distinguish the election in which the deviation ratio is 1 from the one in which it is R, so a certificate that is valid in both must be at least the larger value. □

Proposition G.1 requires pointwise validity: the certificate must be valid for every data set in the common support of the two elections. The same argument also applies to the more usual high-probability notion of validity. As the data have the same distribution in $\mathcal{E}_1$ and $\mathcal{E}_2$, a randomized verification procedure cannot report a large value with high probability in one election without doing so with the same probability in the other. The following corollary makes this precise.

Corollary G.2 (High-probability form of Proposition G.1)**.** *Let a possibly randomized algorithm output $C = C(D, \omega)$ from data D generated by π_b and internal randomness ω independent of D. Suppose the algorithm is uniformly valid at level $1-\delta$, in the sense that*

$$\mathbb{P}_{\mathcal{E}}\left(\mathrm{Dev}_{z^\star}^{\mathcal{E}}(\pi_b;\boldsymbol{\lambda}) \le C\right) \ge 1-\delta$$

for every election $\mathcal{E}$ in the class. Then, for the two elections $\mathcal{E}_1$ and $\mathcal{E}_2$ of Proposition G.1,

$$\mathbb{P}_{\mathcal{E}_1}(C \ge R) = \mathbb{P}_{\mathcal{E}_2}(C \ge R) \ge 1-\delta.$$

Thus, even in $\mathcal{E}_1$, where $\mathrm{Dev}_{z^\star}^{\mathcal{E}_1}(\pi_b;\boldsymbol{\lambda}) = 1$, the algorithm reports a value below R with probability at most δ.

Proof. Let $C = C(D, \omega)$, where ω is independent of D and has the same distribution in both elections. By Proposition G.1*(ii)*, the distribution of D is identical under $\mathcal{E}_1$ and $\mathcal{E}_2$. The joint distribution of (D, ω) is therefore identical as well, and hence

$$\mathbb{P}_{\mathcal{E}_1}(C \ge R) = \mathbb{P}_{\mathcal{E}_2}(C \ge R).$$

Uniform validity in $\mathcal{E}_2$, where the true deviation ratio is R, implies

$$\mathbb{P}_{\mathcal{E}_2}(C \ge R) = \mathbb{P}_{\mathcal{E}_2}\left(\mathrm{Dev}_{z^\star}^{\mathcal{E}_2}(\pi_b;\boldsymbol{\lambda}) \le C\right) \ge 1-\delta.$$

Combining the two displays gives

$$\mathbb{P}_{\mathcal{E}_1}(C \ge R) = \mathbb{P}_{\mathcal{E}_2}(C \ge R) \ge 1-\delta,$$

or equivalently $\mathbb{P}_{\mathcal{E}_1}(C < R) \le \delta$, proving the corollary. □

G.9 Coverage-adaptive certificates

We now allow transition data to be collected adaptively and unevenly across state-alternative pairs. The main difficulty is that the number of observations from a given transition row is then random and may depend on previously observed successor states. The confidence set must therefore remain valid at the realized visit counts without conditioning on those counts.

To formalize the sampling process, index each transition row by a triple $q = (h, s, c)$. For every q, let $(X_{q,j})_{j\ge1}$ be an i.i.d. sequence with distribution P_q, with these sequences independent across transition rows. We call a collection process *admissible* if the successor state observed on the jth visit to q is $X_{q,j}$. The choice of which row to visit next may depend arbitrarily on all previous observations. Equivalently, conditional on the past, each visit to q produces a fresh draw from P_q; the two descriptions give the same distribution by the usual inverse-transform coupling.

Suppose the number of visits to each row is bounded by a deterministic $N_{\max}$. For $k \ge 1$, define

$$\varepsilon(k) = \min\left\{2,\ \sqrt{\frac{2\big(\mathsf{S}\ln 2 + \ln(H\mathsf{S}\mathsf{A}N_{\max}/\delta)\big)}{k}}\right\}.$$

For a row q visited $N_q > 0$ times, let the corresponding component of $\hat{\mathcal{P}}$ be the L_1 ball of radius $\varepsilon(N_q)$ around its empirical transition distribution. At an unvisited row, $N_q = 0$, let the ambiguity set be the entire successor-state simplex. Thus, uncertainty shrinks according to the observed coverage of each transition row, while an unobserved row remains completely unrestricted.

Theorem G.2 (Coverage-adaptive certificate)**.** *Let transition observations be collected by an admissible process, with the number of visits to each transition row bounded by the deterministic constant $N_{\max}$, and construct $\hat{\mathcal{P}}$ as above. Then, with probability at least $1-\delta$, the true kernel satisfies $P \in \hat{\mathcal{P}}$, and the following statements hold without any assumption on which state-alternative pairs were visited:*

(i) Fixed weights. *For every policy π and every $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$, simultaneously, including policies and weights chosen using the same data, and for every state z,*

$$\mathrm{PF}_z^P(\pi;\boldsymbol{\lambda}) \le \mathrm{Cert}_z(\pi;\boldsymbol{\lambda},\boldsymbol{\lambda},\hat{\mathcal{P}}).$$

Consequently, at every state with $B^\pi_{\boldsymbol{\lambda},P}(z) > 0$,

$$\mathrm{Dev}_z^P(\pi;\boldsymbol{\lambda}) \le \mathrm{Cert}_z(\pi;\boldsymbol{\lambda},\boldsymbol{\lambda},\hat{\mathcal{P}}).$$

(ii) Canonical weights. *For a policy π, define*

$$\underline{U}_i = \inf_{Q\in\hat{\mathcal{P}}} U^\pi_{i,Q}, \qquad \overline{U}_i = \sup_{Q\in\hat{\mathcal{P}}} U^\pi_{i,Q}.$$

If $\underline{U}_i > 0$ for every voter i, then

$$\lambda_i^- = \frac{1}{\overline{U}_i}, \qquad \lambda_i^+ = \frac{1}{\underline{U}_i}$$

satisfy $\lambda_i^- \le \lambda_i^P \le \lambda_i^+$, where $\lambda_i^P = 1/U_i^P(\pi)$. Hence, at every state z,

$$\mathrm{PF}_z^P(\pi;\boldsymbol{\lambda}^P) \le \mathrm{Cert}_z(\pi;\boldsymbol{\lambda}^-,\boldsymbol{\lambda}^+,\hat{\mathcal{P}}),$$

and at every state with $B^\pi_{\boldsymbol{\lambda}^P,P}(z) > 0$,

$$\mathrm{Dev}_z^P(\pi;\boldsymbol{\lambda}^P) \le \mathrm{Cert}_z(\pi;\boldsymbol{\lambda}^-,\boldsymbol{\lambda}^+,\hat{\mathcal{P}}).$$

In particular, $B^\pi_{\boldsymbol{\lambda}^P,P}(z_0) = n$.

(iii) Insufficient coverage. *If $\underline{U}_i = 0$ for some voter i, the canonical verification procedure returns* INSUFFICIENT COVERAGE *rather than a numerical bound.*

Proof. **The confidence set.** Fix a transition row q and a deterministic integer $1 \le k \le N_{\max}$, and let $\hat{P}_{q,k}$ be the empirical distribution of the first k entries of its sampling sequence. Since k is deterministic, these are k i.i.d. draws from P_q. Write

$$r(k) = \sqrt{\frac{2\big(\mathsf{S}\ln 2 + \ln(H\mathsf{S}\mathsf{A}N_{\max}/\delta)\big)}{k}}, \qquad \varepsilon(k) = \min\{2, r(k)\}.$$

Here, we need a bound for each fixed number of observations, without conditioning on the realized count N_q. Under adaptive collection, the event $\{N_q \ge k\}$ may itself depend on the first k successor states, so it is not generally valid to argue that these observations remain i.i.d. after conditioning on that event. Instead, we construct a single event on which the desired concentration inequality holds simultaneously for every fixed sample size $k \le N_{\max}$. This is the approach used by Jaksch et al. (2010).

Since the radius is truncated at 2, there are two cases. If $r(k) \ge 2$, then $\varepsilon(k) = 2$ and

$$\|\hat{P}_{q,k} - P_q\|_1 \le 2 = \varepsilon(k)$$

holds with probability one, since both vectors are probability distributions. Thus, the event $\{\|\hat{P}_{q,k} - P_q\|_1 > \varepsilon(k)\}$ is empty. If $r(k) < 2$, then $\varepsilon(k) = r(k)$ and the L_1 concentration inequality of Weissman et al. (2003) gives

$$\mathbb{P}\Big(\|\hat{P}_{q,k} - P_q\|_1 > \varepsilon(k)\Big) \le \mathbb{P}\Big(\|\hat{P}_{q,k} - P_q\|_1 \ge r(k)\Big) \le 2^{\mathsf{S}} e^{-kr(k)^2/2} = \frac{\delta}{H\mathsf{S}\mathsf{A}N_{\max}}.$$

The distinction between these two cases is necessary: when $r(k) > 2$, simply substituting the truncated radius 2 into the exponential bound need not yield the required error probability.

Let E be the intersection, over every transition row q and every $1 \le k \le N_{\max}$, of the events

$$\|\hat{P}_{q,k} - P_q\|_1 \le \varepsilon(k).$$

There are at most $HSAN_{\max}$ such events, so a union bound gives $\mathbb{P}(E) \ge 1 - \delta$. The event E depends only on the sampling sequences and not on any policy or weight vector.

On E, consider the realized count N_q of any row. If $N_q > 0$, the observations collected at that row are precisely $X_{q,1}, \dots, X_{q,N_q}$, and because E contains the concentration inequality for *every* deterministic value of k,

$$\|\hat{P}_q - P_q\|_1 = \|\hat{P}_{q,N_q} - P_q\|_1 \le \varepsilon(N_q).$$

If $N_q = 0$, the corresponding ambiguity set is the full successor-state simplex and therefore contains P_q automatically. Hence $P \in \hat{\mathcal{P}}$ on E.

Part (i). On E, we have $P \in \hat{\mathcal{P}}$. Applying Theorem 7.1 with $\boldsymbol{\lambda}^- = \boldsymbol{\lambda}^+ = \boldsymbol{\lambda}$ gives

$$\mathrm{PF}_z^P(\pi; \boldsymbol{\lambda}) \le \mathrm{Cert}_z(\pi; \boldsymbol{\lambda}, \boldsymbol{\lambda}, \hat{\mathcal{P}})$$

at every state, and the corresponding bound on $\mathrm{Dev}_z^P(\pi; \boldsymbol{\lambda})$ wherever $B^\pi_{\boldsymbol{\lambda},P}(z) > 0$. Since this is a deterministic consequence of $P \in \hat{\mathcal{P}}$, it holds on the same event E simultaneously for every policy and every positive weight vector, including policies and weights chosen from the observed data.

Part (ii). For a fixed policy π, the quantities

$$\underline{U}_i = \inf_{Q \in \hat{\mathcal{P}}} U^\pi_{i,Q}, \qquad \overline{U}_i = \sup_{Q \in \hat{\mathcal{P}}} U^\pi_{i,Q}$$

are exactly the values returned by the pessimistic and optimistic fixed-policy passes $\underline{B}^{\mathbf{e}_i}(z_0)$ and $\overline{B}^{\mathbf{e}_i}(z_0)$ of Lemma G.2. Thus, they require one pass of each type per voter. Since $P \in \hat{\mathcal{P}}$ on E,

$$\underline{U}_i \le U_i^P(\pi) \le \overline{U}_i.$$

If $\underline{U}_i > 0$ for every voter i, then all three quantities are positive, and taking reciprocals gives

$$\lambda_i^- = \frac{1}{\overline{U}_i} \le \frac{1}{U_i^P(\pi)} = \lambda_i^P \le \frac{1}{\underline{U}_i} = \lambda_i^+.$$

The assumptions of Theorem 7.1 therefore hold with $\boldsymbol{\lambda} = \boldsymbol{\lambda}^P$, yielding the stated bound on PF_z^P at every state and on Dev_z^P wherever $B^\pi_{\boldsymbol{\lambda}^P,P}(z) > 0$.

At the initial state,

$$B^\pi_{\boldsymbol{\lambda}^P,P}(z_0) = \sum_i \frac{U_i^P(\pi)}{U_i^P(\pi)} = n > 0,$$

so the deviation ratio is always defined there whenever the canonical weights exist.

Part (iii). Suppose $\underline{U}_i = 0$ for some voter i. Since $\hat{\mathcal{P}}$ is compact and $Q \mapsto U^\pi_{i,Q}$ is continuous, the infimum is attained: there is some $Q_0 \in \hat{\mathcal{P}}$ such that $U^\pi_{i,Q_0} = 0$. Thus, the data provide no strictly positive lower bound on voter i's utility under the reference policy. In particular, no finite upper bound on the reciprocal weight can be justified uniformly over all consistent kernels with positive utility. More explicitly, suppose first that there is some $Q_1 \in \hat{\mathcal{P}}$ with $U^\pi_{i,Q_1} > 0$. For $\eta \in [0,1]$, define

$$Q_\eta = (1-\eta)Q_0 + \eta Q_1.$$

Each component of $\hat{\mathcal{P}}$ is either an L_1 ball intersected with a simplex or the entire simplex, and is therefore convex. Hence, $Q_\eta \in \hat{\mathcal{P}}$. The map $\eta \mapsto U^\pi_{i,Q_\eta}$ is continuous, with value 0 at $\eta = 0$ and a positive value at $\eta = 1$. It therefore takes arbitrarily small positive values, and the corresponding reciprocals $1/U^\pi_{i,Q_\eta}$ are unbounded above. No finite λ_i^+ can therefore bound the canonical weight over all kernels consistent with the data.

If instead $U^{\pi}_{i,Q} = 0$ for every $Q \in \hat{\mathcal{P}}$, then in particular $U^{P}_{i}(\pi) = 0$, because $P \in \hat{\mathcal{P}}$, and the canonical reciprocal weight does not exist. In either case, a canonical numerical certificate using canonical weights is not justified, so the procedure returns INSUFFICIENT COVERAGE.

Finally, the certificates returned need not be finite. When the weight bounds are finite, the optimistic numerator is bounded, but the pessimistic denominator may still be zero. In that case $\mathrm{Cert}_z = +\infty$, as illustrated by Example G.2. □

G.9.1 CANONICAL VERIFICATION PROCEDURE

Given a policy π and data-dependent ambiguity set $\hat{\mathcal{P}}$, define $\underline{U}_i = \inf_{Q\in\hat{\mathcal{P}}} U^{\pi}_{i,Q}$ and $\overline{U}_i = \sup_{Q\in\hat{\mathcal{P}}} U^{\pi}_{i,Q}$ for every voter i.

Definition G.1 (Canonical verification procedure)**.** If $\underline{U}_i > 0$ for every voter, return $\mathrm{Cert}^{\mathrm{can}}_z = \mathrm{Cert}_z(\pi; 1/\overline{\mathbf{U}}, 1/\underline{\mathbf{U}}, \hat{\mathcal{P}})$ at every state; otherwise, return INSUFFICIENT COVERAGE.

Computing $\underline{U}_i$ and $\overline{U}_i$ requires two robust backward passes per voter, and the certificate requires two more, for $2n+2$ passes returning values at all states without enumerating coalitions. INSUFFICIENT COVERAGE may be represented as $+\infty$ for comparisons, but the robust program never uses an infinite weight.

G.10 MONOTONICITY AND ASYMPTOTIC CONSISTENCY

The coverage-adaptive certificate becomes less conservative when its ambiguity set shrinks, but this does not imply that the certificate improves with every additional observation. New data both reduce the confidence radius and move the empirical center, so successive ambiguity sets need not be nested. We first make this distinction precise and then show that, under persistent coverage, the certificate for a fixed policy and fixed weights nevertheless converges to the nominal certificate.

Lemma G.5 (Monotonicity under nested ambiguity sets)**.** *Let $\mathcal{P}^2 \subseteq \mathcal{P}^1$ be (s,a)-rectangular ambiguity sets. Then, $\bar{y}^{\mathbf{w},2} \le \bar{y}^{\mathbf{w},1}$ and $\underline{B}^{\mathbf{w},2} \ge \underline{B}^{\mathbf{w},1}$ for every $\mathbf{w} \in \mathbb{R}^n_{\ge 0}$. Consequently, whenever the corresponding certificates are finite, shrinking an ambiguity set cannot increase the certificate.*

Proof. The result follows by backward induction from Lemma G.2. For the optimistic recursion, taking the supremum over a smaller transition set can only decrease the rowwise continuation value; for the pessimistic policy recursion, taking the infimum over a smaller set can only increase it. Both Bellman recursions are monotone in their continuation arguments, so these inequalities hold at each earlier round. The numerical example below follows directly from the exact L_1 inner problem in Lemma G.1. □

The monotonicity in Lemma G.5 concerns containment of ambiguity sets, not the number of observations. An additional observation changes the empirical center as well as the confidence radius, so ambiguity sets constructed at successive sample sizes need not be nested and the certificate may increase. For example, consider two successor states with continuation-value vector $\mathbf{y} = (0,1)$ and radius $1/\sqrt{k}$. After four observations with empirical distribution $\hat{\mathbf{p}} = (0.75, 0.25)$, the exact optimistic inner value is 0.5. If the fifth observation is the high-value successor, then $\hat{\mathbf{p}} = (0.6, 0.4)$ and the optimistic value increases to approximately 0.6236, despite the smaller radius.

Lemma G.5 rules out a sample-by-sample monotonicity claim, but not convergence. If every transition row continues to be sampled, its empirical center converges to the true row while its confidence radius tends to zero. The next result shows that these two effects yield convergence to the fixed-weight nominal certificate for fixed weights.

Proposition G.2 (Consistency under persistent coverage)**.** *Fix a policy π and weights $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$ independently of the data. For $k \ge 1$, define the anytime radius*

$$\varepsilon^{\mathrm{any}}(k) = \min\{2, r_{\mathrm{any}}(k)\}, \qquad r_{\mathrm{any}}(k) = \sqrt{\frac{2\big(\mathsf{S}\ln 2 + \ln(H\mathsf{S}\mathsf{A}k(k+1)/\delta)\big)}{k}}.$$

Let T denote the total number of collected transition observations, and let $N_q(T)$ be the number of visits to transition row q among the first T observations. For each T, construct the rowwise ambiguity set $\hat{\mathcal{P}}_T$ by placing an L_1 ball of radius $\varepsilon^{\mathrm{any}}(N_q(T))$ around the empirical transition distribution at every row with $N_q(T) > 0$, and using the full successor-state simplex at every unvisited row.

With probability at least $1-\delta$, the true kernel satisfies $P \in \hat{\mathcal{P}}_T$ simultaneously for every T. If, in addition, every transition row is visited infinitely often almost surely, so that $N_q(T) \to \infty$ for every q as $T \to \infty$, then, for every state z satisfying $B^{\pi}_{\boldsymbol{\lambda},P}(z) > 0$,

$$\mathrm{Cert}_z(\pi; \boldsymbol{\lambda}, \boldsymbol{\lambda}, \hat{\mathcal{P}}_T) \to \mathrm{PF}^P_z(\pi; \boldsymbol{\lambda}) \qquad \textit{almost surely as } T \to \infty.$$

No convergence claim is made at states with $B^{\pi}_{\boldsymbol{\lambda},P}(z) = 0$; at such states the robust certificate may remain $+\infty$ for every finite T.

Proof. We first establish simultaneous coverage over all collection times. Fix a transition row q and a fixed number of observations $k \geq 1$, and let $\hat{P}_{q,k}$ denote the empirical distribution of the first k observations from that row. The same two-case argument as in the proof of Theorem G.2 applies.

If $r_{\mathrm{any}}(k) \geq 2$, then $\varepsilon^{\mathrm{any}}(k) = 2$, and

$$\|\hat{P}_{q,k} - P_q\|_1 \leq 2 = \varepsilon^{\mathrm{any}}(k)$$

holds deterministically because both are probability distributions. If $r_{\mathrm{any}}(k) < 2$, then $\varepsilon^{\mathrm{any}}(k) = r_{\mathrm{any}}(k)$, and the L_1 concentration inequality of Weissman et al. (2003) gives

$$\mathbb{P}\Big(\|\hat{P}_{q,k} - P_q\|_1 > \varepsilon^{\mathrm{any}}(k)\Big) \leq \frac{\delta}{H\mathsf{SA}\, k(k+1)}.$$

Since $\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1$, a union bound over all transition rows and all fixed sample sizes shows that, with probability at least $1-\delta$,

$$\|\hat{P}_{q,k} - P_q\|_1 \leq \varepsilon^{\mathrm{any}}(k)$$

simultaneously for every row q and every $k \geq 1$.

On this event, consider any collection time T. If $N_q(T) > 0$, the empirical row used in $\hat{\mathcal{P}}_T$ is $\hat{P}_{q,N_q(T)}$, and the bound that holds for all sample sizes gives

$$\|\hat{P}_{q,N_q(T)} - P_q\|_1 \leq \varepsilon^{\mathrm{any}}(N_q(T)).$$

If $N_q(T) = 0$, the ambiguity set at q is the full successor-state simplex and therefore contains P_q automatically. Hence, $P \in \hat{\mathcal{P}}_T$ simultaneously for every T.

We now establish consistency. Suppose every transition row is visited infinitely often almost surely. Then, for every row q,

$$N_q(T) \to \infty \qquad \text{almost surely as } T \to \infty.$$

By the strong law of large numbers applied to the sampling sequence for row q,

$$\hat{P}_{q,N_q(T)} \to P_q \qquad \text{almost surely as } T \to \infty,$$

while $\varepsilon^{\mathrm{any}}(N_q(T)) \to 0$ almost surely.

For a fixed continuation-value vector $\mathbf{y}$, consider the optimistic inner problem

$$\max\{\langle \mathbf{p}, \mathbf{y}\rangle : \mathbf{p} \in \Delta,\ \|\mathbf{p} - \hat{\mathbf{p}}\|_1 \leq \varepsilon\}.$$

Its value is continuous in $(\hat{\mathbf{p}}, \varepsilon, \mathbf{y})$ by the maximum theorem (Aliprantis & Border, 2006, Theorem 17.31): the feasible-set correspondence is nonempty, compact-valued, and continuous, including at $\varepsilon = 0$, where it reduces to the singleton $\{\hat{\mathbf{p}}\}$, and the objective $(\mathbf{p}, \mathbf{y}) \mapsto \langle \mathbf{p}, \mathbf{y}\rangle$ is continuous. The corresponding pessimistic inner problem is continuous by the same argument.

Let $\bar{y}^{\boldsymbol{\lambda}}_T$ and $\underline{B}^{\boldsymbol{\lambda}}_T$ denote the optimistic optimal value and pessimistic policy value computed from $\hat{\mathcal{P}}_T$. Since the horizon and state spaces are finite, backward induction through the robust Bellman recursions gives, for every state z,

$$\bar{y}^{\boldsymbol{\lambda}}_T(z) \to y^{\boldsymbol{\lambda},P}(z), \qquad \underline{B}^{\boldsymbol{\lambda}}_T(z) \to B^{\pi}_{\boldsymbol{\lambda},P}(z) \qquad \text{almost surely as } T \to \infty.$$

Now, fix a state z with $B^{\pi}_{\boldsymbol{\lambda},P}(z) > 0$. The second convergence implies that $\underline{B}^{\boldsymbol{\lambda}}_{T}(z) > 0$ for all sufficiently large T almost surely. Eventually, the robust certificate is given by the ratio:

$$\mathrm{Cert}_z(\pi; \boldsymbol{\lambda}, \boldsymbol{\lambda}, \hat{\mathcal{P}}_T) = \frac{\bar{y}^{\boldsymbol{\lambda}}_{T}(z)}{\underline{B}^{\boldsymbol{\lambda}}_{T}(z)}.$$

Taking $T \to \infty$ and using continuity of division at a positive denominator therefore yields

$$\mathrm{Cert}_z(\pi; \boldsymbol{\lambda}, \boldsymbol{\lambda}, \hat{\mathcal{P}}_T) \to \frac{y^{\boldsymbol{\lambda},P}(z)}{B^{\pi}_{\boldsymbol{\lambda},P}(z)} = \mathrm{PF}^{P}_{z}(\pi; \boldsymbol{\lambda}) \qquad \text{almost surely,}$$

as claimed. □

The consistency statement applies to each policy π and weight vector $\boldsymbol{\lambda}$ fixed independently of the data. The finite-sample guarantees of Theorems 7.2 and G.2 allow policies and weights to be chosen from the same data because they hold simultaneously on a policy-independent confidence event. That argument does not by itself establish consistency for a sequence of data-dependent policies $\pi_N = \pi_N(D_N)$. Such a result would require uniform convergence over the relevant policy class or a separate argument, and we do not claim it.

H Experimental Details and Additional Results

This appendix gives the full details of the experiments in Section 8. We first describe the synthetic-election generator (Appendix H.1), the reference policies and canonical weights (Appendix H.2), and the computation of the exact deviation measure and robust certificate (Appendices H.3 and H.4). We then give the parameter settings and comparison methods (Appendices H.5 and H.8), followed by the additional results for the certificate comparison, action-dependent transitions, the Nash-path analysis, and verification from sampled transitions (Appendices H.9–H.12). Finally, we report the computational cost, evaluation metrics, statistical methods, implementation checks, and computational environment (Appendices H.13–H.15).

The sampled-transition experiments estimate the transition kernel and evaluate a given policy; they do not learn or optimize that policy from the sampled transitions. The Nash-path analysis in Appendix H.11 is the only analysis that optimizes a policy, and there the election is known: the numerically Nash-optimal policy serves as an endpoint for evaluating how coalitional deviations change as the reference policy approaches the theoretical benchmark.

H.1 Synthetic election generator

We generate finite layered Markov temporal elections following the model of Section 2. An instance has H decision rounds, with a fixed number w of states in each round and k alternatives available at every state. The terminal layer contains a single state and no decision. Selecting alternative c at state s gives voter i utility $a_{i,h}(s,c)$ and draws the next state according to $P_h(\cdot \mid s,c)$. Different alternatives at the same state may therefore provide different immediate utilities and induce different distributions over the next layer. The generator lets us vary these two sources of variation separately while controlling the number of voters, horizon, layer width, and number of alternatives.

Utility values. The default experiments use binary approvals: $a_{i,h}(s,c)$ is drawn independently from $\mathrm{Bernoulli}(p)$, with $p = 0.35$ in the baseline configuration. This produces heterogeneous and relatively sparse approval profiles while leaving positive utility attainable for most voters. To test whether the certificates depend on the discrete values and ties created by binary approvals, one experimental family instead draws utilities independently from $U[0,1]$. The certificates also allow nonnegative cardinal utilities, so no change to their definitions is required.

Transition kernels. For each state-alternative pair, we independently generate a distribution over the states in the next layer. In the default *diffuse* construction, each successor is assigned an unnormalized weight drawn uniformly from $\{1,\dots,5\}$ and the resulting vector is normalized. The *concentrated* construction instead draws the transition row from a symmetric Dirichlet distribution with concentration parameter $\kappa = 0.2$. Since $\kappa < 1$, these rows tend to place most probability on a small number of successors. The two constructions therefore vary how dispersed each state-alternative transition is while leaving the way utilities are generated unchanged.

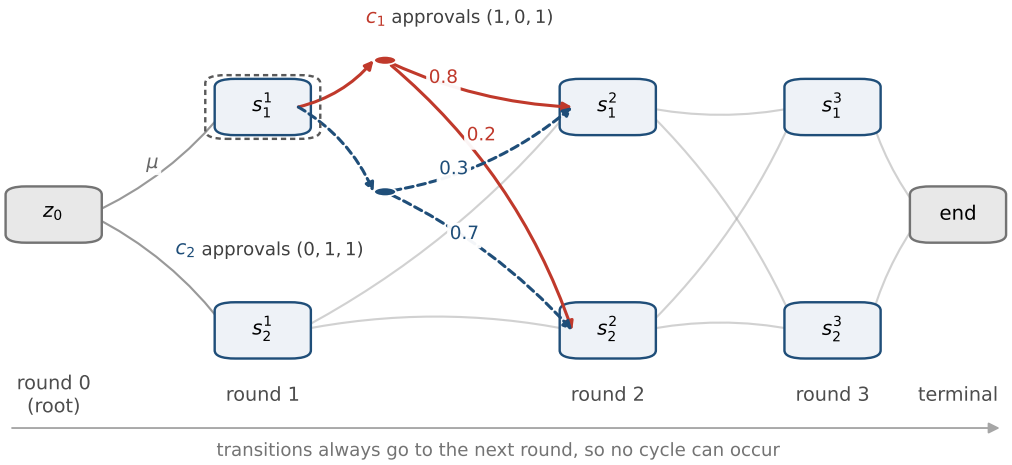


Figure 1: Schematic Markov temporal election with $H = 3$, layer width 2, two alternatives per state, and three voters. Columns represent rounds and rounded boxes represent states; the artificial root z_0 transitions according to the initial distribution μ. At state s_1^1, the two alternatives illustrate action-dependent transitions: c_1 gives approval profile $(1, 0, 1)$ and transitions to (s_1^2, s_2^2) with probabilities $(0.8, 0.2)$, whereas c_2 gives $(0, 1, 1)$ and transitions with probabilities $(0.3, 0.7)$. Thus, the selected alternative determines both current utility and the distribution of future states. All edges connect consecutive layers. The displayed utilities and probabilities are illustrative rather than experimental parameter values.

Layer width and number of alternatives. The layer width w determines the number of possible states in each decision round, while k determines the number of alternatives available at each state. Increasing w enlarges the set of possible successor states, while increasing k enlarges the set of choices through which the next-state distribution can be influenced. The corresponding groups of experiments vary these parameters one at a time relative to the baseline configuration.

Artificial initial state. For experiments that evaluate canonical quantities at the initial state, we use the artificial round-0 state z_0 introduced in Section 2. Its unique alternative gives zero utility and transitions according to the known initial distribution μ. Consequently, $V_i^\pi(z_0) = U_i(\pi)$ and the canonical weights $\lambda_i = 1/U_i(\pi)$ give $b_i(z_0) = 1$ exactly. In the sampled-transition experiments, this initial transition is treated as known rather than estimated.

Figure 1 illustrates the resulting structure on a small example with three decision rounds, two states per round, two alternatives per state, and three voters. The numerical values in the figure are illustrative and do not correspond to an experimental instance or parameter setting.

H.2 Policies, canonical weights, and quantities evaluated

Reference policy. Except in the Nash-path analysis of Appendix H.11, the experiments evaluate a fixed reference policy π. Holding the policy fixed allows each analysis to vary the quantity of interest without simultaneously changing the policy being evaluated. We use the uniform policy whenever it gives every voter strictly positive ex-ante utility. Otherwise, we draw action probabilities at each state from seeded Dirichlet distributions until a positive-utility policy is obtained or the search limit is reached. The construction is deterministic for a given experimental seed.

Specifically, the action-dependence analysis uses the uniform policy, which places equal probability on every alternative available at a state. Under its paired interpolation, averaging the transition rows over the uniformly selected alternatives yields the same policy-induced transition kernel for every value of the interpolation parameter. The reference utilities and canonical entitlements are therefore held fixed while action dependence changes, allowing the experiment to measure the effect on the continuation opportunities available to deviations.

Eligibility and canonical weights. Canonical verification requires $U_i(\pi) > 0$ for every voter because the reciprocal weights are $\lambda_i = 1/U_i(\pi)$. An instance is eligible for an experiment using

canonical weights only if its reference policy satisfies this condition. This criterion is determined before the quantities used for verification are evaluated and does not depend on their values.

For an eligible reference policy, the reciprocal normalization gives $b_i(z_0) = \lambda_i U_i(\pi) = 1$ for every voter. Hence, $B(z_0) = n$ and the initial continuation share of coalition S is $|S|/n$. The normalization therefore ensures that the experimental deviations and certificates use the population shares at the initial state. It does not assume that the reference policy is Nash-optimal or $\boldsymbol{\lambda}$-tight; departures from scalar optimality are what the quantities Dev_z and PF_z are intended to measure.

Nash endpoint. The Nash-path analysis additionally computes a numerically Nash-optimal policy in each known election. We obtain the endpoint using pairwise Frank-Wolfe and verify its numerical near-optimality using the Frank-Wolfe gap. The endpoint is used only to construct the policy path analyzed in Appendix H.11; it is not learned from sampled transitions. That appendix gives the optimization procedure, stopping criteria, and validation checks.

H.3 Exact coalitional deviation computation

The experiments in Appendices H.9 and H.10 compare the scalar certificate with the exact deviation ratio

$$\mathrm{Dev}_z(\pi; \boldsymbol{\lambda}) = \max_{\varnothing \neq S \subseteq A(z)} \max_{\sigma} \alpha_z(S) \min_{i \in S} \frac{V_i^{\sigma}(z)}{V_i^{\pi}(z)},$$

where $A(z) = \{i : V_i^{\pi}(z) > 0\}$ is the set of active voters. We compute this quantity by exhaustive coalition enumeration, solving the continuation problem for each coalition exactly as a linear program.

Coalition enumeration. For each state z with $B(z) > 0$, we enumerate every nonempty coalition $S \subseteq A(z)$. This gives $2^{|A(z)|} - 1$ coalition problems at that state, so the number of problems grows exponentially with the number of active voters.

Continuation linear program. Fix a coalition S. The inner problem $\max_{\sigma} \min_{i \in S} \frac{V_i^{\sigma}(z)}{V_i^{\pi}(z)}$ is a linear program when the continuation is represented by occupancies. We represent a continuation from z by occupancy variables $x_{j,c} \geq 0$ for every state j reachable from z and alternative c available there. The feasible occupancies satisfy

$$\sum_{c} x_{j,c} - \sum_{v,c} P(j \mid v, c) x_{v,c} = \mathbf{1}\{j = z\}$$

at every reachable nonterminal state j. These constraints introduce one unit of flow at z and pass it through the continuation transitions. Voter i's continuation utility is then linear in the occupancy vector: $V_i(x) = \sum_{j,c} a_i(j, c) x_{j,c}$.

To maximize the least improvement ratio among members of S, we introduce an auxiliary variable t and maximize it subject to $V_i(x) \geq t V_i^{\pi}(z)$ for every $i \in S$, together with the occupancy constraints above. Since every member of S is active, $V_i^{\pi}(z) > 0$, and the optimal value is exactly $t_S = \max_{\sigma} \min_{i \in S} V_i^{\sigma}(z)/V_i^{\pi}(z)$. We solve these linear programs with SciPy's `linprog` using the HiGHS simplex/interior-point solver.

For each coalition, we multiply t_S by its continuation share $\alpha_z(S) = b(S)/B(z)$, using the same entitlements $b_i(z) = \lambda_i V_i^{\pi}(z)$ as in the scalar certificate. The exact deviation ratio is the largest value of $\alpha_z(S) t_S$ across the enumerated coalitions. When $B(z) = 0$, Dev_z is undefined by construction and the implementation stores a missing value. All reported statistics involving Dev_z are thus restricted to states with $B(z) > 0$.

This computation requires one linear program for each nonempty coalition of active voters at each evaluated state. By contrast, PF_z is obtained for all states from one policy evaluation and one scalar Bellman pass. This difference in cost is why experiments requiring exact Dev_z use small numbers of voters. Appendix H.13 reports the corresponding runtime measurements.

H.4 Robust certificate computation

We compute the robust certificate of Definition 7.1 using the backward recursions of Lemma G.2. Relative to an ordinary Bellman pass, each transition expectation is replaced by its maximum or

minimum over the corresponding rowwise ambiguity set. The optimistic pass gives the certificate numerator and the pessimistic pass its denominator.

Rowwise robust expectation. For an observed transition row with empirical distribution $\hat{\mathbf{p}}$, the ambiguity set is the L_1 ball of radius ε around $\hat{\mathbf{p}}$, intersected with the simplex over the full next layer. Thus, states assigned zero empirical probability remain possible whenever the L_1 radius permits positive mass to be placed on them. Given the vector $\mathbf{y}$ of continuation values in the next layer, an optimistic Bellman update requires

$$\max_{\mathbf{p}} \{\langle \mathbf{p}, \mathbf{y} \rangle : \mathbf{p} \in \Delta(\mathcal{S}_{h+1}),\ \|\mathbf{p} - \hat{\mathbf{p}}\|_1 \leq \varepsilon\},$$

with the corresponding minimum used in a pessimistic update.

We solve this inner problem exactly using the mass-transport construction of Lemma G.1. For the maximum, probability mass is moved from successors with the smallest continuation values toward a successor with maximal continuation value until either the available L_1 budget or the destination capacity is exhausted; the minimum is symmetric. An L_1 radius ε permits at most $\min\{\varepsilon, 2\}/2$ units of probability mass to be moved. After sorting the continuation values, the required transport is obtained greedily.

The implementation uses this exact robust expectation rather than the bound $\langle \hat{\mathbf{p}}, \mathbf{y} \rangle + (\varepsilon/2)\,\mathrm{span}(\mathbf{y})$, which need not equal the optimum. As an implementation check, Appendix H.14 compares the computed values with those from a generic linear program over the full successor-state simplex.

H.5 Experimental parameters

This subsection gives the parameter settings for each experimental analysis. Unless stated otherwise, elections use the binary approval model and diffuse transition kernels described in Appendix H.1. Experiments using canonical weights retain only instances for which the reference-policy construction satisfies the positive-utility condition of Appendix H.2. The subsections below give the corresponding experimental procedures, aggregation rules, and additional results.

scalar certificate versus exact deviations. The comparison between PF_z and Dev_z uses five families of small elections. For each seed, the number of voters is drawn from $n \in \{2, 3, 4\}$ and the horizon from $H \in \{2, 3\}$. The baseline family uses independent $\mathrm{Bernoulli}(0.35)$ approvals, two states per round, two alternatives per state, and diffuse transition kernels. The remaining four families each change exactly one component of this baseline: the *cardinal* family replaces binary approvals by independent $U[0, 1]$ utilities; the *wide* family increases the layer width from 2 to 4; the *many-alternatives* family increases the number of alternatives per state from 2 to 4; and the *concentrated* family replaces the diffuse transition construction by symmetric $\mathrm{Dirichlet}(0.2)$ transition rows. Holding the other components fixed allows the five families to test different sources of the gap between the certificate and the exact deviation ratio without changing several generator features at once.

Each family is generated from 120 seeds. The positive-utility eligibility condition leaves 115 elections in the baseline family and 115 in the concentrated family, while all 120 elections are eligible in each of the other three families, giving 590 elections in total. We do not replace ineligible seeds. Appendix H.9 gives the state-level eligibility and aggregation rules used when comparing PF_z and Dev_z.

Action-dependent transitions. The action-dependence analysis uses $n = 3$ voters, $H = 4$ decision rounds, three states per round, and three alternatives per state. For each of 40 seeds, we first draw the approval profiles and candidate successor distributions, with the latter drawn independently from a symmetric $\mathrm{Dirichlet}(0.3)$ distribution. These quantities are then held fixed while the degree of action dependence varies through $\theta \in \{0, 0.25, 0.5, 0.75, 1\}$. Thus, comparisons across values of θ are paired within the same generated election rather than made across independently generated instances.

The experiment contains three arms, described in Appendix H.10, which separate the heterogeneous treatment from controls in which steering should have little or no effect. All three use the uniform reference policy. This choice is essential to the paired design: under the interpolation used in the experiment, the policy-averaged transition kernel of the uniform policy is invariant in θ, so its utilities,

canonical weights, and continuation entitlements remain fixed as action dependence changes. The per-election aggregation rule is also fixed before the comparison across θ.

Path to the Nash-welfare endpoint. The Nash-path analysis reuses the 590 eligible elections from the five families used to compare the scalar certificate with the exact deviation ratio. We add the artificial round-0 state and, for each election, compute a numerically Nash-optimal policy in the known model. We then interpolate in occupancy space between the reference policy and this Nash endpoint. The interpolation parameter is evaluated at $t \in \{0, 0.25, 0.5, 0.75, 0.9, 1\}$, with $t = 0$ denoting the reference policy and $t = 1$ the computed Nash endpoint. Since the policy changes along this path, the canonical reciprocal weights are recomputed separately at every evaluated value of t.

The Nash endpoint is obtained using the pairwise Frank-Wolfe procedure described in Appendix H.11. We require a Frank-Wolfe gap of at most 10^{-9} and minimum voter utility greater than 10^{-6}. We also numerically verify the theoretical endpoint conditions using tolerances for PF_z and Dev_z. Appendix H.11 states the precise statewise validation tests and explains how these tolerances are applied. The intermediate policies are points on the occupancy path that we evaluate; they are not produced by a policy-learning algorithm.

Verification under incomplete coverage. The coverage analysis uses $n = 2$ voters, $H = 2$ decision rounds, two states per round, and two alternatives per state. The artificial round-0 state is included, with a known initial distribution concentrated on one state. We generate 30 seeded elections, of which 27 satisfy the positive-utility eligibility condition and are retained. For each retained election, the election and reference policy are held fixed while the transition data are collected under five different methods described in Appendix H.12.

We evaluate total sampling budgets of 40, 200, 800, 4000, and 16000 successor-state observations and use error probability $\delta = 0.05$. The generated election contains nine transition rows in total. The row corresponding to the artificial initial state is the known initial distribution and is therefore not estimated, leaving eight transition rows to be learned from data. Under the sparse behaviour policy, four of these eight rows are never visited, regardless of the total sample budget. Their ambiguity sets therefore remain the full successor-state simplex. This persistent lack of coverage allows the experiment to distinguish the effect of increasing sample count from the effect of observing the state-alternative pairs relevant to possible deviations.

Indistinguishability under on-policy data. To illustrate the coverage limitation of Proposition G.1, we use its construction at $R \in \{2, 3, 4, 5, 6, 8, 10\}$. For each value of R, the two elections agree on every transition observed under the behaviour policy and differ only in the transition following an alternative that the behaviour policy never selects. The experiment therefore does not estimate an average effect over random instances; it implements the theoretical construction at several horizons and verifies the resulting separation in deviation ratios. The full construction and reported quantities are given in Appendix H.12.2.

Moderately larger instances. Finally, we evaluate robust verification on three configurations larger than those for which exact coalition enumeration is practical. The tuples $(n, H, \text{width}, \text{alternatives})$ are $(5, 4, 5, 3)$, $(10, 5, 10, 3)$, and $(5, 4, 10, 5)$. These configurations contain 61, 151, and 201 transition rows, respectively. Each uses 25 seeds, and the total sampling budgets are scaled with the number of transition rows so that larger instances do not receive fewer samples per row simply because they contain more rows. These configurations vary several dimensions simultaneously and are therefore not intended to measure the effect of any one parameter or to establish an asymptotic scaling law. Instead, they test the robust certificate beyond the small configurations used when exact computation of Dev_z is required. Appendix H.12.3 gives the precise budgets and the resulting certificate comparisons.

H.6 Metrics

This subsection defines the quantities used across the empirical analyses. Unless stated otherwise, statewise quantities are first aggregated within an election and then summarized across elections according to the statistical methods of Appendix H.7.

certificate ratio $\mathrm{PF}_z/\mathrm{Dev}_z$. This ratio measures how much the efficiently computable scalar certificate exceeds the exact coalitional deviation ratio at a state. By Theorem 6.1, $\mathrm{PF}_z/\mathrm{Dev}_z \geq 1$

wherever Dev_z is defined, with equality exactly when the certificate is tight. For the certificate-comparison and Nash-path analyses, we take the median of this ratio over the evaluated states of each election before summarizing across elections.

Entitlement-imbalance bound L_z. The full version of Theorem 6.1 gives the upper bound $\mathrm{PF}_z/\mathrm{Dev}_z \le L_z$ when every voter is active. The factor L_z depends only on the continuation-entitlement profile at z and gives the worst-case separation permitted by the theorem for that profile. We report it only over all-active states, where the upper bound applies.

certificate exactness. We report the fraction of evaluated states satisfying $\mathrm{PF}_z = \mathrm{Dev}_z$ up to the numerical tolerance specified below. Unlike the election-level certificate ratio, this is reported as a state-level frequency in Appendix H.9.

Action-dependence effect. For election r, let $\overline{\mathrm{Dev}}_\theta(r)$ denote the average of Dev_z over its eligible states under transition parameter θ. We define $\Delta_r(\theta) = \overline{\mathrm{Dev}}_\theta(r) - \overline{\mathrm{Dev}}_0(r)$. Since the paired construction of Appendix H.10 holds the reference utilities and entitlements fixed across θ, this difference measures the change in the available coalitional deviations as action dependence is introduced. We also compute the change in the scalar certificate.

Measured transition dependence. For each state, we compute the largest pairwise L_1 distance between its alternative-dependent transition rows, $\max_{c,c'} \|P(\cdot \mid s,c) - P(\cdot \mid s,c')\|_1$, and average this quantity over states. This measures the degree of action dependence produced by the interpolation and forms the horizontal axis of Figure 2.

Canonical robust certificate. For sampled-transition experiments, we write $\mathrm{Cert}_z^{\mathrm{can}} = \mathrm{Cert}_z(\pi; \boldsymbol{\lambda}^-, \boldsymbol{\lambda}^+, \hat{\mathcal{P}})$ for the output of the canonical procedure when the reciprocal-weight interval is finite. This quantity is an upper bound on the true PF_z^P on the coverage event. Smaller values, relative to PF_z^P, therefore indicate a more informative robust certificate.

Oracle-weight certificate. For the weight-ablation analysis, we write $\mathrm{Cert}_z^{\mathrm{orc}} = \mathrm{Cert}_z(\pi; \boldsymbol{\lambda}^P, \boldsymbol{\lambda}^P, \hat{\mathcal{P}})$, where $\boldsymbol{\lambda}^P$ denotes the true canonical weights. The sampled transitions and transition ambiguity set are unchanged. Comparing $\mathrm{Cert}_z^{\mathrm{can}}$ with $\mathrm{Cert}_z^{\mathrm{orc}}$ therefore measures the effect of replacing the true canonical weights by their data-dependent interval.

Paired certificate differences. When comparing two verification methods on the same election and sampled data, we form their difference before summarizing across elections. The coverage analysis subtracts the uniform certificate from the sparse certificate for the same election and sample budget. The weight ablation subtracts the oracle-weight certificate from the canonical certificate for the same election, budget, and collection method.

Robust-certificate ratio. On the moderately larger configurations, where exact Dev_z is not computed, we report $\mathrm{Cert}_{z_0}^{\mathrm{can}}/\mathrm{PF}_{z_0}^P$. The denominator is the scalar certificate computed using the true generating kernel and true canonical weights. A ratio of 1 would therefore mean that robust verification introduces no additional conservatism at the initial state.

Reduction from supplying the true canonical weights. To quantify the weight ablation, we use the decomposition

$$\mathrm{Cert}_{z_0}^{\mathrm{can}} - \mathrm{PF}_{z_0}^P = \big(\mathrm{Cert}_{z_0}^{\mathrm{can}} - \mathrm{Cert}_{z_0}^{\mathrm{orc}}\big) + \big(\mathrm{Cert}_{z_0}^{\mathrm{orc}} - \mathrm{PF}_{z_0}^P\big).$$

For elections with $\mathrm{Cert}_{z_0}^{\mathrm{can}} > \mathrm{PF}_{z_0}^P$, we report

$$R_{\mathrm{weights}} = \frac{\mathrm{Cert}_{z_0}^{\mathrm{can}} - \mathrm{Cert}_{z_0}^{\mathrm{orc}}}{\mathrm{Cert}_{z_0}^{\mathrm{can}} - \mathrm{PF}_{z_0}^P}.$$

The numerator is the reduction in the canonical certificate obtained by supplying the true canonical weights while retaining the same transition ambiguity set; the denominator is the canonical certificate's total excess over the known-kernel scalar certificate. Thus, R_{weights} measures what proportion of that excess disappears in the oracle-weight ablation.

Realized share of the certificate gap. Along the Nash path, we additionally report $(\mathrm{Dev}_z - 1)/(\mathrm{PF}_z - 1)$ at states with $\mathrm{PF}_z - 1 > 10^{-7}$. By Theorem 6.1, this quantity lies in $[0,1]$ and compares the exact coalitional deviation above the benchmark 1 with the excess reported by the

scalar certificate. It is first reduced to a within-election median and then summarized across elections.

Absence of all-strict blocks. By Lemma F.1, $\mathrm{Dev}_z = 1$ is equivalent to the absence of an all-strict continuation block at a state with $B(z) > 0$. Numerically, we treat this condition as satisfied when $|\mathrm{Dev}_z - 1| \leq 10^{-7}$. In the Nash-path analysis, we report both the fraction of relevant states satisfying this condition within each election and the number of policies for which it holds at every state with $B(z) > 0$. This condition is weaker than continuation stability, which also excludes weak blocks.

Log-Nash-welfare gap. For a policy π_t on the Nash path, the log-Nash-welfare gap is $\sum_i \log U_i(\pi_{\mathrm{N}}) - \sum_i \log U_i(\pi_t)$, where π_{N} is the numerically Nash-optimal endpoint. As shown in Appendix H.11, this gap is bounded above by $n(\mathrm{PF}_{z_0}(\pi_t) - 1)$.

H.7 Statistical methods

Unit of analysis and aggregation. Elections are the independent experimental units. States within an election share the same generated graph, utility values, and transition kernel and are therefore not treated as independent observations. For statewise quantities, we first compute the specified summary within each election and then summarize those election-level values across independent seeds. The exceptions are quantities explicitly reported as frequencies or summaries over states, such as the certificate-exactness rate and the L_z column of Table 1.

Numbers of instances. The scalar-certificate analysis uses 120 generated seeds per family, yielding 115–120 eligible elections in each of the five families and 590 in total. The Nash-path analysis evaluates these same 590 elections at six points along each occupancy path. The action-dependence analysis uses 40 paired elections. The coverage analysis retains 27 of 30 generated elections and evaluates five collection methods at five sample budgets, giving 675 election-method-budget observations. The larger-instance analysis uses 25 seeds for each of three configurations. The construction attaining the bound uses eight entitlement profiles, and the indistinguishability construction is instantiated at seven values of R.

Medians and spread. We use medians because several reported quantities are ratios with asymmetric or heavy right tails. Unless stated otherwise, spread is reported by the 25th and 75th percentiles of the election-level statistic, which mark its middle 50%. Percentiles are computed using linear interpolation.

Paired comparisons. When two conditions are evaluated on the same election, we form the comparison within each election before summarizing across elections. This applies to the action-dependence comparison with $\theta = 0$, sparse versus uniform collection, canonical versus oracle-weight verification, and comparisons along the occupancy path. When a ratio is reported, we likewise compute the ratio within each election rather than taking a ratio of medians across elections.

Eligibility and missing numerical certificates. Experiments using canonical reciprocal weights require a reference policy with strictly positive utility for every voter. The seeded generation procedure excludes five of the 120 baseline seeds, five of the 120 concentrated-transition seeds, none of the other three small-election families, and three of the 30 coverage seeds. States with $B(z) = 0$ are omitted from statistics involving Dev_z, which is undefined there.

In the coverage analysis, INSUFFICIENT COVERAGE is a defined output of the canonical procedure rather than a numerical error. It occurs in 13 of the 675 observations and is reported separately rather than assigned an arbitrary finite value. Numerical comparisons between canonical certificates therefore use only matched elections for which both quantities being compared are finite: 25 pairs at budget 40, 26 at budget 200, and all 27 from budget 800 onward. For R_{weights}, only elections with a strictly positive denominator enter the statistic, giving 22–24 elections per method-budget cell.

Scope of the conclusions. The experiments are descriptive evaluations of the generated instances. We do not report hypothesis tests, confidence intervals for population parameters, or multiple-comparison-adjusted significance tests. Statements such as all 40 paired action-dependence effects having the same sign are reported as observed counts rather than converted into post hoc significance tests.

Randomness and reproducibility. All random components are seeded, including election generation, fallback reference-policy draws, transition observations, and randomized collection methods. Seeds are consecutive integers beginning at zero and are fixed independently of the resulting experimental outcomes. For these seeds, all outputs other than timings are deterministic. Repeated runs reproduce the saved outputs byte for byte apart from timestamps and elapsed-time fields; we additionally reran the complete Nash-path analysis independently and obtained identical non-timing outputs.

H.8 Reference quantities and experimental comparisons

Different analyses require different reference quantities. Some use exact quantities available only because the generating election is known, while others compare alternative verification procedures or data-collection methods. We describe these roles below.

Exact deviation ratio. In the analyses of the scalar certificate and action-dependent transitions (Appendices H.9 and H.10), the main reference quantity is Dev_z, computed under the true generating kernel by the coalition-enumeration procedure of Appendix H.3. This is the exact coalitional quantity that PF_z is designed to upper-bound. Since computing Dev_z requires enumerating coalitions, these analyses are restricted to elections with few voters.

True scalar certificate. When exact coalition enumeration is impractical, we instead use PF_z^P, computed under the true generating kernel. In particular, the larger-instance analysis in Appendix H.12.3 compares the robust certificate obtained from sampled transitions with this known-kernel value. Unlike Dev_z, this quantity remains efficiently computable, but it is unavailable when the transition kernel is unknown.

Oracle-weight certificate. To measure the effect of uncertainty in the canonical weights, the ablation in Appendix H.12.1 recomputes the robust certificate using the true canonical weights $\lambda_i^P = 1/U_i^P(\pi)$ while leaving the sampled data and transition ambiguity set unchanged. Equivalently, the weight interval is set to $\boldsymbol{\lambda}^- = \boldsymbol{\lambda}^+ = \boldsymbol{\lambda}^P$. Comparing this value with the canonical certificate therefore measures the additional conservatism introduced by placing the endogenous weights in an interval. The oracle-weight certificate is used only for this comparison because the true canonical weights depend on the unknown kernel.

Fixed-weight certificate. The coverage analysis also evaluates $\mathrm{Cert}_z(\pi; \boldsymbol{\lambda}, \boldsymbol{\lambda}, \hat{\mathcal{P}})$ with the specified weights $\boldsymbol{\lambda} = \mathbf{1}$. This certificate is computable from the sampled transitions and does not require estimating reciprocal weights. It therefore separates uncertainty about the transitions from the additional uncertainty about canonical entitlements. Since the fixed weights induce different continuation shares, this quantity is a way to assess the effect of endogenous weights rather than a substitute for the canonical certificate.

Nash-welfare endpoint. In Appendix H.11, the numerically Nash-optimal policy in the known election provides the endpoint of the occupancy path beginning at the reference policy. Its role is to provide the policy benchmark at which the theoretical statewise guarantees apply, allowing us to examine how Dev_z and PF_z change as the reference policy approaches that endpoint.

Data-collection methods and controls. The coverage analysis compares uniform generative, on-policy, sparse behaviour-policy, adaptive, and skewed data collection (Appendix H.12). These methods differ in which transition rows receive observations and are used to study how coverage affects verification. Separately, the action-dependence analysis uses the exact-null and distributional-control constructions of Appendix H.10 to distinguish effects caused by useful steering opportunities from changes that can arise when action-dependent transition rows are introduced without such opportunities.

Finally, Appendix H.13 compares the measured runtime of exact coalition enumeration with that of the scalar certificate. This comparison quantifies the computational difference between the two implementations and is not used as a statistical baseline or reference policy.

Table 1: scalar certificate versus exact coalitional deviations on the five small-election families. "Median" and "IQR" summarize, across elections, the within-election median of $\mathrm{PF}_z/\mathrm{Dev}_z$. "Exact" is the fraction of evaluated states satisfying $\mathrm{PF}_z = \mathrm{Dev}_z$ up to numerical tolerance. The final column reports the median profile-dependent bound L_z over all-active evaluated states.

Family	Factor changed	# elec.	Median	IQR	Exact	L_z
baseline	—	115	1.121	[1.050, 1.200]	28.8%	2.01
cardinal	reward type	120	1.088	[1.064, 1.120]	0.0%	1.95
wide	layer width	120	1.086	[1.046, 1.131]	28.7%	2.06
many_alts	alternatives per state	120	1.122	[1.055, 1.201]	22.8%	1.97
concentrated	kernel concentration	115	1.104	[1.033, 1.249]	29.4%	2.00

H.9 Scalar certificate versus exact deviations

The full comparison in Appendix F.3 gives $\mathrm{Dev}_z \le \mathrm{PF}_z \le L_z \mathrm{Dev}_z$ whenever the corresponding upper bound applies. The first inequality makes PF_z a valid certificate, while the second quantifies how conservative it can be. We evaluate the realized gap between these quantities on the five small-election families described in Appendix H.5. The families share a common baseline generator and each variant changes exactly one generator component. The implementation checks this one-factor-at-a-time design directly before each run.

For each election r, let E_r denote its evaluated states and define $m_r = \mathrm{median}_{z \in E_r} \mathrm{PF}_z/\mathrm{Dev}_z$. We use the election as the unit of analysis: the "Median" and "IQR" columns of Table 1 summarize the distribution of m_r across elections. Here, the IQR is the interval from the 25th to the 75th percentile, and therefore contains the middle 50% of the election-level values. The "Exact" column instead reports the fraction of evaluated states at which $\mathrm{PF}_z = \mathrm{Dev}_z$ up to the numerical tolerance specified in Appendix H.7. Finally, the reported L_z is the median of the profile-dependent worst-case factor over all-active evaluated states, since the upper bound involving L_z requires every voter to be active.

The observed gap is much smaller than the worst-case bound. Across the five families, the median election-level ratio $\mathrm{PF}_z/\mathrm{Dev}_z$ ranges from 1.086 to 1.122. Thus, for the median election, the scalar certificate exceeds the exact deviation ratio by roughly 9–12%. By comparison, the median profile-dependent factor L_z over all-active states ranges from 1.95 to 2.06. The worst-case theory therefore allows substantially more conservatism than we observe on these generated instances. This empirical gap does not indicate that the theoretical factor can be reduced in general: the construction in Proposition F.1 attains L_z exactly, as illustrated below.

Exactness varies across the generated families. The certificate is exactly tight at 22.8–29.4% of evaluated states in the four binary-approval families, but at none of the evaluated states in the cardinal family. This comparison does not establish an effect of binary versus cardinal utilities. In the cardinal family every evaluated state is all-active, whereas the fraction of all-active states ranges from 63.4% to 83.3% in the four binary families. The reward model and set of active voters therefore change together in this comparison, so we report the difference without assigning it to either mechanism.

The worst-case bound is best possible. Proposition F.1 asserts that for every $b \in \mathbb{R}^n_{>0}$ there is an election with canonical entitlements exactly b at a reached state, with $\mathrm{Dev}_z = 1$ and $\mathrm{PF}_z = L(b)$. We construct these elections for eight entitlement profiles, including profiles with both $b_i > 1$ and $b_i < 1$, and confirm all three closed-form values to 10^{-7} in every case. Thus, the upper constant in Theorem 6.1 cannot be improved in general; the looseness observed above is a property of the tested instances rather than of the bound itself.

We also evaluate the variational identity in Proposition F.1 on randomly sampled improvement profiles $r \ge 0$. These are points of an abstract optimization problem, *not* elections: a randomly sampled r need not be the improvement profile of any Markov temporal election, so these values should not be interpreted as typical behavior of the certificate. The sampled ratios reach 85–98% of $L(b)$ at the maximum and 37–66% at the median.

Exactness characterization. Proposition F.2 characterizes $\mathrm{PF}_z = \mathrm{Dev}_z$ by the existence of a scalar-optimal continuation whose improvement ratios take only two values. The condition allows all scalar optima, including randomized ones, so restricting attention to deterministic continuations would give only a sufficient condition. We therefore test the condition directly. For each nonempty coalition S, we solve a linear program over the full continuation occupancy polytope, maximizing ρ subject to $\mathbf{x} \in \mathcal{X}_z$, $V_i(x) = \rho V_i^{\pi}(z)$ for $i \in S$, $V_j(x) = 0$ for $j \notin S$, and $\sum_i \lambda_i V_i(x) = y^{\boldsymbol{\lambda}}(z)$. Over 670 all-active states, 81 satisfy $\mathrm{PF}_z = \mathrm{Dev}_z$, and exactly the same 81 satisfy the structural condition, with no mismatches and errors below 10^{-15}. This is an exact regression check on small instances and an illustration of the characterization, not empirical verification of the proposition, which is proved analytically.

H.10 Action-dependent transitions

The analysis in this subsection measures the effect of action-dependent transitions on coalitional deviations. A direct comparison between independently generated action-dependent and action-independent elections would not measure this effect alone: changing the transition kernel can also change the reference policy's continuation utilities and hence its canonical entitlements. We therefore use a paired construction in which the reference utilities and entitlements remain fixed while the degree of action dependence varies.

Paired construction. For each seed, we fix the state graph, approval vectors, and candidate successor distributions $q_{s,1}, \ldots, q_{s,k}$ at every state. Let $\bar{q}_s = \frac{1}{k}\sum_c q_{s,c}$ and define $P_\theta(\cdot \mid s, c) = (1-\theta)\bar{q}_s + \theta q_{s,c}$ for $\theta \in [0,1]$. At $\theta = 0$, the transition distribution is $\bar{q}_s$ for every alternative, so the election is action-independent. At $\theta = 1$, each alternative uses its own candidate transition row. Intermediate values interpolate continuously between these two cases. Thus, within a seed, the elections differ only in how strongly the selected alternative influences the next-state distribution.

The reference policy is uniform over the k alternatives at every state. Its policy-averaged transition row is therefore $\frac{1}{k}\sum_c P_\theta(\cdot \mid s, c) = \bar{q}_s$ for every θ. Since the approval vectors are also held fixed, backward induction implies that $V_i^{\pi}(z)$ is identical for every θ at every state. Consequently, the ex-ante utilities $U_i(\pi)$, canonical weights $\lambda_i = 1/U_i(\pi)$, continuation entitlements $b_i(z)$, total entitlements $B(z)$, coalition shares $\alpha_z(S)$, and profile-dependent bounds L_z are all unchanged across θ. The implementation verifies this invariance for every seed and evaluated value of θ; the largest numerical discrepancy is 1.8×10^{-15}.

This property is specific to the paired experimental construction. In a general Markov temporal election, changing the action-dependent transition kernel can also change the reference policy's continuation utilities and therefore its entitlements. Here, by holding those quantities fixed, changes in Dev_z can be attributed to changes in the continuation opportunities available to deviating coalitions. The scalar certificate PF_z may change for the same reason, since its numerator optimizes over those continuation opportunities.

Controls. We use an exact-null construction to verify that the measured effect disappears when action-dependent transitions create no additional continuation opportunities. In this arm, all states within a layer offer the same alternatives with the same approval vectors. Let $\mathcal{V}_h$ denote the set of continuation-utility profiles attainable from a state in layer h. Since $\mathcal{V}_{H+1} = \{\mathbf{0}\}$ and $\mathcal{V}_{h+1}$ is convex, mixing continuations across successor states produces another element of $\mathcal{V}_{h+1}$. Conversely, any element of $\mathcal{V}_{h+1}$ can be obtained by using the same continuation after every successor. Hence, $\mathcal{V}_h = \mathrm{conv}\{\, \mathbf{a}_c + \mathcal{V}_{h+1} : c \in [k] \,\}$ regardless of the transition kernel. The attainable continuation set is therefore independent of θ. Since the reference utilities and entitlements are already fixed by the paired design, both Dev_z and PF_z must remain unchanged. Numerically, the largest absolute change in this arm is 8.4×10^{-10}.

A second control uses the remaining horizon. At a state in the final decision round, there are no future states whose distribution can be influenced, so action dependence cannot affect the continuation opportunities and the measured effect must be zero. In the heterogeneous treatment, the median effect is indeed zero with one round remaining and increases to 0.115, 0.169, and 0.181 with two, three, and four rounds remaining, respectively. This pattern is consistent with the effect arising from the ability to steer toward different future opportunities.

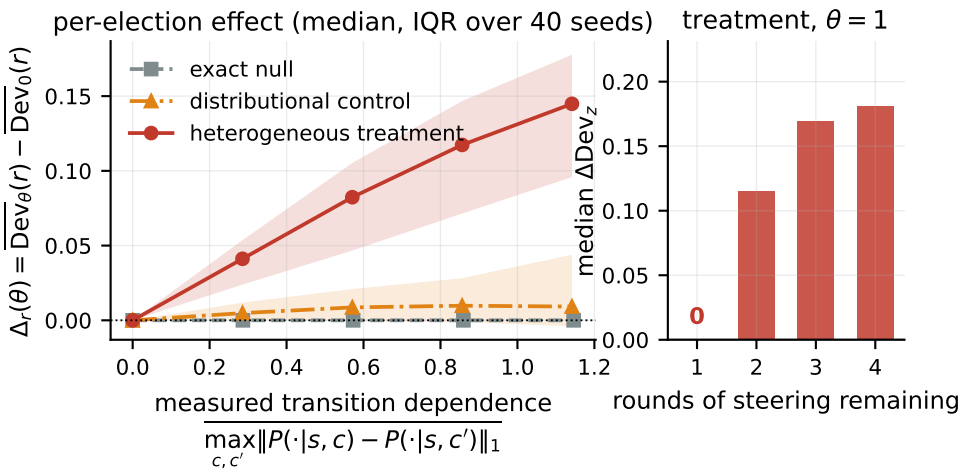


Figure 2: Effect of action-dependent transitions with the reference entitlement profile held fixed. Left: the within-election change $\Delta_r(\theta) = \overline{\mathrm{Dev}}_\theta(r) - \overline{\mathrm{Dev}}_0(r)$, summarized across 40 paired elections, against measured transition dependence. The exact-null construction remains unchanged to solver tolerance, while the distributional control exhibits a much smaller effect than the heterogeneous treatment. Right: within the heterogeneous treatment at $\theta = 1$, the effect is zero when one decision round remains and increases with the number of remaining rounds.

We also include a weaker distributional control in which states have the same ex-ante quality parameter but may realize different approval vectors. Unlike the exact-null construction, this arm does not make steering irrelevant by construction. Its median effect at $\theta = 1$ is 0.009, substantially smaller than in the heterogeneous treatment.

Action dependence enlarges the deviations available to coalitions. For election r, let $\overline{\mathrm{Dev}}_\theta(r)$ be the average of Dev_z over its eligible states under P_θ, and define the paired effect $\Delta_r(\theta) = \overline{\mathrm{Dev}}_\theta(r) - \overline{\mathrm{Dev}}_0(r)$. The left panel of Figure 2 plots this effect against the average, over states, of the largest pairwise L_1 distance between transition rows, $\max_{c,c'} \|P_\theta(\cdot \mid s,c) - P_\theta(\cdot \mid s,c')\|_1$.

In the heterogeneous treatment, moving from $\theta = 0$ to $\theta = 1$ increases the average deviation ratio in every one of the 40 paired elections. The median increase is 0.145, with the middle 50% of the election-level increases ranging from 0.096 to 0.178. The scalar certificate responds similarly, with a median increase of 0.156. Since the paired construction holds the reference utilities, canonical weights, and continuation entitlements fixed, these changes reflect the additional deviation opportunities created by action-dependent transitions rather than a change in the proportional claims used to evaluate them.

Together with the controls, the experiment gives a clear interpretation of the effect. When different successor states offer no different continuation opportunities, action dependence has essentially no effect. When there is no future left to influence, its effect is exactly zero. By contrast, when future states provide different opportunities, allowing the current alternative to influence which state is reached substantially increases the proportional deviations available to coalitions. This is an empirical finding for the generated instances, not a monotonicity result in θ. The theory also does not imply that Dev_z must increase pointwise as action dependence increases.

H.11 From the reference policy to a Nash-welfare optimum

The preceding analyses evaluate a fixed reference policy. Here, we instead vary the policy within each election, following a path from the reference policy to a numerically Nash-optimal policy. The purpose is to examine how the exact deviation ratio and its scalar certificate change as the policy approaches the benchmark for which the theoretical guarantees become exact after completion. Every election is known throughout this experiment; the Nash-welfare optimum is computed only to provide the endpoint of the comparison, not learned from sampled transitions.

Elections and reference policies. We use the same 590 eligible elections from the five families of Appendix H.9, with the artificial round-0 state z_0 added. The reference policy is the one defined in Appendix H.2, extended by the unique action at z_0. Consequently, the canonical reciprocal weights give $B(z_0) = n$. At the reference endpoint, the non-initial-state values of PF_z and Dev_z reproduce those of Appendix H.9 on all 3583 evaluated states.

Computing the Nash endpoint. For each election, we maximize $F(\pi) = \sum_i \log U_i(\pi)$ over the occupancy polytope of Appendix B using pairwise Frank-Wolfe (Lacoste-Julien & Jaggi, 2015). At a positive-utility policy π, the gradient of F with respect to the utility profile has coordinates $\lambda_i = 1/U_i(\pi)$. The Frank-Wolfe gap is the improvement predicted by linearizing the objective at the current policy and optimizing that linearization over the feasible set. In our setting, it is therefore

$$g_{\mathrm{FW}}(\pi) = \max_{\sigma} \sum_i \frac{U_i(\sigma) - U_i(\pi)}{U_i(\pi)}.$$

The maximization in this expression is exactly the scalar MDP with weights $\lambda_i = 1/U_i(\pi)$, so its linear optimization oracle is one scalar Bellman pass. Since $\sum_i \lambda_i U_i(\pi) = n$ and $\max_\sigma \sum_i \lambda_i U_i(\sigma) = y^{\boldsymbol{\lambda}}(z_0)$, the gap can be written as

$$g_{\mathrm{FW}}(\pi) = y^{\boldsymbol{\lambda}}(z_0) - n = n\big(\mathrm{PF}_{z_0}(\pi) - 1\big).$$

This gap also bounds the loss in log-Nash welfare. For every policy σ, the inequality $\log u \le u - 1$ gives

$$\sum_i \log U_i(\sigma) - \sum_i \log U_i(\pi) \le \sum_i \left(\frac{U_i(\sigma)}{U_i(\pi)} - 1\right) \le g_{\mathrm{FW}}(\pi).$$

Thus, the Frank-Wolfe gap is an upper bound on the policy's gap from optimal log-Nash welfare. This gives both the stopping criterion for the optimization and a numerical certificate for the resulting endpoint.

We stop pairwise Frank-Wolfe when $g_{\mathrm{FW}}(\pi) \le 10^{-13}$. The resulting endpoint is randomized in 270 of the 590 elections. We then complete the policy outside its support using actions that are greedy for its reciprocal weights, as in Lemma D.3, and set to zero the numerical probabilities of alternatives whose Bellman value is more than 10^{-9} below the optimum in relative terms.

As a separate validation of the completed endpoint, we require every voter to have utility greater than 10^{-6}, the final Frank-Wolfe gap to be at most 10^{-9}, $|\mathrm{PF}_z - 1| \le 10^{-9}$ and $|B(z) - y^{\boldsymbol{\lambda}}(z)| \le 10^{-9} \max\{1, y^{\boldsymbol{\lambda}}(z)\}$ at every nonterminal state, $|\mathrm{Dev}_z - 1| \le 10^{-7}$ wherever $B(z) > 0$, and the polishing step to change $\sum_i \log U_i$ by at most 10^{-9}. All 590 endpoints satisfy these conditions. Across them, the largest Frank-Wolfe gap is 1.0×10^{-13}, the largest $|\mathrm{PF}_z - 1|$ is 3.0×10^{-13}, and the largest $|\mathrm{Dev}_z - 1|$ is 1.5×10^{-8}. We therefore refer to these endpoints as *numerically Nash-optimal*, rather than exact Nash-welfare optima.

Occupancy path and canonical weights. Let $\mathbf{x}_{\mathrm{ref}}$ and $\mathbf{x}_{\mathrm{N}}$ denote the occupancy vectors of the reference and numerically Nash-optimal policies. For $t \in [0, 1]$, define $\mathbf{x}_t = (1 - t)\mathbf{x}_{\mathrm{ref}} + t\mathbf{x}_{\mathrm{N}}$, and let π_t be the corresponding Markov policy. Thus, $t = 0$ is the reference policy and $t = 1$ is the Nash endpoint. We evaluate $t \in \{0, 0.25, 0.5, 0.75, 0.9, 1\}$. For every $t < 1$, the construction preserves positive occupancy at every state that has positive occupancy under the reference policy.

Utilities are linear in occupancy, so $U_i(\pi_t)$ is affine in t. We recompute the canonical weights at every evaluated point as $\lambda_i(t) = 1/U_i(\pi_t)$. Each policy is therefore evaluated using its own canonical proportional claims, while its initial-state coalition shares remain $|S|/n$.

Two properties of this path follow from the optimization problem itself. First, $\sum_i \log U_i(\pi_t)$ is concave in t and is maximized at $t = 1$, so it is nondecreasing along the segment toward the Nash endpoint. Second, $\mathrm{PF}_{z_0}(\pi_t)$ is nonincreasing in t. Indeed, under the reciprocal normalization, $\mathrm{PF}_{z_0}(\pi_t) = \frac{1}{n} \max_\sigma \sum_i U_i(\sigma)/U_i(\pi_t)$; for each fixed σ, every term $U_i(\sigma)/U_i(\pi_t)$ is convex in t, so the maximum is convex in t. Since this quantity is at least 1 and equals 1 at the exact Nash endpoint, convexity implies that it cannot increase along the segment. No analogous monotonicity result follows for Dev_{z_0}, so its behavior along the path is an empirical question.

Aggregation. Initial-state quantities contribute one observation per election. For statewise quantities, we first aggregate over states with $B(z) > 0$ within each election and then summarize across

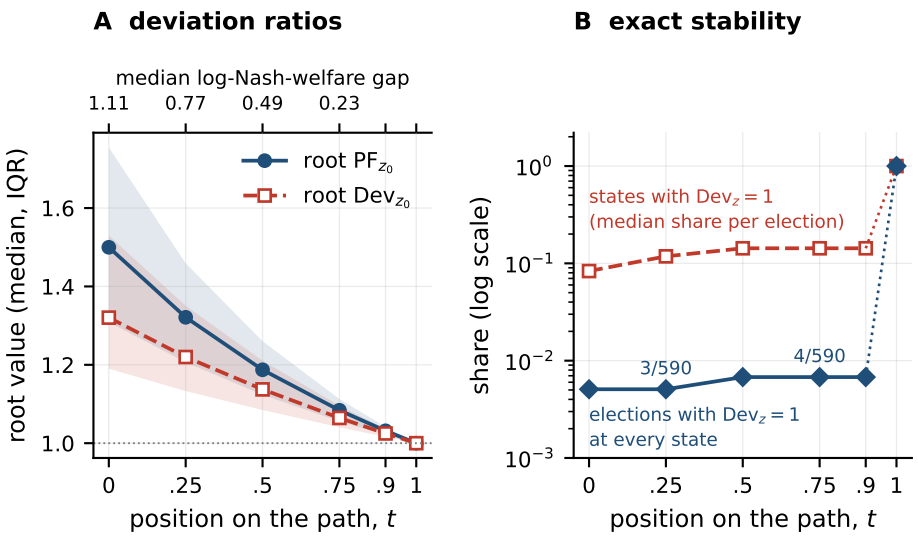


Figure 3: Coalitional deviations and scalar certificates along the occupancy path from the reference policy ($t = 0$) to the numerically Nash-optimal endpoint ($t = 1$), with canonical weights recomputed at every point. **(A)** Initial-state PF_{z_0} and Dev_{z_0}, summarized across elections; the top axis reports the median log-Nash-welfare gap. **(B)** The median within-election fraction of states satisfying $\mathrm{Dev}_z = 1$ and the fraction of elections satisfying $\mathrm{Dev}_z = 1$ at every state with $B(z) > 0$. The latter condition means that no state admits an all-strict continuation block; it is weaker than continuation stability, which also excludes weak blocks. Dotted segments connect $t = 0.9$ to the endpoint; no intermediate values of t were evaluated.

Table 2: Deviation ratios and certificate tightness along the occupancy path over 590 elections. Initial-state quantities are medians across elections. Statewise quantities are first aggregated within each election and then summarized across elections. The final column counts elections with no all-strict continuation block at any state with $B(z) > 0$.

	Initial state		Statewise, aggregated by election			No all-strict block
t	PF_{z_0}	Dev_{z_0}	$\mathrm{PF}_z/\mathrm{Dev}_z$	realized share	States with $\mathrm{Dev}_z = 1$	at every state
0	1.500	1.320	1.096	0.64	0.083	3/590
0.5	1.187	1.137	1.033	0.73	0.143	4/590
0.9	1.032	1.025	1.005	0.78	0.143	4/590
1	1	1	1	—	1	590/590

elections. Specifically, $\mathrm{PF}_z/\mathrm{Dev}_z$ is reduced by its within-election median. The realized share $(\mathrm{Dev}_z - 1)/(\mathrm{PF}_z - 1)$ is formed only at states with $\mathrm{PF}_z - 1 > 10^{-7}$ and is likewise reduced by its within-election median. For the stability calculation, a state is classified as satisfying $\mathrm{Dev}_z = 1$ when $|\mathrm{Dev}_z - 1| \le 10^{-7}$; we report both the fraction of such states within each election and whether the condition holds at every state with $B(z) > 0$.

By Lemma F.1, $\mathrm{Dev}_z = 1$ is equivalent, up to numerical tolerance, to the absence of an all-strict continuation block at that state. This should not be confused with the stronger continuation-stability guarantee of Theorem 5.1, which also rules out deviations that weakly improve every coalition member and strictly improve at least one. Consequently, an election with $\mathrm{Dev}_z = 1$ at every relevant state need not be continuation-stable.

Coalitional deviations become small well before all-strict blocks disappear. At the initial state, the median deviation ratio decreases from 1.320 under the reference policy to 1.025 at $t = 0.9$ (Table 2, Figure 3). Thus, before reaching the Nash endpoint, the median strongest proportional deviation has already fallen from 32% above the no-all-strict-block benchmark of 1 to only 2.5%

above it. Moreover, Dev_{z_0} does not increase between consecutive evaluated values of t in any of the 590 elections. This monotonicity is empirical, not implied by the path construction.

The pattern across states is less uniform. Among the 4173 state sequences tracked across the grid, 47 sequences from 45 elections exhibit an increase in Dev_z somewhere along the path, with a largest increase of 0.047. Thus, the consistent decline at the initial state does not extend to every continuation state.

At the same time, the scalar certificate becomes increasingly tight. The median within-election ratio $\mathrm{PF}_z/\mathrm{Dev}_z$ falls from 1.096 at the reference policy to 1.005 at $t = 0.9$. The close agreement between the two quantities reported in Appendix H.9 is therefore not limited to the particular reference policy: along these paths, the certificate becomes still closer to the exact coalitional deviation ratio as the policy approaches the Nash endpoint. The realized share of the certificate's excess also increases at every evaluated step in each of the five election families.

All-strict blocks disappear much later. Of the 3666 state sequences that do not satisfy $\mathrm{Dev}_z = 1$ at the reference policy, 3542 first satisfy it only at $t = 1$. At $t = 0.9$, only 4 of the 590 policies have $\mathrm{Dev}_z = 1$ at every state with $B(z) > 0$, and 3 of those already have this property at $t = 0$. The main empirical distinction is therefore between the *magnitude* and the *existence* of coalitional deviations: moving toward the Nash-welfare optimum makes the strongest deviations small well before every all-strict block disappears.

H.12 Verification from sampled transitions

This analysis studies how the distribution of transition observations determines how informative the robust certificate is. The relevant notion of coverage is rowwise: observations of a state-alternative pair q shrink the ambiguity set for its transition row according to its visit count N_q, whereas an unvisited row retains the full successor-state simplex. Thus, collecting more transitions elsewhere cannot reduce uncertainty about an alternative that is never selected. This distinction is especially important for verification, because deviations may use alternatives that the behaviour policy itself does not choose.

Design and collection methods. For each of the 27 eligible elections described in Appendix H.5, we hold the election and the policy π being evaluated fixed and vary only the data-collection method and total sample budget. Each transition row receives its own confidence radius $\varepsilon(N_q)$, and the resulting rowwise ambiguity set is used throughout the robust computation, including the optimistic numerator, pessimistic denominator, and utility bounds for each voter used to form intervals for the canonical weights. Unvisited rows are assigned the full successor-state simplex. We write $\mathrm{Cert}_z^{\mathrm{can}}$ for the canonical certificate returned by Definition G.1.

We compare five collection methods. *Uniform generative sampling* distributes observations evenly across the estimated transition rows. *On-policy sampling* generates trajectories under the policy being evaluated π, so rows are observed according to the state-alternative pairs that π visits. *Sparse behaviour-policy sampling* instead uses a policy that assigns zero probability to selected alternatives; in this construction, four of the eight estimated rows are never observed at any sample budget. The *adaptive collector* repeatedly samples a row with the largest current confidence radius, while the *skewed generative collector* samples every row but allocates observations unevenly. The latter two methods distinguish adaptive or imbalanced sampling from the absence of coverage in the sparse sampling method.

Additional samples help only where they improve coverage. At small sample budgets, uniform and sparse collection produce similarly conservative certificates because uncertainty remains large across most rows (Figure 4, right). As the budget grows, the two methods separate. Under uniform sampling, every estimated row receives additional observations and its ambiguity set contracts. Under sparse sampling, four rows remain unobserved and therefore retain the full successor-state simplex at every budget.

For paired comparisons within each election, the sparse canonical certificate exceeds the uniform certificate by a median of 0.419 at 800 samples, 0.646 at 4000, and 0.754 at 16000 (Figure 4, left). At budgets 40 and 200, the corresponding median differences are 0.000 and -0.017. The growing separation does not mean that additional sparse-policy data make the certificate worse; rather, uniform sampling continues to resolve transition rows that the sparse policy never observes.

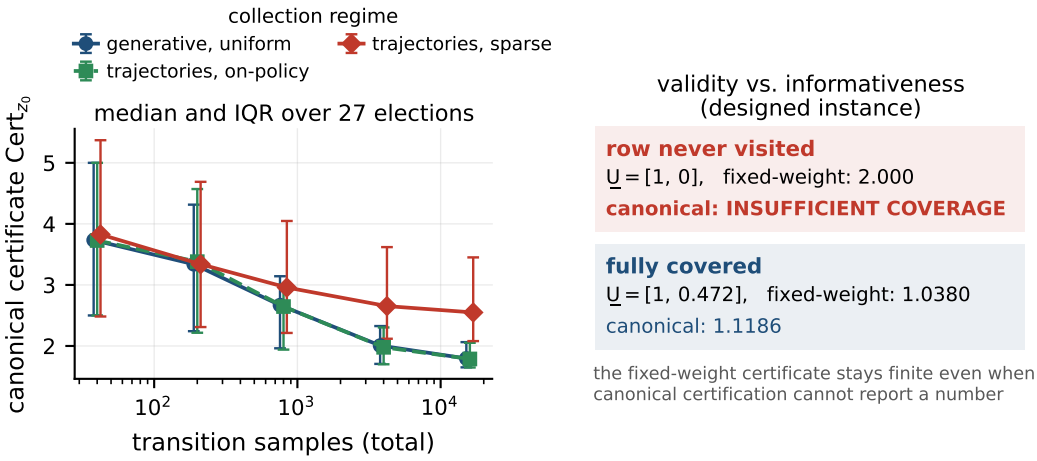


Figure 4: Effect of transition coverage on robust verification. Left: canonical certificates under uniform and sparse data collection as the total sample budget increases, summarized across 27 elections. Four of the eight estimated transition rows remain unobserved under the sparse behaviour policy at every budget. Right: a designed instance in which the canonical procedure cannot establish positive utility for one voter under every kernel in the ambiguity set and therefore returns INSUFFICIENT COVERAGE; the fixed-weight certificate remains finite on the same transition ambiguity set.

Total sample count is therefore not sufficient to determine how informative the data are: what matters is whether observations reach the state-alternative pairs whose transitions affect possible deviations.

The remaining collection methods give similar results. The adaptive collector produces certificates close to those under uniform sampling, while the skewed collector lies between the uniform and sparse methods. For the skewed collector, the paired median excess over uniform sampling is 0.60, 0.62, and 0.39 at budgets 800, 4000, and 16000, respectively. Since every row is eventually observed, this gap begins to close at the largest budget. These two methods are included as additional checks and are not shown in Figure 4.

Validity does not require informative coverage. The right panel of Figure 4 illustrates the distinction between a valid robust certificate and an informative canonical certificate. In this designed instance, the behaviour policy never observes a transition row that matters for the policy being evaluated. Under the resulting ambiguity set, the robust lower utility bounds are $\underline{\mathbf{U}} = (1, 0)$. The data therefore do not establish positive utility for the second voter under every consistent kernel, so Definition G.1 returns INSUFFICIENT COVERAGE rather than a numerical canonical certificate.

This does not invalidate robust assessment itself. With the fixed weights $\boldsymbol{\lambda} = \mathbf{1}$, the certificate on the same ambiguity set is finite, with value 2.000, and retains the guarantee of Theorem G.2(i). When every transition row is observed, the robust lower utility bounds become $\underline{\mathbf{U}} = (1, 0.472)$ and the canonical procedure returns the finite certificate 1.1186. The example therefore illustrates two different consequences of transition uncertainty: it can make the robust deviation bound conservative, and under canonical normalization it can additionally prevent the proportional claims themselves from being identified.

H.12.1 Assessing uncertainty in the canonical weights

Canonical verification must account not only for uncertainty about the transition kernel, but also for uncertainty about the reciprocal weights $\lambda_i^P = 1/U_i^P(\pi)$, since these depend on the policy's utilities under the unknown kernel. To measure the effect of this second source of uncertainty, we compare the canonical certificate with an oracle-weight certificate on exactly the same sampled transitions and transition ambiguity set. The canonical certificate uses the data-dependent reciprocal-weight interval, whereas the oracle-weight certificate is supplied with the true canonical weights. The latter cannot be computed from the sampled data, but removes weight uncertainty without changing transition uncertainty. We use the notation $\mathrm{Cert}_z^{\mathrm{can}}$ and $\mathrm{Cert}_z^{\mathrm{orc}}$ for these two quantities, respectively. Appendix H.6 gives their precise definitions and defines R_{weights}, which measures how much of the

canonical certificate's excess over the known-kernel scalar certificate $\mathrm{PF}^P_{z_0}$ disappears when the true canonical weights are supplied.

Uncertainty about the canonical weights accounts for a substantial part of the certificate gap. For every tested collection method and sample budget, the median R_{weights} lies between 58% and 66%, with 22–24 of the 27 elections contributing to each comparison. Thus, on these instances, supplying the true canonical weights removes more than half of the canonical certificate's excess over $\mathrm{PF}^P_{z_0}$ at the median, even though the transition ambiguity set is left unchanged.

The difference between the canonical and oracle-weight certificates shrinks with improving coverage. Under uniform generative sampling, its median falls from 1.167 at 40 samples to 0.113 at 16000, while the median canonical-to-oracle ratio falls from 1.47 to 1.05. Under sparse behaviour-policy sampling, the corresponding difference falls only from 1.134 to 0.538, and the ratio remains 1.27 at 16000 samples. At the largest budget, the median canonical-minus-oracle differences are 0.107 and 0.111 under on-policy and adaptive collection, respectively, and 0.324 under skewed collection. The contrast is consistent with the coverage results above: when every row continues to receive observations, uncertainty about the canonical weights contracts with the transition ambiguity sets; when some rows remain permanently unobserved, substantial weight uncertainty persists.

Weight uncertainty can prevent a numerical canonical certificate. In 13 of the 675 observations, some robust lower utility bound $\underline{U}_i$ is zero, so the canonical procedure returns INSUFFICIENT COVERAGE. In all 13 cases, supplying the true canonical weights on the same transition ambiguity set produces a finite oracle-weight certificate, between 3.97 and 5.81, while the true $\mathrm{PF}^P_{z_0}$ is also finite, taking value 1.686 or 2.005. The transition ambiguity set therefore still gives a finite robust certificate once the weights are known; what prevents canonical verification in these cases is the inability to obtain a finite reciprocal-weight interval from the data.

The ablation shows that uncertainty in the endogenous proportional claims is not a minor addition to transition uncertainty on these instances. It contributes substantially to the conservatism of finite canonical certificates and, under sufficiently limited coverage, can prevent the canonical procedure from returning a numerical certificate at all. The oracle comparison leaves the transition ambiguity set unchanged, so it measures this effect under the same uncertainty about the transition kernel.

H.12.2 THE INDISTINGUISHABILITY CONSTRUCTION

The preceding experiment shows empirically that permanently unobserved transition rows can keep a valid certificate conservative. The construction of Proposition G.1 establishes a stronger point: no amount of additional data from the same behaviour policy can resolve an alternative that the policy never selects. We use the construction for $R \in \{2, 3, 4, 5, 6, 8, 10\}$. For each R, the two elections differ only in the transition following a single alternative $c^\star$ at a state $z^\star$, and the behaviour policy π_b selects $c^\star$ with probability zero. We verify that the two elections agree on every state-alternative pair used by π_b and separately enumerate their complete trajectory distributions under π_b. The resulting distributions agree exactly at every tested horizon.

Despite these identical observations, the deviation ratios at $z^\star$ are $\mathrm{Dev}^{\mathcal{E}_1}_{z^\star} = 1$ and $\mathrm{Dev}^{\mathcal{E}_2}_{z^\star} = R = H - 1$ (Figure 5). The numerical construction therefore illustrates the inability to distinguish the two elections proved by Proposition G.1: observations generated by π_b contain no information that distinguishes the harmless unchosen alternative in $\mathcal{E}_1$ from the highly profitable deviation in $\mathcal{E}_2$. The impossibility statement follows from theory; the experiment verifies that the implemented instances reproduce the construction.

H.12.3 VERIFICATION ON MODERATELY LARGER INSTANCES

Exact computation of Dev_z restricts the preceding comparisons to small numbers of voters. Since the robust certificate does not enumerate coalitions, we additionally evaluate it on the three moderately larger configurations specified in Appendix H.5: they contain 61, 151, and 201 transition rows, respectively, and each is evaluated over 25 seeds with sample budgets scaled to the number of rows. We do not compute Dev_z in this analysis; the known-kernel reference quantity is $\mathrm{PF}^P_{z_0}$.

The same coverage pattern persists. At the largest tested budgets, the canonical certificate under sparse sampling remains approximately 2.4–2.7 times as large as under uniform sampling. The

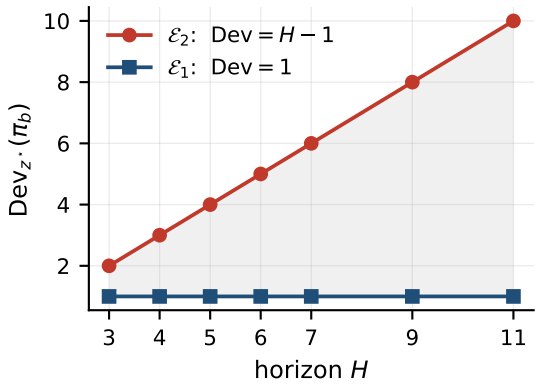


Figure 5: Deviation ratio at $z^\star$ in the two observationally indistinguishable elections of Proposition G.1. The behaviour policy never selects the alternative whose transition differs between the elections, so their trajectory distributions under that policy are identical at every sample size. Nevertheless, the deviation ratios are 1 and $H-1$, respectively. The figure instantiates the deterministic construction; the impossibility claim follows from Proposition G.1.

largest rowwise ambiguity radius under sparse sampling remains at the simplex diameter because some rows are never observed, whereas under uniform sampling it falls to about 0.45. Thus, increasing the overall problem size does not remove the distinction between collecting more samples and covering the transition rows that matter.

The larger configurations also reveal substantially greater conservatism in the robust certificate itself. Under uniform coverage, the median ratio $\mathrm{Cert}^{\mathrm{can}}_{z_0}/\mathrm{PF}^{P}_{z_0}$ is 2.21, 2.75, and 2.24 across the three configurations, compared with 1.11 on the small instances. The larger configurations vary simultaneously in the number of voters, horizon, layer width, and number of alternatives, so this comparison does not identify which dimension is responsible for the increase. It should therefore be read as evidence that the favorable numerical tightness observed on the small instances need not persist unchanged on larger configurations, not as an asymptotic scaling result.

Verification remains computationally inexpensive on these tested configurations, taking 4–21 ms per instance. These timings establish only that the implementation remains practical at the reported sizes; three configurations are insufficient to justify a claim about asymptotic computational scalability.

H.13 Cost of exact deviation computation

The exact computation of Dev_z in Appendix H.3 enumerates every nonempty coalition of active voters and solves one continuation linear program for each. By contrast, PF_z avoids coalition enumeration and is computed by a policy evaluation and a scalar Bellman pass. To illustrate the computational difference between these two procedures, we measure their runtimes on the same instances as the number of voters increases. This experiment complements the NP-completeness result of Theorem F.1, but does not provide a complexity analysis: it measures the runtime of the particular implementations used in our experiments.

For n active voters, exact computation considers 2^n-1 nonempty coalitions. The measured runtime difference grows accordingly. The median runtime of exact Dev_z computation is approximately 35 times that of the scalar certificate at $n=2$, where 3 coalitions are enumerated, and approximately 3000 times as large at $n=8$, where 255 coalitions are enumerated (Figure 7). Timings use warm-up calls followed by batched repetitions and are summarized across instances by their median and the range between the 25th and 75th percentiles. We also verify by call counting that the timed PF_z implementation does not call the coalition-enumeration routine.

The comparison illustrates the reason for computing the scalar certificate: in our implementation, the cost of exact evaluation grows rapidly as more voters introduce more coalitions, whereas computing PF_z requires no coalitional search. This is why the experiments that use exact Dev_z are restricted

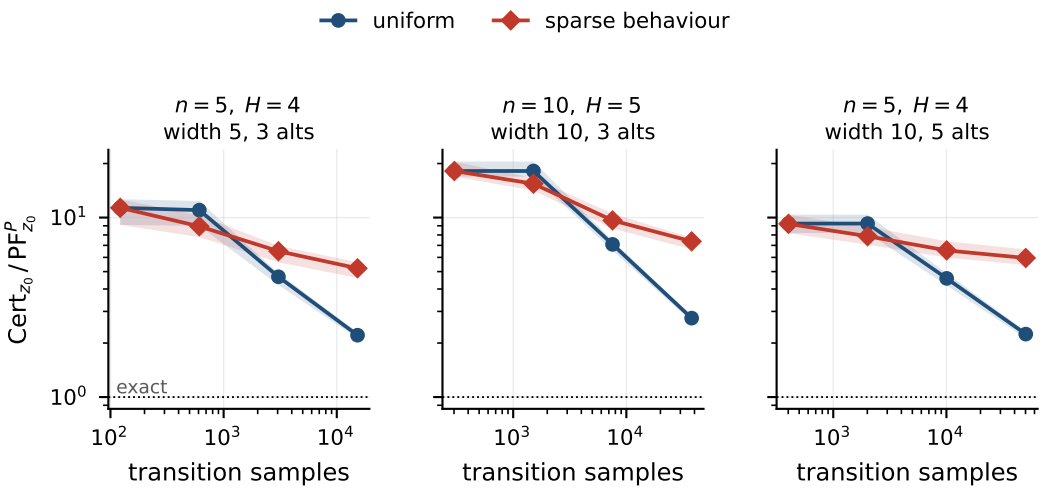


Figure 6: Canonical robust certificate relative to the known-kernel $\mathrm{PF}^{P}_{z_0}$ on three moderately larger configurations, summarized over 25 elections. Sparse and uniform sampling remain separated because some transition rows are never observed under the sparse sampling method. The absolute certificate ratio is also larger than on the small configurations. Since the three configurations change several problem dimensions at once, the comparison does not identify the source of this additional conservatism.

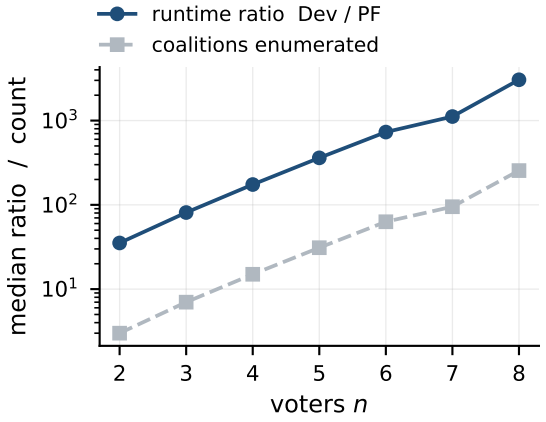


Figure 7: Computational cost of exact deviation evaluation relative to the scalar certificate. The runtime ratio compares coalition enumeration for Dev_z with the computation of PF_z on the same instances as the number of voters increases; the figure also reports the corresponding number $2^n - 1$ of nonempty coalitions. Runtime summaries are medians across instances. The measurements characterize the implementations used in the experiments, rather than the asymptotic complexity of the two problems.

to small numbers of voters. We do not fit a runtime scaling law to these measurements; the formal computational result is the NP-completeness statement of Theorem F.1.

H.14 Implementation validation and regression tests

The experimental implementation includes regression tests that check the numerical routines against analytical identities, independent implementations, and closed-form constructions. These checks are intended to detect implementation errors; the validity of the corresponding mathematical statements follows from the theoretical results proved elsewhere in the paper.

Robust Bellman computation. The robust certificate allows an L_1 ambiguity set to place probability on any state in the full successor-state simplex, including states assigned zero probability by the nominal kernel. We test this implementation in several ways. First, perturbed transition rows are generated by explicitly transferring probability mass between disjoint source and destination sets. This construction ensures membership in the intended ambiguity set, and we additionally verify for all $35{,}760$ generated rows that $\|\mathbf{q} - \mathbf{p}\|_1 \le \varepsilon$, $\mathbf{q} \ge 0$, and $\sum_x q_x = 1$.

Second, we compare the exact rowwise optimization used by the robust Bellman passes with a generic linear program over the full successor-state simplex. The comparison includes empirical kernels with zero coordinates and continuation-value vectors with tied maximizers. Across 6000 such problems, the largest discrepancy is 1.3×10^{-15}. This validates the mass-transport implementation of the inner problem described in Appendix H.4 against an independent optimization routine.

Finally, the regression tests for the robust certificate evaluate all three extended-real cases of Definition 7.1 and checks the conclusion of Theorem 7.1 at every state of every generated test instance. No violation is observed.

Scalar-certificate identities. We separately test the deterministic identity $\mathrm{PF}_z = 1 + \Delta_\pi(z)/B(z)$ wherever $B(z) > 0$. The set of regression tests contains 4760 observations obtained by interpolating from a reference policy toward a policy greedy for a scalarized objective. We run two versions of this check: one keeps $\boldsymbol{\lambda}$ fixed at the reference policy's canonical reciprocal weights, while the other recomputes the canonical weights after the policy changes. The two versions are analyzed separately. Across all observations, the largest error in the identity is 4.4×10^{-16}.

The fixed-weight version also checks that the implementation does not treat fixed scalarization weights as the canonical weights of the changing policy. By the end of the interpolation, the largest discrepancy between the fixed weights and the current policy's reciprocal weights is 3.6. This regression experiment is distinct from the Nash-path analysis of Appendix H.11: its purpose is to test the scalar-certificate identity under controlled changes to the policy, not to study a path toward a Nash-welfare optimum.

Nash-endpoint validation. The optimization routine used in Appendix H.11 is additionally checked on 38 elections drawn from the five small-election families. For each election, we verify that every voter has positive utility and that the Frank–Wolfe gap is at most 10^{-9}. We independently solve the same Nash-welfare optimization problem with an interior-point method and verify that the resulting log-Nash welfare agrees to within 10^{-6}.

After completing and polishing the policy, we also verify the identities predicted at the Nash endpoint: $\lambda_i U_i = 1$ for every voter, $\mathrm{PF}_z = 1$ and $B(z) = y^{\boldsymbol{\lambda}}(z)$ at every nonterminal state, and $\mathrm{Dev}_z = 1$ wherever $B(z) > 0$. We check that polishing changes the log-Nash welfare by at most 10^{-9} and that utilities are affine along the occupancy path. All of these validation checks pass. The full 590-election Nash-path experiment applies the endpoint acceptance criteria reported in Appendix H.11 separately.

Closed-form and cross-method checks. Before running the empirical analyses, we test the core value and occupancy routines on instances for which the relevant quantities can be computed independently. The constructions of Theorems 3.1 and 3.2 are checked against their specified state structure, approval counts, forced lottery, and resulting utilities. Policy utilities computed by dynamic programming are compared with Monte Carlo estimates, and utilities computed from occupancy measures are compared with their dynamic-programming counterparts. Scalar Bellman values are checked against values obtained from sampled policies, and the canonical normalization $\lambda_i U_i = 1$ is verified whenever reciprocal weights are used. We also compare the artificial-root implementation with direct evaluation under the initial distribution μ.

In addition, the statewise identities used by the verification theory are checked on every generated experimental instance. If a checked identity does not hold, the corresponding run terminates and saves the seed and state where the error occurred. All validation and regression tests used for the reported experiments pass.

H.15 COMPUTATIONAL RESOURCES AND REPRODUCIBILITY

Computational environment. All experiments were run on a single Apple M4 Max (arm64) machine under macOS 26.6.2. The main experimental environment uses Python 3.11.15, NumPy 2.4.6, and SciPy 1.17.1. No GPU is used. Computation is single-threaded at the experiment level, apart from any internal parallelism used by the numerical libraries.

The linear programs used for exact deviation computation (Appendix H.3) and the certificate-exactness check (Appendix H.9) are solved through SciPy's `linprog` interface to HiGHS. The Nash endpoint in Appendix H.11 is computed by the pairwise Frank-Wolfe implementation described there, using NumPy and the scalar Bellman routine as its linear optimization oracle. None of the reported experimental results requires a commercial solver or a convex-optimization modelling package. CVXPY 1.9.2 is used only for the independent interior-point cross-check of the Nash endpoint in Appendix H.14. Each saved output includes details of the software environment used to produce it.

Computational cost. The full experimental suite runs in minutes on the machine above. The most expensive analyses are those that compute Dev_z exactly, since they enumerate coalitions and solve a continuation linear program for each one (Appendix H.13). This is why the analyses requiring exact Dev_z use at most four voters. The Nash-path analysis is the most computationally intensive of these experiments: it evaluates exact deviations at every relevant state for six policies in each of 590 elections and takes approximately 96 seconds in total. By comparison, robust verification on the largest configurations tested, containing 201 transition rows, takes 4–21 ms per instance.

Mapping experiments to reported results. Each reported table or figure is generated from a saved experimental output. Table 1 uses the scalar-certificate experiment; Figure 2 uses the action-dependence experiment; Figure 3 uses the Nash-path analysis; Figure 4 uses the coverage experiment; Figure 5 uses the indistinguishability construction; Figure 6 uses the larger-instance analysis; and Figure 7 uses the runtime experiment. The weight ablation in Appendix H.12.1 reproduces the coverage experiment's sampled data from the same seeds and verifies that the resulting rowwise visit counts match before computing the additional certificates. The Nash-path analysis similarly verifies that its $t = 0$ quantities reproduce the corresponding reference-policy results of Appendix H.9.

Each experiment writes its configuration, seeds, and numerical outputs to a result file, and figures and tables are generated from these saved outputs rather than by rerunning the underlying experiment. No external dataset or preprocessing step is needed: all experimental inputs are generated from the specified seeds.